Comparative analysis of novel and common reference genes in adult tissues of the mussel *Mytilus galloprovincialis*

Federica Salatiello¹,², Marco Gerdol², Alberto Pallavicini¹,², Annamaria Locascio¹ and Maria Sirakov¹*

**Abstract**

**Background:** Real-time quantitative PCR is a widely used method for gene expression analyses in various organisms. Its accuracy mainly relies on the correct selection of reference genes. Any experimental plan involving real-time PCR needs to evaluate the characteristics of the samples to be examined and the relative stability of reference genes. Most studies in mollusks rely on reference genes commonly used in vertebrates.

**Results:** In this study, we focused on the transcriptome of the bivalve mollusk *Mytilus galloprovincialis* in physiological state to identify suitable reference genes in several adult tissues. Candidate genes with highly stable expression across 51 RNA-seq datasets from multiple tissues were selected through genome-wide bioinformatics analysis. This approach led to the identification of three genes (*Rpl14*, *Rpl32* and *Rpl34*), whose suitability was evaluated together with 7 other reference genes commonly reported in literature (*Act, Cyp-A, Ef1α, Gapdh, 18S, 28S and Rps4*). The stability analyses performed with geNorm, NormFinder and Bestkeeper identified specific either single or pairs of genes suitable as references for gene expression analyses in specific tissues and revealed the *Act/Cyp-A* pair as the most appropriate to analyze gene expression across different tissues.

**Conclusion:** *Mytilus galloprovincialis* is a model system increasingly used in ecotoxicology and molecular studies. Our transcriptome-wide approach represents the first comprehensive investigation aimed at the identification of suitable reference genes for expression studies in this species.

**Keywords:** *Mytilus galloprovincialis*, Reference genes, Real time RT-PCR, RNAseq

**Background**

Mussels are organisms important both from a commercial and from a scientific standpoint. With a global production of several hundred thousand tons per year, *M. galloprovincialis* is one of the most relevant farmed edible bivalve in the Mediterranean area [1]. The use of mussels as model organism covers a broad range of research fields: from biomonitoring and ecotoxicology [2], to translational medicine [3], to invertebrate immunology [4]. The growing scientific interest in these organisms is progressively revealing the near-complete lack of specific laboratory protocols, whose set-up would undoubtedly improve the quality and reproducibility of scientific research in the field.

Over the years, the study of gene expression has gained importance in many fields of scientific research, providing novel insights about regulatory networks and biological processes. Dye-based Real-time quantitative Polymerase Chain Reaction (RT-qPCR) using dye such as SYBR® Green dye is an economical option for monitoring gene expression which yields reproducible results without the requirement with expensive and labor-intensive
fluorescent probe design. Dye-based real time is a sensitive method for the quantification of mRNA in biological samples [5, 6]. Due to its simplicity, rapidity and high specificity, it has become one of the most popular techniques applied for targeted nucleic acid quantification [6]. Despite its wide use and the multiplicity of available protocols, this technique is affected by several problems related to the intrinsic diversity of RNA populations among samples and to the presence of technical errors and experimental biases [6, 7]. To correct these sources of errors, the data collected are usually normalized with one or more internal controls, defined as reference genes whose mRNAs are stably expressed across a broad range of conditions. Although the use of more than one reference gene is usually recommended, their optimal number and combination need to be adequately selected for each experimental condition [6]. Nevertheless, most gene expression studies carried out in molluscs do not use reference genes validated in the species of interest, often relying on genes whose expression is simply assumed to be stable based on data collected from largely divergent vertebrate model species.

The most popular reference genes, used in a wide variety of model organisms, are β-Actin (Act), Glyceraldehyde-3-phosphate dehydrogenase (Gapdh) and the 18S ribosomal RNA (18S) Act can be considered as one of the first internal standards used in gene expression quantification studies, and it is still one of the most widely used internal standards today [8]. Although Act is often considered as a single-copy gene, it is part of a multi-gene family that includes several nearly-identical members, which usually allows the design of shared PCR primers [9]. Its mRNA is abundantly expressed in most cell types and encodes a ubiquitous cytoskeleton protein [5]. The Gapdh mRNA encodes the glyceraldehyde-3-phosphate dehydrogenase enzyme, which catalyzes an important step in glycolysis. Due to its abundance and ubiquitous expression, it is frequently used as a control in RT-qPCR experiments [5]. Lastly, rRNAs are considered useful internal controls since they are generated by a distinct polymerase (i.e. RNA polymerase 1) [10] and their levels are not expected to vary in a significant manner under conditions affecting the expression of mRNAs [11]. However, a high number of tandemly duplicated identical rRNA genes copies are usually present in eukaryotes, and their abundance depends on genome size [12]. While several studies indicate that 18S rRNA is a more suitable reference gene than Gapdh and Act (e.g. [13–15]), others define its use as inappropriate in some cases, such as mammalian cells (e.g. [16, 17]). Since these genes can undergo regulation in response to specific natural or experimental conditions (e.g. act mRNA has a cell-cycle dependent expression) a validation process is highly recommended before their inclusion in any panel of reference genes in order to avoid measurement errors [5, 7].

Although a few suitable reference genes for RT-qPCR experiments have been previously identified in mussel [18–20], no specific gene set to be used in different tissues and/or experimental conditions has been described yet. A study carried out on M. edulis identified Elongation factor 1 alpha (Efta) and ribosomal 18S and 28S RNAs as the most stable targets across different stages of gametogenesis, while Helicase (Hel) and Act were defined as unstable targets [18]. Different combinations of reference genes have been validated for use in M. galloprovincialis digestive gland and gills (Gapdh, Ribosomal protein S4 - Rps4 and Cytochrome oxidase subunit 1 - Cox1) and mantle (Gapdh, Rps4 and Rps27) following exposure to okadaic acid [19]. A recent study carried out in the golden mussel Limnoperna fortunei, highlighted poor stability of gene expression levels in gonads of individuals of different sexes, suggesting the presence of remarkable intra-specific differences [20]. The recently reported widespread hemizygosity of the mussel genome, associated with gene presence-absence variation phenomena, represent another potential issue in the identification of suitable stable reference genes in this species [21].

Some studies in vertebrates (human, mice, zebrafish) and plants (tomato, seaweed, and kiwi) successfully identified housekeeping or reference genes taking advantage of RNA-Seq data [22–26]. To the best of our knowledge, so far this approach has only been used, among molluscs, in scallop due to the unavailability of complete transcriptomic datasets for most species [27].

The aim of the present study was to identify candidate genes suitable as internal references for RT-qPCR experiments in different tissues of the mussel. The suitability of ribosomal protein L14, L32 and L34 (rpl14, rpl32 and rpl34), selected using a transcriptomic approach, was comparatively assessed with seven additional gene targets commonly used in literature and previously used in gene expression studies targeting mussel. We investigated their stability in digestive gland, gills, mantle, gonads and foot in physiological conditions in order to provide a solid basis for future expression studies. We also used the most stable reference genes to assess their reliability in evaluating the expression of four highly expressed tissue-specific genes of interest (GOI) in the mantle and in the digestive gland.

Results

Identification of the candidate reference genes from RNA-Seq data

As previously reported by Gerdol et al. [21] Rpl32, Rpl14 and Rpl34 were identified as the first, third and fourth most stable core genes across 51 RNA-seq datasets from...
multiple adult tissues and experimental conditions in *M. galloprovincialis*. *Rpl32*, with a stability score equal to 0.24, achieved a mean expression level of 3061.45 TPMs; *Rpl14*, with a stability score equal to 0.25, achieved a mean expression level of 2747.49 TPMs; *Rpl34*, with a stability score equal to 0.27, achieved a mean expression level of 2288.28 TPMs (Figure S1). These three candidate genes were confirmed to lack paralogous gene copies in the *M. galloprovincialis* genome, and in silico primer check ruled out the possibility of non-specific amplification of other genomic regions or transcribed RNAs.

**Expression stability of candidate reference genes**

In this work, we analyzed ten genes, considering three novel candidates selected with a transcriptomic approach (*Rpl14, Rpl32, Rpl34*) and seven genes commonly used as references (*Act, Cyp-A, Ef1α, Gapdh, 18S, 28S, Rps4*), verifying their expression stability in various mussel tissues. The range of cycle threshold (Ct) values was quite similar among samples, with *18S* and *28S* exhibiting the highest range of variation in Ct in all tissues analyzed (Fig. 1, S2).

An overview of the results obtained from the analysis of the selected reference genes using three different algorithms is presented in Fig. 2. In the gills, *Rps4* was identified as the most stable gene by *geNorm* evaluation, whereas *Act* and *Cyp-A* were selected by the *NormFinder* and *BestKeeper*, respectively. However, all these three genes were generally considered stable by the three algorithms but with a different ranking (Table S1). Interestingly, *geNorm* suggested the use of *Gapdh/Rpl34* as the best combination of reference genes, as both showed higher M values than the other analyzed genes (Fig. 2A, Table S1).

In the digestive gland, *Gapdh* and *Rps4* showed the lowest (and identical) M value according to *geNorm*; nevertheless, the algorithm suggested *Rps4/Rpl14* as the best possible combination since the latter gene had an M value slightly higher than *Gapdh* (Table S2). *Gapdh* has ranked first also in the *NormFinder* analysis (Table S2). In addition, both *Rps4* and *Rpl14* displayed good stability values. Conversely, *Gapdh, Rps4* and *Rpl14* were found to be beyond the recommended value (SD < 1) in the *BestKeeper* analysis, while *Cyp-A* ranked first (Table S2). It is important to note that *Cyp-A* was also identified as a stable gene by the *geNorm* and *NormFinder* analysis (Table S2, Fig. 2B).

The comparison among the different algorithms produced less consistent results in the gonads, with *Gapdh* resulting as the most stable gene by *geNorm*, *Rps4* by *NormFinder* and *Rpl34* by *BestKeeper* (Table S4). In addition, the best combination, according to *geNorm* was *Cyp-A/Gapdh*. All the named genes were generally...
considered stable by the three algorithms, except for Rps4 that showed an SD value much higher than recommendations (SD = 1.25) by BestKeeper (Table S4, Fig. 2C).

In the mantle, Cyp-A showed the highest stability according to both geNorm and BestKeeper, in contrast with NormFinder, which indicated Act as the most stable gene (Table S3). Nevertheless, Cyp-A and Act were considered stable by the three analyses, even though they occupied different positions in the rankings. Furthermore, geNorm suggested Cyp-A/Rpl14 as the best combination of reference genes (Table S3, Fig. 2D).

In the foot, Gapdh, Act and Cyp-A had the best stability values in geNorm, NormFinder and BestKeeper ranking, respectively (Table S5). The best pair of reference genes suggested by geNorm was Rpl14/Rpl32. As previously discussed for the other tissues, all these genes were considered relatively stable from the different algorithms, with the exception of act, which had an SD value higher than 1 (SD = 1.29) (Table S5, Fig. 2E).

Finally, we reanalyzed the data by considering all the tissues. As previously mentioned for the analysis of single tissues, 18S and 28S were the most variable candidate reference genes across all the tissues (Fig. 3). NormFinder defined Act and Cyp-A as the best gene pair combination.

Expression level of target genes

We validated the bona fide selection of the previously described reference genes for the interpretation expression data of a set of GOI. To this aim, we analyzed the expression of four genes differentially expressed in adult tissues, selected according to their strong specific expression inferred from RNA-seq data in the mantle and digestive gland, respectively. In detail, MGAL10A005692, specifically expressed at high level (averaging ~ 3900 TPM) in the mantle, encodes a low-complexity protein, which bears distant homology with SCO-spondin in the C-terminal region. The second selected mantle-specific GOI, MGAL10A087091, was expressed at similar levels (averaging 3600 TPM) and was orthologous with M. coruscus MUSP-3, previously reported to be involved in shell mineralization [28].

On the other hand, MGAL10A040115, a meprin A-like metalloprotease, was one of the most distinctive digestive gland-specific genes in mussel, reaching expression values as high as 100,000 TPM in some samples. The second selected digestive gland GOI, which also displayed very high mean expression values (32,000 TPM) was MGAL10A054097, encoding a protein with multiple ShKT domains.

As reported in Fig. 4, the expression profile of MGAL10A005692 and MLGAL10A087091 confirmed
the high mantle-specificity indicated by transcriptome data. In particular, the Ct values were higher than 30 in all the tissues other than mantle and the expression of MGAL10A087091 was almost undetectable in gonad and gills (data not shown). The expression profiles of MGAL10A040115 and MGAL10A054097 are reported in Fig. 5. The specificity of these genes for the digestive gland was strongly confirmed, as in all the tissues other

Fig. 3 Best candidate reference genes across tissues. Stability of the reference genes was evaluated considering their stability in all the tissues based on geNorm analyses of the RT-qPCR experiments. Data is arranged in descending order of comprehensive stability from left to right.

Fig. 4 Relative expression of GOI mantle specific. Relative expression was calculate according to the ΔCt methods using Act (A, C) or the Geometric mean of Act-Cyp-A (B, D) as reference. N = 3.
than digestive gland the Ct values were higher than 30 (data not shown).

We tested a few other specific GOIs in the digestive gland (MGAL10A065261, MGAL10A066134), in the mantle (MGAL10A006232, MGAL10A059842, MGAL10A079192) and in the foot (MGAL10A033477, MGAL10A064866, MGAL10A070506). However, possibly owing to the low sequence complexity of these gene targets (which often encoded highly repetitive proteins), the efficiency of amplification observed by RT-qPCR data in these cases was very low (data not shown).

For all analyzed GOIs, no significant differences could be evidenced between the use of Act as a single reference gene and the use of the Act-Cyp-A pair (Figs. 4 and 5). As reported in Figure S3, the use of the last ranking genes, such as 18S and the couple 18S and 28S still allow to detect the tissue specific of the GOIs (i.e MGAL10A087091 and MGAL10A040115, mantle and digestive gland, respectively) but with an higher standard error.

**Discussion**

In this work, we analyzed the stability of ten genes suitable as references, including three novels (Rpl14, Rpl32, Rpl34) and several commonly used (Act, Cyp-A, Ef1a, Gapdh, 18S, 28S, Rps4) genes within and across different mussel tissues. We first looked at the Ct values, revealing that Act was the reference gene showing the highest Ct value among those considered in this study. With this regard, it needs to be specified that primer design targeted one out of the several paralogous actin genes found in the *M. galloprovincialis* genome (sequence ID: AF157491.1), which may not correspond to the most abundant cytoskeletal isoform.

The stability of the suitable reference genes was evaluated using specific algorithms such as geNorm [29], NormFinder [30] and BestKeeper [31]. Despite differences in the ranking of the genes, the results obtained from the three algorithms were quite similar in the different tissues. Most discrepancies were founded in the BestKeeper ranking, but all methods were concordant in the identification of unstable genes. Notably, 28S and 18S were the targets with the lowest score in all the tissues according to both geNorm and NormFinder, and 18S was one the most stable reference only in the DG according to BestKeeper (Fig. 2). The poor stability of rRNA expression may have multiple explanations, which include the presence of inter-individual variability in the number of rRNA gene copies present in individual genomes and a different level of DNA methylation of rRNA gene clusters. Although this has been previously demonstrated in human [32, 33], to the best of our knowledge no study

---

**Fig. 5** Relative expression of GOI digestive gland specific. Relative expression was calculate according to the ΔCt methods using Act (A,C) or the Geometric mean of Act-Cyp-A (B,D) as reference. N = 3
has ever been carried in bivalves to clarify this aspect. Nevertheless, the widespread hemizygosity and gene presence/absence variation phenomena found in mussel and other bivalves [34] may provide an explanation for the poor stability of expression of 18S and 28S rRNA genes in M. galloprovincialis.

As shown in Tables S1, S2, S3, S4 and S5, all the other genes were considered relatively stable with geNorm and NormFinder, even with differences in the specific ranking. On the other hand, the BestKeeper analysis calculated a SD value below the desired threshold for some suitable references. These include Act (in gills, digestive gland, gonads and foot), Ef1a (in gills, digestive gland and gonad), 18S (in gills, digestive glands, gonads, foot), Rsp4 (in digestive gland and gonads), Gapdh (in digestive gland and mantle), Rpl32 (in digestive gland and gonad), Rpl14 (in digestive gland) and Cyp-A (in gonads).

Finally, the analyses carried out revealed some discrepancies between RNA-seq and RT-qPCR expression data, such as the low performance of the three newly selected reference genes compared to those traditionally used in literature. This could definitely become a point of interest, not only for mussels studies but possibly for many other non-model marine organisms for which no specific reference genes have been identified and validated yet. Although RT-qPCR approaches are commonly used to confirm the results of differential gene expression analyses carried out on an “omic” scale, the data obtained with the two methodologies are not always concordant. Such differences can be explained by multiple technical and biological factors, which include, among the others, the different dynamic range of the two methods and the dependence of RT-qPCR on appropriate primer design [35]. One crucial point that could mitigate these potential issues, in particular in non-model organisms with a complex genomic background such as M. galloprovincialis, could be the choice of reference genes showing high stability in the biological context of interest. Other “bench tips”, such as those part of the MIQE guidelines [6], may further improve the application of RT-qPCR analyses to non-model marine species and allow their comparison with the results obtained with “omic” approaches.

**Conclusions**

In the present study, we analyzed the expression profiles of ten candidate genes from M. galloprovincialis in physiological state to assess their potential use as references in RT-qPCR experiments. Normalization strategies are required to correct experimental errors such as variations in extraction yield, reverse-transcription yield and efficiency of amplification [6]. Although the calculation of the expression levels of genes of interest commonly relies on the use of reference genes as internal controls for normalization, their reliability must be experimentally validated for each experimental design [6]. The analyses conducted with geNorm [29], NormFinder [30] and BestKeeper [31] allowed us to identify the best candidate genes to be used in different adult mussel tissues, as well as the most suitable pair of references to be used for comparisons among tissues (Act/Cyp-A). Remarkably, the 18S and 28S rRNA genes, which use as references is widespread in literature, were in all cases the more unstable among the ten candidate reference genes tested. We therefore suggest that RT-qPCR analyses in M. galloprovincialis should rely on the use of different nuclear reference genes, to be carefully selected based on the tissue of interest and the experimental conditions.

**Methods**

**Sample collection**

For the validation of reference genes, adult M. galloprovincialis mussels (5.0 ± 1.0 cm shell length) were sampled from the Bay of Naples (Italy) in autumn 2019. In the laboratory, specimens were dissected to separate digestive gland, gills, mantle, gonad and foot. Tissue samples were immediately weighed and stored at −80°C until further analysis.

**In silico identification of candidate stable reference genes and GOI**

“A total of 51 different RNA-seq dataset were mapped to the M. galloprovincialis reference assembly. These datasets derive from multiple previously published studies, carried out by different research teams and taking into account mussels sampled in different geographical locations [36–40]. These transcriptomic data refers to multiple tissues, subject to different experimental conditions (see Table S6)”. Only core genes (i.e. those invariably present in all individuals) were considered, whereas dispensable genes (i.e. those subject to presence-absence variation) were discarded. Mapping was carried out with the CLC Genomics Workbench v. 20 (Qiagen, Hilden, Germany), setting the length fraction and similarity fraction parameters to 0.75 and 0.98, respectively. Digital read counts were used to computing gene expression levels as Transcript Per Million (TPM) [41]. The selection of candidate stable reference gene followed an approach similar to the one described below for BestKeeper (see Results paragraph). The mean expression value was calculated for each core gene across all 51 samples. All TPM expression values, for each gene and dataset, were subsequently transformed by dividing them by the mean value, obtaining a distribution of all records centred on 1. The calculation of standard deviation of the standardized gene expression values allowed to evaluate gene expression stability, with the most and least stable genes...
being identified by low and high SD values, respectively. The four GOIs (two for the digestive gland and two for the mantle tissue) were selected based on the following criteria: (i) inclusion in the *M. galloprovincialis* core gene set, indicating presence in all individuals; (ii) absence of paralogous genes, to avoid any chance of non-specific cross-amplification; (iii) high expression in the tissue of interest (i.e. \( \geq 500 \) TPM), as evaluated by the analysis of the available RNA-seq datasets (Table S6); (iii) highly specific tissue expression, defined as a fold change \( \geq 10 \) in the pairwise comparisons between the expression levels observed in digestive gland/mantle and all the other tissues; (iv) a gene architecture compatible with the design of primer pairs able to discriminate between specific cDNA amplification signals and genomic DNA contamination.

### Total RNA extraction and cDNA synthesis

Up to 100 mg of each tissue were lysed and total RNA was processed with *SV Total RNA Isolation System* kit (Promega) according to the manufacturer’s instructions. For all RNA samples, the absence of DNA contamination was tested by performing PCR with *ef1a* primers designed within an exon (for sequence see Table 1 for *ef1a-g*). Concentration and purity of all RNA samples were measured using a Nanodrop ND1000 spectrophotometer. RNA integrity was evaluated through 1.5% agarose gel electrophoresis. Total RNA (0.3 μg) from each tissue was reverse transcribed using the *iScript RT Supermix* (Bio-rad) following the manufacturer’s protocol and stored at \(-20^\circ C\). Subsequently cDNA was diluted with \( H_2O \) prior to use in RT-qPCR experiments.

### Table 1  Primer sequences

| Gene     | Primer sequence 5′-3′ | Amplicon | Intron | Refs | E  |
|----------|-----------------------|----------|--------|------|----|
| Ef1α-g   | F TTTCTGGATGGCAGCGAGAC | 150 bp   | No     |      |    |
|          | R CTGTGGCTTGTGATGTTGGG |          |        |      |    |
| Rpl34    | F ACACAAAGAAGGAGGTGCA | 136 bp   | Yes    |      |    |
|          | R GGCCCTCAAGAAGCTGTTAC |          |        |      |    |
| Rpl32    | F CACCTGATGCTGATGTTCC | 143 bp   | Yes    |      |    |
|          | R GCAGCTCAAGATTCTGTTTC |          |        |      |    |
| Rpl14    | F CAATGCGCAAGATCCCTGTA | 167 bp   | Yes    |      |    |
|          | R AGCTTGGCTGCGCTCATCAG |          |        |      |    |
| Act      | F ATGCTCCAAGGGCCTGGTTC | 114 bp   | Yes    |      |    |
|          | R CTCCTCTGCTGCGCTCATC |          |        |      |    |
| Cyp-A    | F GGAGTGTGACGCTAGTGTTAC | 100 bp   | Yes    |      |    |
|          | R TGGTTGGAATTGTGTCTTG |          |        |      |    |
| Ef1α     | F CCAGTGGCAACGATTTTGAG | 183 bp   | No     |      |    |
|          | R TGGCTGGCGCAAAGTTAACAC |          |        |      |    |
| 18S      | F AGAAAGGCTACACACATCCA | 160 bp   | Yes    |      |    |
|          | R TGCCCTCAAGATATCTCTG |          |        |      |    |
| 28S      | F AGTGACTCTTCTGCGTGTAG | 123 bp   | Yes    |      |    |
|          | R GACGAGTCAACACTAGAGG |          |        |      |    |
| Gapdh    | F AGGAATAGCCCTCCAGGAC | 114 bp   | Yes    |      |    |
|          | R TCGATGCTGCTTTTAATGGCT |          |        |      |    |
| Rps4     | F TGCCGTAGGGCGCTAG | 121 bp   | Yes    | [19] | 1.9|
|          | R TCCCTTAAATTGTGTAGGCACTCTG |          |        |      |    |
| MGA10A040115 | F TGGCACAAGAACGACATTC | 167 bp   | Yes    |      | 1.9|
|          | R CTGTGGGTGGTGAAAGTCA |          |        |      |    |
| MGA10A054097 | F AACAACCCCGAGATCCACCT | 212 bp   | Yes    |      | 1.9|
|          | R TCCTATTGGGCAAGCTGGGC |          |        |      |    |
| MGA10A087091 | F GCAACGGCGCAAGAATTACG | 141 bp   | Yes    |      | 2.0|
|          | R TTGCTCTTGCAGCACCACCTG |          |        |      |    |
| MGA10A005692 | F GCAGGTGTGGTTGTGTAC | 156 bp   | Yes    |      | 2.1|
|          | R AGAGAGATGCTGTGGTGCT |          |        |      |    |

*F* Forward, *R* Reverse, amplicon size in bp, presence of an intron within the amplified region and calculated reaction efficiencies (E) for RT-qPCR experiments
Primer design
Ten putative reference genes: Act, Cyp-A, Ef1α, 18S, 28S, Rpl34, Rpl32, Rpl14, Gapdh and Rps4 and four GOI (MGAL10A005692, MGAL10A087091, MGAL10A040115 and MGAL10A054097) were selected. Primers are listed in Table 1.

Genomic sequences were retrieved from NCBI or from the reference genome assembly [21] and gene structure was predicted using Splilign (https://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi). RT-qPCR primers were manually designed in adjacent exons and verified using Primer blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome).

The sequences of Gapdh and Rsp4 primers were retrieved from the work of Martinez-Escauriaza et al. (2018) and the presence of a spanning intron was verified as described above.

Previously generated M. galloprovincialis whole genome sequencing data [21] were used to confirm that all selected target genes were not subject to presence-absence variation. At the same time, primer specificity was assessed in silico by BLASTn against the reference genome assembly, verifying the absence of paralogous genes and significant non-specific matches between the forward and reverse primers and non-target genomic regions.

The specificity of amplicons was confirmed by the presence of a single band of the expected size on 2% agarose gel. Finally, the PCR products were purified using the illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare) following the manufacturer’s instructions and their identity was confirmed by sequencing.

Real time qPCR
Diluted cDNA from the different tissues was used as template in a reaction containing a final concentration of 0.5μM for each primer and 1× FastStart SYBR Green master mix (total volume of 10μl). PCR amplifications were performed in triplicate using the Viia™ 7 Real-Time PCR System (Applied Biosystems) applying the following thermal profile: 95°C for 10 min, one cycle for cDNA denaturation; 95°C for 35s, 58°C for 30s and 72°C for 40s, 40 cycles for amplification; one cycle for melting curve analysis, to verify the presence of a single product. Each assay included a no-template control for each primer pair. We used tissues from four and three different animals to test reference genes stability and GOI expression level, respectively.

The corresponding real-time PCR efficiency (E) reported in Table 1 can be obtained in two ways. It can be computed either as sample-specific [42], or as factor specific [43] metric according to the eq. \( E = 10^{-1/slope} \).

The slope of linear regression model fitted over log-transformed data of five serially diluted input cDNA concentrations plotted against the corresponding CT values [43, 44]. The maximal efficiency of PCR is \( E = 2 \), in the case where every single template is replicated in each cycle; the minimal value is \( E = 1 \), corresponding to no replication.

Analysis of gene expression stability
The gene expression stability of the candidate genes was evaluated using geNorm [29], NormFinder [30] and BestKeeper [31]. For each algorithm, a stepwise exclusion method was applied in order to rank the selected genes according to their expression stability.

The geNorm-Based Analysis of candidate reference genes aims to select suitable references considering those displaying minimal variation across different biological conditions. Ideally, the normalized expression of GOI is obtained using the geometric mean of the optimal number of references selected in the analysis. This analysis renders a ranking of the tested genes based on their stability measure (M), determining the two most stable references or even a combination of multiple stable genes to be used for normalization. The M value results from the mean pairwise variation between every single gene compared to all other tested candidates. The algorithm discards the gene with the highest M and recalculates a new score for the remaining genes in a stepwise manner. Thus, references rank according to their M, from the most (lowest M values) to the least stable (highest M values) [29].

The NormFinder approach identifies the optimal reference among a set of candidates according to their expression stability in a given sample set and a given experimental design. This approach is based on a two-way ANOVA and provides a stability value for each reference analyzed, which is a direct measure of the estimated expression variation, which allows the user to evaluate the systematic error introduced when using the gene for normalization [30].

BestKeeper performs the descriptive statistic for each candidate, calculates the expression levels of the candidate references and estimates their expression stability based on the inspection of calculated variations (SD values). The candidates are then ordered from the most stably expressed, with the lowest variation, to the least stable one, with the highest mobility, according to the variability observed. Any evaluated gene with an SD higher than 1 (= starting template variation by the factor 2) is considered inconsistent [31].
Expression analysis of target genes
The relative expression level of each GOI was calculated using the most stable reference gene/gene identified across all tissues by NormFinder. In detail, relative expression levels were calculated as previously described [45], applying the ΔCt method, considering the E = 2 and using either a single reference or the geometric mean of the best pair of references for normalization. Whenever no amplification signal could be obtained for a GOI in a given tissue to allow calculation, we arbitrarily assigned to these samples a Ct value = 40.

Abbreviations
Ct: Cycle threshold; GOI: Gene Of Interest; M: Stability Measure; TPM: Transcript Per Million.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12864-022-08553-1.

Additional file 1.

Acknowledgements
Special thanks to Dr. Pasquale De Luca for critical reading of the manuscript and Alberto Macina for animal husbandry.

Authors’ contributions
FS performed all the experiment, carried out the data analysis and helped to draft the manuscript. AL participated in the design of the study and revised the draft manuscript. AP and MG provided gene sequences, carried out RNAseq data analysis and helped to draft and revise the manuscript. MS drafted and revised the manuscript, conceived and supervised the study. All authors read and approved the final manuscript.

Funding
No funding was received for conducting this study.

Availability of data and materials
All data analysed during this study are included in the published article: Gerdol M, Moreira R, Cruz F, Gomez-Garrido J, Vlasova A, Rosani U, Venier P, Alfonso LF, et al. Identification of suitable reference genes for qPCR analyses of mussel immune response to exogenous estrogen exposure. Environ Sci Pollut Res. 2018;25:2711–21. https://doi.org/10.1007/s11356-018-2008-x.

Declarations
Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1Stazione Zoologica Anton Dohrn, Department of Biology and Evolution of Marine Organisms, Villa Comunale 1, 80121 Naples, Italy. 2Department of Life Sciences, Università degli Studi di Trieste; Via Licio Giorgieri 5, 34127 Trieste, Italy.

References
1. FAO fisheries and aquaculture department, cultured aquatic species information Programme. Mytilus galloprovincialis. Cultured Aquatic Species Information Programme. FAO Fish. Aquac. Dep. 2019.http://www.fao.org/fishery/culturedspecies/Mytilus.
2. Serafini A, Lopes B, Company R, Cravo A, Gomes T, Sousa V, et al. A multi-biomarker approach in cross-transplanted mussels Mytilus galloprovincialis. Ecotoxicology. 2011;20:1959–74.
3. Tascetta F, Malagoli D, Accorsi A, Rigillo G, Blom JMC, Ottaviani E. Molluscs as models for translational medicine. Med Sci Monit Basic Res. 2015;21:96–9.
4. Gerdol M, Venier P. An updated molecular basis for mussel immunity. Fish Shellfish Immunol. 2015;46:17–38. https://doi.org/10.1016/j.fsi.2015.02.013.
5. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol. 2000;25:1969–93.
6. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009;55:611–22.
7. Huggett J, Dheda K, Bustin S, Zumbi A. Real-time RT-PCR normalisation; strategies and considerations. Genes Immun. 2005;6:279–84.
8. Kreuzer KA, Lass U, Landt O, Nitsche A, Laser J, Ellerbrok H, et al. Highly sensitive and specific fluorescence reverse transcription-PCR assay for the pseudogene-free detection of beta-actin transcripts as quantitative reference. Clin Chem. 1999;45:297–300.
9. Perrin BI, Ervasti JM. The actin gene family: function follows isoform. Cytoskeleton. 2010;67:630–4.
10. Paule MR. SURVEY AND SUMMARY transcription by RNA polymerases I and III. Nucleic Acids Res. 2000;28:1283–98.
11. Barbu V, Dautry F. Northern blot normalisation with a 28S rRNA oligonucleotide probe. Nucleic Acids Res. 1989;17:7115.
12. Prokopowich CD, Gregory TR, Crease TJ. The correlation between rDNA copy number and genome size in eukaryotes. Genome. 2003;46:29–80.
13. Selvey S, Thompson EW, Matthaei K, Lea RA, Irving MG, Griffiths LR. β-Actin – an unsuitable internal control for RT-PCR. Mol Cell Probes. 2001;15:307–11.
14. Godin D, Marnessier A, Staquet MJ, Schmitt D, Berthier-Vergnes O. Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate dehydrogenase and β-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations. Anal Biochem. 2001;295:17–21.
15. Bas A, Forsberg G, Hammarström S, Hammarström ML. Utility of the housekeeping genes 18S rRNA, β-actin and glyceraldehyde-3-phosphate-dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human T lymphocytes. Scand J Immunol. 2004;59:566–73.
16. Solanas M, Moral R, Escrich E. Unsuitability of using ribosomal RNA as loading control for northern blot analyses related to the imbalance between messenger and ribosomal RNA content in rat mammary tumors. Anal Biochem. 2001;288:99–102.
17. Raaijmakers MHGP, Van Emst L, De Witte T, Mensink E, Raymakers RAP. Quantitative assessment of gene expression in highly purified hematopoietic cells using real-time reverse transcriptase polymerase chain reaction. Exp Hematol. 2002;30:481–7.
18. Cubero-Leon E, Giocan CM, Miner C, Rochell JM. Reference gene selection for qPCR in mussel, Mytilus edulis, during gametogenesis and exogenous estrogen exposure. Environ Sci Pollut Res. 2012;19:2728–33.
19. Martínez-Escauriaza R, Lozano V, Pérez-Parallé ML, Pazos AJ, Sánchez JL. Validation of reference genes in mussel Mytilus galloprovincialis tissues under the presence of Okadaic acid. J Shellfish Res. 2018;37:93–101.
20. Americo JA, Guerreiro LTA, Gondim TS, da Cunha YR, Wajsenzon IR, Alfonso LF, et al. Identification of suitable reference genes for qPCR expression analysis on the gonads of the invasive mussel Limnoperna fortunei. bioRxiv. 2019. https://doi.org/10.1101/835447.
21. Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, et al. Massively gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. Genome Biol. 2020;21:275.

22. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. Trends Genet. 2013;29:569–74. https://doi.org/10.1016/j.tig.2013.05.010

23. Zeng J, Liu S, Zhao Y, Tan X, Aljohi HA, Liu W, et al. Identification and analysis of house-keeping and tissue-specific genes based on RNA-seq data sets across 15 mouse tissues. Gene. 2016;576:560–70. https://doi.org/10.1016/j.gene.2015.11.003.

24. Xu H, Li C, Zeng Q, Agraval I, Zhu X, Gong Z. Genome-wide identification of suitable zebrabfish Danio rerio reference genes for normalization of gene expression data by RT-qPCR. J Fish Biol. 2016;88:2095–110.

25. Pombo MA, Zheng Y, Fei Z, Martin GB, Rosli HG. Use of RNA-seq data to identify and validate RT-qPCR reference genes for studying the tomato-Pseudomonas pathosystem. Sci Rep. 2017;7:1–11.

26. Gao D, Kong F, Sun P, Bi G, Mao Y. Transcriptome-wide identification of optimal reference genes for expression analysis of Pyropia yezoensis responses to abiotic stress. BMC Genomics. 2018;19:1–15.

27. Li Y, Zhang L, Li R, Zhang M, Li Y, Wang H, et al. Systematic identification and validation of the reference genes from 60 RNA-Seq libraries in the scallop Mizuhopecten yessoensis. BMC Genomics. 2019;20:1–13.

28. Marie B, Le Roy N, Zanella-Cléon I, Becchi M, Marin F. Molecular evolution of mollusc shell proteins: insights from proteomic analysis of the edible mussel mytilus. J Mol Evol. 2011;72:531–46.

29. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3:1–2.

30. Andersen CL, Jensen JL, Ørntoft TF. 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. 2004;64:5245–50.

31. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - excel-based tool using pair-wise correlations. Biotechnol Lett. 2004;26:509–15.

32. Malinovskaya EM, Estrhova ES, Golimbet VE, Porokhovnik LN, Lyapunova NA, Kutsev SI, et al. Copy number of human ribosomal genes with aging: Unchanged mean, but narrowed range and decreased variance in elderly group. Front Genet. 2018;9:1–11.

33. Gibbons JG, Branco AT, Yu S, Lemos B. Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans. Nat Commun. 2014;5:1–2.

34. Calcino AD, Kenny NJ, GM. Single individual structural variant detection uncovers widespread hemizygosity in mussels. Phil Trans R Soc. 2021;B 376:20200153.

35. Ashvin Ganesh M, Ganti G, Choudhury A, Raja SY. Design and optimization of dish for solar thermal incineration. Appl Mech Mater. 2016;852:675–80.

36. Gerdol M, De Moro G, Manfrin C, Milandri A, Riccardi E, Beran A, et al. RNA sequencing and de novo assembly of the digestive gland transcriptome in Mytilus galloprovincialis fed with toxicogenic and non-toxic strains of Alexandrium minutum. BMC Res Notes. 2014;7:1–16.

37. Rey-Campos M, Moreira R, Gerdol M, Pallavicini A, Novoa B, Figueras A. Immune tolerance in Mytilus galloprovincialis hemocytes after repeated contact with vibrio splendidus. Front Immunol. 2019;10:1–15.

38. Moreira R, Pereira F, Canchaya C, Posada D, Figueras A, Novoa B. RNA-Seq in Mytilus galloprovincialis: comparative transcriptomics and expression profiles among different tissues. BMC Genomics. 2015;16:1–18. https://doi.org/10.1186/s12864-015-1817-5.

39. Romiguer J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature. 2014;515:261–3.

40. Bjärrnmark NA, Yarra T, Churcher AM, Felix RC, Clark MS, Power DM. Transcriptomics provides insight into Mytilus galloprovincialis (Mollusca: Bivalvia) mantle function and its role in biomineralisation. Mar Genomics. 2016;27:37–45.

41. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 2012;131:281–5.

42. Tichopad A, Dilger M, Schwarz G, Pfaffl MW. Standardized determination of real-time PCR efficiency from a single reaction set-up. Nucleic Acids Res. 2003;31:2–7.

43. Rasmussen R. Quantification on the LightCycler. In: Meuer S, Wittwer C, Nakagawara K-I, editors. Rapid Cycle Real-Time PCR: Methods and Applications. Berlin: Springer Berlin Heidelberg; 2001. p. 21–34. https://doi.org/10.1007/978-3-642-59254-0_4.

44. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001;29:2002–7.

45. Sirakov M, Zarrella I, Borra M, Rizzo F, Biffali E, Arnone M, et al. Selection and validation of a set of reliable reference genes for quantitative RT-PCR studies in the brain of the cephalopod Mollusc Octopus vulgaris. BMC Mol Biol. 2009;10:1–8.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.