

RESEARCH ARTICLE

# Identification and validation of reference genes for qPCR in the terrestrial gastropod *Cepaea nemoralis*

Susanne Affenzeller, Nicolas Cerveau, Daniel John Jackson\*

Department of Geobiology, Georg-August University of Göttingen, Göttingen, Germany

\* [djackso@uni-goettingen.de](mailto:djackso@uni-goettingen.de)



## Abstract

Identifying and understanding mechanisms that generate phenotypic diversity is a fundamental goal of evolutionary biology. With a diversity of pigmented shell morphotypes governed by Mendelian patterns of inheritance, the common grove snail *Cepaea nemoralis* (Linnaeus, 1758) has been a model for evolutionary biologists and population geneticists for decades. However, the genetic mechanisms by which *C. nemoralis* generates this pigmented shell diversity remain unknown. An important first step in investigating this pigmentation pattern is to establish a set of validated reference genes for differential gene expression assays. Here we have evaluated eleven candidate genes for reverse transcription quantitative polymerase chain reaction (qPCR) in *C. nemoralis*. Five of these were housekeeping genes traditionally employed as qPCR reference genes in other species, while six alternative genes were selected *de novo* from *C. nemoralis* transcriptome data based on the stability of their expression levels. We tested all eleven candidates for expression stability in four sub-adult tissues of *C. nemoralis*: pigmented mantle, unpigmented mantle, head and foot. We find that two commonly employed housekeeping genes (*alpha tubulin*, *glyceraldehyde 3-phosphate dehydrogenase*) are unsuitable for use as qPCR reference genes in *C. nemoralis*. The traditional housekeeping gene *UBIquitin* on the other hand performed very well. Additionally, an *RNA-directed DNA polymerase (RNAP)*, a *Potassium Channel Protein (KCHP)* and a *Prenylated Rab acceptor protein 1 (PRAP)*, identified *de novo* from transcriptomic data, were the most stably expressed genes in different tissue combinations. We also tested expression stability over two seasons and found that, although other genes are more stable within a single season, *beta actin (BACT)* and *elongation factor 1 alpha (EF1a)* were the most reliable reference genes across seasons.

## OPEN ACCESS

**Citation:** Affenzeller S, Cerveau N, Jackson DJ (2018) Identification and validation of reference genes for qPCR in the terrestrial gastropod *Cepaea nemoralis*. PLoS ONE 13(8): e0201396. <https://doi.org/10.1371/journal.pone.0201396>

**Editor:** Christian Schönbach, Nazarbayev University, KAZAKHSTAN

**Received:** March 8, 2018

**Accepted:** July 13, 2018

**Published:** August 29, 2018

**Copyright:** © 2018 Affenzeller et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

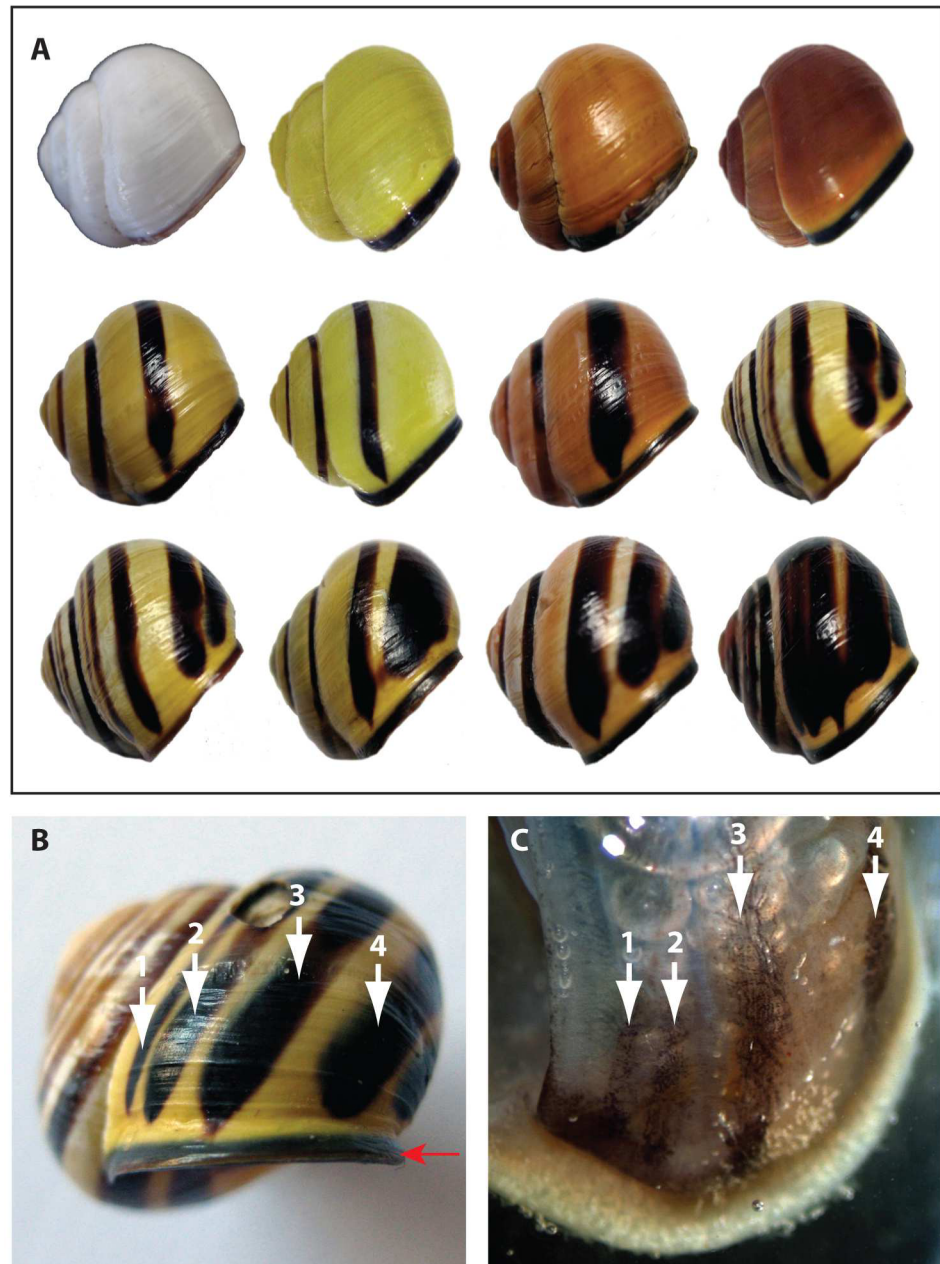
**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files. In addition the 11 sequences we report have been submitted to GenBank and have the following access numbers: MH035488 - MH035498.

**Funding:** This work was supported by Deutsche Forschungsgemeinschaft grants (JA 2108/2-1 and JA 2108/6-1) to Daniel J. Jackson. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Introduction

Variation in the pigmentation and patterning of the shell of the common grove snail *Cepaea nemoralis* (Linnaeus, 1758), has been studied by ecologists and evolutionary biologists for decades [1, 2, 3]. Shell background colour can range from a very light yellow to dark pink, orange and brown with up to five dark brown bands. The shell itself and any patterns of pigmentation are produced by the underlying mantle tissue (Fig 1).

**Competing interests:** The authors have declared that no competing interests exist.



**Fig 1. Pigmentation and banding patterns display marked variation in *C. nemoralis*.** A. Representative shell colour morphs of *C. nemoralis*. Shell background colour can range from white to yellow and pink to dark orange or brown. Banding patterns include zero to five bands that can merge or be distinct. B. A shell with a yellow background and a typical banding pattern. The lateral pigmented band at the leading edge of the shell that indicates a fully differentiated snail is indicated with a horizontal red arrow. C. Banding patterns in the shell are reflected in the pigmentation of the underlying mantle tissue, and facilitate the isolation of populations of cells for separate RNA extractions and gene expression profiling. The mantle tissue presented here was responsible for both constructing and pigmenting the shell shown in B. The individual pigmented bands in the mantle are numbered and correspond to those indicated on the shell.

<https://doi.org/10.1371/journal.pone.0201396.g001>

Early studies of *C. nemoralis* focused on the distribution of colour morphs and attempted to identify the evolutionary pressures that may have generated this variation. Hypotheses such as crypsis and body temperature control have been investigated [4, 5, 6, 7, 8], however no single

mechanism can yet convincingly account for the observed distributions of variably pigmented populations of *C. nemoralis*. Nonetheless the variable banding patterns and background colours make it an excellent system to study the underlying molecular mechanisms of complex pigmentation patterns [9, 10, 11]. The shell polymorphism present in *C. nemoralis* is thought to be controlled by a supergene, a region of genomic DNA containing tightly linked and neighbouring loci controlling different aspects of the observed variation, such as shell background colour and the number of dark pigmented bands [12, 13]. A variety of next generation sequencing (NGS) and proteomic approaches have been applied to identify this supergene in recent years [13, 14, 15]. However the precise genomic loci and mechanism responsible for these variable pigmented phenotypes remain unknown.

Partly hampering these efforts is a lack of established methods in *C. nemoralis* to independently verify candidate pigmentation genes. Reverse transcription quantitative polymerase chain reaction (qPCR) is a commonly used molecular technique to measure gene expression levels between biological samples. When employing qPCR to measure gene expression levels between samples it is critical to normalise values using reference genes that must be stably expressed across the conditions being investigated [16]. Concerningly, previous studies have shown that commonly employed reference genes are often far from stably expressed [17, 18, 19]. It is therefore important to validate any panel of reference genes intended for qPCR prior to conducting analyses of gene expression.

To date, only a handful of molluscan species have had reference genes tested for stable expression [20, 21, 22, 23, 24], and in *C. nemoralis* this has never been performed. Our aim was therefore to validate a set of reference genes for qPCR from a selection of traditionally used reference genes and alternative genes identified *de novo* from transcriptome data. These genes could then be used for future shell pigmentation and patterning studies in this organism.

## Material and methods

### Transcriptome sequencing

Three sub-adult individuals of *C. nemoralis* (i.e. individuals actively laying down shell material without the terminal pigmented lateral band present in all mature individuals shown in Fig 1) were collected at the University of Göttingen, Germany (51°33'24.0"N 9°57'27.3"E). These three individuals comprise three biological replicates. Two RNA extractions were performed for each individual from the most distal edge of the mantle; one from a dark pigmented region and one from an adjacent unpigmented region (see Fig 1C). Total RNA was extracted using Trizol (Molecular Research Center, #TR118) according to the manufacturer's instructions. Library preparation and sequencing was carried out at the Transcriptome and Genome Analysis Laboratory, University Medical Center Göttingen, Germany. A TrueSeq RNA Sample Kit (Illumina, Cat. N°RS-122-2002) was used to generate an mRNA library using 500 ng of total RNA. During TrueSeq Illumina library construction rRNA were depleted with oligo dT beads. Quantitation of cDNA libraries was performed using the QuantiFluor™ dsDNA System (Promega) and the size range of the final cDNA libraries was determined using a Bioanalyzer 2100 from Agilent (280 bp). Stranded sequencing was conducted on the Illumina HiSeq 2000 platform to generate 100 bp paired-end reads. More than 470 million paired reads were generated. This raw data is the basis of our ongoing efforts to identify genes involved in shell pigmentation in *C. nemoralis* and will be reported in full elsewhere, however all sequence data relevant to the analyses performed here have been made publicly available (see below). Trimomatic [25] was used to remove low quality reads and adapter sequences. Reads from the six samples were pooled and treated together for the assembly process. Pooled reads were assembled *de novo* using our assembly pipeline [26]. Briefly, we employed three assembly packages

with unique assembly strategies: Trinity V2.0.3 [27], CLC Genomics Workbench ([www.qiagenbioinformatics.com](http://www.qiagenbioinformatics.com)) and IDBA-tran V1.1.1 [28]. The CLC assembler was run using a word size of 20 base pairs (bp), a bubble size of 50 bp, with reads mapped back to the transcriptome using default parameters. IDBA\_tran was run with kmer values ranging from 20 to 100 bp with a step size of 10 bp. Trinity assemblies were run with default parameters and a k-mer value of 25. Contigs with coding regions larger than 100 amino acid residues were merged into a concatenated assembly, and redundancy was removed as described [26].

Differential gene expression analysis was performed using CLC Genomics Workbench ([www.qiagenbioinformatics.com](http://www.qiagenbioinformatics.com)). A matched pairs experimental design (1 matched pair from each individual replicate) was established to compare pigmented mantle and unpigmented mantle samples. Trimmed paired-end reads from each condition were mapped to the reference assembly using total gene read counts.

### Selection of reference gene candidates

Candidate reference genes were selected from our transcriptome data in two fundamentally different ways. First, a set of five commonly employed housekeeping genes were selected from the literature and subsequently identified in our transcriptome data. These included *Alpha Tubulin* (*ATUB*; Accession MH035488), *Beta Actin* (*BACT*; Accession MH035489), *Elongation Factor 1 alpha* (*EF1α*; Accession MH035491), *Glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH*; Accession MH035493), *UBIquitin* (*UBI*; Accession MH035498). Second, six candidates were chosen based on the low standard deviation in their total read counts across all six transcriptomes. These six alternative candidates were annotated using SmartBLAST against SwissProt, and tBLASTx against ref-seq. These candidates were identified as *DNA Repair Protein* (*DNARP*; Accession MH035490), *Fibronectin type III domain containing protein* (*FIB3*; Accession MH035492), *GTP-binding protein* (*GTPB*; Accession MH035494), *Potassium Channel Protein* (*KCHP*; Accession MH035495), *Prenylated Rab acceptor protein 1* (*PRAP*; Accession MH035496) and *RNA-directed DNA polymerase* (*RNAP*; Accession MH035497). Primers to all of these genes were designed using Primer3 [29, 30]. Table 1 lists all primer sequences and their resulting amplicon sizes. Primer specificity was verified with melt curve analysis and agarose gel electrophoresis.

### RNA extraction and cDNA synthesis

Total RNA was extracted from six sub-adult individuals of *C. nemoralis* (independent of those used for transcriptome sequencing) collected at the University of Göttingen (51° 33' 24.0" N 9°

**Table 1. Primer sequences and amplicon sizes for all used reference genes.**

Gene	Forward primer (5' to 3')	Reverse primer (5' to 3')	Amplicon size (bp)
<i>ATUB</i>	ATGGTATTCAGCCTGACGGT	AACATGTTTGCCAGCTCCTG	104
<i>BACT</i>	CAGAAGCAATGTTCCAGCCA	TGAGCCACCAGACAAGACAA	137
<i>DNARP</i>	GACCACAAGAGGACAGCTCT	GTTATCCTGCAGCACACCAC	105
<i>EF1α</i>	GTACCGGAGAGTTTGAGGCT	GAGTAAGGTGGAGTGGTGCT	133
<i>FIB3</i>	TGGATGACATCGGGGAACAA	ATCACTATCGCAACAACGGC	170
<i>GAPDH</i>	CATGTTTGCTGGGTGTCA	TCAGGCCCTCAACAATTCCA	137
<i>GTPB</i>	ATTCAGCACCCGGAGAGAT	TAAGTGCCAGAACAGACGGT	107
<i>KCHP</i>	GTTGTTGCCTCAATCGTGGT	AAGCAGTCTCCAAGTCCCTC	138
<i>PRAP</i>	ATCTTGACATCTCCGCTGCT	GATTTACGACCCATCACGG	111
<i>RNAP</i>	CGCGTGACAATGGACAAAGA	CGCCTCTCTATCCGTATCCC	102
<i>UBI</i>	AGAATGCCCAACAATGCT	AGAATCAGCCTCTTCTCCGG	121

<https://doi.org/10.1371/journal.pone.0201396.t001>

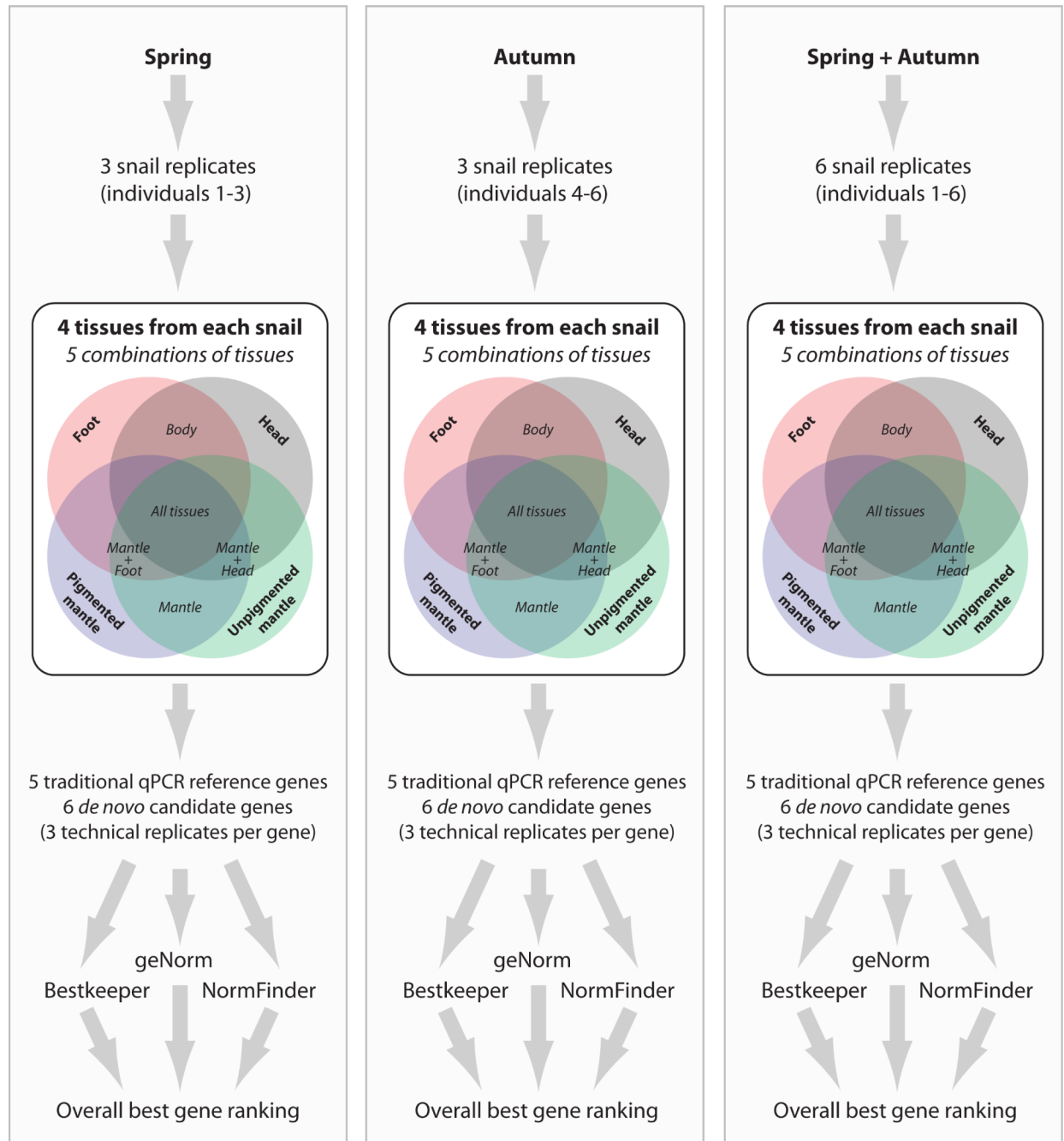
57°27.3"E). Three individuals were collected in spring (May, 2017), while three were collected in autumn (September, 2017). These seasons encompass periods of relatively rapid and depressed shell growth respectively [31]. All individuals possessed a yellow background shell colour with two or more bands (Fig 1). Tissue from pigmented mantle (PM) and unpigmented mantle (UPM), foot (F) and head (H) were carefully dissected and used for a total of 24 RNA extractions using Qiazol (Qiagen, #79306) according to the manufacturer's instructions (Fig 2). Following successful RNA extraction a DNase treatment (RQ1 RNase-free DNase, Promega, #M6101) was carried out according to the manufacturer's instructions. Briefly, 1  $\mu$ L DNase and 1  $\mu$ L buffer solution were used on 8  $\mu$ L of extracted RNA. The reaction was stopped with 1  $\mu$ L of Promega STOP solution. The quality and integrity of RNA was verified with a Nanodrop and agarose gel electrophoresis, and a standard PCR against the *COI* barcoding fragment was carried out to verify that the DNase treatment was effective. For each sample, 1  $\mu$ g of total extracted RNA was used in a cDNA synthesis reaction containing 10  $\mu$ L oligo dT (10 $\mu$ M), 2  $\mu$ L M-MLV reverse transcriptase (RT) H-(Promega #M3682), 10  $\mu$ L 5x M-MLV RT buffer (Promega), 2  $\mu$ L dNTPs (10mM) and x  $\mu$ L nuclease free water to a total volume of 50  $\mu$ L. RNA, oligo dTs and nuclease free water to a volume of 15  $\mu$ L were heated to 70°C for 5 min, then cooled to 42°C. The remaining reaction components were added at 42°C and the reaction was incubated for a further 75 min. The RT was inactivated at 70°C for 15 min and the cDNA was stored at -20°C until further use.

## qPCR

All qPCR runs followed a maximum sample layout for each gene (all tissue samples included in one run) and comply with the MIQE guidelines [32]. Each run also included no template controls (NTC) for each primer pair and three inter run calibrators (IRC). *EF1 $\alpha$* , *RNAP* and *UBI* were randomly chosen as IRCs and were run on mixed tissue cDNA. Samples were run in triplicate, NTC and IRCs were run in duplicate. qPCR reactions contained 5  $\mu$ L 2x Rotor-Gene SYBR Green PCR Master Mix (Qiagen, #204074), 0.4  $\mu$ L cDNA, 0.1  $\mu$ L of each primer (1 $\mu$ M final concentration) and 4.4  $\mu$ L ddH<sub>2</sub>O to a final volume of 10  $\mu$ L. Amplification and detection were performed on a Rotor-Gene Q (Qiagen) running Rotor-Gene Q software (version 2.0.2) with the following temperature profile: 5 min initial activation and denaturation at 95°C; 40 cycles of 5 sec denaturation at 95°C, 10 sec annealing and extension at 60°C (data collection at this step); a final melt curve analysis from 60°C to 95°C at a rate of 5 sec/1°C. PCR amplicons were also assessed on agarose gels to qualitatively visualise amplification efficiency (S1 Fig).

## Data analyses

Raw fluorescence data was exported from the Rotor-Gene software for further analysis. Baseline correction and amplification efficiency correction were performed using LinRegPCR [33]. Inter run correction was performed using Factor-qPCR [34]. The resulting corrected cycle threshold (Cq) values were used to calculate the geometric means of technical replicates. The expression stability of each candidate reference gene was calculated for each of the six individuals independently, before being averaged across individuals. We also analysed each candidate reference gene in all four tissues independently (Pigmented Mantle (PM), Unpigmented Mantle (UPM), Foot (F) and Head (H), and also after pooling all four tissues (All). Finally we tested different combinations of tissues: Body (F + H), Mantle (PM + UPM), Mantle and Foot (PM + UPM + F) and Mantle and Head (PM + UPM + H). These sample combinations were also run for spring animals, autumn animals and a season independent 'total' (spring + autumn) set (Fig 2). This allowed us to identify the most appropriate reference genes for our research interests (shell pigmentation), and to also assess the broader stability of these genes across



**Fig 2. A schematic representation of our sampling and experimental design and data analysis.** RNA was extracted from four different tissues derived from six snail replicates across two seasons. The four independent tissues (and five combinations thereof) and three algorithms were used to rank gene performance, and these results were then combined to find the overall best ranked genes.

<https://doi.org/10.1371/journal.pone.0201396.g002>

diverse tissue types and seasons. We tested gene expression stability using three different algorithms: BestKeeper (which calculates efficiency corrected descriptive statistics) [35], NormFinder (which is based on inter- and intra-group variation calculations for reference gene stability testing) [36] and geNorm (which estimates stability of reference genes based on pairwise comparison of gene specific variation) [37]. Data was transformed with the following formula for

use in NormFinder and geNorm:

$$\text{Transformed data} = \text{sample efficiency}^{(\min Cq(\text{Gene}) - Cq(\text{Sample}))}$$

Where sample efficiency is the amplification efficiency of the reaction,  $\min Cq(\text{Gene})$  is the lowest observed cycle threshold per gene, and  $Cq(\text{Sample})$  is the observed cycle threshold for the sample [35]. To summarise the results of all three algorithms, each candidate reference gene was ranked from the most (1) to the least (11) stably expressed and then averaged across algorithm (see the Supplementary Material for the results of each individual algorithm). Data preparation for BestKeeper, NormFinder and geNorm, as well as ranking, and descriptive statistics and visualisation of results were carried out in Excel 2016 MSO (v16.0.8326.2096, Microsoft).

## Results

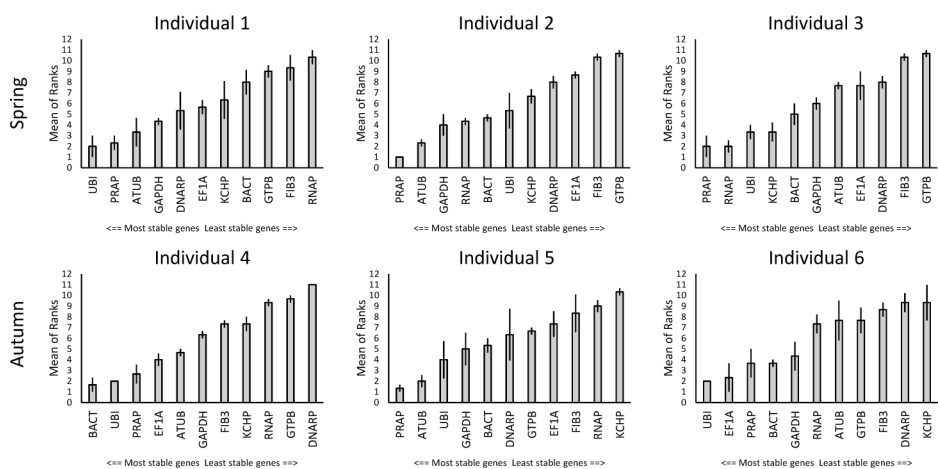
### Variation between individual snails influences the ranking of candidate reference genes

The ranking stability of candidate reference genes on the pooled tissue data from each individual snail revealed some differences. However the overall mean of ranks across all three algorithms revealed *PRAP* to be among the top three most stably expressed genes in all individuals. Only one other gene, *UBI* performed just as well, except in individual 2. Additionally *ATUB*, *GAPDH*, *RNAP*, *BACT* and *EF1 $\alpha$*  showed good overall ranks in single individuals (Fig 3).

### Tissue type and season influence the ranking of candidate reference genes

Within Foot tissue *KCHP*, *UBI* and *PRAP* were identified as the best candidate reference genes for the spring subset (individuals 1–3). For animals sampled in autumn (individuals 4–6) *BACT*, *GTPB* and *ATUB* performed the best in Foot tissue. In combining all Foot tissue seasons (spring + autumn) *BACT*, *EF1 $\alpha$*  and *UBI* were identified as the best reference genes.

In Head tissue *BACT*, *RNAP* and *DNARP* showed the best performance in spring samples, while *BACT*, *PRAP* and *GAPDH* performed best in autumn samples. In the spring + autumn Head tissue set *BACT* and *EF1 $\alpha$*  were revealed as the best genes.



**Fig 3. Expression stability ranking across all tissues within each individual snail.** Means of ranked data generated by three algorithms (BestKeeper, geNorm and NormFinder). Error bars indicate standard error.

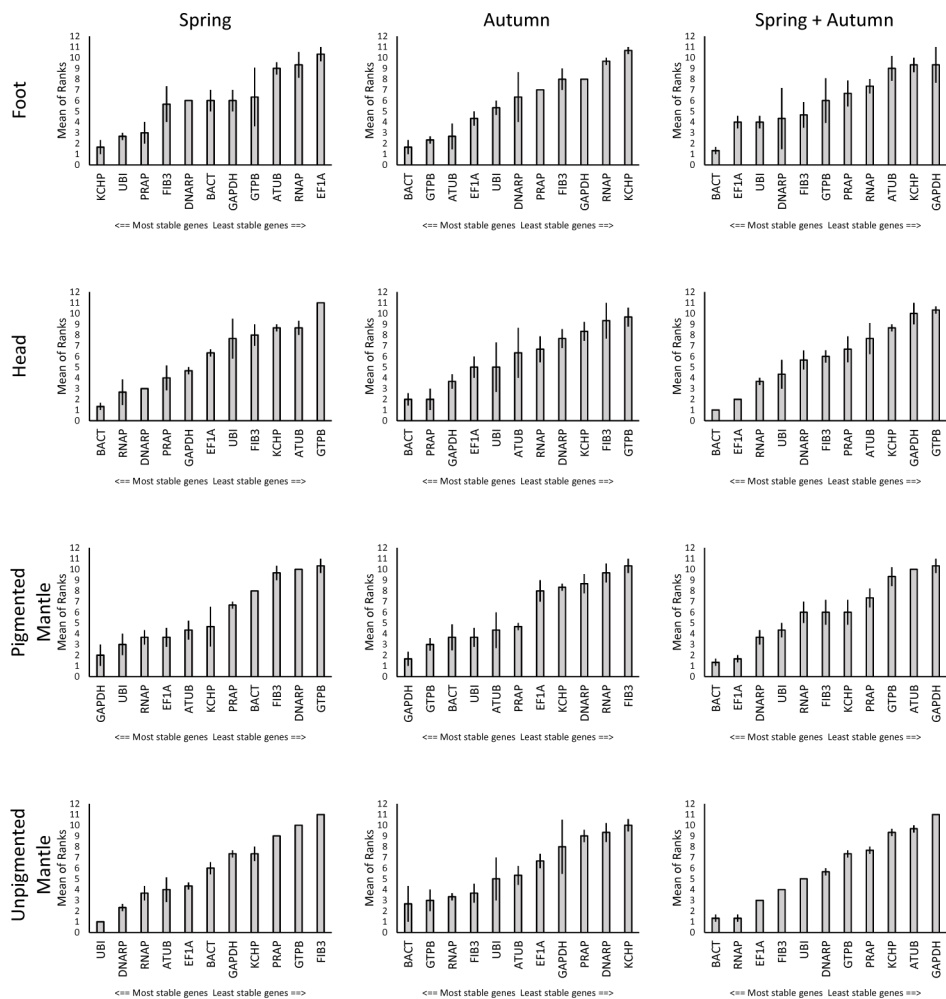
<https://doi.org/10.1371/journal.pone.0201396.g003>

*GAPDH* and *UBI* performed best in Pigmented Mantle tissue in spring, whereas *GAPDH* and *GTPB* outperformed all other genes in autumn. In the spring + autumn Pigmented Mantle sample set *BACT* and *EF1 $\alpha$*  performed best.

In Unpigmented Mantle tissue extracted in spring *UBI* and *DNARP* showed the best results, while in the autumn subset *BACT* and *GTPB* performed best. Combining both seasonal subsets of Unpigmented Mantle identified *BACT* and *RNAP* as the best reference genes in all three algorithms (Fig 4).

### Tissue type combinations influence the ranking of candidate reference genes

When we combined more than one tissue type and repeated our analyses *UBI* emerged as the most stably expressed gene for all three algorithms in spring Mantle (PM + UPM), spring and autumn Mantle and Foot (PM + UPM + F) and spring All (PM + UPM + F + H). Considering the spring subset of individuals *RNAP* was revealed as a second candidate gene in Mantle, Mantle and Foot, as well as in All. *KCHP* performed second best with *PRAP* performing equally well in All. For Body tissue (F + H) *PRAP* and *BACT* performed best overall. *RNAP*,



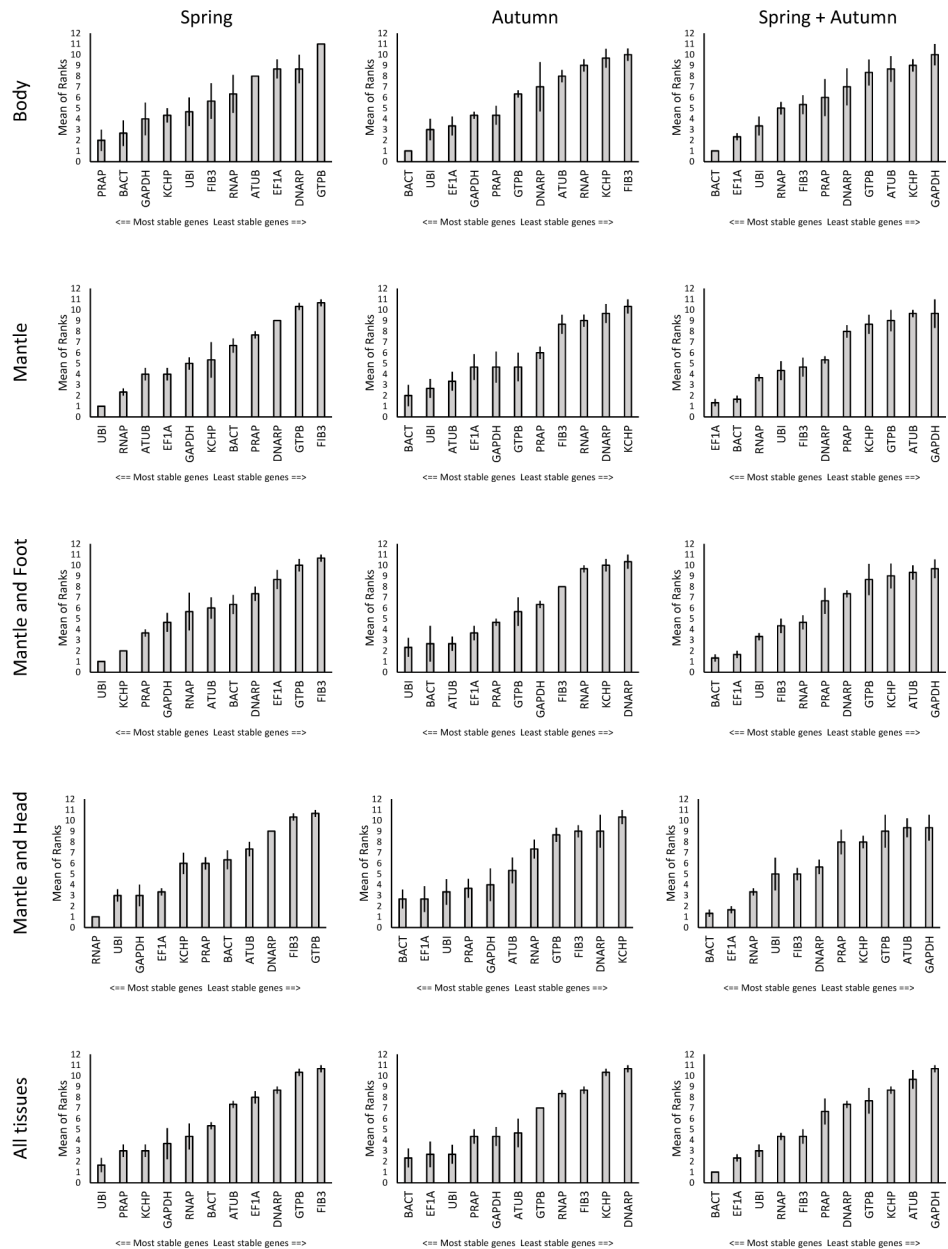
**Fig 4. Expression stability ranking within each tissue type.** Means of ranked data generated by three algorithms (BestKeeper, geNorm and NormFinder). Error bars indicate standard error.

<https://doi.org/10.1371/journal.pone.0201396.g004>



*UBI* and *GAPDH* tested as best candidate genes in the data set Mantle and Head (PM + UPM + H). In autumn animals *BACT* was revealed as the best overall candidate in all remaining multiple tissue combinations (all except M + F), followed by *EF1α* in All and in M + H, and *UBI* in Body and Mantle tissue sets. In the combined analyses of all extractions *BACT* and *EF1α* performed best in all multiple tissue sets across the three algorithms (Fig 5).

Results for all three algorithms over all data sets, as well as the mean of ranks over all three algorithms results are summarized in S1 Table. The results generated by each of the three



**Fig 5. Expression stability ranking of combined tissue types.** Body (Foot and Head); Mantle (Mantle pigmented and unpigmented); M+F (Mantle and Foot); M+H (Mantle and Head); All (all tissue sets). Means of ranked data generated by three algorithms (BestKeeper, geNorm and NormFinder). Error bars indicate standard error.

<https://doi.org/10.1371/journal.pone.0201396.g005>

individual algorithms on all datasets are presented and discussed in the Supplementary Material (S2–S7 Figs).

## Discussion

*C. nemoralis* holds great promise as a model for investigating the molecular mechanisms that regulate gastropod shell pigmentation. As a first step towards this goal we have assayed a total of eleven candidate reference genes for their suitability in qPCR experiments. Five of these were genes commonly employed in other model systems as reference genes, and six were alternative genes we identified *de novo* as having potentially stable expression profiles across pigmented and unpigmented mantle tissue. In our analyses one of the major challenges to obtaining robust measures of gene expression was highlighted; expression levels of specific genes can vary significantly between individuals, across tissue types and across seasons. One prominent example is our single tissue analyses of *GAPDH*. Although this gene is stably expressed within spring and autumn in Pigmented Mantle, it is ranked as the least stable gene across seasons (Fig 4). This is because *GAPDH* seems to have a very uniform expression within a given season, but expression levels change drastically between seasons. This highlights the importance of biological replicates, experimental design and reference gene choice to any qPCR experiment when samples across tissue types and time points are to be compared.

Despite this variability some patterns emerged, perhaps most importantly that *PRAP* consistently tested within the top three candidates for all six individuals across seasons, tissue type and algorithm (Fig 3). The often employed house keeping genes *BACT*, *UBI* and *EF1 $\alpha$*  tested well in different sub sets of tissues and tissue combinations. The traditionally employed house keeping genes *ATUB* and *GAPDH* however did not perform consistently. Although this is the first time the genes *PRAP*, *RNAP* and *KCHP* have been considered as potential reference genes for qPCR, they performed very well in specific data sets and across all algorithms, and we propose they should be considered in all qPCR studies in *C. nemoralis*. We suggest they may also be applicable for qPCR in other gastropod species, and perhaps even more widely, and should therefore be included in reference gene testing.

Because the literature concerning the identification of stable reference genes in molluscs is extremely small we can compare our results to these previous efforts. In a study testing eight reference genes in the pond snail (*Bellamya aeruginosa*) the highly expressed *ribosomal protein L7* gene was identified as the most stably expressed [24]. However several studies in diverse animals have demonstrated that ribosomal genes are not reliable candidates for stable expression [20, 21, 38], therefore we did not consider them for investigation here. *BACT* performed well in our study of *C. nemoralis*, both in single and multiple tissue sets, as well as in both seasons separately. It is also one of the genes that displayed stable expression across seasons. However it did not behave ideally in all data sets, a result that agrees with differences in *BACT* expression between tissues and treatments in other molluscs [20, 21, 22, 23, 24]. Our results also identified the traditional housekeeping gene *UBI* as a good candidate reference gene in multiple data sets in both spring, autumn and spring + autumn (e.g. spring All tissues and Mantle and Foot, autumn Body and Mantle, spring + autumn All tissues, Body and Mantle and Foot). This agrees with results reported in larval and adult cephalopods [20, 23]. In *Mytilus edulis* and *Haliotis discus* *EF1 $\alpha$*  could be identified as a stable reference gene [21, 22]. Our data agrees with this when considering comparisons across seasons and in some autumn tissue sets. However our results for tissue samples taken in spring did not support *EF1 $\alpha$*  as a good candidate, agreeing with the reported unstable expression of this gene across different adult tissues in *O. vulgaris* [20]. *GAPDH* performed well in Pigmented Mantle, but showed variable expression in other tissue combinations and especially across seasons, an instability already observed

in *Haliothis* [21]. Another frequently employed housekeeping gene *ATUB* displayed very tissue dependent results in our analyses and did not in general rank well under many conditions. We therefore cannot recommend *GAPDH* or *ATUB* as reference genes for qPCR in *C. nemoralis*.

To summarise, we recommend different sets of reference genes for qPCR studies in *C. nemoralis* depending on the research question. By mining digital gene expression profiles we identified *PRAP*, *RNAP* and *KCHP de novo* as candidate reference genes for qPCR that performed well under a range of conditions. For our particular research interests in shell pigmentation *EF1 $\alpha$*  and *BACT* across seasons, *BACT* and *UBI* in autumn and *UBI* and *RNAP* in spring appear to be the most appropriate reference genes. Three of these genes (*UBI*, *EF1 $\alpha$*  and *BACT*) can also be used for studies including more diverse tissue samples of *C. nemoralis*. Our results clearly demonstrate that when choosing reference genes for qPCR, the selection must not only be adapted to the target species, but also to the intended experimental design and should therefore consider variables such as seasonal variation, tissue specificity, environmental factors and the method of analysis.

## Supporting information

**S1 Table. Expression stability indices of reference genes tested in all data sets.** Results for Individuals 1 to 6 (Ind1–6), Foot (F), Head (H), Pigmented Mantle (PM), Unpigmented Mantle (UPM), Body (F + H), Mantle (PM + UPM), Mantle and Foot (PM + UPM + F), Mantle and Head (PM + UPM + H) and All tissues (PM + UPM + F + H) spring, autumn and all year data sets for Mean of Ranks, BestKeeper, geNorm and NormFinder. Mean of Ranks values represent the ranked means of all three algorithms, where 1 is the most stable gene and 11 is the least stable gene. geNorm values represent M values as calculated by the algorithm. NormFinder values are Stability values as calculated by the algorithm. BestKeeper values represent Standard Deviation of C<sub>q</sub> as calculated by the algorithm. (XLSX)

**S1 Fig. Qualification of qPCR amplification specificity by gel electrophoresis.** Representative images of agarose gel electrophoresis after qPCR amplification for each of the tested reference genes. Size ladders on the left indicate fragments from 100 to 1000 base pairs. (TIF)

**S2 Fig. qPCR results of reference gene tests for each individual.** Detailed qPCR results for all six individuals (Individual 1–6). Standard deviation (SD) of cycle thresholds (C<sub>q</sub>) as calculated by BestKeeper. geNorm results in M values (suggested upper threshold indicated by dashed line) as calculated by the algorithm. NormFinder results in Stability value as calculated by the algorithm. (EPS)

**S3 Fig. qPCR results of reference gene tests for foot and head tissue.** Detailed qPCR results for Foot tissue set and Head tissue set. Standard deviation (SD) of cycle thresholds (C<sub>q</sub>) as calculated by BestKeeper. geNorm results in M values (suggested upper threshold indicated by dashed line) as calculated by the algorithm. NormFinder results in Stability value as calculated by the algorithm. First column represents spring animals (Individuals 1–3), second column represents autumn animals (Individuals 4–6), third column represents all individuals (Individuals 1–6). (EPS)

**S4 Fig. qPCR results of reference gene tests for pigmented and unpigmented mantle tissue.** Detailed qPCR results for Pigmented Mantle tissue set and Unpigmented Mantle tissue set.

Standard deviation (SD) of cycle thresholds (Cq) as calculated by BestKeeper. geNorm results in M values (suggested upper threshold indicated by dashed line) as calculated by the algorithm. NormFinder results in Stability value as calculated by the algorithm. First column represents spring animals (Individuals 1–3), second column represents autumn animals (Individuals 4–6), third column represents all individuals (Individuals 1–6). (EPS)

**S5 Fig. qPCR results of reference gene tests for body (F + H) and mantle (PM + UPM) tissue.** Detailed qPCR results for Body tissue set and Mantle tissue set. Standard deviation (SD) of cycle thresholds (Cq) as calculated by BestKeeper. geNorm results in M values (suggested upper threshold indicated by dashed line) as calculated by the algorithm. NormFinder results in Stability value as calculated by the algorithm. First column represents spring animals (Individuals 1–3), second column represents autumn animals (Individuals 4–6), third column represents all individuals (Individuals 1–6). (EPS)

**S6 Fig. qPCR results of reference gene tests for mantle and foot (PM + UPM + F) and mantle and head (PM + UPM + H) tissue.** Detailed qPCR results for Mantle and Foot tissue set and Mantle and Head tissue set. Standard deviation (SD) of cycle thresholds (Cq) as calculated by BestKeeper. geNorm results in M values (suggested upper threshold indicated by dashed line) as calculated by the algorithm. NormFinder results in Stability value as calculated by the algorithm. First column represents spring animals (Individuals 1–3), second column represents autumn animals (Individuals 4–6), third column represents all individuals (Individuals 1–6). (EPS)

**S7 Fig. qPCR results of reference gene tests for all tissues (PM + UPM + F + H).** Detailed qPCR results for all tissue types combined. Standard deviation (SD) of cycle thresholds (Cq) as calculated by BestKeeper. geNorm results in M values (suggested upper threshold indicated by dashed line) as calculated by the algorithm. NormFinder results in Stability value as calculated by the algorithm. First column represents spring animals (Individuals 1–3), second column represents autumn animals (Individuals 4–6), third column represents all individuals (Individuals 1–6). (EPS)

## Acknowledgments

Gabriela Salinas-Riester and her team at the TAL laboratory (University Medical Centre Göttingen) are acknowledged for NGS sequencing. This work was supported by DFG grants (JA 2108/2-1 and JA 2108/6-1) to DJJ.

## Author Contributions

**Conceptualization:** Daniel John Jackson.

**Formal analysis:** Susanne Affenzeller, Nicolas Cerveau.

**Funding acquisition:** Daniel John Jackson.

**Methodology:** Susanne Affenzeller.

**Project administration:** Daniel John Jackson.

**Resources:** Daniel John Jackson.

**Software:** Nicolas Cerveau.

**Supervision:** Daniel John Jackson.

**Writing – original draft:** Susanne Affenzeller.

**Writing – review & editing:** Susanne Affenzeller, Nicolas Cerveau, Daniel John Jackson.

## References

1. Fisher R, Diver C. Crossing-over in the Land Snail *Cepaea nemoralis*, L. *Nature*. 1934. 133: 834–835.
2. Cain AJ, Sheppard PM. Selection in the polymorphic land snail *Cepaea nemoralis*. *Heredity*. 1950. 4: 275–294. PMID: [14802986](#)
3. Cain AJ, Sheppard PM. The effects of natural selection on body colour in the land snail *Cepaea nemoralis*. *Heredity*. 1952. 6: 217–231.
4. Cain AJ, Sheppard PM. Natural selection in *Cepaea*. *Genetics*. 1954. 39: 89–116. PMID: [17247470](#)
5. Lamotte M. Polymorphism of Natural Populations of *Cepaea nemoralis*. *Cold Spring Harb Symp Quant Biol*. 1959. 24: 65–86. <https://doi.org/10.1101/SQB.1959.024.01.009> PMID: [14413686](#)
6. Greenwood J. Visual and other selection in *Cepaea*: A further example. *Heredity*. 1974. 33: 17–31.
7. Jones JS, Leith B, Rawlings P. Polymorphism in *Cepaea*: a problem with too many solutions?. *Annu Rev Ecol Syst*. 1977. 8: 109–143. <https://doi.org/10.1146/annurev.es.08.110177.000545>
8. Surmacki A, Ożarowska-Nowicka A, Rosin ZM. Color polymorphism in a land snail *Cepaea nemoralis* (Pulmonata: Helicidae) as viewed by potential avian predators. *Naturwissenschaften*. 2013. 100: 1–8. <https://doi.org/10.1007/s00114-012-1000-7>
9. Cain AJ, King J, Sheppard P. New data on the genetics of polymorphism in the snail *Cepaea nemoralis* L. *Genetics*. 1960. 45: 393–411. PMID: [17247932](#)
10. Cook L. The genetics of *Cepaea nemoralis*. *Heredity*. 1967. 22: 397–410. <https://doi.org/10.1038/hdy.1967.49>
11. Cuthill IC, Allen WL, Arbuckle K, Caspers B, Chaplin G, Hauber ME, et al. The biology of color. *Science*. 2017. 357. <https://doi.org/10.1126/science.aan0221> PMID: [28774901](#)
12. Schwander T, Libbrecht R, Keller L. Supergenes and complex phenotypes. *Curr Biol*. 2014. 24: R288–R294. <https://doi.org/10.1016/j.cub.2014.01.056> PMID: [24698381](#)
13. Richards PM, Liu MM, Lowe N, Davey JW, Blaxter ML, Davison A. RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Mol Ecol*. 2013. 22: 3077–3089. <https://doi.org/10.1111/mec.12262> PMID: [23496771](#)
14. Mann K, Jackson DJ. Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. *BMC Genomics*. 2014. 15: 249. <https://doi.org/10.1186/1471-2164-15-249> PMID: [24684722](#)
15. Kerkvliet J, de Boer T, Schilthuizen M, Kraaijeveld K. Candidate genes for shell colour polymorphism in *Cepaea nemoralis*. *PeerJ*. 2017. 5: e3715. <https://doi.org/10.7717/peerj.3715> PMID: [28948095](#)
16. Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A. Guideline to reference gene selection for quantitative real-time PCR. *Biochem Biophys Res Commun*. 2004. 313: 856–862. <https://doi.org/10.1016/j.bbrc.2003.11.177> PMID: [14706621](#)
17. Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, Zumla A. Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques*. 2004. 37(1): 112–119. <https://doi.org/10.2144/04371RR03> PMID: [15283208](#)
18. Gubern C, Hurtado O, Rodríguez R, Morales JR, Romera VG, Moro MA, et al. Validation of housekeeping genes for quantitative real-time PCR in in-vivo and in-vitro models of cerebral ischaemia. *BMC Mol Biol*. 2009. 10(1): 57.
19. Chapman JR, Helin AS, Wille M, Atterby C, Järhult JD, Fridlund JS, et al. A panel of stably expressed reference genes for real-time qPCR gene expression studies of mallards (*Anas platyrhynchos*). *PloS One*. 2016. 11(2): e0149454. <https://doi.org/10.1371/journal.pone.0149454> PMID: [26886224](#)
20. Sirakov M, Zarrella I, Borra M, Rizzo F, Biffali E, Arnone MI, 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): 70.
21. Wan Q, Whang I, Choi CY, Lee JS, Lee J. Validation of housekeeping genes as internal controls for studying biomarkers of endocrine-disrupting chemicals in disk abalone by real-time PCR. *Comp*

- Biochem Physiol C Toxicol Pharmacol. 2011. 153(3): 259–268. <https://doi.org/10.1016/j.cbpc.2010.11.009> PMID: 21168524
22. Cubero-Leon E, Ciocan CM, Minier C, Rotchell JM. Reference gene selection for qPCR in mussel, *Mytilus edulis*, during gametogenesis and exogenous estrogen exposure. *Environ Sci Pollut Res Int*. 2012. 19(7): 2728–2733.
  23. García-Fernández P, Castellanos-Martínez S, Iglesias J, Otero JJ, Gestal C. Selection of reliable reference genes for RT-qPCR studies in *Octopus vulgaris* paralarvae during development and immune-stimulation. *J Invertebr Pathol*. 2016. 138: 57–62. <https://doi.org/10.1016/j.jip.2016.06.003> PMID: 27267177
  24. Liu Q, Lei K, Ma Q, Qiao F, Li ZC, An LH. Ribosomal protein L7 as a suitable reference gene for quantifying gene expression in gastropod *Bellamya aeruginosa*. *Environ Toxicol Pharmacol*. 2016. 43: 120–127. <https://doi.org/10.1016/j.etap.2016.03.007> PMID: 26991845
  25. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*. 2014. 30: 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
  26. Cerveau N, Jackson DJ. Combining independent de novo assemblies optimizes the coding transcriptome for nonconventional model eukaryotic organisms. *BMC Bioinformatics*. 2016. 17: 525. <https://doi.org/10.1186/s12859-016-1406-x> PMID: 27938328
  27. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013. 8: 1494–1512. <https://doi.org/10.1038/nprot.2013.084> PMID: 23845962
  28. Peng Y, Leung HCM, Yiu S-M, Lv M-J, Zhu X-G, Chin FY. IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. *Bioinformatics*. 2013. 29: i326–i334. <https://doi.org/10.1093/bioinformatics/btt219> PMID: 23813001
  29. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics*. 2007. 23(10): 1289–1291. <https://doi.org/10.1093/bioinformatics/btm091> PMID: 17379693
  30. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res*. 2012. 40(15): e115. <https://doi.org/10.1093/nar/gks596> PMID: 22730293
  31. Williamson P. Size-Weight Relationships and Field Growth Rates of the Land snail *Cepaea nemoralis* L. *J Anim Ecol*. 1976. 45: 875–885.
  32. 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(4): 611–622. <https://doi.org/10.1373/clinchem.2008.112797> PMID: 19246619
  33. Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, Van den Hoff MJB, et al. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res*. 2009. 37(6): e45–e45. <https://doi.org/10.1093/nar/gkp045> PMID: 19237396
  34. Ruijter JM, Villalba AR, Hellemans J, Untergasser A, van den Hoff MJ. Removal of between-run variation in a multi-plate qPCR experiment. *Biomol Detect Quantif*. 2015. 5: 10–14. <https://doi.org/10.1016/j.bdq.2015.07.001> PMID: 27077038
  35. 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(6): 509–515. PMID: 15127793
  36. Andersen CL, Ledet-Jensen J, Ørntoft T. Normalization of real-time quantitative RT-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: 5245–5250. <https://doi.org/10.1158/0008-5472.CAN-04-0496> PMID: 15289330
  37. 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(7): research0034. <https://doi.org/10.1186/gb-2002-3-7-research0034> PMID: 12184808
  38. Thorrez L, Van Deun K, Tranchevent LC, Van Lommel L, Engelen K, Marchal K, et al. Using ribosomal protein genes as reference: a tale of caution. *PLoS One*. 2008. 3(3): e1854. <https://doi.org/10.1371/journal.pone.0001854> PMID: 18365009