A new set of endogenous control genes for use in quantitative real-time PCR experiments show that formin *Ldia2*^{dex} transcripts are enriched in the early pond snail embryo

Harriet F. Johnson* and Angus Davison

School of Life Sciences, University Park, University of Nottingham, NG7 2RD. UK

Correspondence: Harriet Johnson; e-mail: harrietfjohnson@gmail.com

Running head: control genes for qRT-PCR in pond snails

ABSTRACT

Although the pond snail *Lymnaea stagnalis* is an emerging model organism for molecular studies in a wide variety of fields, there are a limited number of verified endogenous control genes for use in quantitative real-time PCR (qRT-PCR). As part of larger study on snail chirality, or left-right asymmetry, we wished to assay gene expression in pond snail embryos. We therefore evaluated six candidate control genes, by comparing their expression in three tissues (ovotestis, foot, and embryo) and across three programs (geNorm, Normfinder and Bestkeeper). The specific utility of these control genes was then tested by investigating the relative expression of six experimental transcripts, including the formin *Ldia2*, a gene that has been associated with chiral variation in *L. stagnalis*. All six control genes were found to be suitable for use in the three tissues tested. Of the six experimental genes, it was found that all were relatively depleted in the early embryo compared with other tissues, except the formin gene *Ldia2.* Instead, transcripts of the wild type $Ldia2^{dex}$ were enriched in the embryo, whereas a non-functional frameshifted version $Ldia2^{sin}$ was severely depleted. These differences in $Ldia2^{sin}$ expression were less evident in the ovotestis and not evident in the foot tissue, possibly because nonsense-mediated decay is obscured in actively transcribing tissues. This work therefore provides a set of control genes that may be useful to the wider community, and illustrates how they may be used to assay differences in expression in a variety of tissues.

INTRODUCTION

The pond snail *Lymnaea stagnalis* is a hermaphrodite, pulmonate snail which is increasingly used in a wide range of research areas including ecology, evolution, development, neuroscience, behaviour, parasitology and sexual selection. Due to the species perhaps predominant prior use as a model system in neuroscientific studies, many of the earlier published molecular studies were confined to the central nervous system (e.g. Feng *et al.*, 2009). More recently, molecular studies have come from different fields, including especially ecotoxicology and biomineralisation (Bouetard *et al.*, 2012; Hohagen & Jackson, 2013). An unannotated draft genome sequence is available (Davison *et al.*, 2016), and there is a collaborative effort underway to produce a publically available, high-quality, annotated genome sequence (Genoscope-CEA, de la Recherche à l'Industrie, France).

In the past few years, *L. stagnalis* snails have also become an important organism in the study of left-right asymmetry, because the species exhibits genetically tractable variation in chirality (Kuroda *et al.*, 2009; Shibazaki, Shimizu & Kuroda, 2004). This recently culminated in our finding that a disabling frameshift mutation in one copy of a diaphanous-related formin *Ldia2* is associated with early symmetry breaking in the developing embryo (Davison*, et al.*, 2016).

In preparing that work, we decided that it was necessary to design a new set of control genes to use with quantitative real-time PCR (qRT-PCR) in *L. stagnalis*. Specifically, as we needed to measure the expression of cytoskeletal genes in *L. stagnalis*, then this precluded the use of genes such as actin and tubulin as appropriate endogenous controls, because they are themselves cytoskeletal genes. Unfortunately, many of the previously published qRT-PCR studies on *L. stagnalis*

either used ribosomal RNA (rRNA), actin or tubulin genes as endogenous controls (Bavan *et al.*, 2012; Bouetard *et al.*, 2013; Carter *et al.*, 2015; Hatakeyama *et al.*, 2013; Lu & Feng, 2011; Ribeiro *et al.*, 2010; van Kesteren *et al.*, 2006; van Nierop *et al.*, 2006). rRNA genes are potentially problematic because the over-abundance of rRNAs relative to the target mRNA sequence can lead to problems in accurate normalisation, and in any case, rRNA is transcribed through an independent pathway from mRNA and therefore not regulated in the same manner (Radonic *et al.*, 2004). More generally, it is now widely accepted that there are no universal endogenous control genes, and each gene intended for use as an endogenous control should ideally be validated as consistently expressed across all experimental conditions.

Therefore, we aimed to develop and test a new set of endogenous genes as controls, for use in our study, but also for subsequent use by the wider community, just as has been the case in some other species (Hibbeler, Scharsack & Becker, 2008; Li *et al.*, 2017; Olias *et al.*, 2014; Sirakov *et al.*, 2009). These new control genes were then used to compare expression between different chiral genotypes of snail, and between different tissues.

METHODS

Sample preparation

Three separate tissues were used: single-cell embryo, ovotestis (hermaphrodite gonad) and foot, all from laboratory reared individuals of *L. stagnalis*. Total RNA was extracted from pooled embryos from a single individual using the RNeasy micro kit (Qiagen), including DNAse (Qiagen) treatment, yielding approximately 0.5 ng total RNA per embryo (Johnson, 2016). Ovotestis and foot tissue samples were removed from individual adult snails and snap frozen using a dry ice/ethanol slurry. Total RNA was immediately extracted from them using TRI Reagent® solution (Applied Biosystems).

RNA quality and concentration were immediately assessed using both agarose gel electrophoresis and a NanoDrop spectrophotometer, with samples then stored at -80°C. To establish a set of endogenous gene controls, it would be advisable to use precisely the same quantities of mRNA in each complementary DNA (cDNA) synthesis reaction. We therefore aimed to use 500 ng total RNA for each sample, using Superscript III reverse transcriptase (Invitrogen) and random primer mix (NEB). There was limited variation (Supplementary Table 1 for full details) for the foot tissue (n = 10, mean = 500 ng, S.D. = 2, range = 497-503 ng – beyond the precision of measurement) and for the ovotestis (n = 9, mean = 499 ng, S.D. = 1, range 497-500). Less RNA was available from the embryo, because 500 ng of embryo RNA would have required ~1000 eggs from a single snail. Therefore, instead of reducing the RNA to that of the lowest yielding (and possibly poorest quality sample), we allowed 2.5 fold variation (n = 12, mean = 121 ng, S.D. 30, range 75-189 ng) for this tissue.

Including individuals that were also used to compare expression of experimental genes, there was slightly wider variation in initial staring RNA quantity for both ovotestis (n = 36, mean = 505 ng, S.D. 31, range 375-572; only one sample less than 463 ng) and embryo (n = 17, mean = 135 ng, S.D. 40, range 75-233 ng) (Supplementary Table 1).

Serial dilutions were performed independently for each standard curve experiment. Aliquots were then made of the experimental working concentration dilutions of cDNA to reduce freeze-thaw cycles. All cDNA samples were gently vortexed before use and prior to each serial dilution step.

Primer design

Using transcriptomic resources of 1-2 cell stage *L. stagnalis* embryos (Liu *et al.*, 2014), six genes were selected as potential endogenous controls, all with well-characterised gene function. These were short-chain specific acyl-CoA dehydrogenase (*Lacads*), elongation factor 1-alpha (*Lef1a*), histone protein, H2A (*Lhis2a*), 60S ribosomal protein L14 (*Lrpl14*), ubiquitin-conjugating enzyme E2 (*Lube2*), and 14-3-3 protein zeta (*Lywhaz*). Primer pairs were then designed using Primer 3 (Rozen & Skaletsky, 2006), aiming for a Tm range within 2°C, and amplicon product sizes between 110-130bp, including GC clamps where possible. All primer pairs were intron-spanning, with the primers on exon/intron boundaries, where possible, to minimise problems with accidental genomic DNA carry over. To initially verify the primers, produced by IDT, a standard PCR was used alongside a genomic DNA control sample and the products visualised on an agarose gel. Additionally, the specificity of the amplicons of all six primer pairs was verified through Sanger sequencing.

Primer specificity and amplification efficiency

Primer efficiencies for each primer pair were calculated via standard curve qRT-PCR experiments using the Applied Biosystems 7500 fast system v2.3 and the same cycling parameters (below). Five standardised concentrations were used with an additional negative control (PCR grade water). Five-step serial cDNA dilutions were performed using molecular grade water and a dilution factor of 1:5. Primer efficiencies for all six endogenous control gene primer pairs were estimated using the same reference sample, created from pooling cDNA samples. Average primer efficiencies for each primer pair were then calculated via the arithmetic mean of a minimum of two successful standard curve experiments. A standard curve experiment was considered successful if it produced a R^2 value of >0.98. Values from the lowest concentration dilutions were omitted if they dramatically reduced the amplification efficiency or R^2 value of an experiment. The range of dilutions included in the standard curve experiment indicates the limits of acceptable working concentration/dilution factor for an experimental comparative qRT-PCR assessment.

Cycle threshold (Cq) values were obtained from qRT-PCR experiments using the ABI 7500 fast system v2.3. Each reaction contained 5 µl of Primer Design's fast SYBR® green master mix, 0.5 µl forward and reverse primer (4 µM), 1.5 µl PCR grade water and 3 µl of cDNA. All samples were used at a 1:30 dilution of the original cDNA concentration. Mastermixes were prepared for each target gene experiment and a temperature melt curve step was included at the end of all qRT-PCR reactions. Thermocycling parameters were as follows: 95°C for 20, 95°C for 3 seconds, 60°C for 30s (data collection, Cq), then 39 cycles between steps 2 and 3; this was followed by 95°C for 15 seconds, 60°C for 60 seconds, a slow temperature ramp 1% (data collection; temperature melt curve), 95°C for 15 seconds, 60°C 15

seconds; temperature melt curves indicated that a single specific product was produced in all cases.

Normalising control software

Three methods were used to assess the same qRT-PCR data, all of which run as macros within Microsoft Excel 2003. BestKeeper used raw Cq values (Pfaffl *et al.*, 2004), whereas NormFinder (Andersen, Jensen & Orntoft, 2004) and geNorm (Vandesompele *et al.*, 2002) required linearised Cq values. Efficiency-corrected linearised relative Cq values were calculated for each sample using the Pfafll method (Hellemans *et al.*, 2007). BestKeeper ran entirely from raw Cq values and corrected for amplification efficiency via the inbuilt formulas within the macro, via the manually-input amplification efficiency values.

Snail lines and tissues

Variation in the left-right asymmetry of snails, or chirality, is under the control of a single maternally expressed locus. In *L. stagnalis*, maternal *D* alleles dominantly determine a clockwise ("dextral") twist in offspring. For our experiments we created a single near-isogenic line of snails (>99% inbred) that was still variable for the chirality locus, by repeated backcrossing (Davison*, et al.*, 2016). From this line, separate homozygous dextral (*DD*) and sinistral (*dd*) lines were produced. Heterozygote (*Dd*) snails were then derived by crossing individuals from the near-isogenic lines. qRT-PCR data was generated for *DD*, *Dd* and *dd* genotypes using both embryo and ovotestis tissue; however as resources were limited, only *DD* and *dd* genotypes were sampled for foot tissue.

Relative expression of cytoskeletal genes

Previously, we reported finding that tandemly duplicated, diaphanous-related formin genes, *Ldia1* and *Ldia2*, are perfectly associated with variation in chirality of the pond snail, and that the sinistral-derived version of *Ldia2* contains a disabling frameshift mutation, which results in much reduced levels of *Ldia2* mRNA in the embryo (Davison*, et al.*, 2016). To further explore changes in expression between genotypes (*DD*, *Dd*, *dd*) and tissues (single-cell embryo, foot, ovotestis), the relative expression of *Ldia1* and *Ldia2* was tested against the validated endogenous control genes.

As above, primer pairs were designed using Primer 3 with the same conditions (Table 1). However, because of the high sequence identity between *Ldia1* and *Ldia2*, it was not possible to design intron-spanning PCRs for these loci. Instead, primer pairs were designed in the 3'UTR, because this region was most variable between copies, *Ldia1 3'UTR* and *Ldia2 3'UTR*, in addition to a primer pair in the open reading frame, *Ldia2 ORF*. Four other genes were also tested, including the cytoskeletal genes furry *Lfry* and fat-like cadherin *Lfat,* both tightly linked to *Ldia1/2* on the same chromosome, as well as unlinked actin-related proteins subunits 1a and 3, *Larp2/3-1a* and *Larp2/3-3*.

Relative expression of these six genes (seven primer pairs) was tested against the endogenous controls by calculating the Normalised Relative Quantity (NRQ) values from the average Cq value of each sample using the Pfaffl method (Hellemans*, et al.*, 2007; Pfaffl, 2001; Pfaffl*, et al.*, 2004), relative to a single standard *DD* snail. Experimental samples were performed in triplicate repeat and negative controls in duplicate repeat for each of the six genes.

For each sample, first the relative quantity per target gene (ΔCq target) was calculated by subtracting the average Cq value of the sample from that of the calibrator sample. This ΔCq value was then corrected for amplification efficiency (E) by multiplying ΔCq to the base percentage amplification efficiency (represented as a value between 1 and 2). The efficiency-corrected relative quantities were then normalised to the endogenous control genes by dividing by the geometric mean (geoM) of the efficiency corrected delta Cq values calculated for each of the control genes (ΔCq ref) in the same manner as described above.

To measure the relative expression of cytoskeletal genes between genotypes *within* the same tissue, NRQ values were normalised to the geometric mean of the three endogenous control genes, using *Lhis2a*, *Lube2* and *Lywhaz* for embryo and foot, and *Lhis2a*, *Lube2* and *Lrpl14* for ovotestis. The standard was cDNA made from the same tissue, using a single *DD* snail. To make comparisons *across* different tissues, NRQ values were normalised to the geometric mean of the two endogenous control genes quantified in all tissues, *Lhis2a* and *Lube2*. The standard was the exactly the same sample for each analysis, cDNA made from pooled ovotestis RNA of mixed genotype.

RESULTS

Primer specificity and amplification efficiency

All control primer pairs demonstrated amplification efficiencies between 1.906 and 2.115 with R^2 values exceeding 0.98 (Table 1). All primers demonstrated acceptable amplification efficiencies in dilutions up to 1:150 (0.67%) of the full concentration. The working concentration of a 1:30 dilution that was used in the subsequent qRT-PCR experiments fell well within these limits.

Comparing normalising control software

Summaries of the top genes to use for each tissue are shown in Table 2. In fact, all genes and all combinations were acceptable for use. *Lef1a* was consistently found to be the least stable gene in all tissues but was still recommended in some analyses (Table 2).

In full detail (Table 3), in the embryo geNorm placed *Lhis2a* and *Lube2* as the most stable pair of genes, with a combined stability score of 0.196. The inclusion of any number of the genes provided a V score of <0.15, indicating that the combination of genes will provide a reliable normalisation factor (PrimerDesign 2014). The lowest (best) V score was achieved with the inclusion of the five genes *Lhis2a*, *Lube2*, *Lrpl14*, *Lacads* and *Lywhaz*. In the foot, geNorm placed *Lywhaz* and *Lube2* as the most stable pair of genes with a combined stability score of 0.217. The inclusion of any number of the genes provided a V score of <0.15, although the lowest V score was achieved with the inclusion of the four genes *Lywhaz*, *Lube2*, *Lhis2a* and *Lacads*. In the ovotestis, geNorm placed *Lrpl14* and *Lube2* as the most stable pair of control genes with a combined score of 0.250. *Lhis2a* bore the lowest M score of all the target genes, at 0.360, yet it was placed fourth in the combined

stability score. Again, the inclusion of any number of the genes provided a V score of <0.15, although the lowest V score was achieved with the inclusion of the three genes; *Lrpl14*, *Lube2* and *Lywhaz*.

NormFinder outputs the most stable pair and most stable individual gene. *Lhis2a* was the most stable gene in the embryo (stability value of 0.058; Table 3), but the best combined pair was *Lacads* and *Lube2* (0.047). In the foot, *Lywhaz* was identified as the most stable gene (0.074) and was paired with *Lube2* (combined stability 0.066). In the ovotestis, *Lhis2a* was most stable (0.124), but the best combined pair of genes was *Lef1a* and *Lywhaz* (0.083), despite the fact that *Lef1a* presented the poorest (highest) individual gene stability value (0.243). As with geNorm, the embryo analyses yielded the least variable scores. *Lef1a* was found to be the least stable or second least stable individual gene in all tissues. In all analyses, the stability value of the best combined pair of genes was lower than that of any individual gene stability score.

The BestKeeper program provides two measures of gene stability, with a low SD (<1) and a high r value indicating a more stable control gene; additionally a statistically significant correlation with the BK index (generated from all data), is used to shows that the data is in keeping with that expected across samples/low variability. In the embryo, the gene ranked as most stable according to SD was *Lhis2a* (0.408; Table 3), whereas the least stable gene was *Lef1a* (0.577). Every gene in the embryo analysis resulted in a highly significant positive correlation with the BestKeeper index (P = 0.001). *Lhis2a* demonstrated the highest correlation, with an r value of 0.979, and *Lywhaz* the lowest with an r value of 0.900. In the foot, *Lrpl14* was ranked as most stable according to SD (0.500), whereas the least stable

gene was *Lhis2a* (0.947). With the exception of *Lef1a* in the ovotestis, every gene/tissue combination showed a significant correlation with the BK index.

Relative expression of cytoskeletal genes

Relative expression of *Ldia2* transcripts depends upon the genetic background of the mother. Thus, levels of Ldia2 transcripts in embryos derived from a genetically sinistral mother *dd* were 0.006 (*Ldia2 3'UTR,* 0.6%) or 0.03 (*Ldia2 ORF,* 3%) relative to embryos from a wild-type *DD* mother (Figure 1; Tables 4 and 5); levels of the same transcripts in offspring of a heterozygote mother *Dd* were about half that of wild-type, 0.56 (*Ldia2 3'UTR,* 0.56%) or 0.48 (*Ldia2 ORF,* 0.48%). Notably, the relative differences in expression of *Ldia2* transcripts were much less striking in ovotestis, though still significantly lower, 0.81 (*Dd*) and 0.69 (*dd*) using *Ldia2 3'UTR* and 0.80 (*Dd*) and 0.62 (*dd*) using *Ldia2 ORF*; in foot tissue, there were no significant differences in expression between *Ldia2* transcripts from snails of different genotype (Figure 1; Tables 4 and 5). In comparison, there were few significant differences in the expression of the other genes with the exception of *Larp2/3-3* (*DD*:*Dd*, and *Dd*:*dd*) and *Larp2/3-1a* (*Dd*:*dd*) in the embryo.

Using homozygous dextral (DD) snails, all of the tested gene transcripts were relatively depleted in the single-cell embryo compared against ovotestis and foot, except *Ldia2*, which was enriched (Figure 2; Table 6). *Larp*, *Lfat* and *Lfry* transcripts were reduced to ~0.03 to 0.27 of the level in embryo compared to ovotestis, and ~0.11 to 0.38 when comparing foot to single-cell embryo. In comparison, levels of *Ldia2* expression were ~1.27 to 2 times higher in the single-cell embryo compared to the ovotestis and ~2.8 times higher when compared to the foot tissue.

DISCUSSION

Individually, all six gene targets were found to provide stable endogenous controls across all tissues. However, the best individual gene and combination of genes differed between tissues used and analysis program (Table 2). As it is recommended to use more than one control gene in combination in an experiment, then a tissue specific analysis is advisable prior to the experiment proper. Whether adding a third gene is worth the additional time and resources will depend on the individual experiment and the extent of the increase in stability gained.

Genes to use as endogenous controls in different tissues

Within the embryo, all three algorithms ranked *Lhis2a* as the most stable single gene, but there was less consensus for the rankings of the remaining endogenous controls. Generally, *Lhis2a*, *Lrpl14* and *Lube2* were in the top three most stable genes across software and tissue (Tables 2 and 3). For the foot tissue analyses, due to the agreement of GeNorm and Normfinder, *Lywhaz* and *Lube2* might be recommended as endogenous normalising controls. In comparison, for the ovotestis both geNorm and BestKeeper showed that the use of *Lrpl14* and *Lube2* might be recommended, with the inclusion of a third gene, *Lywhaz*, indicating the most stable combination of genes.

Lef1a was consistently ranked least stable, interesting because it has been a common choice by others as an endogenous control (Foster, Lukowiak & Henry, 2015; van Nierop*, et al.*, 2006). However, we found that it is still acceptable for use, just not necessarily the gene of choice. The reason for the relatively poor performance may be due to a low level of expression rather than variable expression, indicated in the amplification efficiency experiments (Table 1). *Lef1a*

may thus provide a reliable endogenous control gene when using an increased cDNA concentration.

Compared to the other tissues assessed, the embryo was found to be least variable (Figure 2). There are many reasons why some tissues may be more variable than others. In our experiments, it was difficult to temporally control the extraction of the ovotestis (e.g. time since egg-laying), and especially to make sure that it was free of contaminating hepatopancreas. In comparison, the embryos were from a clean and temporally controlled sample.

All three analytical programs used here provided a unique aspect of the data analysis. geNorm provided a measure of the optimum number of genes to include in the analysis and an advised cut-off value (V, <0.15) for an acceptable endogenous control gene combination. BestKeeper output a quotable measure of SD for each gene and a statistical measure of the relatedness of gene expression. Finally, NormFinder provided valuable information on the experimental design; calculating variation created both within and between experimental groups and importantly provides an alternative to pairwise comparison methods.

Comparing expression in different tissues

Previously, we used qRT-PCR to show the formin *Ldia2* shows significant fold-change differences in the quantity of mRNA transcripts between different chirality-associated genotypes (Davison*, et al.*, 2016). Here we showed that the cytoskeletal genes, including *Ldia1*, are substantially depleted in the embryo, except for dextral-derived alleles of the formin, $Ldia2^{dex}$, which are substantially enriched (Figure 2). In comparison, the frameshifted version, $Ldia2^{sin}$*,* was severely depleted in single-cell embryos (Figure 1), but these differences were less evident in the ovotestis and not

evident in the foot tissue (Figure 2). This may partly due to lower levels of expression, especially in the foot tissue, but mainly because the ability to detect the dynamics of nonsense mediated decay of RNA may be obscured in actively transcribing tissues.

As the cellular processes associated with variations in *L. stagnalis* chirality are predominantly cytoskeletal (Davison*, et al.*, 2016; Shibazaki*, et al.*, 2004; Tee *et al.*, 2015), this work further emphasises the potential pitfalls of using the commonly employed endogenous control genes, actin or tubulin, without adequate testing of their expression stability. It also suggests that *Ldia1*/*Ldia2* have different roles during development, despite the close sequence similarity, and that *Ldia2* may be particularly critical in early development, given the relatively enriched levels of transcript present in the single cell embryo.

Comprehensive studies of nonsense-mediated decay have not been performed in molluscs. However, in nonsense-mediated decay studies, from yeast to mammals, decay has been observed in both a 5' to 3' and 3' to 5' direction of the mRNA, originating from either the 3' end or exon-exon boundaries (Karousis, Nasif & Mühlemann, 2016; Lykke-Andersen & Jensen, 2015). The variation in starting position of nonsense-mediated decay limits interpretation of the differences between the reduction *Ldia2* in the 3' UTR and ORF. However, the frameshift in $Ldia2^{sin}$ transcripts should be present in all tissues, and therefore the resulting nonsense-mediated decay would be expected to be evident in all tissues. The lack of significant quantitative differences in the foot tissue suggests that nonsense-mediated decay may be obscured in actively transcribing tissues. As transcription does not begin in *L. stagnalis* before the 8-cell stage (Liu*, et al.*, 2014), so a single-cell embryo only contains maternal mRNAs that are transcribed prior to oviposition; ovotestis is

presumably enriched for this same material. Further experiments with later stage embryos would presumably confirm this hypothesis.

Limitations

In seeking to establish a set of verified endogenous control genes for use in quantitative real-time PCR (qRT-PCR), it would be advisable to use a consistent starting quantity of mRNA in each cDNA synthesis reaction. In this study, our aim was to use 400-500 ng total RNA in each starting reaction, but reality meant that there was some variation, minor for the foot and ovotestis tissue but more significant for the embryo, also using less RNA for the latter due to the limited yield of RNA from a single cell embryo.

While the results show that the primer-pairs are efficient over several orders of magnitude, the relative ratios of the endogenous controls and tested genes should not differ, irrespective of the starting quantity of RNA that was used as input. Any minor variation should not therefore impact upon measured relative expression of experimental genes. However, in terms of establishing the endogenous controls in the first place, it would have been preferable to use samples with less variation, especially for the embryo, and to use a more accurate method to measure RNA over the NanoDrop, such as a Qubit fluorimeter.

The stability analyses using foot and ovotestis may be more reliable than those using embryo, because of the greater variation in input RNA for embryo. Nonetheless, it is worth noting that the tissue with the lowest starting quantity of RNA and the most variation, embryo, showed the least variation in relative transcript levels.

Conclusions

It was established that any of the six genes would provide acceptable endogenous controls to standardise gene expression between chiral genotypes within any of the three different tissues. These primers should therefore permit rapid verification of endogenous controls suitable for use in qRT-PCR experiments assessing ovotestis, foot and embryo tissue within and between chiral variants of *L. stagnalis*, which was lacking previously.

SUPPORTING INFORMATION

Supplementary Information 1. Fasta format alignments of genomic sequences against transcriptome and primer sequences for endogenous control genes and cytoskeletal genes.

Supplementary Information 2. Raw data for the qRT-PCR experiments.

ACKNOWLEDGEMENTS

The authors would like to thank Maureen Liu and Sheila Keeble for assistance with some research planning and laboratory work. This work was mainly funded by a BBSRC DTP studentship to Harriet Johnson. The work presented here is adapted from a chapter in her PhD thesis, available electronically at http://eprints.nottingham.ac.uk/33183/

REFERENCES

ANDERSEN, C.L., JENSEN, J.L. & ORNTOFT, 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 Research*, **64**: 5245-5250.

BAVAN, S., STRAUB, V.A., WEBB, T.E. & ENNION, S.J. 2012 Cloning and characterization of a P2X receptor expressed in the central nervous system of *Lymnaea stagnalis*. *PLoS One*, **7**.

BOUETARD, A., BESNARD, A.L., VASSAUX, D., LAGADIC, L. & COUTELLEC, M.A. 2013 Impact of the redox-cycling herbicide diquat on transcript expression and antioxidant enzymatic activities of the freshwater snail *Lymnaea stagnalis*. *Aquatic Toxicology*, **126**: 256-265.

BOUETARD, A., NOIROT, C., BESNARD, A.L., BOUCHEZ, O., CHOISNE, D., ROBE, E., KLOPP, C., LAGADIC, L. & COUTELLEC, M.A. 2012 Pyrosequencing-based transcriptomic resources in the pond snail *Lymnaea stagnalis*, with a focus on genes involved in molecular response to diquat-induced stress. *Ecotoxicology*, **21**: 2222-2234.

CARTER, C.J., RAND, C., MOHAMMAD, I., LEPP, A., VESPRINI, N., WIEBE, O., CARLONE, R. & SPENCER, G.E. 2015 Expression of a retinoic acid receptor (RAR)-like protein in the embryonic and adult nervous system of a protostome species. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, **324**: 51-67.

DAVISON, A., MCDOWELL, G.S., HOLDEN, J.M., JOHNSON, H.F., KOUTSOVOULOS, G.D., LIU, M.M., HULPIAU, P., VAN ROY, F., WADE, C.M., BANERJEE, R., YANG, F., CHIBA, S., DAVEY, J.W., JACKSON, D.J., LEVIN, M. &

BLAXTER, M.L. 2016 Formin is associated with left-right asymmetry in the pond snail and the frog. *Current Biology*, **26**: 654-660.

FENG, Z.P., ZHANG, Z., VAN KESTEREN, R.E., STRAUB, V.A., VAN NIEROP, P., JIN, K., NEJATBAKHSH, N., GOLDBERG, J.I., SPENCER, G.E., YEOMAN, M.S., WILDERING, W., COORSSEN, J.R., CROLL, R.P., BUCK, L.T., SYED, N.I. & SMIT, A.B. 2009 Transcriptome analysis of the central nervous system of the mollusc *Lymnaea stagnalis*. *BMC Genomics*, **10**: 15.

FOSTER, N.L., LUKOWIAK, K. & HENRY, T.B. 2015 Time-related expression profiles for heat shock protein gene transcripts (HSP40, HSP70) in the central nervous system of *Lymnaea stagnalis* exposed to thermal stress. *Communicative & integrative biology*, **8**: e1040954.

HATAKEYAMA, D., OKUTA, A., OTSUKA, E., LUKOWIAK, K. & ITO, E. 2013 Consolidation of long-term memory by insulin in Lymnaea is not brought about by changing the number of insulin receptors. *Communicative & integrative biology*, **6**: e23955.

HELLEMANS, J., MORTIER, G., DE PAEPE, A., SPELEMAN, F. & VANDESOMPELE, J. 2007 qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol*, **8**.

HIBBELER, S., SCHARSACK, J.P. & BECKER, S. 2008 Housekeeping genes for quantitative expression studies in the three-spined stickleback Gasterosteus aculeatus. *BMC Molecular Biology*, **9**: 18.

HOHAGEN, J. & JACKSON, D.J. 2013 An ancient process in a modern mollusc: early development of the shell in *Lymnaea stagnalis*. *BMC Developmental Biology*, **13**.

JOHNSON, H.F. 2016 Left-right asymmetry variation in the pond snail, *Lymnaea stagnalis*: exploring patterns of gene expression. PhD thesis, School of Life Sciences, University of Nottingham.

KAROUSIS, E.D., NASIF, S. & MÜHLEMANN, O. 2016 Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdisciplinary Reviews: RNA*, **7**: 661-682.

KURODA, R., ENDO, B., ABE, M. & SHIMIZU, M. 2009 Chiral blastomere arrangement dictates zygotic left-right asymmetry pathway in snails. *Nature*, **462**: 790-794.

LI, T., DIAO, H., ZHAO, L., XING, Y., ZHANG, J., LIU, N., YAN, Y., TIAN, X., SUN, W. & LIU, B. 2017 Identification of suitable reference genes for real-time quantitative PCR analysis of hydrogen peroxide-treated human umbilical vein endothelial cells. *BMC Molecular Biology*, **18**: 10.

LIU, M.M., DAVEY, J.W., JACKSON, D.J., BLAXTER, M.L. & DAVISON, A. 2014 A conserved set of maternal genes? Insights from a molluscan transcriptome. *International Journal of Developmental Biology*, **58**: 501-511.

LU, T.Z. & FENG, Z.-P. 2011 A sodium leak current regulates pacemaker activity of adult central pattern generator neurons in *Lymnaea stagnalis*. *PLoS One*, **6**: e18745.

LYKKE-ANDERSEN, S. & JENSEN, T.H. 2015 Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nature Reviews Molecular Cell Biology*, **16**: 665-677.

OLIAS, P., ADAM, I., MEYER, A., SCHARFF, C. & GRUBER, A.D. 2014 Reference genes for quantitative gene expression studies in multiple avian species. *PLoS One*, **9**: e99678.

PFAFFL, M.W. 2001 A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, **29**.

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. *Biotechnology Letters*, **26**.

RADONIC, A., THULKE, S., MACKAY, I.M., LANDT, O., SIEGERT, W. & NITSCHE, A. 2004 Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications*, **313**: 856-862.

RIBEIRO, M., SCHOFIELD, M., KEMENES, I., BENJAMIN, P.R., O'SHEA, M. & KORNEEV, S.A. 2010 Atypical guanylyl cyclase from the pond snail *Lymnaea stagnalis*: cloning, sequence analysis and characterization of expression. *Neuroscience*, **165**: 794-800.

ROZEN, S. & SKALETSKY, H.J. 2006 Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols*: (Krawetz, S., Misener, S. and Totowa, N.J., eds). Methods in Molecular Biology, Humana Press.

SHIBAZAKI, Y., SHIMIZU, M. & KURODA, R. 2004 Body handedness is directed by genetically determined cytoskeletal dynamics in the early embryo. *Current Biology*, **14**: 1462-1467.

SIRAKOV, M., ZARRELLA, I., BORRA, M., RIZZO, F., BIFFALI, E., ARNONE, M.I. & FIORITO, G. 2009 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 Molecular Biology*, **10**: 70.

TEE, Y.H., SHEMESH, T., THIAGARAJAN, V., HARIADI, R.F., ANDERSON, K.L., PAGE, C., VOLKMANN, N., HANEIN, D., SIVARAMAKRISHNAN, S., KOZLOV,

M.M. & BERSHADSKY, A.D. 2015 Cellular chirality arising from the self-organization of the actin cytoskeleton. *Nature Cell Biology*, **17**: 445-457.

VAN KESTEREN, R.E., CARTER, C., DISSEL, H.M.G., VAN MINNEN, J., GOUWENBERG, Y., SYED, N.I., SPENCER, G.E. & SMIT, A.B. 2006 Local synthesis of actin-binding protein beta-thymosin regulates neurite outgrowth. *Journal of Neuroscience*, **26**: 152-157.

VAN NIEROP, P., BERTRAND, S., MUNNO, D.W., GOUWENBERG, Y., VAN MINNEN, J., SPAFFORD, J.D., SYED, N.I., BERTRAND, D. & SMIT, A.B. 2006 Identification and functional expression of a family of nicotinic acetylcholine receptor subunits in the central nervous system of the mollusc *Lymnaea stagnalis*. *Journal of Biological Chemistry*, **281**: 1680-1691.

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 Biology*, **3**.

Figure 1. Boxplots showing Log scale NRQ values (LOG10 NRQ) for four different genes in in three genotypes, *DD*, *Dd* and *dd* across three tissues, embryo, foot and ovotestis tissue, relative to a single standard *DD* snail. *Ldia2 3'UTR* transcripts are almost absent from the embryo in *dd* individuals, and also show reduced expression in the ovotestis. This effect is not seen in the foot tissue. The graphs also show that, in general, between-sample variation is least in the embryo.

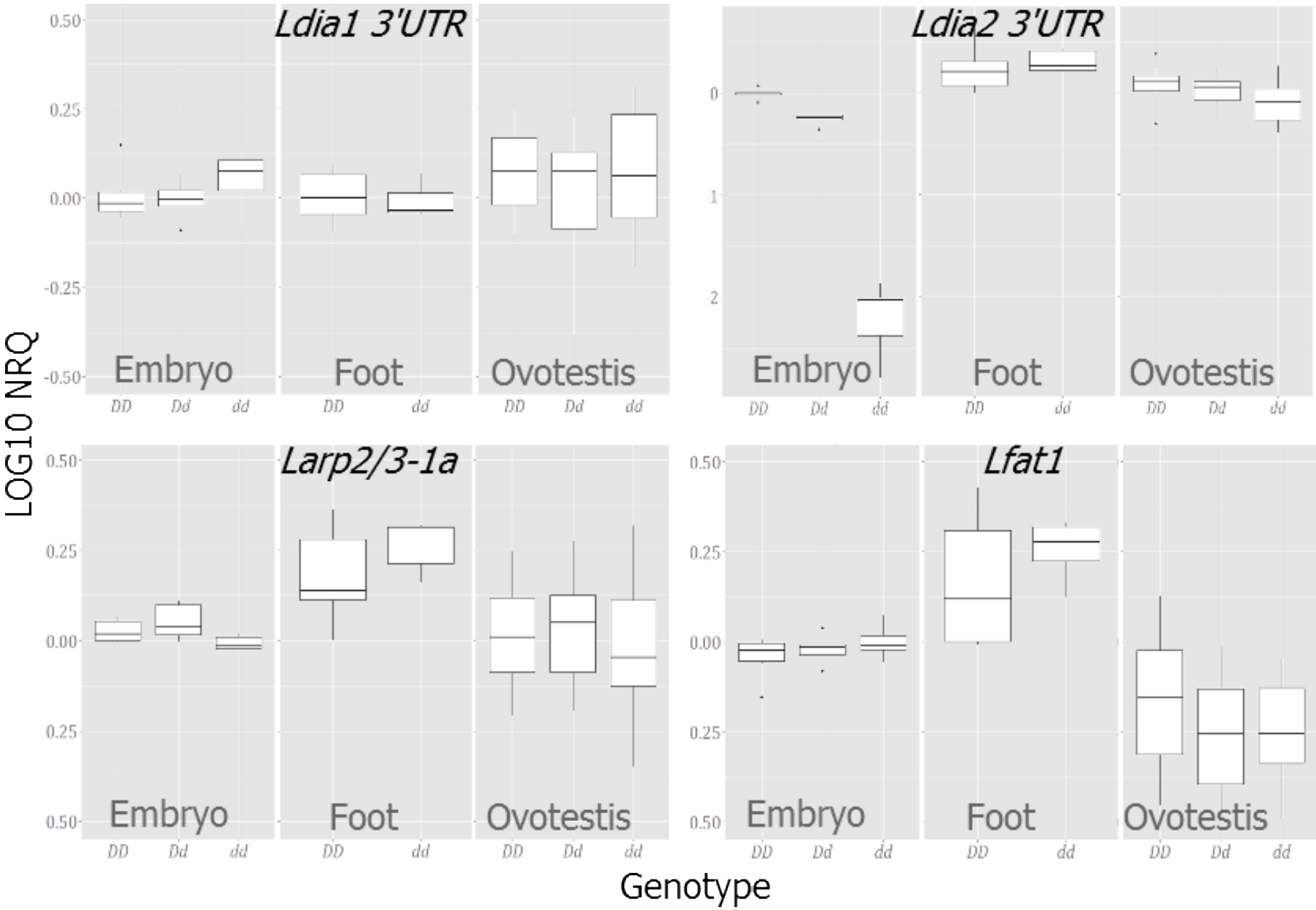
Ldia1 3'UTR
Ldia2 3'UTR
Larp2/3-1a
Lfat1
Embryo
Foot
Ovotestis
DD
Dd
dd
LOG10 NRQ
Genotype

Figure 2. Boxplots showing Log scale NRQ values (LOG10 NRQ) for six different genes in three different tissue), using a single genotype *DD*, and relative to a single standard. The expression of five genes is depleted in the embryo, with the exception of *Ldia2*.

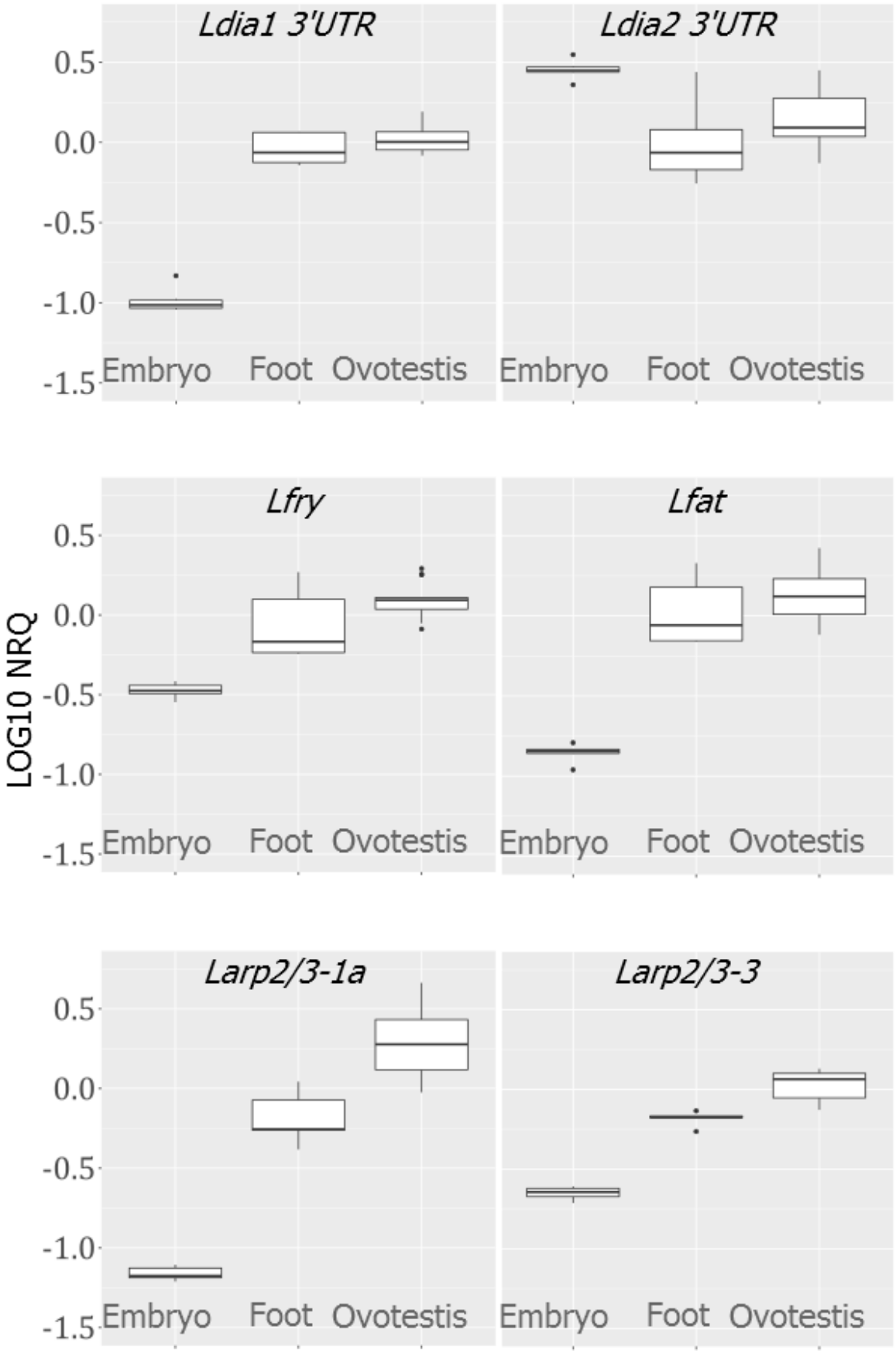
Ldia1 3'UTR
Ldia2 3'UTR
Lfry
Lfat
Larp2/3-1a
Larp2/3-3
LOG10 NRQ
0.5
0.0
-0.5
-1.0
-1.5
Embryo
Foot
Ovotestis

Table 1. Primers used for the amplification of endogenous control genes (top) and the tested cytoskeletal genes (bottom), including the estimated intron size and the minimum concentration of sample cDNA (as a percentage of full concentration) required to achieve the amplification efficiency. * indicates primer crosses exon/intron boundary.

ID	**Endogenous controls**	Primers	Tm (°C)	Intron (bp)	Efficiency	Min cDNA conc (%)
Lacads	acyl-CoA dehydrogenase	TGCACTCTCTAAACGAACTTCC	58.4	866	1.912	0.27
		TCCCTTGATTGTGCTGTTGAC	58.8			
Lef1	elongation factor 1-alpha	CGTCACAACCAGCATATCCC	58.7	663	2.115	0.67
		AGAGTTCGAGGGCTGCTTAC*	59.5			
Lhis2a	histone H2A	TCAGAGGAGATGAGGAGTTGG	58.3	785†	1.943	0.03
		CCCCAAGTTATGCTGCCTTC	58.9			
Lrpl14	60S ribosomal protein L14	TAATAAGTCGGTTGCGCGC*	59.0	2254	1.906	0.03
		GGGAACAGTCTACTTGGGC*	57.5			
Lube2	ubiquitin-conjugating E2	GCGGATCCTCTTGCAATCTT*	58.3	3224	1.923	0.03
		TCTGTGGACTGCATATCACTCT	58.6			
Lywhaz	14-3-3 protein zeta	GGAGGAGCTGAAGTCAATATGC	58.9	711	1.918	0.03
		AGTCACCCTGCATTTTGAGG	58.1			
ID	**Cytoskeletal genes**	Primers	Tm (°C)	Intron (bp)	Efficiency	Min cDNA conc (%)
Ldia1 3'UTR	diaphanous-related formin	AGTGGTGTGGGCAAAAGATG	58.7	n/a	1.986	0.27
		TATTCTGTTGATGCACGGCC	58.6			
Ldia2 3'UTR	diaphanous-related formin	GGGAGTTCAAGTTCAAGCCTATC	59.1	n/a	1.912	0.27
		GGCAAGCTACGACTCTTCTC	58.1			
Ldia2 ORF	diaphanous-related formin	GGGTGACAATGAAGTGGACC	58.5	n/a	1.948	1.33
		ACATGCATCTGTAACATCTGCC	59.1			
Lfry	furry	ACTTACCCTGCTCAAATGCC	58.2	717	1.876	0.59
		ATGTTTCTTGTGCTGCCGTC	59.4			
Lfat	fat-like cadherin	TGCCCATGTTGCTAAGTTCAG	58.8	1347	1.838	0.59
		CCTCTATCCCAGTTCGACGG	59.9			
Larp2/3-1a	actin-related protein 1a	CTGAAAATAGCCTTGTTGCAGC	58.8	343	1.847	0.67
		CCAGACTCCTTTTCCTGGGAC	60.0			
Larp2/3-3	actin-related protein 3	AGCCAGCTAACAAGGGAGAAG	59.7	520	1.775	0.67
		AGCATAGCCACCATTTGCTTG*	59.5			

Table 2. Summary of recommended single genes and gene pairs to use for each tissue and according to the normalising control software method. Although there is little consistency in the top genes to use across tissues and methods, in reality, all six endogenous control genes are acceptable for use.

Tissue	GeNorm	NormFinder	BestKeeper, SD
Embryo	*Lhis2a/Lube2*	*Lhis2a* or *Lacads/Lube2*	*Lhis2a/Lywhaz*
Foot	*Lyhwaz/Lube2*	*Lywhaz* or *Lywhaz/Lube2*	*Lrpl14/Lacads*
Ovotestis	*Lrpl14/Lube2*	*Lhis2a or Lef1/Lywhaz*	*Lrpl14/Lube2*

Table 3. Gene expression stability results per tissue, using three different normalising control methods. geNorm provides the best paired combination of genes, with additional V scores indicating the best accumulative combination and individual M scores giving a measure of individual expression stability. Normfinder provides the best combined pair of genes with a separate associated stability score. Bestkeeper results are presented as both their correlation with the BestKeeper index (r), with associated probability values (*P*), and the standard deviation (SD) associated with the average Cq per gene.

Tissue	geNorm				NormFinder				BestKeeper, SD		BestKeeper, r		
	Target	Stability score	V score	M score	Target	Stability score	Best Pair	Stability Value	Target	SD	Target	r	*P* value
Embryo, n=12	*Lhis2a*/ *Lube2*	0.196	0.061	0.246/ 0.259	*Lhis2a*	0.058	*Lacads*/ *Lube2*	0.047	*Lhis2a*	0.408	*Lhis2a*	0.979	0.001
					Lrpl14	0.076			*Lywhaz*	0.457	*Lef1a*	0.969	0.001
	Lrpl14	0.204	0.064	0.267	*Lube2*	0.086			*Lrpl14*	0.476	*Lube2*	0.962	0.001
	Lacads	0.242	0.052	0.282	*Lacads*	0.104			*Lacads*	0.493	*Lrpl14*	0.957	0.001
	Lywhaz	0.262	0.049	0.324	*Lef1a*	0.122			*Lube2*	0.514	*Lacads*	0.949	0.001
	Lef1a	0.285	n/a	0.330	*Lywhaz*	0.124			*Lef1a*	0.577	*Lywhaz*	0.900	0.001
Foot, n=10	*Lywhaz*/ *Lube2*	0.217	0.092	0.325/ 0.401	*Lywhaz*	0.074	*Lube2*/ *Lywhaz*	0.066	*Lrpl14*	0.500	*Lywhaz*	0.998	0.001
					Lube2	0.133			*Lacads*	0.638	*Lube2*	0.993	0.001
	Lhis2a	0.269	0.091	0.461	*Lacads*	0.151			*Lef1a*	0.695	*Lhis2a*	0.984	0.001
	Lacads	0.327	0.082	0.407	*Lrpl14*	0.176			*Lywhaz*	0.754	*Lacads*	0.981	0.001
	Lrpl14	0.376	0.088	0.489	*Lhis2a*	0.215			*Lube2*	0.867	*Lrpl14*	0.964	0.001
	Lef1a	0.444	n/a	0.579	*Lef1a*	0.298			*Lhis2a*	0.947	*Lef1a*	0.907	0.001
Ovotestis, n=9	*Lrpl14*/ *Lube2*	0.250	0.097	0.367/ 0.363	*Lhis2a*	0.124	*Lef1a*/ *Lywhaz*	0.083	*Lrpl14*	0.176	*Lywhaz*	0.894	0.001
					Lrpl14	0.147			*Lube2*	0.313	*Lube2*	0.877	0.002
	Lywhaz	0.292	0.070	0.384	*Lube2*	0.153			*Lhis2a*	0.330	*Lacads*	0.876	0.002
	Lhis2a	0.309	0.079	0.360	*Lywhaz*	0.171			*Lywhaz*	0.366	*Lrpl14*	0.853	0.003
	Lacads	0.360	0.077	0.473	*Lacads*	0.206			*Lef1a*	0.369	*Lhis2a*	0.831	0.005
	Lef1a	0.409	n/a	0.507	*Lef1a*	0.243			*Lacads*	0.500	*Lef1a*	0.655	0.056

Table 4. Normalised relative quantities (NRQ) of each gene, presented as a geometric mean per genotypic group (Geno), relative to different genotypes. Heterozygote snails, *Dd*, were not used with foot tissue.

	Genotype	N	*Larp2/3-1a*	*Larp2/3-3*	*Ldia1 3'UTR*	*Ldia2 3'UTR*	*Ldia2 ORF*	*Lfat*	*Lfry*
Embryo	*DD*	6	1	1	1	1	1	1	1
	Dd	5	1.059	0.735	0.969	0.563	0.476	1.052	0.984
	dd	6	0.926	0.960	1.111	0.006	0.029	1.103	0.960
Foot	*DD*	5	1	1	1	1	1	1	1
	dd	5	1.218	1.053	0.973	1.154	1.425	1.217	1.041
Ovotestis	*DD*	14	1	1	1	1	1	1	1
	Dd	8	1.087	0.758	0.843	0.800	0.809	0.832	0.910
	dd	14	0.965	0.877	1.017	0.619	0.686	0.847	0.878

Table 5. Wilcoxon rank test results for pairwise comparisons between genotypes *DD*, *Dd* and *dd* within embryo, foot and ovotestis tissue for cytoskeletal genes. The Wilcoxon rank value (W) is presented with the associated probability value (*P*). Statistical significance is highlighted via * <0.05, ** <0.01.

		DD versus *dd*					*DD* versus *Dd*					*Dd* versus *dd*				
Gene	Tissue	N, *DD*	N, *dd*	W	*P*		N, *DD*	N, *Dd*	W	*P*		N, *Dd*	N, *dd*	W	*P*	
	Embryo	6	6	30	0.065		6	5	10	0.429		5	6	27	0.03	*
Larp2/3-1a	Foot	5	5	7	0.31		n/a									
	Ovotestis	14	14	106	0.735		14	8	50	0.714		8	14	65	0.57	
	Embryo	6	6	23	0.485		6	5	29	0.009	**	5	6	3	0.03	*
Larp2/3-3	Foot	5	5	6	0.222		n/a									
	Ovotestis	14	14	117	0.401		14	8	81	0.095		8	14	50	0.714	
	Embryo	6	6	12	0.394		6	5	15	1		5	6	7	0.178	
Ldia1 3'UTR	Foot	5	5	13	1		n/a									
	Ovotestis	14	14	95	0.91		14	8	61	0.764		8	14	47	0.57	
	Embryo	6	6	36	0.002	**	6	5	30	0.004	**	5	6	30	0.004	**
Ldia2 3'UTR	Foot	5	5	8	0.421		n/a									
	Ovotestis	14	14	155	0.008	**	14	8	73	0.267		8	14	74	0.238	
	Embryo	6	6	36	0.002	**	6	5	30	0.004	**	5	6	30	0.004	**
Ldia2 ORF	Foot	5	5	7	0.31		n/a									
	Ovotestis	14	14	152	0.012	*	14	8	74	0.238		8	14	68	0.441	
	Embryo	6	6	11	0.31		6	5	14	0.931		5	6	12	0.662	
Lfat	Foot	5	5	8	0.421		n/a									
	Ovotestis	14	14	123	0.265		14	8	70	0.365		8	14	53	0.868	
	Embryo	6	6	25	0.31		6	5	16	0.931		5	6	19	0.537	
	Foot	5	5	11	0.841		n/a									
Lfry	Ovotestis	14	14	112	0.541		14	8	69	0.402		8	14	56	1	

Table 6. Normalised relative quantities (NRQ) of each gene, not log transformed, presented as a geometric mean per genotypic group and tissue, relative to the same single reference standard. Heterozygote snails, *Dd*, were not used with foot tissue.

Tissue	Genotype	N	*Larp2/3-1a*	*Larp2/3-3*	*Ldia1 3'UTR*	*Ldia2 3'UTR*	*Ldia2 ORF*	*Lfat*	*Lfry*
Embryo	*DD*	6	0.069	0.224	0.103	2.835	1.973	0.14	0.336
	Dd	5	0.077	0.175	0.105	1.676	0.988	0.154	0.348
	dd	6	0.065	0.219	0.118	0.019	0.058	0.158	0.33
Foot	*DD*	5	0.652	0.656	0.91	1.012	0.708	1.062	0.88
	dd	5	0.784	0.682	0.875	1.153	0.997	1.277	0.905
Ovotestis	*DD*	14	1.935	1.068	1.049	1.418	1.553	1.339	1.239
	Dd	8	2.256	0.869	0.95	1.218	1.348	1.196	1.21
	dd	14	1.82	0.914	1.04	0.855	1.039	1.105	1.06