

## A new reliable reference gene *UBA52* for quantitative real-time polymerase chain reaction studies in pyloric cecal tissues of the starfish *Asterias rubens*

A.F. Sadritdinova<sup>1</sup>, A.A. Dmitriev<sup>2</sup>, A.V. Snezhkina<sup>1</sup>, M.S. Belenikin<sup>1</sup>, G.S. Krasnov<sup>2</sup>, O.G. Manylov<sup>3</sup>, A.A. Kudryavtsev<sup>3</sup>, N.V. Melnikova<sup>1</sup>, A.S. Speranskaya<sup>1,4</sup>, M.V. Darii<sup>1</sup>, V.A. Lakunina<sup>1</sup>, L.A. Uroshlev<sup>1</sup>, A.O. Smurov<sup>5</sup>, O.A. Stepanov<sup>6</sup> and A.V. Kudryavtseva<sup>1</sup>

<sup>1</sup>Group of Post-Genomic Research, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

<sup>2</sup>Laboratory of Structural and Functional Genomics, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

<sup>3</sup>Department of Invertebrate Zoology, Saint Petersburg State University, Saint Petersburg, Russia

<sup>4</sup>Department of Higher Plants, Lomonosov Moscow State University, Moscow, Russia

<sup>5</sup>Laboratory of Brackish Water Hydrobiology, Zoological Institute of Russian Academy of Sciences, Saint Petersburg, Russia

<sup>6</sup>Department of Molecular and Biological Physics, Moscow Institute of Physics and Technology (State University), Moscow Region, Dolgoprudnyi, Russia

Corresponding author: A.V. Kudryavtseva  
E-mail: rhizamoeba@mail.ru

Genet. Mol. Res. 13 (2): 3972-3980 (2014)

Received May 28, 2013

Accepted September 19, 2013

Published May 23, 2014

DOI <http://dx.doi.org/10.4238/2014.May.23.8>

**ABSTRACT.** The starfish *Asterias rubens* is one of the most abundant echinoderm species in the White, Barents, North, and Baltic Seas.

This species is an important component of marine ecosystems and a model object for certain biological studies, in particular those requiring quantitative estimation of gene expression. As a rule, expression at the transcriptional level is estimated by real-time qPCR using the  $\Delta\Delta C_t$  method, which allows the comparison of the copy number of target gene transcripts in samples with unknown mRNA/cDNA concentration. Application of this method requires normalization of the results relative to genes with stable expression levels (reference genes). The identification of reference genes is still a challenging task since data of this kind are missing for certain taxa, whereas the use of “standard” endogenous control genes without additional tests might lead to erroneous conclusions. We performed a preliminary analysis of the expression of many housekeeping genes in the pyloric ceca of *A. rubens* by high-throughput sequencing under normal and heat shock conditions. For one of them, the ubiquitin gene *UBA52*, low variation of expression (not greater than 2-fold) was shown using real-time qPCR. Tissues of pyloric ceca of normal adults and underyearlings and of adults after heat shock were used. The data obtained suggest that the *UBA52* gene may be used as reference for normalization of gene expression at the mRNA level in the starfish *A. rubens* and probably in closely related species.

**Key words:** Real-time qPCR; Reference gene; Gene expression; *UBA52*; *Asterias rubens*; High-throughput sequencing

## INTRODUCTION

Correct normalization of gene expression data is very important for analyzing any functional change in an organism. Differential gene expression can be observed during ontogenetic events, stress, and various pathological conditions (Hoogewijs et al., 2008; Krasnov et al., 2011; Senchenko et al., 2011; Oparina et al., 2012). Since disturbance of gene expression is caused by functional changes in different systems of an organism, identifying differentially expressed genes might be essential for understanding the mechanisms responsible for such changes.

Real-time quantitative polymerase chain reaction (qPCR) is the principal method for quantitative assessment of gene expression, which is also commonly applied for the clarification of data obtained by microarrays and high-throughput sequencing. The main prerequisites of reliable and reproducible results include standardization of all the stages of research, assessment of RNA quality, selection of primers and probes, normalization of data, calculation of amplification efficiency, and processing of statistical data. Data normalization allows the comparison of gene expression levels in samples with unknown mRNA/cDNA concentration. Normalization might be performed with respect to the number of cells, mass or volume of the sample/solution, concentration of total or ribosomal RNA, and mRNA level of the genes characterized by stable expression in the specimens studied - the so-called endogenous control genes, or reference genes.

Normalization by the mRNA level of reference genes is most commonly used when studying gene expression in different tissues under normal and experimental conditions. The common method of relative quantification ( $\Delta\Delta Ct$ ) is based on the comparison of the levels of the gene of interest and the reference gene in normal tissues and tissues in some pathological state or subjected to an experimental influence (Livak and Schmittgen, 2001; Pfaffl, 2001). In the case of several reference genes, the geometric mean of the values obtained is used (Vandesompele et al., 2002).

No universal reference genes showing a stable level of expression in all tissues have yet been found (Gene Quantification, 2013). Such “standard” reference genes such as *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase), *ACTB* (beta-actin), *RN18S1*, and *RN28S1* (ribosomal RNA genes), *RPLP0* (a ribosomal protein), *MHC I* (major histocompatibility complex I), *B2M* (beta-2 microglobulin), and *GUSB* (glucuronidase) have been previously used to study human and animal tissues (Huggett et al., 2005; Hoogewijs et al., 2008). However, according to many studies, even the “traditional” housekeeping genes might be differentially expressed in some cases (Jung et al., 2007).

When selecting reference genes, they should be tested for expression stability in the particular tissue of the species in question, under the particular experimental conditions. As a rule, the use of 2 or more reference genes increases preciseness and reproducibility of the results and represents a *de facto* standard for the assessment of changes in mRNA levels (Gene Quantification, 2013). The best reference genes can presently be determined using several computer programs: GeNorm, qBase, BestKeeper, and Normfinder, which are capable of selecting the most stable genes among a considerable number of candidate genes whose mRNA levels should be known in all the specimens studied. However, these programs are designed not for identifying new reference genes but merely for selecting among those already known.

Very few studies have focused on mRNA quantification of various genes in most animals, especially invertebrates. One of the main reasons for this is the absence of data on expression stability even for the principal reference genes that are routinely used in human studies. According to the published data, researchers investigating invertebrates and protists typically use only one of the traditional reference genes (*GAPDH*, *ACTB*, *RN18S1*) without additionally testing the validity of this gene for the species studied and the particular research goals (Dondero et al., 2004). Studies focusing on the identification of reference genes for these organisms have been published only recently, whereas those investigating their inducibility are almost completely lacking (Bagnall and Kotze, 2010). Thus, identifying new reference genes with stable expression and testing the traditional ones using different animal tissues from representatives of different taxa under different conditions is now a critical task.

There is almost no data on the molecular biology of echinoderms, even though this taxon is a common component of marine ecosystems. Among echinoderms, there are mass species inhabiting the littoral and sublittoral zones. The starfish *Asterias rubens* was selected as the object of our research. This choice was determined for many reasons: the species inhabits shallow areas of the Barents, North, White Seas, and the western Baltic Sea; it is an abundant and ecologically significant species; the starfish is a convenient model object for histological and cytological studies since it is large enough to be manually dissected, can be easily maintained in an aquarium, and remains viable after experiments.

This study aimed to identify reference genes needed for the normalization of quantitative data on gene expression in the starfish *A. rubens*, a common echinoderm species inhabiting the northern seas.

## MATERIAL AND METHODS

### Tissue sampling

Sampling of *A. rubens* was carried out in September 2009 and June 2012, at the White Sea Biological Station of St. Petersburg State University (Srednii Island, Kandalaksha Bay of the White Sea). The specimens were collected in the sublittoral zone and kept in seawater aquaria with salinity equal to that of their original habitat at 10°C. The tissues were sampled immediately after collection. Some specimens were subjected to experimental stress in order to induce gene expression. The pyloric cecal tissues were placed in RNAlater Solution (Ambion, Life Technologies, USA), transported to the laboratory within 2 days at room temperature, and then stored at -20°C. In 2009, tissue samples were obtained from 50 adults and 30 underyearlings (the offspring of that year) and maintained under normal conditions. In 2012, samples were obtained from 50 adults maintained under normal conditions and 30 adults exposed to different temperatures for 30 min.

### Temperature stress experiments

During the temperature stress experiments, the organisms were placed in aerated aquaria with different water temperatures (25°, 27°, and 29°C) for 30 min. The water salinity was 22-24 ppm. Each experimental group consisted of 10 specimens.

### RNA isolation and cDNA synthesis

Total RNA was isolated from homogenized tissue samples after storage in RNAlater Solution by using the RNeasy Mini Kit (Qiagen, Netherlands). The RNA concentration was determined using a NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies Inc., USA) and a Qubit® 2.0 fluorometer (Invitrogen, Life Technologies, USA). The  $A_{260}/A_{280}$  ratio of the RNA samples was 1.8-2.0. The integrity of the isolated RNA (RNA integrity number, RIN) was determined using the Bioanalyzer Agilent 2100 (Agilent Technologies, USA). Only the samples with an RIN value not less than 8.0 were used. Single-strand cDNA was synthesized using 1 µg of total RNA pretreated with DNase I (Fermentas, Lithuania), hexanucleotide primers, and M-MuLV reverse transcriptase (Fermentas, Lithuania) by the following scheme: 10 min at 25°C, 60 min at 42°C, 10 min at 50°C, and 10 min at 70°C.

### High-throughput sequencing

High-throughput transcriptome sequencing of 5 samples from the pyloric ceca of *A. rubens* was carried out on an Illumina GAIIx sequencer (Illumina, USA). The first and second samples were obtained from adult starfish under normal conditions; the third sample, after a 30-min exposure at a temperature of 25°C; the fourth, after 30-min exposure at 27°C; and the fifth, after 30-min exposure at 29°C. Paired-end reads (75 nucleotides long) were used. About 50 million reads were obtained for each sample. The reads were assembled into contigs by using the SOAPdenovo software. Differential expression was analyzed using the edgeR software package.

## Real-time PCR

Real-time PCR was carried out on a LightCycler 480 (Roche, Switzerland) by using modified short 6-carboxyfluorescein (FAM)-labeled probes from the Universal Probe Library (UPL, Roche, Switzerland). Two pairs of primers were used and selected for different areas of the transcript. The primer and probe sequences are listed in Table 1. The primers were selected on the basis of transcript sequence of the gene *UBA52* determined using the high-throughput sequencing method. Set No. 1 contained specific primers not binding to vertebrate cDNA to prevent contamination. Set No. 2 was used to confirm the results obtained. Each reaction was run 3 times with 20  $\mu$ L mix, as described earlier (Kudriavtseva et al., 2009). The following parameters were determined: the median  $C_q$  value (the 50th percentile), i.e., the value dividing the distribution so that 50% of the values are below and 50% are above it; the range containing 50% of all the values (between the 25th and 75th percentiles); and also the maximum and minimum  $C_q$  values. The amplification efficiency values are shown in Table 1.

**Table 1.** Primers and probes used for assessment of *UBA52* gene expression in the pyloric cecal tissues of *Asterias rubens* by the qPCR method, and the PCR efficiency values.

No.	Sequences of the forward (F) and reverse primers (R) and the probe (P), 5'→3'	Amplification efficiency (E, %)
1	F: GCAGATCTTTGTGAAGACCCTAA R: TGGTATCTGAAGGTTTCGACCTAA P: CCACCACC (No. 77)*	95 ± 7%
2	F: CCCCTGATCAGCAGAGA R: TGGATGTTGTAGTCTGACAGTGTG P: GGAGGATG (No. 88)*	93 ± 9%

\*UPL reference numbers (Universal Probe Library, Roche, Switzerland) are given in parentheses.

## Analysis of qPCR products

The PCR products were analyzed by electrophoresis on 1.8% agarose TBE gel. The nucleotide composition of amplification products was confirmed by sequencing on an 3730 DNA Analyzer (Applied Biosystems, USA) by using DYEnamic ET Terminator Cyclor Sequencing Kit (GE Healthcare, UK). The primers and probes proved to be specific; the size and sequences of amplification products were as expected.

## RESULTS

### Identification of the candidate reference genes

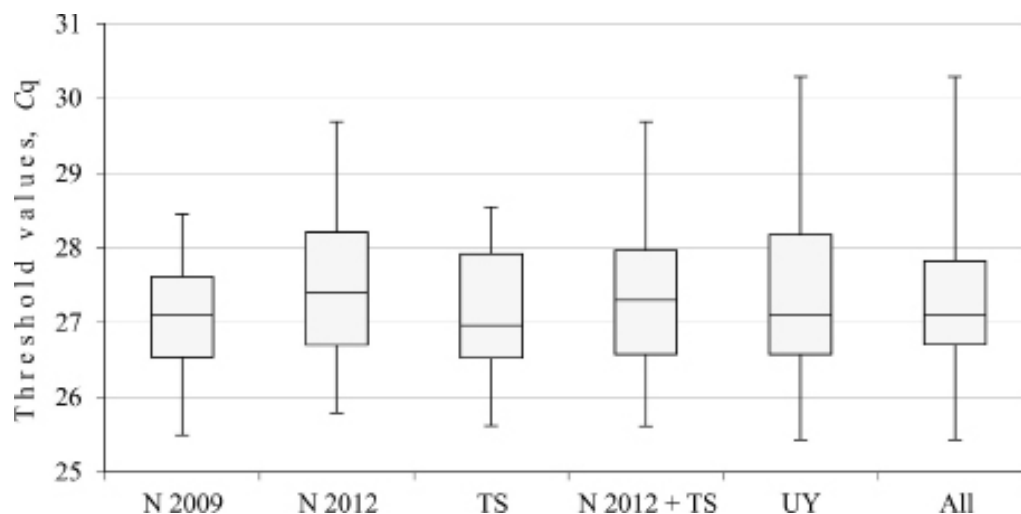
The housekeeping genes that could potentially be used for the normalization of qPCR data for *A. rubens* tissues were selected on the basis of the available publications. A preliminary analysis of expression of these genes in the pyloric cecal tissues was performed using the high-throughput transcriptome sequencing method on the Illumina platform. RNA was isolated from 5 specimens of *A. rubens*, of which 2 were intact (control) and 3 had been subjected to different temperature stresses. One of the most stably expressed genes was *UBA52*, coding for ubiquitin-60S ribosomal L40 fusion protein. The preliminary data were then verified by real-time PCR.

## Real-time PCR

The *de novo* sequencing of the *A. rubens* transcriptome allowed the selection of primers and probes specific to the *UBA52* gene. Therefore, performing real-time PCR was possible using fluorescent-labeled oligonucleotides instead of intercalating dyes, which yielded much more precise and reliable data. Since the standard (long) TagMan™ probe could not be selected, the unique Roche 8-base hydrolysis probes were used. Thus, normalized cDNA samples were used to estimate the cycle threshold value by real-time qPCR for different groups of experimental and control samples. Variability of expression of the *UBA52* gene was estimated in adult starfish under normal conditions, in underyearlings, and in adults subjected to different temperatures.

## Expression of the *UBA52* gene

Variability of the *UBA52* mRNA level was estimated using real-time qPCR. The  $C_q$  values, reflecting the initial concentration or the transcript copy number with regard to the amplification efficiency, are shown in Figure 1. The mean  $C_q$  values of the *UBA52* gene were considerably similar, with 50% of the values (between the 25th and 75th percentiles) within a narrow range from 26.5 to 28.2. The median values for the experimental groups varied from 27.0 to 27.4. The expression level after stress was insignificantly higher than that in the norm. The total variation for all the samples was not greater than 1.1 cycles. Even though this value allowed the *UBA52* gene to be used as endogenous control, the best estimates of expression of the target genes might be obtained using several different reference genes.



**Figure 1.** Threshold values ( $C_q$ ) for two sets of primers and probes to the *UBA52* gene of *Asterias rubens*. Rectangles correspond to the ranges containing 50% of the values (between the 25th and 75th percentiles); the horizontal line inside the rectangle is the median value (the 50th percentile); the bars are the maximum and minimum  $C_q$  values. N = samples from adults kept under normal conditions, TS = from adults exposed to different temperatures, UY = samples from underyearlings.



## DISCUSSION

Recent studies have shown that the choice of reference genes for quantitative data normalization remains a pressing problem, especially for invertebrates. This problem has to be addressed even in human research, in particular for studying various pathological states such as cancer in various tissues (Krasnov et al., 2011; Gene Quantification, 2013). No universal reference gene, which might be used for any species or best suited for a particular tissue or pathology, has yet been found. In previous studies, the use of a single “standard” reference gene was considered to be sufficient for adequate assessment of changes in target gene expression. However, this approach had to be changed since real-time qPCR, microarray hybridization, and computational methods were introduced. None of the traditionally used reference genes were found to ensure correct analysis of gene expression during ontogenesis under the conditions of stress, inflammation, tumor progression, and other pathologies. Thus, gene expression data should be correctly normalized using a set of several reference genes, including the most stable ones (Gene Quantification, 2013).

The studies focusing on suitable reference genes for invertebrates are scarce. Different reference genes have been proposed for different species (Hoogewijs et al., 2008; Ponton et al., 2011). Most of the studies involve mollusks, arthropods, and nematodes, whereas other taxa are only occasionally considered. The studies of gene expression in invertebrates often involve one of the traditional reference genes without testing its suitability, which may considerably affect the results.

The identification of reference genes in little-studied organisms, especially invertebrates, poses many difficulties. As a rule, the reference gene suitable for studying a particular pathological state in humans or some model organism is selected according to several criteria: 1) stability of its expression level, according to both expressed sequence tag and microarray data; 2) low frequency of mutations and abnormal splicing events; 3) localization in the chromosomal region not subjected to genomic transformations or methylation; 4) necessity that the gene should belong to the housekeeping group; and 5) the absence of similar functions with genes already accepted as “good” reference genes for the given tissue (Krasnov et al., 2011). Since high-throughput sequencing and microarray hybridization data as well as those on the frequency of mutations, genomic transformations, methylation, and abnormal splicing are practically absent for invertebrates, preliminary selection of candidate genes by bioinformatic methods is impossible. Therefore, our study mainly focused on the known housekeeping genes. Although the group of “housekeeping genes” has no strict definition, it conventionally includes genes showing similar mRNA levels in the majority of normal tissues, which implies the absence of distinct tissue specificity and possibly the presence of constitutive expression. Over 1500 housekeeping genes are presently known due to the extensive transcriptome data (Gene Quantification, 2013). The sets of relatively stably expressed genes vary depending on the species, tissue, and type of pathology. Whether some universal reference genes can be used for a wide range of organisms or at least for a certain phylogenetic group remains unknown. If identified, such genes might be used for semi-quantitative comparison of different transcriptomes. The available data on the specific structure of housekeeping genes, which differ from genes with tissue-specific expression in some characters besides the mRNA levels, suggest that such genes might indeed be found. In particular, they are conservative, affected by strong negative selection, and characterized by promoter regions lacking TATA-boxes. Other characters, in particular the specific structure of CpG islands near the transcription start site, were

considered in the literature (Ganapathi et al., 2005). In view of these results, the candidate reference genes can be selected by the comparative genomic methods. Due to evolutionary conservatism of many reference genes, the sets of such genes suitable for different taxa might be quite similar.

Ubiquitin is a small conservative protein. During post-translational modifications, one or several monomers of ubiquitin are bound by ubiquitin ligases to lateral amino groups of the target protein. This process, known as ubiquitination, affects the intracellular localization and functions of proteins. The first result of ubiquitination to be discovered was degradation of proteins tagged with multiple ubiquitin chains by the 26S proteasome. However, the ubiquitin system is also involved in such important processes such as cell proliferation, development, and differentiation; response to stress and pathogens; and DNA reparation. An additional function of ubiquitin includes the regulation of the transcriptional apparatus by modification of the RNA polymerase complex. In mammals, ubiquitin is encoded by 4 genes: the *UBA52* and *RPS27A* genes code for single copies of ubiquitin that fuse with the ribosomal proteins L40 and S27a, respectively, while the *UBB* and *UBC* genes code for precursors of polyubiquitin. The structure of ubiquitin in *A. rubens* has not been completely studied; however, one of the contigs obtained in our material was highly homologous to the mammalian *UBA52* gene.

All the results obtained in this studies - transcriptome sequencing of 5 samples from the pyloric ceca of the starfish *A. rubens*; selection of potential reference genes; and experimental assessment of mRNA level variability by qPCR in individuals of different ages collected in different years, both intact and subjected to temperature stress - suggest that the *UBA52* gene might be used as a reference for normalization of transcriptome data in this species. Additional research is necessary to determine the suitability of this gene for studying other echinoderm species and also other organs and tissue types.

## ACKNOWLEDGMENTS

The high-throughput sequencing and real-time PCR were performed at the EIMB RAS "Genome" Center. Research supported by the Russian Foundation for Basic Research (grant #10-04-01505-a) and the Ministry of Education and Science of the Russian Federation (state contract #16.552.11.7069).

## REFERENCES

- Bagnall NH and Kotze AC (2010). Evaluation of reference genes for real-time PCR quantification of gene expression in the Australian sheep blowfly, *Lucilia cuprina*. *Med. Vet. Entomol.* 24: 176-181.
- Dondero F, Cavaletto M, Ghezzi AR, La Terza A, et al. (2004). Biochemical characterization and quantitative gene expression analysis of the multi-stress inducible metallothionein from *Tetrahymena thermophila*. *Protist.* 155: 157-168.
- Ganapathi M, Srivastava P, Das Sutar SK, Kumar K, et al. (2005). Comparative analysis of chromatin landscape in regulatory regions of human housekeeping and tissue specific genes. *BMC Bioinformatics* 6: 126.
- Gene Quantification (2013). Available at [<http://www.gene-quantification.info>]. Accessed February 20, 2013.
- Hoogewijs D, Houthoofd K, Matthijssens F, Vandesompele J, et al. (2008). Selection and validation of a set of reliable reference genes for quantitative sod gene expression analysis in *C. elegans*. *BMC Mol. Biol.* 9: 9.
- Huggett J, Dheda K, Bustin S and Zumla A (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun.* 6: 279-284.
- Jung M, Ramankulov A, Roigas J, Johannsen M, et al. (2007). In search of suitable reference genes for gene expression studies of human renal cell carcinoma by real-time PCR. *BMC Mol. Biol.* 8: 47.
- Krasnov GS, Oparina NYu, Dmitriev AA, Kudriavtseva AV, et al. (2011). RPN1, a new reference gene for quantitative



- data normalization in lung and kidney cancer. *Mol Biol.* 45: 211-220.
- Kudriavtseva AV, Anedchenko EA, Oparina NYu, Krasnov GS, et al. (2009). Expression of FTL and FTH genes encoding ferretin subunits in lung and renal carcinomas. *Mol. Biol.* 43: 972-981.
- Livak KJ and Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)) method. *Methods* 25: 402-408.
- Oparina NYu, Sadritdinova AF, Snezhkina AV, Dmitriev AA, et al. (2012). Increase in NETO2 gene expression is a potential molecular genetic marker in renal and lung cancers. *Russ. J. Genet.* 48: 506-512.
- Pfaffl MW (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29: e45.
- Ponton F, Chapuis MP, Pernice M, Sword GA, et al. (2011). Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in *Drosophila melanogaster*. *J. Insect Physiol.* 57: 840-850.
- Senchenko VN, Krasnov GS, Dmitriev AA, Kudryavtseva AV, et al. (2011). Differential expression of CHL1 gene during development of major human cancers. *PLoS One* 6: e15612.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3: RESEARCH0034.