Validation of Zebrafish (*Danio rerio***) Reference Genes for Quantitative Real-time RT-PCR Normalization**

Rongying TANG¹, Andrew DODD^{1&}, Daniel LAI¹, Warren C. MCNABB², and Donald R. LOVE^{1*}

1 *School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand;* 2 *Food and Health Group, AgResearch Grasslands, Private Bag 11008, Tennant Drive, Palmerston North, New Zealand*

Abstract The normalization of quantitative real time RT-PCR (qRT-PCR) is important to obtain accurate gene expression data. The most common method for qRT-PCR normalization is to use reference, or housekeeping genes. However, there is emerging evidence that even reference genes can be regulated under different conditions. qRT-PCR has only recently been used in terms of zebrafish gene expression studies and there is no validated set of reference genes. This study characterizes the expression of nine possible reference genes during zebrafish embryonic development and in a zebrafish tissue panel. All nine reference genes exhibited variable expression. The β -*actin*, $EFI \alpha$ and $Rpl13\alpha$ genes comprise a validated reference gene panel for zebrafish developmental time course studies, and the $EFI\alpha$, $Rp113\alpha$ and 18S rRNA genes are more suitable as a reference gene panel for zebrafish tissue analysis. Importantly, the zebrafish *GAPDH* gene appears unsuitable as reference gene for both types of studies.

Key words zebrafish; quantitative real-time RT-PCR; housekeeping genes; *GAPDH* gene; GeNorm

The application of labeling fluorescence techniques and novel instrumentation has led to the development of quantitative real-time RT-PCR (qRT-PCR) methods that allow the real-time quantification of transcript levels[1]. Unlike traditional PCR, which detects amplification products at the end of the reaction, qRT-PCR allows amplification and detection to proceed simultaneously. It offers a rapid, automated method for the detection of multiple transcript levels with high sensitivity, reproducibility and a broad dynamic range.

A common method for the normalization of qRT-PCR data is the simultaneous amplification of an endogenous reference, or a housekeeping gene [1,2]. Ideally, this reference gene should be expressed at the same level in all samples, for example, samples from different tissues, during all developmental stages, and before and after

This work was supported by the grants from the University of Auckland Research Grants Committee, the Maurice and Phyllis Paykel Trust and the Lottery Grants Board of New Zealand

experimental manipulation[1]. However, there is emerging evidence that common housekeeping genes can significantly vary in expression over different conditions. For example, the transcript level of β-actin appears to vary widely in response to experimental treatments, and *GAPDH* gene expression also varies during development [1,3]. As a consequence, it has been suggested that a set of reference genes should be used for normalization and the genes comprising the set should be validated for each type of experiment [1,2].

qRT-PCR has only recently been used for zebrafish gene expression studies. A PubMed search of "real time PCR" and "zebrafish" showed there are 64 papers published in the period from 2001 to 2006 on gene expression analysis in zebrafish using qRT-PCR. Of the reference genes described in these publications, the β*-actin* and *GAPDH* genes were the most common. However, previous studies have suggested that caution should be exercised when using these two reference genes without validation [1]. To date, there has been no validated set of reference genes for qRT-PCR described in the zebrafish.

The aim of the study presented here was to evaluate a

Received: December 13, 2006 Accepted: March 5, 2007

[&]amp;Present address: Alan Wilson Center for Molecular Ecology and Evolution, Institute of Molecular Biosciences, Massey University, Private Bag 102 904, North Shore Mail Center, Auckland, New Zealand

^{*}Corresponding author: Tel, 64-9-3737599-87228; Fax, 64-9- 3737417; E-mail, d.love@auckland.ac.nz DOI: 10.1111/j.1745-7270.2007.00283.x

set of reference genes for the normalization of qRT-PCR data in zebrafish. Candidate reference genes were tested for their expression stability during embryonic development and in tissue samples from adult zebrafish.

Materials and Methods

Zebrafish

Wild type zebrafish (*Danio rerio*) were purchased from a commercial supplier (Hollywood Fish Farms, Auckland, New Zealand) and were maintained in a dedicated zebrafish facility. The zebrafish facility was maintained on light control of 14 h of light in a day at 26−27 ºC. Adult zebrafish were kept in 2.75 L tanks on a water recirculation rack system with a male to female ratio of 1:2. Adult fish were fed a range of dry fish food and artemia. Adult male and female fish were separated for one week prior to breeding. Embryos were harvested by breeding four males and three females.

RNA extraction

Ten developmental stages were selected for the zebrafish embryonic developmental time course study: sphere (4 hpf), germ ring (5.7 hpf), 75% epiboly (8 hpf), bud (10 hpf), 3-somite (11 hpf), 6-somite (12 hpf), 10-somite (14 hpf), 18-somite (18 hpf), prim-16 (31 hpf) and protruding mouth (72 hpf). Zebrafish embryos were derived from a single spawning and 20 zebrafish embryos were pooled for RNA extraction at each developmental timepoint. A zebrafish tissue panel was constructed by extracting RNA from the eye, heart, liver, intestine, muscle, skin and ovaries of adult zebrafish. The time course and tissue panel studies were carried out in duplicate. Zebrafish embryos or dissected zebrafish tissues were quickly frozen in liquid nitrogen, followed by thorough homogenization in 1 ml Trizol reagent (Invitrogen, Carlsbad, USA) using a homogenizer (Pro Scientific, Model PRO200 with 5 mm×75 mm flat bottom generator) at maximum speed. Chloroform (250 μl) was added to homogenized embryo/tissue followed by vortexing for 15 s and incubating at room temperature for 3 min. The samples were then centrifuged at 12,000 *g* for 5 min. The upper aqueous phase containing RNA was carefully transferred to a new tube without disturbing the interface. RNA was precipitated by the addition of an equal volume of 70% ethanol and loaded onto a spin column from an RNeasy mini kit (Qiagen, Valencia, USA) according to manufacturer's instructions.

Analysis of RNA concentration and quality

RNA concentrations were determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, USA); each RNA sample was assayed three times and an average value determined. The quality of RNA samples was determined using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, USA) in accordance with the manufacturer's instructions. The Bioanalyzer provides an RNA integrity number (RIN), with 0 corresponding to fully degraded RNA and 10 corresponding to intact RNA. For all qRT-PCRs, only RNA samples with RIN of at least 7.5 were used, with the vast majority of samples having a RIN of at least 8.0. These values satisfy one of the requirements of a satisfactory qRT-PCR experiment [4].

First strand cDNA synthesis

Total RNA (1 μg) was reverse transcribed to produce cDNA using Superscript III reverse transcriptase (Invitrogen) primed with random hexamers essentially as described previously [5]. In all cases, a reverse transcriptase negative control was used for testing genomic DNA contamination.

Primer design

Initially, thirteen candidate reference genes were selected from the literature for real time PCR [1−3,6−11]. To minimize the effects of gene co-regulation, the reference genes were selected from as many different biological pathways as possible (**Table 1**). For the 18S rRNA, a generic Taqman probe supplied by Applied Biosystems was used. For the remaining twelve reference genes, zebrafish orthologues of mammalian gene transcripts were retrieved from the NCBI (http://www.ncbi.nlm.nih.gov/Genbank) and Ensembl databases (http://www.ensembl.org/ Danio rerio). Exons for each zebrafish gene were identified by a BLAST search of the zebrafish genome DNA sequence database (http://www.ensembl.org/Danio_rerio) using the mammalian sequence data. Peptide alignments of human and zebrafish reference genes were performed to confirm the reading frame of each exon and splice sites were manually annotated. To optimize the design of primers across exon boundaries, primers were initially manually designed for each reference gene (**Table 1**), and these were subsequently analyzed using the primer analysis software NetPrimer (http://www.PremierBiosoft.com/). NetPrimer analyzes primer quality by scanning the primer sequence for the formation of secondary structures, and predicts annealing temperatures. Except for the β*-actin*

©Institute of Biochemistry and Cell Biology, SIBS, CAS

and *HPRT* genes, all primer pairs had at least one primer crossing an exon-exon boundary. The primers were designed to have similar melting temperatures and to give similar amplicon sizes. Amplified products were electrophoresed in 1% agarose gels and a single fragment was obtained for all but the *TBP*, *tubulin* and *YWHAZ* amplicons; these latter three genes were excluded from further analysis. The apparently unique amplicons exhibited a single peak by dissociation curve analysis using an Applied Biosystems 7900 HT real-time PCR platform. Later experiments showed that primers for the *PBGD* gene had a low PCR efficiency, so this gene was also excluded from further analysis.

Quantitative real-time RT-PCR

For all reference genes except 18S rRNA, qRT-PCRs were performed using SYBR green. Standard reactions (10 μl) were assembled using an Eppendorf epMotion 5075 liquid handling robot as follows: 5 μl of Platinum SYBR green qPCR supermix-UDG with Rox (Invitrogen), 0.2 μl of forward primer (10 μM), 0.2 μl of reverse primer (10 μM), 2 μl of template and 2.6 μl DEPC water. Templates were 1:10 diluted cDNA samples, and in the case of negative controls, cDNAs were replaced by DEPC water. All real time assays were carried out in triplicate using an Applied Biosystems 7900 HT real-time PCR platform. Forty amplification cycles were performed, with each cycle consisting of 94 ºC for 15 s followed by 59 ºC for 1 min. Amplification and dissociation curves generated by the SDS 2.2 software were used for gene expression analysis.

In the case of the 18S rRNA Taqman probe, 10 μl of standard reactions comprised 1 μl of probe, 6 μl of platinum qPCR mastermix, 1 μl of ROX dye and 2 μl of cDNA template. Cycling conditions were the same as stated above. In all cases, a reverse transcriptase negative control was included.

Statistical analysis

Ct values were obtained for each reference gene. Following the removal of outliers, raw fluorescence data were exported to the programme LinReg to determine the PCR amplification efficiency [12]. All amplifications had a PCR efficiency value of approximately 1.9; PCR efficiency values close to 2 have been taken to suggest efficient amplification [13].

The PCR efficiency of each primer pair, together with Ct values, was used to calculate a relative gene expression value for each transcript, according to the equation E ^ΔCt (Min Ct−Ct sample), where E refers to PCR efficiency, Min Ct is the lowest Ct value for the relevant primer pair, and Ct sample is the Ct value for each amplification. A standard error for each relative gene expression value was calculated as a measure of data variation**.**

The relative stability of the nine reference genes was calculated using GeNorm[2]. This program calculates a gene expression stability measure (M) for each reference gene and the reference genes were ranked in the order of increasing expression stability in an output file.

Results

Expression of reference genes during zebrafish embryonic development

The raw expression levels for the nine reference genes that were assessed during zebrafish embryonic development showed that the genes fell into three categories based on transcript abundance: (1). High transcript abundance (average Ct value below 20): β*-actin*, *EF1*α, *Rpl13*α and 18S rRNA gene; (2). Median transcript abundance (average Ct value 20−25): *HPRT*, *RNAP* and *SDHA*; and (3). Low transcript abundance (average Ct value 25−30): β*2mic* and *GAPDH* (**Table 2**).

The relative expression levels of the nine reference genes for the developmental time course study were entered into GeNorm to calculate gene expression stability values, designated M (**Fig. 1**). The GeNorm package relies on the principle that the expression ratio of two ideal housekeeping genes should be identical in all samples, regardless of cell type or condition. As a consequence, any variation in the expression ratio of two housekeeping genes would suggest that one or both of the housekeeping genes is (are) not stably expressed; increasing variation correlates with decreasing expression stability. The value M represents an average pairwise variation of a test gene with all other tested genes and low M values correspond to stable gene expression.

Table 2 Average Ct values for nine reference genes during the zebrafish developmental timecourse

Reference gene	B-actin	B2mic	$EFI\alpha$	GAPDH	HPRT	<i>RNAP</i>	$Rpl13\alpha$	<i>SDHA</i>	18S rRNA
Average Ct value	17.41	26.34	16.19	27.79	22.44	24.97	19.35	23.49	17.47

Fig. 1 GeNorm output of zebrafish developmental timecourse M value (y-axis) is defined as a measure of gene expression stability, with an increasing M value correlating with less stability. The least stable genes are displayed on the left and the most stable genes are displayed on the right.

As shown in **Fig. 1**, the reference genes with the lowest M values, and thus highest expression stability, were the β -*actin*, *Rpl13* α and *EF1* α genes. A normalization factor (NF) was calculated as the geometric mean of the relative expression levels of these three genes. To test for consistency of the normalization factor, a comparison of relative expression levels of the *RNAP* and *GAPDH* genes was undertaken (**Fig. 2**).

The normalized expression data show that *RNAP* gene expression is high during early development and is at its highest at the sphere stage (4 hpf). This conclusion is consistent with developmental processes occurring at this stage. During zebrafish embryonic development, the midblastula transition (MBT) occurs at cell cycle 10 (about 2−3 hpf), which marks the initiation of transcription of zygotic genes. The work of Kane and Kimmel [14] reported a burst of RNA synthesis activity in zebrafish embryos at the MBT, which increases progressively for several cycles. This observation may account for the high level of *RNAP* gene expression at the sphere stage (cell cycle 13). Unnormalized data, however, did not show the high level of RNAP gene expression at the early stages of zebrafish embryonic development.

The GeNorm package rated the *GAPDH* gene as the gene with the lowest expression stability, which is evident in unnormalized and normalized data (**Fig. 2**). Almost insignificant *GAPDH* gene expression occurs during the early zebrafish developmental stages including the blastula, gastrula and segmentation periods, but an increase in

Fig. 2 Comparison of relative expression of the zebrafish *RNAP* **and** *GAPDH* **genes during development**

Raw expression levels and normalized levels are shown for the *RNAP* and *GAPDH* genes in panels A and B, respectively. Developmental stages: S, sphere; G, Germ ring; 75% epi, 75% epiboly; B, bud; 3-S, 3 somite; 6-S, 6 somite; 10-S, 10 somite; 18-S, 18 somite; P16, prim 16; PM, protruding mouth. Error bars represent standard errors.

expression occurs during the pharyngula period, with a sharp increase in the hatching period. These data are consistent with the study of Rauch *et al*. [15], who performed whole mount *in situ* hybridzation of early zebrafish embryos and showed no detectable *GAPDH* gene expression until the prim 5 stage, which occurs in the pharyngula period. A separate study also showed no detectable *GAPDH* gene expression at early stages of zebrafish embryonic development [16]. Taken together, there is strong evidence that the *GAPDH* gene is differentially regulated during zebrafish embryogenesis.

Expression of reference genes in zebrafish tissue panel

Raw expression levels for the nine reference genes across a panel of seven zebrafish tissues are shown in **Table 3**. The relative gene expression data were submitted to GeNorm and gene stability values were calculated for each reference gene (**Fig. 3**). GeNorm analysis identified *Rpl13* α *, EF1* α and 18S rRNA genes as the most stable reference genes across the tested tissues.

Fig. 3 GeNorm output of zebrafish tissue panel study M value (y-axis) is defined as a measure of gene expression stability, with an increasing M value correlating with less stability. The least stable genes are displayed on the left and the most stable genes are displayed on the right.

Discussion

The study described here shows that the β*-actin*, *EF1*^α and $Rpl13\alpha$ genes provide a collection of validated reference genes for a developmental timecourse study of zebrafish, but that the $EFI\alpha$, $Rpl13\alpha$ and 18S rRNA genes are more suitable for tissue analysis. Importantly, the *GAPDH* gene appears unsuitable as reference gene for both types of study in the zebrafish.

The $EFI \alpha$ and $Rpl13\alpha$ genes are common to both panels of validated genes. The expression products of these two genes are both involved in translation and hence it is not surprising that they are among the most stable reference genes. This finding is similar to the conclusions reported in three publications characterizing reference gene expression in salmon. The study by Jorgensen *et al*. [10] showed that accurate normalization of transcript data could be obtained by combining 18S rRNA, *EF1α* and the *RPL1* genes for studying gene expression in Atlantic salmon. Olsvik *et al.* [9] validated the *EF1AA* and *EF1AB* genes as suitable reference genes for transcript studies of tissues in the same fish species, and Ingerslev *et al.* [11] confirmed the $EFI\alpha$ gene as a suitable reference gene among three that were evaluated.

Previously, the 18S rRNA gene has been considered an ideal reference control for qRT-PCR analysis as the level of rRNA appears to vary considerably less than mRNA expression [1]. This conclusion is generally supported by this study, as the raw expression data of 18S rRNA during the developmental timecourse and tissue panel studies exhibited the least variation among the reference genes. However, when analyzed by GeNorm, the 18S rRNA gene ranks behind the $EFI \alpha$ and $Rpl13\alpha$ genes; a similar finding was reported in Atlantic salmon [9]. One of the major limitations of using the 18S rRNA gene as a reference control is that an imbalance of rRNA and mRNA fractions can occur between samples, which makes 18S rRNA less suitable as a normaliser in calculating relative mRNA levels [1−3]. Hence caution must be exercised when using 18S rRNA as a reference control. In preference, mRNA coding reference genes are generally used for data normalization. As an additional caveat, while the $EFI \alpha$ gene is a validated reference gene for zebrafish timecourse analysis, its use in studies that might invoke a stress response should be assessed in view of the up-regulation of this gene in stressed human cells [17].

To date, β-actin has been the most commonly used reference gene for normalizing qRT-PCR data in zebrafish. Time-course analysis has shown that the β*-actin* gene is one of the three most suitable reference genes, which is compatible with its fundamental role in cell motility, which occurs throughout embryogenesis. In contrast, expression of the gene is not stable in terms of a tissue panel analysis. Previous publications characterizing tissuespecific β*-actin* gene expression have shown that it is differentially expressed in post-mortem brain specimens,

and the evaluation of human heart tissue has provided evidence against the use of this gene as a reference gene [18,19]. The papers published on studies of Atlantic salmon also showed that β*-actin* gene expression varied greatly between tissue samples [9−11].

Finally, the *GAPDH* gene is the second most commonly used reference gene in zebrafish qRT-PCR analysis. However, in terms of both zebrafish development and tissue panel analysis, *GAPDH* gene expression was highly variable. Similar results were obtained from studies of Atlantic salmon, which showed that the *GAPDH* gene exhibits significant variation in expression levels among tissue samples [9,10]. Together, our data describe a validated set of reference genes for zebrafish qRT-PCR experiments under a limited but practical set of experimental conditions.

References

- 1 Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 2000, 25: 169−193
- 2 Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002, 3: research0034. 1−research0034.11
- 3 Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, Grisar T *et al*. Housekeeping genes as internal standards: Use and limits. J Biotechnol 1999, 75: 291−295
- 4 Fleige S, Walf V, Huch S, Prgomet C, Sehm J, Pfaffl MW. Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative realtime RT-PCR. Biotechnol Lett 2006, 28: 1601−1613
- 5 Dodd A, Chambers SP, Love DR. Short interfering RNA-mediated gene targeting in the zebrafish. FEBS Lett 2004, 561: 89−93
- 6 Frost P, Nilsen F. Validation of reference genes for transcription profiling in the salmon louse, *Lepeophtheirus salmonis*, by quantitative real-time PCR. Vet Parasitol 2003, 118: 169−174
- 7 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
- 8 Radonic A, Thulke S, Bae HG, Muller MA, Siegert W, Nitsche A. Reference gene selection for quantitative real-time PCR analysis in virus infected cells: SARS corona virus, Yellow fever virus, Human Herpesvirus-6, Camelpox virus and Cytomegalovirus infections. Virol J 2005, 2: 7
- 9 Olsvik PA, Lie KK, Jordal AE, Nilsen TO, Hordvik I. Evaluation of potential reference genes in real-time RT-PCR studies of Atlantic salmon. BMC Mol Biol 2005, 6: 21
- 10 Jorgensen S, Kleveland E, Grimholt U, Gjoen T. Validation of reference genes for real-time polymerase chain reaction studies in Atlantic salmon. Mar Biotechnol 2006, 8: 398−408
- 11 Ingerslev HC, Pettersen EF, Jakobsen RA, Petersen CB, Wergeland HI. Expression profiling and validation of reference gene candidates in immune relevant tissues and cells from Atlantic salmon (*Salmo salar* L.). Mol Immunol 2006, 43: 1194− 1201
- 12 Ramakers C, Ruijter JM, Deprez RH, Moorman AF. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 2003, 339: 62−66
- 13 Skillman LC, Toovey AF, Williams AJ, Wright AD. Development and validation of a real-time PCR method to quantify rumen protozoa and examination of variability between entodinium populations in sheep offered a hay-based diet. Appl Environ Microbiol 2006, 72: 200−206
- 14 Kane DA, Kimmel CB. The zebrafish midblastula transition. Development 1993, 119: 447−456
- 15 Rauch GJ, Lyons DA, Middendorf I, Friedlander B, Arana N, Reyes T, Talbot WS. Submission and curation of gene expression data. ZFIN Direct Data Submission 2003
- 16 Thisse B, Pflumio S, Fthauer M, Loppin B, Heyer V, Degrave A, Woehl R *et al*. Expression of the zebrafish genome during embryogenesis (NIH R01 RR15402). ZFIN Direct Data Submission 2001
- 17 Talapatra S, Wagner JD, Thompson CB. Elongation factor-1 alpha is a selective regulator of growth factor withdrawal and ER stress-induced apoptosis. Cell Death Differ 2002, 9: 856−861
- 18 Gutala RV, Reddy PH. The use of real-time PCR analysis in a gene expression study of Alzheimer's disease post-mortem brains. J Neurosci Methods 2004, 132: 101−107
- 19 Yperman J, De Visscher G, Holvoet P, Flameng W. Beta-actin can not be used as a control for gene expression in ovine interstitial cells derived from heart valves. J Heart Valve Dis 2004, 13: 848−853

Edited by **Rongjia ZHOU**