Selection of proper reference genes for the cyanobacterium Synechococcus PCC 7002 using real-time quantitative PCR

Edina Szekeres¹, Cosmin Sicora², Nicolae Dragoș¹,³ & Bogdan Druga¹,⁴

¹Institute of Biological Research, Cluj-Napoca, Romania; ²Center of Biological Research, Jibou, Romania; ³Faculty of Biology and Geology, Babe - Bolyai University, Cluj-Napoca, Romania; and ⁴Swiss Federal Institute of Aquatic Science and Technology (Eawag), Dübendorf, Switzerland

Correspondence: Bogdan Druga, Institute of Biological Research, Cluj-Napoca, Romania.
Tel.: +41(0)787669773; e-mail: bogdan.druga@icbcluj.ro

Received 14 July 2014; revised 7 August 2014; accepted 7 August 2014. Final version published online 8 September 2014.

DOI: 10.1111/1574-6968.12574

Editor: David Studholme

Keywords
Synechococcus sp. PCC 7002; real-time quantitative PCR; reference genes; microaerobiosis; UV-B irradiation.

Abstract
Synechococcus sp. PCC 7002 is known to be tolerant to most of the environmental factors in natural habitats of Cyanobacteria. Gene expression can be easily studied in this cyanobacterium, as its complete genome sequence is available. These properties make Synechococcus sp. PCC 7002 an appropriate model organism for biotechnological applications. To study the gene expression in Cyanobacteria, real-time quantitative PCR (qPCR) can be used, but as this is a highly sensitive method, data standardization is indicated between samples. The most commonly used strategy is normalization against internal reference genes. Synechococcus sp. PCC 7002 has not yet been evaluated for the best reference genes. In this work, six candidate genes were analyzed for this purpose. Cyanobacterial cultures were exposed to several stress conditions, and three different algorithms were used for ranking the reference genes: geNorm, NormFinder, and BestKeeper. Moreover, gene expression stability value M and single-control normalization error E were calculated. Our data provided a list of reference genes that can be used in qPCR experiments in Synechococcus sp. PCC 7002.

Introduction
Synechococcus sp. PCC 7002 is a unicellular, photoheterotrophic, and euryhaline cyanobacterium, which can tolerate various light intensities, exposure to oxidative stress, changes in nutrient supply, temperature, and salinity, making it an ideal system for experiments that would be difficult to be accomplished in less robust Cyanobacteria (Rippka et al., 1979; Nomura et al., 2006a, b; Zhu et al., 2010). This strain has the highest growth rate among all Cyanobacteria investigated so far, with a doubling time of 2.6 h under certain conditions (Frigaard et al., 2004; Ludwig & Bryant, 2012). The complete genome of Synechococcus sp. PCC 7002 is available in GenBank – NCBI (accession number CP000951) and encodes around 3200 proteins; transcripts were identified for most of the open reading frames. It is naturally transformable, and a system for genetic complementation and gene overexpression is available, making it a perfect candidate for studies in biotechnological or industrial application, being also a model organism for studies of cyanobacterial metabolism (Ludwig & Bryant, 2011; Xu et al., 2011). To study the global expression patterns in Cyanobacteria, microarrays have been extensively used (Postier et al., 2003; Foster et al., 2007; Ostrowski et al., 2010). However, if the expression pattern of a specific set of genes is desired, the use of real-time quantitative PCR (qPCR) is more suitable, as it also involves high accuracy, specificity, and reproducibility at a lower cost (Ludwig & Bryant, 2011). qPCR is a widely used, sensitive, and reproducible technique for gene expression measurements. This fluorescence-based method is used in molecular biology for quantification of transcripts expressed at low levels in many different samples for a various number of genes (Silver et al., 2006; Lee et al., 2007; Bustin et al., 2010; Zhang et al., 2013).

In qPCR, many factors can affect the results (technical and biological variation), thus standardization is needed to reveal significant changes in mRNA levels (Nolan et al., 2006; Derveaux et al., 2010). To enable the comparison of transcript concentrations among different samples, data normalization is required. The most commonly used method is normalizing to an endogenous reference gene, because both gene of interest and reference gene are measured during the same PCR reaction. This involves
reporting the transcript ratios of the target gene to those of the reference gene (Thellin et al., 1999).

An ideal reference gene is essential for cell function and shows relatively constant expression between samples, experimental conditions, or time points (Bustin et al., 2009; Guénin et al., 2009). Gene pairs that have stable expression patterns relative to each other are considered as suitable control genes (Vandesompele et al., 2002). The use of inappropriate internal reference genes can lead to biased expression profiles especially when a single gene is used for normalization (Pinto et al., 2012). Statistical models with different normalization algorithms are available to identify genes with minimal variance: geNorm, NormFinder, or BestKeeper (Vandesompele et al., 2002; Andersen et al., 2004; Pfaffl 2004; Pfaffl et al., 2004).

In this work, we have evaluated the expression stability of six typical candidate reference genes in Synechococcus sp. PCC 7002 by qPCR under three stress treatments which are often used in experiments: microaerobiosis, incubation in darkness, and UV-B stress. Our results provide a valuable starting point for the selection of reference genes that can be used in qPCR experiments concerning Synechococcus sp. PCC 7002.

Materials and methods

Strains and culture conditions

Synechococcus sp. strain PCC 7002 was acquired from the Pasteur Culture Collection of Cyanobacteria (PCC) and grown at 38 °C under air bubbling in liquid medium A+ supplemented with vitamin B12 (final concentration of 0.04 mg L⁻¹) under constant cool-white fluorescent illumination at an intensity of 250 μmol of photons m⁻² s⁻¹ (Stevens et al., 1973).

For sample collection, four independent cultures were grown under three experimentally induced stress conditions, until they reached exponential phase (OD₆₀₀ nm = 0.7–0.9), microaerobiosis, incubation in darkness, and UV-B stress. The absorption spectra were measured using a Shimadzu UV-1700 spectrophotometer (Shimadzu Corporation, Kyoto, Japan).

Microaerobiosis was induced by bubbling the culture flasks with argon for up to 120 min. To assure the continuous dark condition, the cultures were placed into a Sanyo Versatile Environmental Test Chamber (Sanyo, Osaka, Japan) and grown in standard conditions for 120 min, but without any light source. For the UV-B treatment, cells were exposed for 120 min to 1.3 W m⁻² UV-B light provided by a Philips TL 20W/12 RS fluorescent lamp, with an emission spectrum of 275–380 nm, and a peak at 310–315 nm (Philips Lighting, Stockholm, Sweden). Culture samples (12 mL of cell suspension) were collected at 15, 30, 60, and 120 min after the stress conditions were initiated. Cells were rapidly chilled on ice and centrifuged (2 min, 12,000 g, 4 °C), and the pellets were suspended in 250 μL TRIzol reagent (Invitrogen, Carlsbad, CA) and stored at −20 °C until further use.

RNA extraction and cDNA synthesis

For total RNA extraction, the TRIzol reagent was used, according to the manufacturer’s instruction. RNA concentration was measured with a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA). To remove any trace of genomic DNA, samples were treated with 1 U DNase (Ambion Turbo DNase, Austin, TX). Synthesis of first strand cDNA was performed using a First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA) using 1.5 μg of purified RNA as template and random hexamer primers.

Primer design and transcription analysis by RT-qPCR

The sequences of the six candidate reference genes were obtained from CYANOBASE (accession numbers in Table 1), and primers were designed using the GENEFISHER tool from the BIBISERV server (Giegerich et al., 1996; Nakao et al., 2010). The length of the amplicons for the designed primers was set around 150 bp (Table 2). The size of the PCR products for each primer pair was verified on 1% agarose gel electrophoresis using 1x TAE buffer. qPCR was performed with a BioRad iQ5 System (BioRad, Berkeley, CA) using domed cap PCR tubes. For the reac-

<table>
<thead>
<tr>
<th>Table 1. Candidate reference genes tested in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene symbol</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>rimM</td>
</tr>
<tr>
<td>rpaA</td>
</tr>
<tr>
<td>perB</td>
</tr>
<tr>
<td>16S</td>
</tr>
<tr>
<td>ppc</td>
</tr>
<tr>
<td>secA</td>
</tr>
</tbody>
</table>
tions, SensiFast SYBR & Fluorescein Kit (Bioline, London, UK) was used, with the reaction mix containing both forward and reverse primers at a concentration of 400 nM each, 10 μL of SensiFAST SYBR & Fluorescein Mix to a final concentration of 1x, 4 μL of 5x diluted cDNA and H₂O added up to a final volume of 20 μL. The PCR profile contained: one cycle of 95 °C for 3 min, followed by 45 cycles of 95 °C for 15 s, 55 °C for 30 s, and 72 °C for 30 s. Negative controls (reactions without cDNA template) were included to detect the presence of any DNA contamination. All the reactions were performed in triplicate for each cDNA sample. Ct (crossing thresholds) values were exported and gene-stability measure (M value) and single-control normalization error (E value) were calculated according to Vandesompele et al. (2002). The equation model for E value calculation was extended, taking into account the three abiotic stress conditions that were used to verify expression stability of the candidate genes. Moreover, the gene stability was approached using three different algorithms: geNorm, NormFinder, and BestKeeper (Vandesompele et al., 2002; Andersen et al., 2004; Pfaffl et al., 2004).

Results and discussion

Choice of candidate reference genes

The candidate reference genes were selected based on previously reported genes that have been used as controls in qPCR and microarray studies in Synechococcus sp. strain PCC 7002 and other Cyanobacteria (Sakamoto et al., 1997; Engelbrecht et al., 1999; Hihara et al., 2001; Price et al., 2004; Balasubramanian et al., 2006; Nomura et al., 2006b; Woodger et al., 2007; Cumino et al., 2010; McNeely et al., 2011; Pinto et al., 2012). Six candidate reference genes belonging to independent metabolic pathways (to minimize the effect of co-regulation) were selected for further investigation: rimM (GI: 170077861, locus tag: SYNPRC7002_A1245), rnpA (GI: 169885279, locus tag: SYNPRC7002_A0989), petB (GI:169885133, locus tag: SYNPRC7002_A0842), 16S (GI: 169884305, locus tag: SYNPRC7002_A2788), ppc (GI: 169885692, locus tag: SYNPRC7002_A1414), and secA (GI: 169885544, locus tag: SYNPRC7002_A1259).

qPCR amplification specificity, efficiency and analysis of Ct values

A standard PCR amplification has confirmed that all six primer pairs target single DNA fragments for each gene. In the electrophoresis gel, single bands of the desired molecular weight were observed (Supporting Information, Fig. S1). qPCR was used to measure RNA transcription variations of the candidate genes, in all the samples. After 45 cycles of amplification, melting curves were analyzed. They displayed distinct peaks, suggesting specific melting temperature and unique amplicons (data not shown). Three replicates for each stress condition were examined, with abnormal plots being excluded from further data analysis. Control samples lacking cDNA did not produce amplification, indicating that the template was not contaminated.

Raw Ct values were extracted from BioRad iQ5 System and analyzed. Ct values varied between 10 and 37, the lowest value being observed in case of the microaerobic stress, and the highest in the samples irradiated with UV-B. The gene coding for 16S rRNA consistently displayed the lowest median Ct values (10–18), while the highest median value was obtained for the ppc gene, although it did not show a significant difference when compared to other median Ct values. The raw expression values of candidate genes across different experimental conditions were calculated using the comparative Ct method (2−ΔΔCt) (Livak & Schmittgen, 2001), and they can be observed in Table S1.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Primer name</th>
<th>Primers sequence 5′–3′</th>
<th>Tm (°C)</th>
<th>Amplicon length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rimM</td>
<td>7002_rimM_F</td>
<td>GATGCCGCCCCGAACTCGAAC</td>
<td>58</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>7002_rimM_R</td>
<td>TCTGTTGGCGATCGGTGACTTC</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>rnpA</td>
<td>7002_rnpA_F</td>
<td>GCCCAACAAATTGGCATCAG</td>
<td>58</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td>7002_rnpA_R</td>
<td>TACCGACAGCAATGACGATGTG</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>petB</td>
<td>7002_petB_F</td>
<td>GGGTGTTGATCGTGCTGCCAG</td>
<td>58</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>7002_petB_R</td>
<td>CGTGATCGTCGTTAGGATGAG</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>16S</td>
<td>7002_16S_F</td>
<td>CGGGTTTGAATGAGATCCGCTGC</td>
<td>58</td>
<td>157</td>
</tr>
<tr>
<td></td>
<td>7002_16S_R</td>
<td>AGTGGGCACTCTGAGGACAG</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>ppc</td>
<td>7002_ppc_F</td>
<td>CACCCCTGGCGGAATTTATGATAC</td>
<td>58</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>7002_ppc_R</td>
<td>CCACGTAACCTGACGACAG</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>secA</td>
<td>7002_secA_F</td>
<td>GCGGAAATGAGAACCAGGGAAG</td>
<td>58</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>7002_secA_R</td>
<td>GAAACGCTGTACCTGCCCCATC</td>
<td>58</td>
<td></td>
</tr>
</tbody>
</table>
Ranking and determination of optimal reference genes

For each stress condition the $C_t$ values were used to analyze gene expression stability. To minimize bias, three different algorithms were used for this purpose: geNorm (Vandesompele et al., 2002), NormFinder (Andersen et al., 2004), and BestKeeper (Pfaffl et al., 2004). They rank the candidate reference genes based on distinct statistical algorithms using stability value – $M$ value (geNorm and NormFinder); or Pearson’s correlation coefficient – $r$ value (BestKeeper). This resulted in some discrepancies between the rankings, but usually the same genes were identified as having the most stable expression under the tested conditions. Moreover, the $M$ and $E$ parameters (Vandesompele et al., 2002) were separately calculated to verify data generated by the three algorithms and also for an easier understanding and interpretation of the results.

The gene-stability parameter ($M$) is based on geometric averaging of multiple control genes and it is calculated through pairwise comparison and stepwise exclusion of least stable gene (Jain et al., 2006; Pinto et al., 2012). GeNorm determined a default limit $M = 1.5$, below which candidate reference genes can be classified as stably expressed, while lower value indicates an even greater stability of the gene expression (Vandesompele et al., 2002; Ohl et al., 2005; Han et al., 2012). $M$ values for most of the expression stability calculated were under the limit 1.5 attesting the possibility of using them for normalization. The only gene with the $M$ value above 1.5 was petB: 1.53 under dark incubation in one of the three replicates. The $M$ values varied, from 0.37 for secA gene in UV-B irradiation conditions to 1.53 for petB in dark incubation. The most constant values were observed for secA ($M$ value varying between 0.38 and 0.84) and ppC ($M$ value varying between 0.45 and 0.84). Under microaerobiosis, the rimM gene displayed the lowest $M$ value (0.45) (Fig. 1a). In dark incubation, the lowest $M$ value was 0.57 in case of ppC gene (Fig. 1b), while under irradiation with UV-B, this parameter was 0.45 for the secA gene (Fig. 1c).

Under microaerobiosis, the three algorithms ranked the rimM gene in the top position, suggesting that it could be used for normalization under this particular stress condition (Table 3). Under dark incubation, the ppC, rnpA, and secA were shown to be the most suitable reference genes, with ppC being ranked the first according to two of three algorithms (Table 3). For UV-B irradiation, the algorithms ranked secA, ppC, and rnpA as the most stable genes to be used for normalization (Table 3). For more detailed information concerning the results acquired with the three algorithms, see Fig. S2.

According to the results acquired by all three algorithms and to the $M$ values, the most suitable reference genes for *Synechococcus* sp. strain PCC 7002 gene expression stability under all the tested conditions are ppC, secA, and rnpA. Under microaerobiosis, secA and rimM genes are recommended to be used for RT-qPCR normalization. In dark incubation, the most stable reference genes are ppC and rnpA, while for UV-B treatment, secA, rnpA, and ppC are also stable (Table 3). To display small variance among conditions, the use of two reference genes is recommended. The $E$ value (single-control normalization error) was calculated for all the 15 pair combinations of the six analyzed genes, to determine the most suitable
reference gene combination. For two ideal control genes, the E parameter equals 1, although in practice it is generally larger than 1 (Vandesompele et al., 2002). E values were calculated for all two-by-two combinations of candidate genes (Table 4). According to these calculations, ppC + secA is the best choice as reference for normalizing gene expression data under all three stress treatments, as it had the best E values ranging from 1.4 to 1.5 (Table S2). rnpA + ppC and rnpA + secA were also found to be optimal combinations to be used in pairwise normalization with E values of 1.5–1.9. In general, combinations including 16S or petB genes were ranked the lowest, as the highest E values for every condition was calculated for rimM + petB, with values ranging between 2.7–8.6. This underlines our previous indications that 16S and petB genes and their combinations are not the best choice as reference genes in our own experiment.

As the experimental procedure included three stress conditions, we adapted the E equation which calculates the value of different gene combinations under two different conditions, for the calculation of pairwise variation under three conditions. The results were consistent, confirming the previous findings, the combinations with the best E values being, in this order, rnpA + ppC (1.74), ppC + secA (1.77), and rnpA + secA (2.13) (Table 4). The exact E values are presented in Table S2.

16S rRNA gene has been used in several qPCR studies for data normalization so far (Sakamoto et al., 1997; Engelbrecht et al., 1999; Hihara et al., 2001; Schafer et al., 2006; Pinto et al., 2012). However, our results showed that the expression rate of 16S rRNA gene is not stable, and it has significantly higher copy number than other genes. Therefore, it can be concluded that this gene is not the best choice to be used for normalization in experiments with Synechococcus sp. strain PCC 7002.

It should be noted that there are concerns with using rRNA genes as control as rRNA may not be influenced by the degradation machinery in a matter similar to mRNA. rRNA genes are not always a good internal control as the rRNA is not representative of the mRNA and there is a significant imbalance between the rRNA and mRNA fractions (Solanas et al., 2001; Ludwig & Bryant, 2011). It had been suggested that this is not associated with a certain condition but to a more general phenomenon that would affect all the mRNA populations. For example, a differential regulation of the RNA polymerases
I and II, responsible for transcription of genes encoding rRNA and mRNA, could produce an rRNA/mRNA imbalance (Solanas et al., 2001). For accurate quantification by qPCR, it is important to choose a reference target whose transcription is regulated in a similar manner (Radonic et al., 2004). Certain experiments on Synechocystis sp. PCC 6803 showed that stress conditions could induce the expression of genes coding for proteins involved in translation of other proteins. In Synechocystis, genes for ribosomal proteins are located in a putative ribosomal-protein operon, thus stress can enhance their expression, and this could explain the low \( C_i \) values observed during our experiments. Moreover, the crystallographic structure of the large subunit of ribosomes has revealed that ribosomal proteins are located near the peptidyltransferase center, suggesting that some stress conditions (e.g. salt stress) might destabilize ribosomes and that de novo synthesis of these proteins might be necessary to maintain the activity of ribosomes (Kanesaki et al., 2002).

The petB gene was previously recommended as a reference gene in Cyanobacteria as it displayed stability in microarray studies (Hihara et al., 2001; Pinto et al., 2012). However, according to our findings, this gene is not suitable for qPCR analyses in Synechococcus sp. strain PCC 7002, as in this study, the \( M \) values and also the three algorithms have placed it among the weakest reference genes. The diversity of these results suggests the importance of validating reference genes under multiple experimental conditions. Experiments demonstrate that statistical significance is based on correct reference gene selection, even when changes in reference gene expression are minor, while use of unsuitable references can lead to over- or underestimation of relative transcript abundance. Even though a ‘universal’ set is difficult to be identified, our experiments have shown that the most suitable reference genes for qPCR in Synechococcus sp. PCC 7002 are \( ppC \), \( secA \), and \( rnpA \). However, it is highly recommended the use of gene pairs, the most stable combination being \( ppC + secA \). Genes \( petB \) and 16S rDNA are not recommended to be used as reference genes in this organism.

Thus, we conclude that certain genes should be used with caution in studies on Cyanobacteria, and only after verifying their quality as reference gene.

### Acknowledgements

Funding for this research was provided by the Romanian Ministry of National Education, project PN 09-360201, and by a grant of the Romanian National Authority for Scientific Research, CNCS – UEFISCDI, project number PN-III-P1-113602-2014-3-0765.

### References


**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

**Fig. S1.** Agarose gel (1%) showing amplification of specific PCR products of expected size for the genes tested in this study.

**Fig. S2.** Comprehensive gene stability value graphic obtained by combining values calculated by the three algorithm (geNorm, NormFinder, BestKeeper) used for accurate ranking and normalization under: (a) microaerobiosis (b) dark incubation (c) UV-B irradiation.

**Table S1.** Average *Ct* values of the candidate reference genes under the three stress condition.

**Table S2.** Average *E* values for the 15 pairs of genes.