Plant Physiology and Biochemistry 83 (2014) 20–25

Contents lists available at ScienceDirect

Plant Physiology and Biochemistry

journal homepage: www.elsevier.com/locate/plaphy

Short communication

Selection and validation of reference genes for transcript normalization in gene expression studies in *Catharanthus roseus*

Jacob Pollier [a, b], Robin Vanden Bossche [a, b], Heiko Rischer [c], Alain Goossens [a, b, *]

[a] Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Gent, Belgium
[b] Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052 Gent, Belgium
[c] VTT Technical Research Centre of Finland, P.O. Box 1000, 02044-VTT, Finland

ARTICLE INFO

Article history:
Received 19 June 2014
Accepted 3 July 2014
Available online 10 July 2014

Keywords:
BestKeeper
Catharanthus roseus
geNorm
NormFinder
qPCR
Reference gene
Transcript normalization

ABSTRACT

Quantitative Real-Time PCR (qPCR), a sensitive and commonly used technique for gene expression analysis, requires stably expressed reference genes for normalization of gene expression. Up to now, only one reference gene for qPCR analysis, corresponding to *40S Ribosomal protein S9* (*RPS9*), was available for the medicinal plant *Catharanthus roseus*, the only source of the commercial anticancer drugs vinblastine and vincristine. Here, we screened for additional reference genes for this plant species by mining *C. roseus* RNA-Seq data for orthologs of 22 genes known to be stably expressed in *Arabidopsis thaliana* and qualified as superior reference genes for this model plant species. Based on this, eight candidate *C. roseus* reference genes were identified and, together with *RPS9*, evaluated by performing qPCR on a series of different *C. roseus* explants and tissue cultures. NormFinder, geNorm and BestKeeper analyses of the resulting qPCR data revealed that the orthologs of *At2g28390* (SAND family protein, *SAND*), *At2g32170* (N2227-like family protein, *N2227*) and *At4g26410* (Expressed protein, EXP) had the highest expression stability across the different *C. roseus* samples and are superior as reference genes as compared to the traditionally used *RPS9*. Analysis of publicly available *C. roseus* RNA-Seq data confirmed the expression stability of *SAND* and *N2227*, underscoring their value as reference genes for *C. roseus* qPCR analysis.

1. Introduction

The medicinal plant *Catharanthus roseus* (Madagascar periwinkle) is a species of flowering plants native to Madagascar. Because of its long flowering period and its tolerance to drought and nutrient deficiency, *C. roseus* is widely cultivated as an ornamental plant in parks and gardens in regions with a (sub)tropical climate (Kumar et al., 2012). As a medicinal plant, *C. roseus* synthesizes over 130 different monoterpenoid indole alkaloids (MIAs), including the pharmaceutically important molecules ajmalicine and serpentine. Furthermore, it is the only source of the commercial drugs vinblastine and vincristine, MIAs that are commonly used in the treatment of several types of cancer, including leukemia and lymphoma (El-Sayed and Verpoorte, 2007; Van Der Heijden et al., 2004; Verma et al., 2012). The combination of complex MIA mixtures naturally occurring in *C. roseus* leaves, low production amounts of the individual MIAs and environmental instability in *C. roseus* cultivating countries leads to high market prices of the purified compounds and makes them excellent targets for breeding, metabolic engineering or synthetic biology programs that aim to enhance or alter their production (Miettinen et al., 2014; Van Der Heijden et al., 2004).

For any metabolic engineering effort or study of *C. roseus*, it may be necessary to investigate the expression of a target gene by quantitative Real-Time PCR (qPCR), a sensitive technique for gene expression analysis that is used in many research fields. The most appropriate normalization strategy for qPCR is the use of experimentally validated reference genes (Bustin et al., 2009), i.e., genes that are stably expressed among the different analyzed samples and that are unaffected by an experimental treatment. Furthermore, it is often mandatory to use more than one reference gene to achieve accurate normalization of gene expression (Vandesompele et al., 2002). For several studies with *C. roseus*, qPCR was used to assess the expression of a particular gene. For most of them, normalization was done using a single reference gene, *40S Ribosomal protein S9* (RPS9), originally described as a control gene for Northern blot analysis (Menke et al., 1999). Hence, this gene may not be the most suitable reference gene for normalization of

* Corresponding author. Technologiepark 927, B-9052 Gent, Belgium. Tel.: +32 9 3313851; fax: +32 9 3313809.
E-mail addresses: alain.goossens@psb.vib-ugent.be, algoo@psb.vib-ugent.be (A. Goossens).

http://dx.doi.org/10.1016/j.plaphy.2014.07.004
0981-9428/

C. roseus qPCR data. In this study, we used the geNorm, NormFinder and BestKeeper algorithms to evaluate the expression stability of *RPS9* and eight other candidate *C. roseus* reference genes that were identified by mining of *C. roseus* RNA-Seq data for orthologs of a set of 22 known reference genes for transcript normalization from the model plant *Arabidopsis thaliana* (Czechowski et al., 2005).

2. Materials and methods

2.1. Plant materials, cultivation and treatments

Ten different *C. roseus* plant materials were generated to use for the qPCR analysis. Cultivation of *C. roseus* plants and generation of calli and hairy roots were carried out as described (Häkkinen et al., 2012; Van Moerkercke et al., 2013). Elicitation of *C. roseus* shoots with 1 mM methyl jasmonate (MeJA) was performed as described (Van Moerkercke et al., 2013). An overview of the different *C. roseus* tissues and treatments is given in Table 1.

2.2. RNA extraction and cDNA synthesis

Total RNA was extracted from 100 mg of homogenized *C. roseus* plant material using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). To avoid genomic DNA contamination, an on-column DNase digestion was performed during RNA extraction according to the manufacturer's instructions. The concentration and purity of the obtained RNA was determined using a NanoDrop® ND-1000 UV–Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and the RNA integrity was assessed by gel electrophoresis on a 1.2% agarose gel stained with SYBR Safe™ (Life Technologies, Carlsbad, CA, USA). cDNA was synthesized from 1 μg of total RNA using the iScript cDNA synthesis kit (Biorad, Hercules, CA, USA) according to the manufacturer's instructions in a final volume of 20 μL. The iScript cDNA synthesis kit uses a blend of oligo(dT) and random hexamer primers for cDNA synthesis. After synthesis, the template for qPCR was prepared by diluting the obtained cDNA eightfold with ultrapure water.

2.3. qPCR

qPCR primers were designed using Beacon Designer version 4.0 (Premier Biosoft International, Palo Alto, CA, USA) with default parameters. qPCR reactions were carried out in triplicate in 384-well plates in a final volume of 5 μL. Each 5 μL reaction contained 0.5 μL qPCR template, 2.5 μL SYBR Green QPCR Master Mix (Stratagene, La Jolla, CA, USA) and 2.0 μL of primer mix (1.25 μM for each primer, leading to a final concentration of 0.5 μM for each primer in the 5 μL reaction). The reactions were prepared and dispensed in the 384-well plates using a JANUS Automated Workstation (Perkin Elmer, Waltham, MA, USA) and qPCR was carried out with a LightCycler 480 (Roche Applied Science, Penzberg, Germany). The qPCR reaction profile consisted of an initial activation of 10 min at 95 °C, followed by 45 cycles of 95 °C for 10 s, 60 °C for 15 s and 72 °C for 15 s. Finally, a melting curve was generated by increasing the temperature from 65 °C to 95 °C with continuous monitoring of the SYBR Green fluorescence. For each reaction, the crossing point (Cp) value was determined using the Second Derivative Maximum method with the LightCycler 480 software with default parameters. After converting the data to the required input files, stable reference genes were selected using the geNorm (Vandesompele et al., 2002), NormFinder (Andersen et al., 2004) and BestKeeper (Pfaffl et al., 2004) algorithms.

Table 1
C. roseus plant materials used for analysis of candidate reference genes.

Sample	Tissue	Cultivar	Treatment
1	Callus	Würzburg	–
2	Callus	Leiden	–
3	Hairy roots	Würzburg	–
4	*In vitro* plants	Würzburg	–
5	Roots	Würzburg	–
6	Buds	Würzburg	–
7	Shoots	Würzburg	6 h DMSO
8	Shoots	Würzburg	6 h MeJA
9	Shoots	Würzburg	24 h MeJA
10	Shoots	Würzburg	24 h DMSO

3. Results and discussion

3.1. Selection of candidate reference genes

Previously, we described the assembly of CathaCyc, a *C. roseus* metabolic pathway database built from Illumina HiSeq2000 RNA-Seq data (Van Moerkercke et al., 2013). The *C. roseus* reference transcriptome used to create this pathway database contains a set of 31,450 contigs that are accessible for analysis through the web interface of the ORCAE database (http://bioinformatics.psb.ugent.be/orcae). These contigs were assembled by combining RNA-Seq data from two research consortia, SmartCell (http://www.smartcell.org/) and the Medicinal Plant Genomics Resource (MPGR) consortium (http://medicinalplantgenomics.msu.edu/). The SmartCell dataset comprises two independent experiments in which *C. roseus* suspension cells and shoots treated or not with MeJa were used (Van Moerkercke et al., 2013). The MPGR dataset comprises RNA-Seq data from 23 different *C. roseus* tissues and cultures grown in different conditions (Góngora-Castillo et al., 2012).

To screen for candidate qPCR reference genes, TBLASTX searches using the nucleotide sequences of 22 genes from *A. thaliana*, shown to be superior reference genes in this model plant (Czechowski et al., 2005), were performed in the *C. roseus* reference transcriptome, revealing orthologs for 20 of them. Subsequently, candidate qPCR reference genes were withheld based on three criteria: (1) they were stably expressed in both the suspension cells and shoots datasets from SmartCell; (2) only one unique ortholog is present in the *C. roseus* reference transcriptome; and (3) the orthologs show >99% sequence identity between the SmartCell and MPGR datasets. Based on these criteria, eight candidate reference genes were retained.

The eight genes that were retained are the orthologs of *At5g46630* (clathrin adaptor complex subunit, *CACS*), *At4g26410* (expressed protein with unknown function, *EXP*), *At5g12240* (expressed protein with unknown function, *EXPR*), *At4g33380* (unknown protein F17M5, *F17M5*), *At*5g15710 (F-box domain containing protein, *Fbox*), *At2g32170* (N2227-like family protein, *N2227*), *At2g28390* (SAND family protein, *SAND*), and *At4g34270* (TIP41-like family protein, *TIP41*). An overview of the names and the *C. roseus* reference transcriptome (Caros) accession numbers of the retained genes is provided in Table 2.

3.2. Analysis of expression stability of the candidate reference genes

3.2.1. Primer design and qPCR analysis

By performing qPCR on ten different *C. roseus* samples (Table 1), including callus, hairy roots, in vitro plants, roots, buds and shoots treated or not with MeJa, Cp values were obtained for the eight candidate reference genes and for *RPS9* (Menke et al., 1999), that was previously used as a reference gene for *C. roseus* expression analysis. Analysis of the melting curves and agarose gel electrophoresis indicated that for each specifically designed primer pair