



RESEARCH ARTICLE

Selection and validation of reference genes for quantitative gene expression studies in *Erythroxylum coca* [v1; ref status: indexed, <http://f1000r.es/y1>]

Teresa Docimo^{1,2*}, Gregor W Schmidt^{1,3*}, Katrin Luck¹, Sven K Delaney⁴, John C D'Auria¹

¹Max-Planck-Institut für Chemische Ökologie, Jena, D-07745, Germany

²Current address: Istituto Biologia e Biotecnologia Agraria (CNR), Milan, 20133, Italy

³Current address: Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule Zurich, Basel, 5048, Switzerland

⁴School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, 2052, Australia

* Equal contributors

v1 First Published: 08 Feb 2013, 2:37 (doi: 10.12688/f1000research.2-37.v1)
Latest Published: 08 Feb 2013, 2:37 (doi: 10.12688/f1000research.2-37.v1)

Abstract

Real-time quantitative PCR is a powerful technique for the investigation of comparative gene expression, but its accuracy and reliability depend on the reference genes used as internal standards. Only genes that show a high level of expression stability are suitable for use as reference genes, and these must be identified on a case-by-case basis.

Erythroxylum coca produces and accumulates high amounts of the pharmacologically active tropane alkaloid cocaine (especially in the leaves), and is an emerging model for the investigation of tropane alkaloid biosynthesis. The identification of stable internal reference genes for this species is important for its development as a model species, and would enable comparative analysis of candidate biosynthetic genes in the different tissues of the coca plant. In this study, we evaluated the expression stability of nine candidate reference genes in *E. coca* (*Ec6409*, *Ec10131*, *Ec11142*, *Actin*, *APT2*, *EF1 α* , *TPB1*, *Pex4*, *Pp2aa3*). The expression of these genes was measured in seven tissues (flowers, stems, roots and four developmental leaf stages) and the stability of expression was assessed using three algorithms (geNorm, NormFinder and BestKeeper). From our results we conclude that *Ec10131* and *TPB1* are the most appropriate internal reference genes in leaves (where the majority of cocaine is produced), while *Ec10131* and *Ec6409* are the most suitable internal reference genes across all of the tissues tested.

Article Status Summary

Referee Responses

Referees	1	2
v1 published 08 Feb 2013	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

1 **Sheila McCormick**, University of California, Berkeley USA

2 **Sarah O'Connor**, John Innes Centre UK

Latest Comments

No Comments Yet

Corresponding author: John C D'Auria (dauria@ice.mpg.de)

How to cite this article: Docimo T, Schmidt GW, Luck K *et al.* (2013) Selection and validation of reference genes for quantitative gene expression studies in *Erythroxylum coca* [v1; ref status: indexed, <http://f1000r.es/y1>] *F1000Research* 2013, 2:37 (doi: 10.12688/f1000research.2-37.v1)

Copyright: © 2013 Docimo T et al. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: This work was supported by the Max Planck Society and an Alexander von Humboldt Foundation postdoctoral fellowship to JCD.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: No competing interests were disclosed.

First Published: 08 Feb 2013, 2:37 (doi: 10.12688/f1000research.2-37.v1)

First Indexed: 15 Feb 2013, 2:37 (doi: 10.12688/f1000research.2-37.v1)

Introduction

Erythroxylum coca has been cultivated by humans for more than 8000 years and has been selected for high-level production of cocaine, a pharmacologically active tropane alkaloid. Cocaine and other tropane alkaloids such as atropine and scopolamine act on the nervous system, and their activity is largely due to their common chemical backbone (the tropane nucleus)¹. Despite the socioeconomic importance of cocaine and other tropane alkaloids, the molecular basis for the biosynthesis of the tropane nucleus remains unknown. *E. coca* is emerging as a model for the investigation of tropane alkaloid synthesis²⁻⁴, and shows high-level, localized tropane alkaloid production and storage in its leaf tissue^{3,4}.

We have performed metabolic and enzymatic studies to identify the molecular and biochemical basis of tropane alkaloid biosynthesis in *E. coca*, and have developed a number of genomic tools such as expressed sequence tag (EST) libraries and 454 sequence databases²⁻⁴. Quantitative real-time reverse-transcription PCR (qRT-PCR) would be a further source of information on candidate tropane alkaloid biosynthesis genes in the different tissues of the coca plant.

qRT-PCR is widely used to quantify and compare levels of gene transcription⁵. Variables such as RNA quality and the efficiencies of reverse transcription and PCR may compromise the accuracy and reliability of qRT-PCR, and so results are typically 'normalized' by comparison with one or more internal reference genes⁶. The internal reference genes must be stably expressed, and the most stable reference genes vary widely in different species, tissues and sets of experimental conditions. Therefore, the identification of stable reference genes is a crucial step in the design of qRT-PCR experiments.

Traditionally, 'Housekeeping' genes such as actin, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and ubiquitin were used for data normalization^{7,8}. These genes were widely assumed to have a uniform level of expression due to their involvement in fundamental cellular processes. However, evaluation of the expression stability of classical housekeeping genes in many species including *Arabidopsis thaliana*, *Oryza sativa*, *Zea mays* and *Linum usitatissimum*⁹⁻¹² has revealed unstable expression of these genes under a range of experimental conditions. In addition, several novel reference genes have been shown to be more stably expressed than classical housekeeping genes¹³. Hence there is a need for systematic validation of internal reference genes in each organism and experiment^{9,14}.

The stability of candidate internal reference genes may be assessed using a number of models, including geNorm¹⁵, NormFinder¹⁶ and BestKeeper¹⁷. These models differ significantly in their assumptions, and so candidate genes are often assessed with several of these algorithms¹⁸. geNorm iteratively calculates an expression stability value (M) for each candidate gene. This is based on the mean pairwise variation between the gene and the other candidate genes across all samples. Genes with lower M values are more stably expressed, and less stable genes (with higher M) are progressively excluded from the analysis. The optimal number of reference genes for qRT-PCR normalization may also be determined by identifying the smallest number of genes needed to minimize mean variation. By contrast, NormFinder estimates the standard deviation for each gene relative

to the global expression of all genes included in the analysis, and genes with lower standard deviations are considered better reference genes. BestKeeper uses a third approach involving the calculation of a stability index (the 'BestKeeper index' or BKI), which is assumed to represent the highest level of stability because it includes all genes across all samples. The stability of each reference gene is assessed by its correlation with the BKI, with a high correlation indicating a more stable reference gene¹⁵⁻¹⁷.

In this study, we evaluate the stability of nine candidate reference genes (*Ec6409*, *Ec10131*, *Ec11142*, *Actin*, *APT2*, *EF1 α* , *TPB1*, *Pex4* and *Pp2aa3*) in a variety of *E. coca* tissues (four developmental leaf stages, stems, roots and flowers). We then identify the most stable internal reference genes using the geNorm, NormFinder and BestKeeper algorithms and present guidelines for transcript analysis in different tissues of *E. coca* by qRT-PCR.

Materials and methods

Plant material

Erythroxylum coca was obtained from the Bonn Botanical Garden. Plants were grown at 22°C under a photoperiod of 12 h light/12 h dark with relative humidities of 65% and 70% for light and dark conditions respectively (and fertilized once a week with Fertyl 3 (15-10-15) and Wuxal Top N (Planta, Regenstauf, Germany)).

The organs used for RNA extraction and qRT-PCR analysis were obtained from four-month old *E. coca* plants grown from rooted cuttings. Leaves in four developmental stages, roots, stems and flowers were analysed. The leaf developmental stages were: leaf buds; young expanding leaves in a rolled state (Stage 1); young expanded (unrolled) leaves (Stage 2); and fully mature leaves (Stage 3) (see [Figure 1](#)).

RNA extraction and cDNA synthesis

Total RNA was extracted from 100 mg of fresh plant tissue using a total RNA extraction kit (Invitex, Berlin, Germany). Genomic DNA was removed by treatment with RNase-free DNase I (Qiagen, Hilden, Germany). RNA quality was assessed on an Agilent Bioanalyzer 2100 using a RNA 6000 Nano Kit (Agilent, Böblingen, Germany). RNA concentration was determined using a NanoDrop 2000 c spectrophotometer (NanoDrop Technologies, Wilmington, USA). cDNA was synthesized using a Super Script III First Strand Kit (Invitrogen, Karlsruhe, Germany) according to the manufacturer's instructions. In brief, random hexamer primers and deoxyribonucleoside 5'-triphosphates (dNTPs) were added to 5 μ g total RNA and the mixture was incubated at 65°C for 5 min before brief chilling on ice. The first strand was then reverse transcribed by adding First Strand Buffer, 20 mM dithiothreitol and Super Script III reverse transcriptase to a final volume of 20 μ l and incubating the mixture at 42°C for 1h. The resulting cDNA was diluted 1:20 (vol:vol) with deionized water and stored at -20°C.

Reference gene selection

Candidate reference genes were selected from an *E. coca* 454 sequence library² based on their homology to previously reported reference genes in *A. thaliana*⁹. Nine candidate reference genes with an E-value higher than $2e^{-72}$ were identified by BlastN comparison as orthologues to *Arabidopsis* genes: Expressed protein (*Ec6409*),

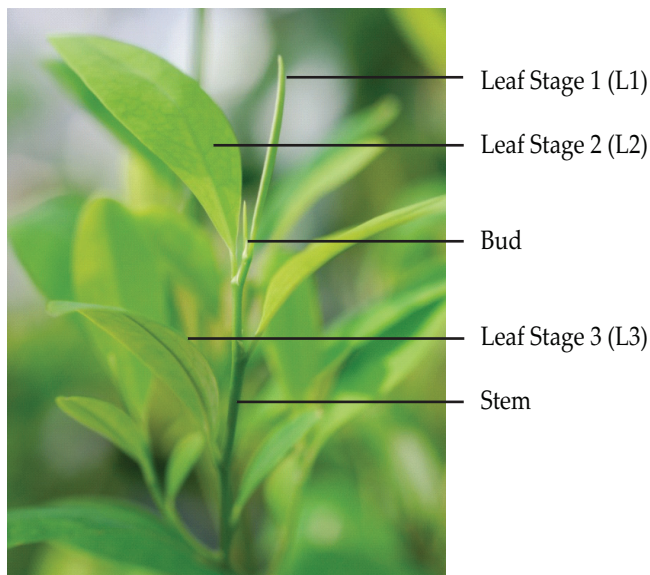


Figure 1. Developmental leaf stages of *Erythroxylum coca* plant. Leaf Stage I (L1) young rolled leaves, Leaf Stage II (L2) young expanded leaves, Leaf Stage III (L3) fully mature leaves.

Expressed protein (Ec10131), Clathrin adaptor complex subunit (Ec11142), Actin (ACT), Adenine phosphoribosyl transferase 2 (APT2), Elongation factor 1 alpha (EF1 α), Protein tyrosine phosphatase 1B (TPB1), Peroxin 4 (Pex4) and Pp2aa3-like protein. Primers for qRT-PCR were designed using Primer Express 3.0 (Applied Biosystems) and their sequences are shown in [Table 1](#).

All primer pairs were validated prior to their use in gene expression analysis. PCR reactions were performed with each primer pair and the products were visualised by gel electrophoresis to confirm the presence of a single PCR product of the expected size. The sequence specificity of the PCR products was also verified by sequencing.

Quantitative real-time PCR

All PCR reactions were performed on a Stratagene Mx3000P (La Jolla, USA). Each reaction contained 12.5 μ l Brilliant Sybr Green (Agilent/Böblingen, Germany), 0.375 μ l Rox, 0.4 μ M primers and 1 μ l cDNA in a final volume of 20 μ l. All samples were run in triplicate. The thermocycling conditions were denaturation at 95°C for 10 min; followed by 40 cycles of denaturation (95°C, 15 s) and annealing/extension (60°C, 1 min). A melting curve analysis protocol was performed after completion of the PCR reaction to confirm the absence of multiple amplicons and/or primer dimers. A no template control (NTC) was included to ensure the absence of contamination. In addition, the presence of genomic DNA contamination was excluded by performing reactions without reverse transcriptase. PCR efficiency was determined using a standard curve based on between five and seven different four-fold dilutions of a cDNA cloned amplicon.

Data analysis

Cycle threshold (Ct) values were exported from the MxPro software (Stratagene) to Microsoft Excel using the qBASE v1.3.5 macro¹⁹. PCR efficiencies and regression coefficients were calculated in qBASE and

are reported in [Table 1](#). The expression stability of the nine reference genes in *E. coca* tissues was evaluated with geNorm v3.5¹⁵, NormFinder¹⁶ and Bestkeeper v1¹⁷. Relative expression quantities were exported from qBASE and analyzed in Microsoft Excel using the geNorm v3.5 and NormFinder macros. For analysis using the BestKeeper macro, Ct values from the MxPro Software and PCR efficiencies calculated by qBASE were utilized.

Results

Selection and expression profiling of candidate reference genes

A similarity search (BlastN) between previously identified reference genes from *A. thaliana*⁹ and an *E. coca* 454 sequence library² was conducted to identify orthologous sequences. Nine *E. coca* genes with high similarity to *A. thaliana* were selected and PCR primers targeting these sequences were developed (see [Table 1](#)). To confirm the specificity of the primers and identity of the amplicons, RT-PCR was performed on cDNA from four developmental leaf stages, stems, roots and flowers. Primer specificity was investigated by electrophoresis and a single amplicon of the expected size was obtained for each primer pair ([Supplementary Figure 1](#)). Sequence analysis of ten cloned amplicons revealed that the amplified fragments were identical to the targeted sequences in the 454 sequence database. All primer pairs achieved amplification in fewer than 35 cycles in all samples, demonstrating that all of the candidate reference genes are expressed at experimentally useful levels. The C_t between samples and no template controls (NTCs) was always greater than five cycles, showing that contamination during the set-up of the experiment was negligible²⁰. All RNA samples were tested for contamination with genomic DNA by performing qPCR analysis on negative control reverse transcriptase reactions in which the reverse transcriptase was omitted. No amplification product could be detected in these control reactions.

The gene-specific amplification efficiency was calculated by linear regression analysis of the standard curve and ranged between 79% (Ec10131) and 97% (Actin). The coefficient of correlation (r^2) of the linear regression analysis was always greater than 0.986 as shown in [Table 1](#), indicating a linear relationship between C_t values and log-transformed transcript quantities in the range of the standard curve.

To ensure that the primer pairs are specific for the desired sequence in all samples and do not target homologous transcripts in some sample subsets, a melting curve analysis of each sample was performed after PCR amplification ([Supplementary Figure 2](#)). A single peak in the melting curve specific for each primer pair was obtained for all samples, and no peak could be observed in the melting curves of the control reactions (NTC and negative control reverse transcription reactions).

Expression stability of candidate reference genes

The expression stability of the candidate genes were evaluated with the geNorm, NormFinder and BestKeeper algorithms ([Table 2](#)). C_t values were transformed to relative quantities using qBASE prior to analysis with geNorm and NormFinder, while C_t values and PCR efficiencies were used in BestKeeper. The cDNA samples were considered as either a single, diverse set derived from all organ

Table 1. Description of *Erythroxylum coca* candidate reference genes. GenBank accession numbers are given for each gene used in this study. The orthologous locus in *A. thaliana* is referred to by its AGI (Arabidopsis Genome Initiative) designation. Similarity values are represented by E-values for the pairwise comparison of the coca gene with its Arabidopsis ortholog. PCR amplification efficiencies and the regression coefficients for their standard curves are reported for each primer pair.

Gene	Genbank accession number	Ortholog locus in <i>A. thaliana</i>	Similarity (E-value)	PCR efficiency	R ² of standard curve	Primer sequence (forward/reverse)
<i>Actin</i>	JN020155	AT5G09810	2e ⁻⁴⁰	97%	0.9974	GGATTTCCAAAGGTGAATACGATG/ TTGAACCAGCAAAGTTGAATAAGC
<i>APT2</i>	JN020149	AT5G11160	1e ⁻¹⁶	88%	0.9947	ACTCAGAGAGCGAGAGAGGATGTT/ TCAACTCCAGCAACCACAGAAATG
<i>EF1α</i>	JN020156	AT5G60390	0.00	84%	0.9981	TGGAGGTATTGACAAGCGTGTGATTGAGAG/ TTTGACACCAAGAGTGAAAGCAAGAAGAGC
<i>Ec11142</i>	JN020151	AT5G46630	2e ⁻⁷²	83%	0.9967	ACATTACCAAAGCAGGCTCATACG/ TACATCTTCTCACCACCAACACAGG
<i>Ec10131</i>	JN020153	AT2G32170	8e ⁻⁴⁵	79%	0.9916	TGGAAGGGTAGTGGGGTAACAATG/ GAGCGTAGTCGTCAGAGAAGGC
<i>Ec6409</i>	JN020150	AT4G26040	0.013	92%	0.9984	GAAGAGACAAGTGGTGGGGTGAG/ AGAAGAGAGCAAAGAGGAAGAGTGG
<i>Pp2aa3</i>	KC189827	AT1G13320	e ⁻¹⁴⁴	88%	0.9860	TGCTCCTGTTATGGGTCCTGAAG/ TGCTCCTGTTATGGGTCCTGAAG
<i>Pex4</i>	JN020157	AT5G25760	4e ⁻³⁴	88%	0.9968	GTCGGTCTTTAGCAAGGTCAGTG/ CGTGGTGGCGGTGGTTGG
<i>TPB1</i>	JN020152	AT3G01150	e ⁻¹⁰⁴	93%	0.9996	CCGATTGAAGCCATAACAGGAGAC/ CCCACAGGACCAGCACCAG

samples; or as two subsets derived from leaf buds and leaves (leaf buds, Stage 1, Stage 2 and Stage 3 leaves) or mature organs (Stage 3 leaves, flowers, roots and stems).

geNorm calculates the average expression stability value (M) for each candidate gene on the basis of the average pair-wise variation between all genes analyzed. geNorm analysis indicated that *Ec10131* and *Ec6409* are the most stable candidate reference genes across all of the *E. coca* tissues tested (Table 2). In the leaf bud/leaf sample subset, *Ec10131*, *TPB1* and *Ec6409* were ranked as the three most stable genes (in that order) (Supplementary Table 1), while in the mature organ subset *Ec10131* and *Ec6409* were again ranked as the most stable. In contrast, *Pex4* and *APT2* were consistently ranked as the least stable in all sample subsets (Table 2 and Supplementary Table 1 and Supplementary Table 2). The ‘house-keeping’ genes *Actin* and *EF1α* were relatively unstable and were ranked at positions six and seven (respectively) in all sample sets.

The optimal number of reference genes required for accurate normalization in the respective sample sets (all samples, leaf bud/leaf and mature tissues) was determined by calculating the mean variation in each normalisation factor (V) and then observing the effect of iterative addition of the next most stable reference gene (V_n/V_{n+1}) (as detailed in Vandesompele *et al.* 2002¹⁵). In each case, the two most stable reference genes were sufficient for accurate normalization, since inclusion of a third gene had little impact on the calculation of the normalization factor (V_n/V_{n+1} below 0.15).

BestKeeper ranks gene stability by calculating the correlation coefficient (*r*) between the expression of each candidate gene and the BestKeeper index (BKI; calculated using all genes across all samples). Across all of the samples tested, BestKeeper indicated that *Actin* (*r* = 0.784) and *Ec6409* (*r* = 0.768) were the most stable, while *Ec10131* was ranked as the least stable (*r* = 0.638). In the leaf bud/leaf sample subset, *Actin* (*r* = 0.869) and *APT2* (*r* = 0.868) had the highest correlation with the BKI, and *Ec10131* again showed the lowest correlation (*r* = 0.385). In the mature organs sample subset, *Pex4* and *APT2* were strongly correlated with the BestKeeper index (*r* = 0.767 and *r* = 0.724, respectively), whereas *Ec10131* showed low correlation (*r* = 0.309) (Supplementary Table 1 and Supplementary Table 2).

To provide a further ranking of gene stability, the results were also evaluated with NormFinder, in which candidate reference genes are ranked according to the variance of their expression relative to the expression variance within a defined group of samples¹⁶. *Pp2aa3* was the most stably expressed gene with the lowest expression variance (stability value of 0.291), followed by *Ec6409* and *Ec11142*, when all samples were included in the calculation. When the leaf bud/leaf and mature organ subsets of samples were considered, the rankings varied considerably (Supplementary Table 1 and Supplementary Table 2). *Actin*, *APT2* and *Pex4* were always ranked as the seventh, eighth and ninth most stable reference genes (respectively), but there was no consistent order of ranking for the other reference genes. The NormFinder rankings were also distinct from the geNorm rankings, although both algo-

Table 2. Ranking of *Erythroxylum coca* reference gene stability in all *Erythroxylum coca* tissues according to the geNorm, BestKeeper and NormFinder algorithms.

Gene rank	geNorm (M^* , $V_{n/n+1}$)	BestKeeper (correlation coefficient, r)	NormFinder (stability value)
1	Ec10131 / 6409 (0.28)	Actin (0.784)	Pp2aa3 (0.291)
2		Ec6409 (0.768)	Ec6409 (0.294)
3	Pp2aa3 (0.30; 0.095)	APT2 (0.765)	Ec11142 (0.300)
4	TPB1 (0.34; 0.083)	Pp2aa3 (0.737)	EF1 α (0.304)
5	Ec11142 (0.38; 0.080)	EF1 α (0.73)	TPB1 (0.339)
6	EF1 α (0.50; 0.115)	Pex4 (0.715)	Ec10131 (0.350)
7	Actin (0.62; 0.125)	TPB1 (0.688)	Actin (0.483)
8	APT2 (0.72; 0.120)	Ec11142 (0.661)	APT2 (0.596)
9	Pex4 (0.88; 0.147)	Ec10131 (0.638)	Pex4 (0.904)

*M indicates stability values listed from most stable to least stable.

gorithms identified *Actin*, *APT2* and *Pex4* as having the least stable expression profiles.

Raw Ct values and relative quantities for *Erythroxylum coca* reference genes

<http://dx.doi.org/10.6084/m9.figshare.154973>

Discussion

Real time RT-PCR has become a central technique for the evaluation of quantitative changes in gene expression^{21–24}. Reliable and accurate expression data can only be obtained by normalization with stably expressed reference genes. Normalization is an essential prerequisite for the correct measurement of gene expression changes in different plant tissues, organs, developmental stages or treatments of a given plant species and is highly influenced by the choice of reference genes. Traditional reference genes (e.g. actin and ubiquitin) are useful as stable reference genes in some experiments^{9,25}, but their expression is often highly variable^{26–28}, and is often inferior to the stability of less-commonly used genes⁸. Therefore it is important to assess the expression stability of several candidate reference genes before gene expression studies are performed. Several models including geNorm, NormFinder and BestKeeper have been developed to rank candidate reference genes on the basis of their expression stability. These methods often vary in their stability rankings^{18,25} and so expression data is commonly analysed using several approaches.

In this study, we report the identification and validation of nine candidate reference genes in *E. coca* (*Ec6409*, *Ec10131*, *Ec11142*, *Actin*, *APT2*, *EF1 α* , *TPB1*, *Pex4* and *Pp2aa3*). These genes were identified by analysing a 454 *E. coca* sequence library for sequences with homology to the top 100 reference genes of Arabidopsis⁹, on the assumption that homologous genes are likely to have similar expression patterns. Primer pairs specifically targeting the *E. coca* transcripts were successfully developed and evaluated: all primer pairs produced only the expected amplicon and were highly efficient (Table 1 and Supplementary Figure 1). The relative sta-

bilities of the candidate reference genes were then assessed using geNorm, BestKeeper and NormFinder (Table 2 and Supplementary Table 1 and Supplementary Table 2).

geNorm produced similar results in all sample sets. *Ec6409* and *Ec10131* were always identified as two of the three most stably expressed reference genes (although *Ec10131* and *TPB1* were most stable in the leaf bud/leaf sample subset), and *Actin*, *APT2* and *Pex4* were always identified as the least stable. geNorm may identify co-regulated genes as stable reference genes¹⁶. However, exclusion of either *Ec10131* or *Ec6409* did not change the gene rankings (not shown), suggesting that their high ranking is not attributable to co-regulation.

BestKeeper yielded very different rankings to geNorm, and these varied according to the sample subset. The inconsistent results with BestKeeper may be explained by several features of the BestKeeper algorithm. Calculation of the BestKeeper index excludes genes with a standard deviation of more than one C_t value, which results in the exclusion of different genes in different sample sets¹⁷. Extensive variation in C_t values is to be expected in a non-normalized data set, and so the algorithm may not be able to effectively distinguish between stable and unstable reference genes. In our experiments, the candidate *E. coca* reference genes showed very similar correlations with the BestKeeper index, suggesting that the algorithm could not distinguish between the genes to produce useful stability rankings. NormFinder produced a third ranking of gene stability that differed from both BestKeeper and geNorm. *Pp2aa3* and *Ec6409* were ranked as the most stably expressed genes when all samples were considered (Table 2). geNorm also identified *Ec6409* as one of the most stable genes in the entire sample set. However, only *Pp2aa3* was consistently ranked by Normfinder, geNorm and BestKeeper as one of the most stable genes in the leaf bud/leaf and mature organs sample sets, and there was no consistency between the algorithms in the order of ranking for the most stable genes (Table 2). The ranking of the least stable genes was more consistent: NormFinder identified *Actin*, *APT2* and *Pex4* as the least stable genes in all of the sample sets, and geNorm ranked these genes in the same order.

The NormFinder, BestKeeper and geNorm models have been shown to produce conflicting stability rankings in many studies^{18,29}. The rankings produced by one or more of the models may be combined to produce a hybrid ranking¹⁸, but this complicates the analysis by merging models with very different underlying assumptions. Hence, we favour using a single model when possible.

geNorm produced a consistent gene ranking across all of our samples, and provides a clear rationale for determining the minimum number of genes required for accurate normalization. We therefore recommend the use of *Ec10131* and *Ec6409* as internal reference genes for most *E. coca* sample sets. If leaves and leaf buds are the primary organs of interest, then we recommend the use of *Ec10131* and *TPB1*. These results provide a foundation for qRT-PCR studies in *E. coca*, and will further its development as a model of tropane alkaloid biosynthesis.

Author contributions

TD, GWS and JCD designed the research; TD, GWS, KL, and JCD performed the research; TD, GWS, and JCD analyzed the data; TD, GWS, SKD and JCD wrote the paper. All authors have approved the final manuscript for publication.

Competing interests

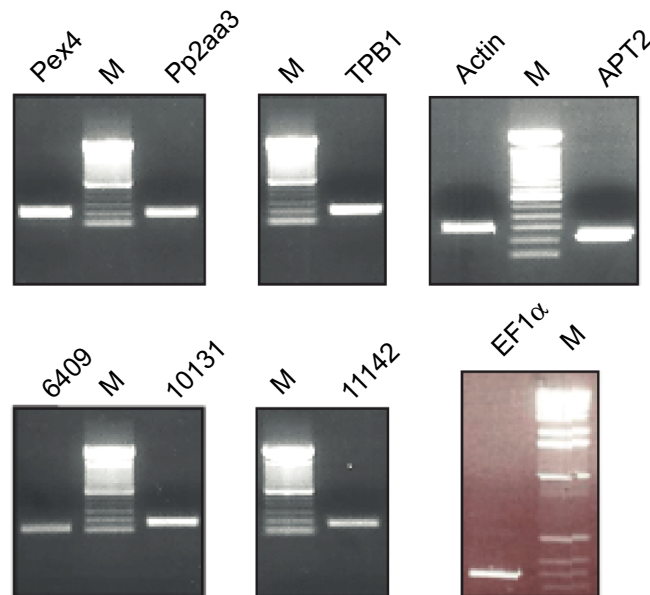
No competing interests were disclosed.

Grant information

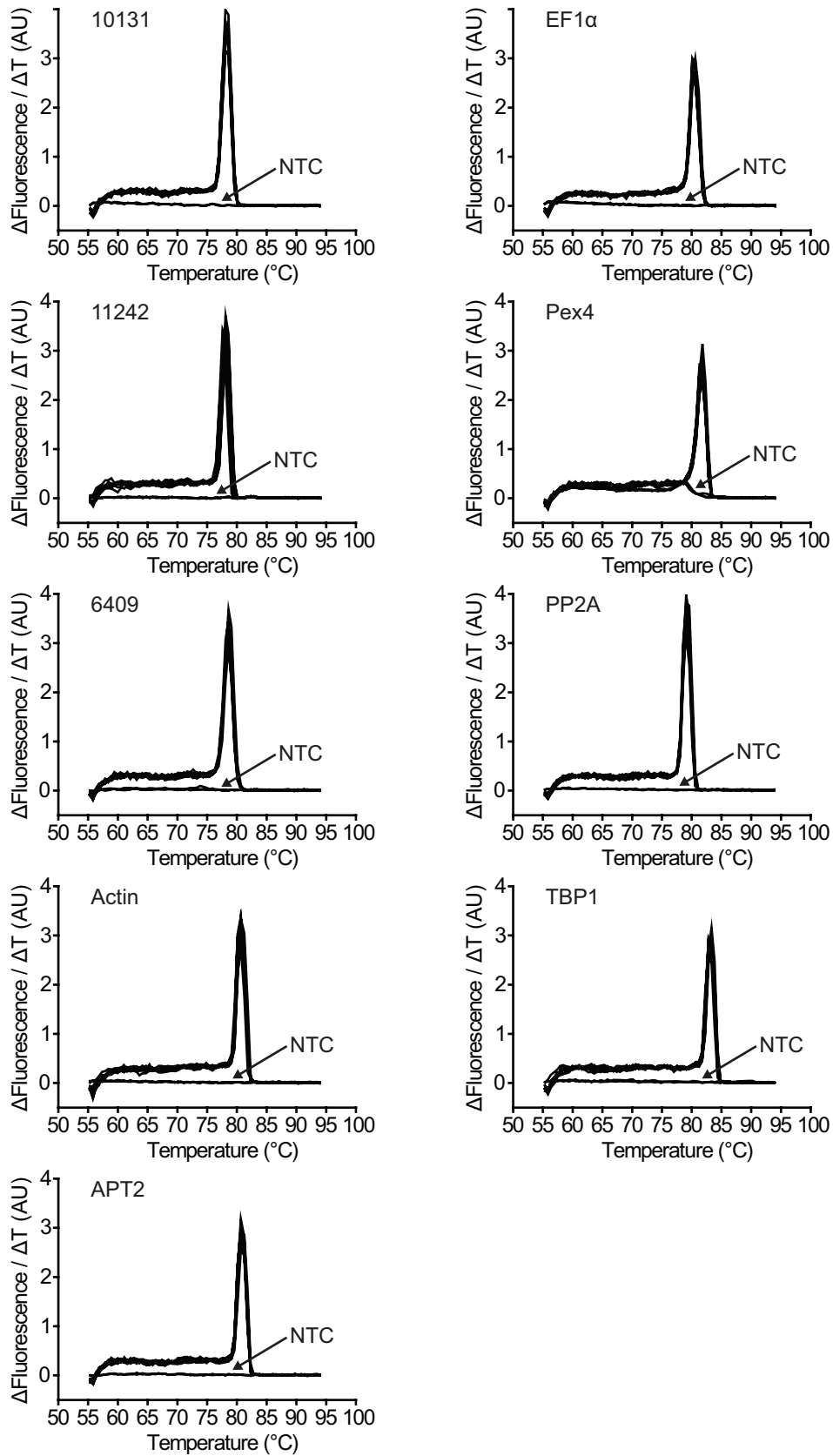
This work was supported by the Max Planck Society and an Alexander von Humboldt Foundation postdoctoral fellowship to JCD.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplementary figures



Supplementary Figure 1. Specificity of qRT-PCR primers and amplicon length. The lane marked M represents a 1 Kb ladder (Invitrogen, California) used for size comparisons.



Supplementary Figure 2. Melting curve analysis of RT-PCR products. NTC indicates: no template control.

Supplementary tables

Supplementary Table 1. Ranking of *Erythroxylum coca* reference gene stability in a sample subset containing only leaf tissues (Buds, Leaf Stage I-III). Analysis was performed using the geNorm, BestKeeper and NormFinder algorithms.

Gene rank	geNorm (M^* , $V_{n/n+1}$)	BestKeeper (correlation coefficient, r)	NormFinder (stability value)
1	Ec10131/TPB1 (0.26)	Actin (0.869)	Ec6409 (0.176)
2		APT2 (0.868)	EF1 α (0.264)
3	Ec6409 (0.30; 0.096)	Ec6409 (0.837)	Pp2aa3 (0.306)
4	Pp2aa3 (0.34; 0.089)	EF1 α (0.805)	TPB1 (0.318)
5	Ec11142 (0.40; 0.088)	Pex4 (0.762)	Ec11142 (0.366)
6	EF1 α (0.48; 0.100)	TPB1 (0.733)	Ec10131 (0.415)
7	Actin (0.68; 0.161)	Pp2aa3 (0.652)	Actin (0.642)
8	APT2 (0.78; 0.128)	Ec11142 (0.554)	APT2 (0.664)
9	Pex4 (0.90; 0.133)	Ec10131 (0.385)	Pex4 (0.815)

*M indicates stability values listed from most stable to least stable.

Supplementary Table 2. Ranking of *Erythroxylum coca* reference gene stability in a sample subset containing only mature organs (Leaf stage III, Flowers, Roots, Stems). Analysis was performed using the geNorm, BestKeeper and NormFinder algorithms.

Gene rank	geNorm (M^* , $V_{n/n+1}$)	BestKeeper (correlation coefficient, r)	NormFinder (stability value)
1	Ec10131/6409 (0.17)	Pex4 (0.767)	EF1 α (0.226)
2		APT2 (0.724)	Pp2aa3 (0.250)
3	Ec11142 (0.24; 0.088)	EF1 α (0.693)	Ec11142 (0.273)
4	Pp2aa3 (0.27; 0.066)	Actin (0.576)	Ec6409 (0.332)
5	TPB1 (0.30; 0.064)	Pp2aa3 (0.538)	Ec10131 (0.395)
6	EF1 α (0.45; 0.121)	TPB1 (0.452)	TPB1 (0.436)
7	Actin (0.61; 0.136)	Ec6409 (0.442)	Actin (0.522)
8	APT2 (0.74; 0.133)	Ec11142 (0.441)	APT2 (0.638)
9	Pex4 (0.97; 0.189)	Ec10131 (0.309)	Pex4 (1.170)

*M indicates stability values listed from most stable to least stable.

References

1. Lounasmaa M, Tamminen T: **The tropane alkaloids**. The Alkaloids, ed Cordell GA (Academic, New York) 1993; **44**: 1–114.
2. Docimo T, Reichelt M, Schneider B, *et al.*: **The first step in the biosynthesis of cocaine in *Erythroxylum coca*: the characterization of arginine and ornithine decarboxylases**. *Plant Mol Biol*. 2012; **78**(6): 599–615.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Jirschitzka J, Schmidt GW, Reichelt M, *et al.*: **Plant tropane alkaloid biosynthesis evolved independently in the Solanaceae and Erythroxylaceae**. *Proc Natl Acad Sci U S A*. 2012; **109**(26): 10304–10309.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Torre JC, Schmidt GW, Paetz C, *et al.*: **The biosynthesis of hydroxycinnamoyl quinate esters and their role in the storage of cocaine in *Erythroxylum coca***. *Phytochemistry*. 2013; **91**: 177–186.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Xu Y, Zhu X, Gong Y, *et al.*: **Evaluation of reference genes for gene expression studies in radish (*Raphanus sativus* L.) using quantitative real-time PCR**. *Biochem Biophys Res Commun*. 2012; **424**(3): 398–403.
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Willems E, Leys L, Vandesompele J: **Standardization of real-time PCR gene expression data from independent biological replicates**. *Anal Biochem*. 2008; **379**(1): 127–129.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Guénin S, Mauriat M, Pelloux J, *et al.*: **Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references**. *J Exp Bot*. 2009; **60**(2): 487–493.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Gutierrez L, Mauriat M, Guénin S, *et al.*: **The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants**. *Plant Biotechnol J*. 2008; **6**(6): 609–618.
[PubMed Abstract](#) | [Publisher Full Text](#)
9. Czechowski T, Stitt M, Altmann T, *et al.*: **Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis**. *Plant Physiol*. 2005; **139**(1): 5–17.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Wang L, Xie W, Chen Y, *et al.*: **A dynamic gene expression atlas covering the entire life cycle of rice**. *Plant J*. 2010; **61**(5): 752–766.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Sekhon RS, Lin H, Childs KL, *et al.*: **Genome-wide atlas of transcription during maize development**. *Plant J*. 2011; **66**(4): 553–563.
[PubMed Abstract](#) | [Publisher Full Text](#)
12. Huis R, Hawkins S, Neutelings G: **Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.)**. *BMC Plant Biol*. 2010; **10**: 71.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Mafra V, Kubo KS, Alves-Ferreira M, *et al.*: **Reference genes for accurate transcript normalization in citrus genotypes under different experimental conditions**. *PLoS One*. 2012; **7**(2): e31263.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Remans T, Smeets K, Opendakker K, *et al.*: **Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations**. *Planta*. 2008; **227**(6): 1343–1349.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. Vandesompele J, De Paepe A, Speleman F: **Elimination of primer-dimer artifacts and genomic coamplification using a two-step SYBR green I real time RT-PCR**. *Anal Biochem*. 2002; **303**(1): 95–98.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Andersen CL, Jensen JL, Orntoft TF: **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 Res*. 2004; **64**(15): 5245–5250.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Pfaffl MW, Tichopad A, Prgomet C, *et al.*: **Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: bestKeeper–Excel-based tool using pair-wise correlations**. *Biotechnol Lett*. 2004; **26**(6): 509–515.
[PubMed Abstract](#) | [Publisher Full Text](#)
18. Expósito-Rodríguez M, Borges AA, Borges-Pérez A, *et al.*: **Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process**. *BMC Plant Biol*. 2008; **8**: 131.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Hellemans J, Mortier G, De Paepe A, *et al.*: **qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data**. *Genome Biol*. 2007; **8**(2): R19.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Nolan T, Hands RE, Bustin SA: **Quantification of mRNA using real-time RT-PCR**. *Nat Protoc*. 2006; **1**(3): 1559–1582.
[PubMed Abstract](#) | [Publisher Full Text](#)
21. Giulietti A, Overbergh L, Valckx D, *et al.*: **An overview of real-time quantitative PCR applications to quantify cytokine gene expression**. *Methods*. 2001; **25**(4): 386–401.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Gachon C, Mingam A, Charrier B: **Real-time PCR: what relevance to plant studies?** *J Exp Bot*. 2004; **55**(402): 1445–1454.
[PubMed Abstract](#) | [Publisher Full Text](#)
23. VanGuilder HD, Vrana KE, Freeman WM: **Twenty-five years of quantitative PCR for gene expression analysis**. *Biotechniques*. 2008; **44**(5): 619–26.
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Cruz F, Kalaoun S, Nobile P, *et al.*: **Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR**. *Mol Breed*. 2009; **23**(4): 607–616.
[Publisher Full Text](#)
25. Schmidt GW, Delaney SK: **Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress**. *Mol Genet Genomics*. 2010; **283**(3): 233–41.
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Barsalobres-Cavallari CF, Severino FE, Maluf MP, *et al.*: **Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions**. *BMC Mol Biol*. 2009; **10**: 1.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Radonić A, Thulke S, Mackay IM, *et al.*: **Guideline to reference gene selection for quantitative real-time PCR**. *Biochem Biophys Res Commun*. 2004; **313**(4): 856–62.
[PubMed Abstract](#) | [Publisher Full Text](#)
28. Tong Z, Gao Z, Wang F, *et al.*: **Selection of reliable reference genes for gene expression studies in peach using realtime PCR**. *BMC Mol Biol*. 2009; **10**: 71.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Artico S, Nardeli SM, Brilhante O, *et al.*: **Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data**. *BMC Plant Biol*. 2010; **10**: 49.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Current Referee Status:

Referee Responses for Version 1



Sarah O'Connor

Biological Chemistry Department, John Innes Centre, Norwich, UK

Approved: 15 February 2013

Referee Report: 15 February 2013

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.



Sheila McCormick

Plant Gene Expression Center, University of California, Berkeley, Berkeley, CA, USA

Approved: 11 February 2013

Referee Report: 11 February 2013

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.
