


SCIENTIFIC REPORTS

OPEN

Selection of stable reference genes for RT-qPCR in *Rhodococcus opacus* PD630

Drew M. DeLorenzo & Tae Seok Moon 

Rhodococcus opacus PD630 is a gram-positive bacterium with promising attributes for the conversion of lignin into valuable fuels and chemicals. To develop an organism as a cellular factory, it is necessary to have a deep understanding of its metabolism and any heterologous pathways being expressed. For the purpose of quantifying gene transcription, reverse transcription quantitative PCR (RT-qPCR) is the gold standard due to its sensitivity and reproducibility. However, RT-qPCR requires the use of reference genes whose expression is stable across distinct growth or treatment conditions to normalize the results. Unfortunately, no in-depth analysis of stable reference genes has been conducted in *Rhodococcus*, inhibiting the utilization of RT-qPCR in *R. opacus*. In this work, ten candidate reference genes, chosen based on previously collected RNA sequencing data or literature, were examined under four distinct growth conditions using three mathematical programs (BestKeeper, Normfinder, and geNorm). Based on this analysis, the minimum number of reference genes required was found to be two, and two separate pairs of reference genes were identified as optimal normalization factors for when ribosomal RNA is either present or depleted. This work represents the first validation of reference genes for *Rhodococcus*, providing a valuable starting point for future research.

Rhodococcus opacus PD630 (hereafter *R. opacus*) is a gram-positive, oleaginous bacterium that possesses beneficial traits for the conversion of lignin into valuable fuels and chemicals^{1–4}. Some of these features include a number of catabolic pathways for lignin-derived aromatic compound consumption and a high native tolerance towards these and other inhibitory lignocellulosic biomass breakdown products^{2–5}. Furthermore, *R. opacus* can consume multiple types of carbon sources simultaneously⁶, thereby facilitating higher rates of feedstock conversion. Additionally, this organism can direct a large fraction of its cellular resources to the production of biofuel precursors (up to ~78% triacylglycerol [TAG] of cell dry weight)¹. *R. opacus* has been previously engineered to facilitate lignocellulose conversion^{3,4,7,8} and a substantial genetic toolbox has recently been developed^{9,10}. However, a deep understanding of this organism's metabolism and any heterologous pathways being expressed is required to maximize its potential.

A number of technologies exist for examining an organism's gene expression (i.e. the transcriptome), which is the first step to a systems level understanding. One such technology is the microarray, which allows for gene expression profiling¹¹. RNA sequencing (RNA-Seq) is a newer technology that has become the default method for examining the entire transcriptome of an organism¹². However, it can add additional costs if only several genes are of interest, is limited when mRNA concentrations are low (although this is changing with the advent of single cell sequencing¹³), and generally still requires corroboration via additional quantitative methods¹². One such complimentary method is reverse transcription quantitative PCR (RT-qPCR), which is considered the gold standard of mRNA quantification due to its high sensitivity, reproducibility, speed, ability to examine numerous samples simultaneously, and large dynamic range^{14,15}. Both microarrays and RT-qPCR require the use of an internal standard, optimally a gene that is stably expressed across the tested growth or treatment conditions, to normalize expression data between samples and conditions¹⁶.

Unfortunately, no in-depth analysis of stable reference genes (RGs) has been performed in *Rhodococcus*, limiting the ability to quantitatively analyze gene expression. One qPCR study in *Rhodococcus equi* even stated that no reference gene was included in their experiment and that such an inclusion could have improved their work¹⁷. We could find only two examples of reference genes previously reported in *Rhodococcus*. The first reference gene was

Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri, 63130, USA. Correspondence and requests for materials should be addressed to T.S.M. (email: tsmoon@wustl.edu)

Reference gene (RG)	Gene number	Gene annotation	Amplicon size (bp)	PCR efficiency	Min C _T	Max C _T	C _T Std Dev	NTC C _T
RG1	PD630_RS22865	Pup-protein ligase	157	94%	23.44	25.23	0.40	N.d.
RG2	PD630_RS20570	Hypothetical protein	87	94%	25.86	28.20	0.67	N.d.
RG3	PD630_RS03840	23S rRNA	87	95%	8.74	10.16	0.33	N.d.
RG4	PD630_RS15810	Polyribonucleotide nucleotidyltransferase	115	101%	22.93	27.41	1.37	N.d.
RG5	PD630_RS25785	L,D-transpeptidase Mb0493	90	101%	24.56	25.96	0.37	N.d.
RG6	PD630_RS13910	NAD(P)H dehydrogenase	82	100%	25.47	29.01	1.13	N.d.
RG7	PD630_RS25530	ATP-dependent Clp protease ATP-binding subunit ClpX	94	92%	24.93	25.77	0.21	N.d.
RG8	PD630_RS01395	16S rRNA	101	92%	10.48	11.74	0.29	N.d.
RG9	PD630_RS37755	rRNA small subunit methyltransferase G	102	92%	28.32	30.33	0.46	N.d.
RG10	PD630_RS27310	DNA polymerase IV	71	94%	28.34	30.61	0.61	N.d.

Table 1. List of candidate reference genes (RGs). The amplicon size, PCR efficiency percentage, minimum and maximum threshold cycle (C_T) values observed across all tested growth conditions, the standard deviation of C_T values across all tested growth conditions (determined by BestKeeper), and the C_T value of the no template control (NTC) are listed for each RG. See Supplementary Figure 1 for standard curves used to calculate PCR efficiency. See Supplementary Figure 2 for melting curve analysis. See Supplementary Table 1 for oligonucleotide sequences and melting and annealing temperatures. N.d. = no amplification detected.

a gene encoding sigma factor A (*sigA*) in *Rhodococcus* sp. RHA1¹⁸, although no justification for this choice was provided. The second reference gene was a gene encoding DNA Polymerase IV, which was used in *Rhodococcus* sp. RHA1, *Rhodococcus jostii*, and *Rhodococcus erythropolis*, and this choice was justified based on a microarray experiment performed in *Rhodococcus* sp. RHA1^{19,20}. Both of these reference genes were used in isolation and their characterization was incomplete, which fails to satisfy the current minimum information guidelines for publication of quantitative PCR experiments (i.e. MIQE guidelines stating that the minimum number of reference genes needs to be quantitatively determined and that one gene is not generally sufficient for normalization)^{21,22}.

In this work, we identified ten candidate reference genes (RGs) and examined the stability of their expression in *R. opacus* across four distinct growth conditions using three mathematical models (BestKeeper²³, NormFinder²⁴, and geNorm^{16,25}). Additionally, the minimum number of required reference genes was identified. Two different sets of genes were identified as optimal normalization factors (NFs) depending on whether ribosomal RNA (rRNA) is either present or depleted. This work facilitates the utilization of RT-qPCR in *R. opacus*, in addition to providing a valuable preliminary set of reference genes for further future validation in other *Rhodococcus* spp.

Results and Discussion

Choice of candidate reference genes. Two methods were utilized for the selection of candidate reference genes (RGs). The primary approach used our previously published transcriptomic data collected from *R. opacus* grown in a minimal salts medium with either glucose or phenol to identify stably expressed genes³. We selected nine genes as candidates (RG1 to RG9; Table 1) whose DeSeq 2 normalized transcript level did not vary significantly between the two growth conditions, whose DeSeq 2 expression value was greater than 750, and whose coding region is at least 350 bp in length^{3,26}. The secondary approach utilized a literature review which found that *sigA* had been previously used as a RG in *Rhodococcus* sp. RHA1¹⁸ and that a DNA Polymerase IV gene has been previously used in *Rhodococcus jostii*, *Rhodococcus erythropolis*, and *Rhodococcus* sp. RHA1 as a RG^{19,20}. *sigA* was ruled out due to no justification for its selection as a RG being provided¹⁸. As PD630_RS27310 is annotated as a DNA Polymerase IV gene in *R. opacus*, it was selected as the tenth candidate RG (RG10; Table 1).

RT-qPCR primer characterization and data collection. RT-qPCR primers were designed based on previously published suggestions²², in addition to the specific criteria discussed in the Supplementary Materials. The size and specificity of each amplicon was confirmed using agarose gel electrophoresis and melting curve analysis (Fig. 1 and Supplementary Figure 2), and the no template controls (NTCs) demonstrated non-detectable levels of amplification (Table 1 and Supplementary Figure 2). To calculate the PCR amplification efficiency, a ten-fold serial dilution of template DNA (PCR amplified product) was performed and followed by qPCR (Supplementary Figure 1). A linear regression analysis was performed on the resultant C_T values to confirm the linearity of each serial dilution and to calculate the PCR amplification efficiency (Table 1 and Supplementary Figure 1). As all serial dilutions had an acceptable R² of at least 0.97 and PCR efficiencies ranged from 92 to 101%, all primer sets were deemed suitable for qPCR (Table 1 and Supplementary Figure 2).

R. opacus was cultured in four distinct growth conditions, and RNA was extracted from biological triplicate cultures for candidate RG expression stability analysis. The growth conditions are described in full in the Materials and Methods, but in summary they consist of a minimal salts medium with glucose and either high nitrogen (HN) or low nitrogen (LN), a rich tryptic soy broth medium (TSB), or a minimal salts medium with phenol and high nitrogen (PHE). Each medium was selected based on growth conditions that *R. opacus* is likely to experience during general research endeavors and based on the diverse predicted changes in metabolic topology required for catabolism of each respective feedstock. Glucose was chosen as a representative sugar feedstock as it is frequently provided in *R. opacus* cell cultures^{3,9,27}. High and low nitrogen concentrations were examined

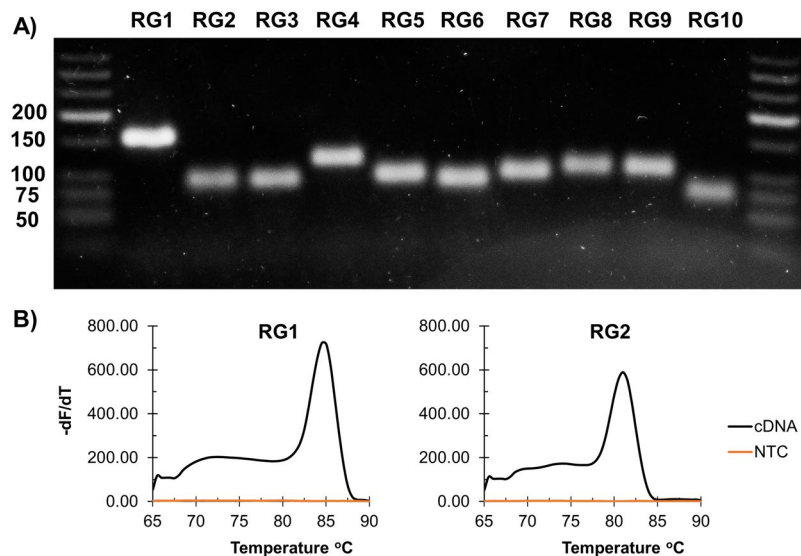


Figure 1. Confirmation of specificity of primers used in RT-qPCR analysis. **(A)** Contrast enhanced image of electrophoresis gel confirming amplicon size and primer specificity using RT-qPCR amplification product. The sizes for nucleotide ladder are indicated to left of bands (50 to 200 bp). The size of each amplicon is denoted in Table 1. The original non-enhanced gel image, in addition to the corresponding gel image of no template controls (NTCs), can be found in Supplementary Figure 5. **(B)** Representative melt curves post RT-qPCR for RG1 and RG2 as well as the corresponding NTCs (see Supplementary Figure 2 for all melt curves).

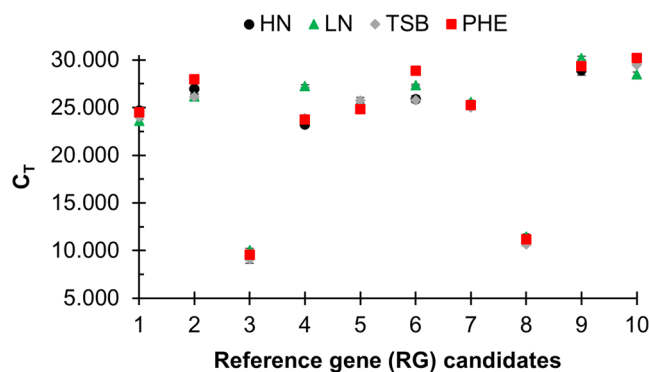


Figure 2. C_T values for ten candidate reference genes in *R. opacus*. RNA was extracted from biological triplicates of *R. opacus* grown in four distinct growth conditions (HN [circle], LN [triangle], TSB [diamond], and PHE [square]; see Materials and Methods for full description). RT-qPCR was performed in technical triplicate on each biological replicate. Each point represents the average threshold cycle (C_T) value of all replicates for the listed gene and growth condition. Error bars represent one standard deviation.

because *R. opacus* undergoes a metabolic flux shift to produce large quantities of industrially relevant lipids during nitrogen deprivation^{1,28,29}. TSB was selected as it contains an array of amino acids, which are utilized by a diverse set of metabolic pathways³⁰. Phenol was selected as a representative aromatic compound that may be found in depolymerized lignin^{3,4} and because it requires a different subset of metabolic pathways than glucose and amino acids⁶.

RT-qPCR was then performed on the cDNA generated from the triplicate RNA samples collected from the HN, LN, TSB, and PHE growth conditions (Fig. 2). The averaged raw C_T values ranged from 9.2 to 30.2 for RG1 through RG10 across all growth conditions. The minimum and maximum C_T values, in addition to the standard deviation, observed for each candidate RG across all growth conditions are listed in Table 1. A visual appraisal of Fig. 2 can provide initial insight into candidate RG expression stability, as some RGs demonstrated a tight clustering of C_T values (e.g. RG1, RG3, RG5, RG7, RG8, and RG9) while others demonstrated a spread in C_T values (e.g. RG2, RG4, RG6, and RG10) across growth conditions.

Expression stability of candidate reference genes across distinct growth conditions. Candidate RG expression stability was quantitatively examined using three different statistical programs: BestKeeper²³, NormFinder²⁴, and geNorm^{16,25}. Each of these programs generates a gene expression stability coefficient (r-value

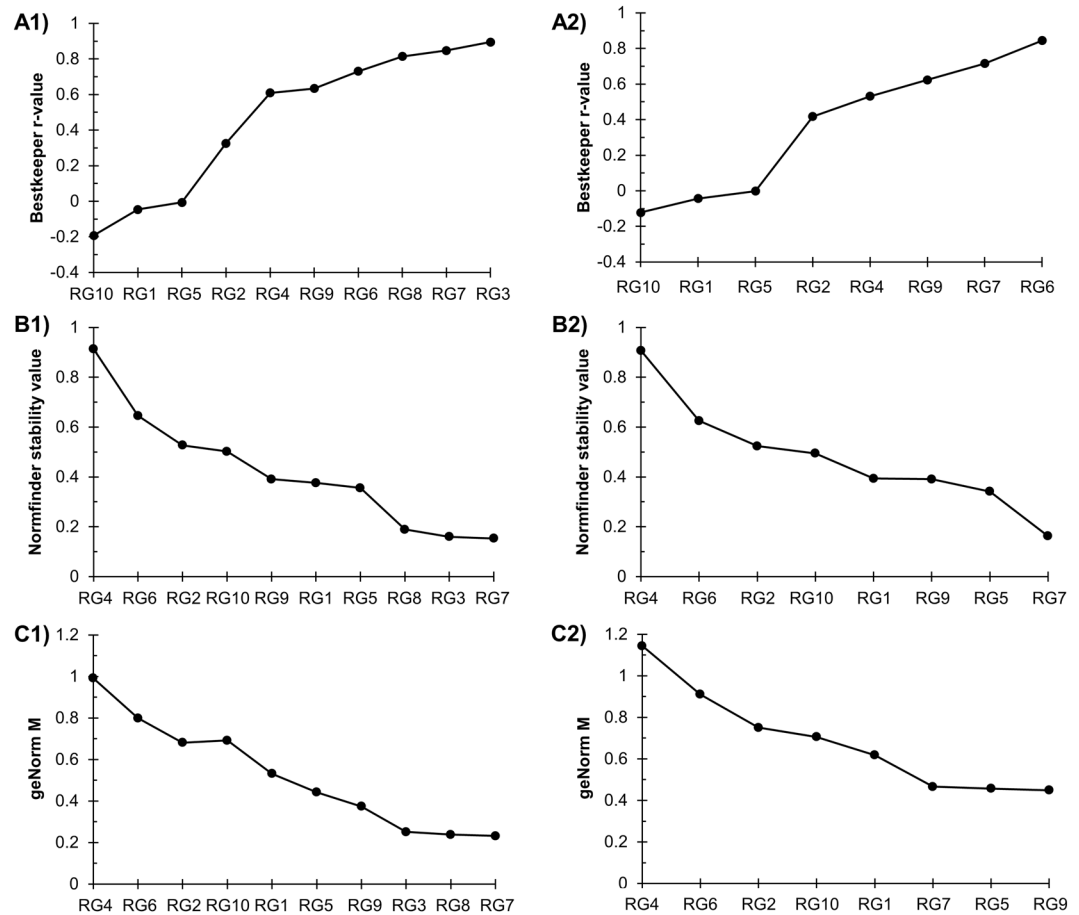


Figure 3. Rankings of candidate reference genes. Genes were ranked from the least stable (on the left) to the most stable (on the right). Analysis was performed after pooling C_T data across all four growth conditions. The designation of 1 means that the analyses were performed on all ten candidate RGs. The designation of 2 means that the analyses were performed on the eight non-rRNA candidate RGs. (A) Genes were ranked according to their BestKeeper r-value. BestKeeper r-value significance can be found in Supplementary Tables 3 and 4. (B) Genes were ranked according to their Normfinder stability value. (C) Genes were ranked according to their geNorm M value.

for BestKeeper, stability value for Normfinder, and M value for geNorm) and rank order for the candidate RGs (see Materials and Methods for full description; Fig. 3). Additionally, candidate RGs were ranked based on their C_T standard deviation calculated by BestKeeper (Supplementary Table 2). A C_T standard deviation greater than 1 is considered unstable²³.

All analyses identified the same three candidate RGs, albeit in different rank orders, as being the most stably expressed across all four growth conditions. According to BestKeeper, the top three most stable RGs from 1st to 3rd, based on the smallest standard deviation of C_T values, were RG7, RG8, and RG3 (Supplementary Table 2). The top three RGs with the lowest BestKeeper r-value from 1st to 3rd were RG3, RG7, and RG8 (Fig. 3A1). Normfinder determined that the top three most stable RGs from 1st to 3rd, based on its stability value calculation, were RG7, RG3, and RG8 (Fig. 3B1). Finally, geNorm determined that the top three most stable RGs from 1st to 3rd, based on its M value, were RG7, RG8, and RG3 (Fig. 3C1).

Two of the top three most stable candidate RGs encode rRNAs (RG3 and RG8), which is consistent with previous works in other organisms that identified an rRNA gene as a stable RG^{31,32}. However, rRNA is often depleted from mRNA samples being prepared for RNA-Seq³³. If RT-qPCR is to be used to corroborate RNA-Seq results, other non-rRNA RGs need to be identified. All analyses were repeated on the non-rRNA candidate RGs, and two genes (RG7 and RG9) appeared in the top 3 rankings of three analyses, while another (RG5) appeared in the top 3 rankings of two analyses (Fig. 3). The BestKeeper r-value identified the top three non-rRNA RGs from 1st to 3rd as RG6, RG7, and RG9 (Fig. 3A2). As RG6 had a C_T standard deviation greater than 1 (Supplementary Table 2), it was ruled out as a candidate. Normfinder identified the top three non-rRNA RGs from 1st to 3rd as RG7, RG5, and RG9 (Fig. 3B2). geNorm identified the top three non-rRNA RGs from 1st to 3rd as RG9, RG5, and RG7 (Fig. 3C2).

Minimum required number of reference genes. Analyses of candidate RG expression stability identified several consistently transcribed genes across the four distinct growth conditions, but it was still unknown how many RGs were required for optimal normalization of expression data. The geNorm software is also capable of producing a V value (see Materials and Methods), which produces a quantitative suggestion for the ideal

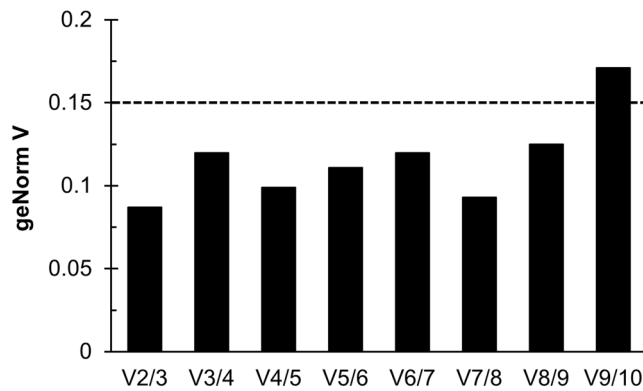


Figure 4. Minimum number of reference genes. The pair-wise variation V_n/V_{n+1} , where n represents the number of RGs used in the normalization factor (NF), was calculated by geNorm to determine the minimum number of RGs required for normalization. A geNorm V value below 0.15 signifies that no additional benefit is gained from increasing the number of reference genes from n to $n + 1$. The dashed line represents a V value of 0.15.

number of required RGs. By calculating the pairwise variation of V_n/V_{n+1} (V value), where n is the number of RGs, the benefit of using n versus $n + 1$ RGs for normalization can be quantified. A V value below 0.15 signifies that there is no additional benefit from adding another RG to the normalization factor (NF)^{16,25}. A V value was calculated for all ten candidate RGs and the non-rRNA RG subset, and both had a V_2/V_3 value below 0.15 (0.087 and 0.145, respectively), meaning that there is no benefit of using three RGs over two RGs in the NF (Fig. 4 and Supplementary Figure 3).

Validation of the selected reference genes. To examine the effect of utilizing different combinations of RGs in the NF, a normalized expression analysis was performed using expression data from PD630_RS15810 (RG4) grown under the four distinct growth conditions (Fig. 5). A box plot visualizing the non-normalized expression data for PD630_RS15810 was generated to show the general trends in expression level changes across the four growth conditions (Fig. 5A). The normalized relative fold changes in expression going from the HN condition to either the LN, TSB, or PHE condition were calculated using REST 2009 with one of six NFs, including RG6, RG10, RG3/RG7, RG3/RG8, RG7/RG8, and RG3/RG7/RG8 (Fig. 5). Confidence intervals of 95% were calculated by REST 2009 and used for comparison between different NFs. RG6 was chosen as an example of a poor RG candidate, while RG10 was chosen as DNA Polymerase IV had been used as a RG in other *Rhodococcus* spp.^{19,20}. All two-component combinations (NF₂) of RG3, RG7, and RG8 were examined as it was unclear whether there were differences between them, in addition to a NF comprising all three to confirm that there was no significant difference between a NF₂ and a NF₃.

The results of the normalized expression analysis revealed that there was no significant difference in the normalized relative fold change ratio going from the HN condition to either the LN, TSB, or PHE condition when using a NF₂ comprising any two selections of RG3, RG7, and RG8 or a NF₃ comprising all three genes. As the NF₂ comprising RG7 and RG8 (RG7/RG8) had the smallest 95% confidence interval, it was chosen as the optimal pair of RGs for future use in *R. opacus* under the tested growth conditions. RG7/RG8 normalization revealed that PD630_RS15810 expression with 95% confidence was downregulated between 0.054 to 0.088-fold when going from HN to LN, 0.355 to 0.575-fold when going from HN to TSB, and 0.494 to 0.916-fold when going from HN to PHE (Fig. 5).

The effect of non-optimal RGs was also examined in comparison to the NF₃ comprising RG3, RG7 and RG8. When PD630_RS15810 expression was normalized with RG6, there was a significantly different 95% confidence range ($p < 0.05$) of 0.120 to 0.240-fold downregulation when going from HN to LN, a comparable 0.430 to 0.824-fold downregulation going from HN to TSB, and a significantly different ($p < 0.05$) 4.009 to 7.703-fold upregulation when going from HN to PHE (Fig. 5). When PD630_RS15810 expression was normalized with the literature-based selection RG10, there was a significantly different 95% confidence range ($p < 0.05$) of 0.013 to 0.032-fold downregulation when going from HN to LN, a comparable 0.294 to 0.490-fold downregulation going from HN to TSB, and a comparable 0.480 to 0.971-fold downregulation when going from HN to PHE. These results demonstrate that a poorly selected RG can substantially alter the observed change in gene expression (e.g. RG6), and that RGs should be examined in all relevant growth conditions prior to use as they may be stable in some instances but not in others (e.g. RG10).

To determine an optimal NF for a scenario where rRNA has been depleted, a normalized expression analysis was performed again on PD630_RS15810 using the top three non-rRNA candidate RGs (RG5, RG7, and RG9). The normalized expression ratio of all NF₂ combinations of these three genes, in addition the NF₃ comprising all of them, was examined in comparison to the best NF₂ comprising RG7 and RG8 (Supplementary Figure 4). The NFs including RG5/RG7, RG5/RG9, and RG5/RG7/RG9 all produced 95% confidence normalized expression fold change ranges that were significantly different ($p < 0.05$), compared to the expression fold change range produced by RG7/RG8 when going from HN to TSB (Supplementary Figure 4C). The NF comprising RG7 and RG9 was comparable to RG7/RG8 in all scenarios, with 95% confidence normalized expression ranges of

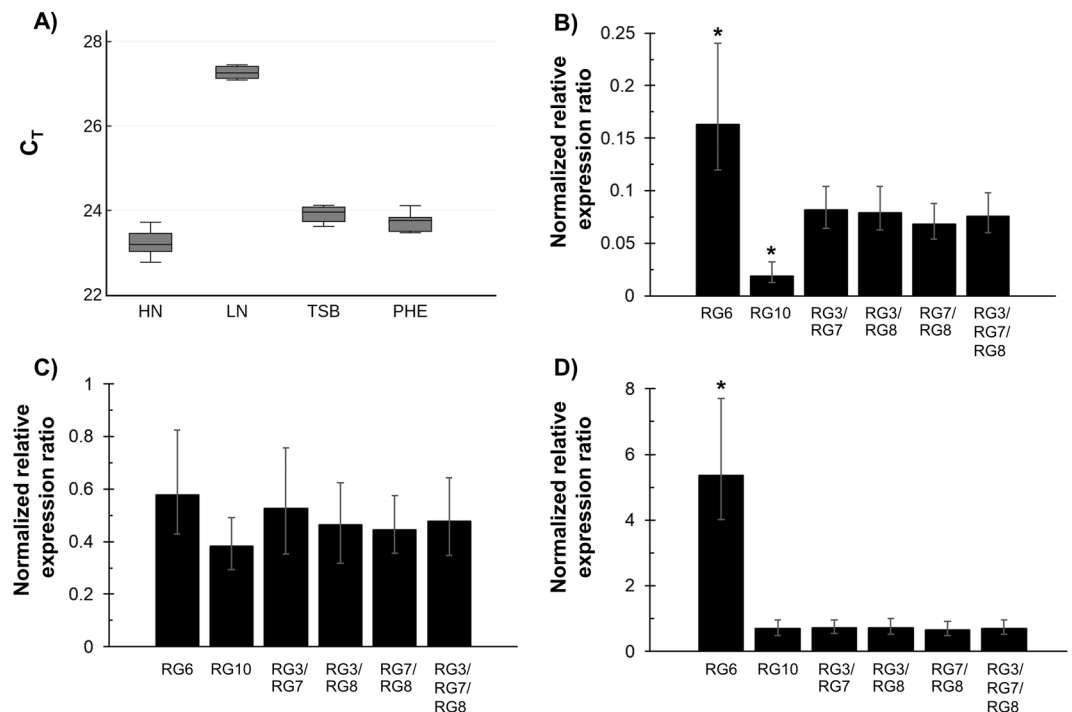


Figure 5. Effect of reference gene choice on RT-qPCR normalization. **(A)** Box plots of averaged PD630_RS15810 expression data (C_T) for all four growth conditions (HN, LN, TSB, and PHE). Each gray box represents the first through third quartiles, the solid black line represents the median, and the whiskers represent the minimum and maximum values. **(B–D)** The normalized relative expression ratio of PD630_LPD05540 going from HN to either LN **(B)**, TSB **(C)**, or PHE **(D)**. The expression data was normalized with a NF, including RG6, RG10, RG3/RG7, RG3/RG8, RG7/RG8, and RG3/RG7/RG8 using REST 2009. Error bars represent the 95% confidence interval (CI) as calculated by REST 2009. Stars indicate that a 95% CI range falls outside of the 95% CI range of the RG3/RG7/RG8 (and RG7/RG8) normalized ratio.

PD630_RS15810 similarly changing 0.072 to 0.127-fold when going from HN to LN, 0.534 to 0.909-fold when going from HN to TSB, and 0.550 to 1.096-fold when going from HN to PHE.

Conclusion

This study represents the first in-depth analysis of stable reference genes in *Rhodococcus* and identified two comparable normalization factors in *R. opacus* for samples with or without rRNA. The expression of ten candidate reference genes was examined across four distinct growth conditions (HN, LN, TSB, and PHE) using three mathematical models (BestKeeper, Normfinder, and geNorm) to identify and rank gene expression stability. Based on the geNorm V value and corroboration via a normalized gene expression analysis using REST 2009, only two reference genes are required for an optimal normalization factor in *R. opacus*. For samples containing rRNA, the two best reference genes were RG7 (PD630_RS25530; ATP-dependent Clp protease ATP-binding subunit ClpX) and RG8 (PD630_RS01395; 16S rRNA). For samples depleted in rRNA, the best two reference genes were RG7 and RG9 (PD630_RS37755; rRNA small subunit methyltransferase G). The diversity of growth conditions tested in this work bestows confidence in our selections of reference genes for *R. opacus*, in addition to providing a valuable starting point in the choice of reference genes when studying other *Rhodococcus* spp.

Materials and Methods

Strain and culture conditions. *Rhodococcus opacus* PD630 (DSMZ 44193) was grown in either a minimal salts medium as previously described (minimal media recipe B)⁹, with one of three distinct combinations of carbon and nitrogen sources (discussed in detail in the Gene expression studies section), or a rich tryptic soy broth (TSB) medium. Cultures were incubated at 30 °C and 250 rotations per minute (rpm). For the growth experiment, all cultures were grown in triplicate 10 mL cultures with an initial optical density at 600 nm (OD_{600}) of 0.2. Chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise noted. MIQE guidelines were applied as appropriate to the design and execution of experiments^{21,22}.

Candidate reference gene selection and primer design. The selection of candidate reference genes was based on either previously published transcriptomic data for *R. opacus*³ or literature for other *Rhodococcus* spp.^{19,20}. RG1 through RG9 were selected as potentially stable genes based on analysis of RNA-Seq data performed on *R. opacus* grown in a minimal salts medium with either glucose or phenol³. RG10 was chosen based on its gene annotation as DNA Polymerase IV, as other *Rhodococcus* spp. studies have used a DNA Polymerase IV gene as a

reference gene^{19,20}. RT-qPCR primers (Integrated DNA Technologies) were designed based on literature suggestions^{21,22} and guidelines discussed in the Supplementary Materials.

Primer amplification efficiency and specificity. To calculate primer amplification efficiency, a serial dilution of template DNA and a linear regression analysis of the resultant qPCR results were required. PCR product was obtained by using 0.5 μ L of GoTaq G2 polymerase (Promega), 10 μ L of GoTaq buffer, 1.5 μ L of a mixture containing the forward and reverse primers (5 μ M each), 0.5 μ L of genomic DNA (gDNA), 1 μ L of 10 mM dNTPs (GBiosciences), and 36.5 μ L of H₂O; and thermocycler conditions of 95 °C for 2 min, followed by 35 cycles of 95 °C for 15 s, 62 °C for 30 s, and 72 °C for 30 s. PCR products were gel extracted using the ZymoClean DNA Gel Recovery kit (Zymo Research) and prepared for qPCR using the DNA Clean and Concentrator kit (Zymo Research). Five rounds of a 10-fold serial dilution were performed on the purified PCR product. qPCR was performed in duplicate on the serially diluted PCR product using a Bio-Rad CFX96 real time thermocycler, TempPlate 96-well semi-skirt 0.1 mL PCR plates (USA Scientific), and Power SYBR Green PCR Master Mix (Applied Biosystems). 10 μ L of Power SYBR Green PCR Master Mix, 0.5 μ L of a mixture containing the forward and reverse primers (5 μ M each), 1 μ L of PCR product, and 8.5 μ L of H₂O were used for each reaction. Cycling conditions were 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s, 62 °C for 20 s, 72 °C for 20 s, and a fluorescence measurement. A fluorescence threshold value of 750 was set for all reactions. To confirm that no off-target amplification had occurred, a melting curve analysis was performed at the end of each qPCR program and a single melting peak was verified for each amplicon. A linear regression was performed on the threshold cycle (C_T) values observed from each serial dilution to confirm linearity (R^2) and to calculate the equation of the line, which was used to estimate the efficiency of the PCR (Equation 1; Supplementary Figure 1).

$$PCR \text{ Efficiency Percentage} = 100 * \left(\left[\frac{-1}{10^{\text{standard curve slope}}} \right] - 1 \right) \quad (1)$$

Gene expression studies. To examine gene expression stability, *R. opacus* was grown in four distinct growth conditions. The first condition was a rich TSB medium. The other three conditions were minimal salts media with carbon and nitrogen sources as follows: glucose with high nitrogen (HN; 2 g/L glucose and 1 g/L ammonium sulfate), glucose with low nitrogen (LN; 2 g/L glucose and 0.05 g/L ammonium sulfate), and phenol with high nitrogen (PHE; 0.75 g/L phenol + 1 g/L ammonium sulfate). A 5 mL seed culture in a 50 mL glass tube containing the minimal salts medium with 2 g/L glucose and 1 g/L ammonium sulfate was used for the HN, LN, and TSB growth conditions, while the PHE seed culture also had 0.3 g/L phenol added to acclimate the cells to the inhibitory aromatic. To remove remaining glucose and ammonium sulfate from the seed cultures, samples were centrifuged at 3500 relative centrifugal force (rcf), washed with minimal media containing no carbon or nitrogen sources, centrifuged again at 3500 rcf, and re-suspended in their final growth condition media. When the cultures reached mid-exponential phase, the cells were centrifuged at 3500 rcf and re-suspended in DNA/RNA Shield (Zymo Research), per the instructions. Samples were stored at –80 °C until needed for further use.

RNA extraction and cDNA synthesis. RNA was extracted from the triplicate biological samples stored at –80 °C in the DNA/RNA Shield using the Quick-RNA MiniPrep Plus kit (Zymo Research), per the instructions. The optional ZR BashingBead Lysis Tubes (0.1 and 0.5 mm beads; Zymo Research) were used to lyse the cells. All samples were treated with the TURBO DNA-free kit (Ambion) to remove any gDNA present in the samples and then purified using the RNA Clean and Concentrator kit (Zymo Research). To confirm that all gDNA was depleted, PCR was performed by using the previously described GoTaq protocol with primers targeting the genome. Gel electrophoresis was performed to confirm that no PCR amplification bands existed. Any samples with persisting gDNA were re-treated with the TURBO DNA-free kit, purified again, and examined via PCR and gel electrophoresis. RNA concentration and purity were quantified using a NanoDrop 2000c (all samples had a 260 nm/280 nm absorbance ratio of 2 to 2.1). 1 μ g of RNA per sample was converted to cDNA using the AffinityScript QPCR cDNA synthesis kit (Agilent Technologies).

RT-qPCR. RT-qPCR was performed in technical triplicate on the biological triplicate RNA extracts using a Bio-Rad CFX96 and Power SYBR Green PCR Master Mix (Applied Biosystems). 10 μ L of Power SYBR Green PCR Master Mix, 0.5 μ L of a mixture containing the forward and reverse primers (5 μ M each), 1 μ L of cDNA, and 8.5 μ L of H₂O were used for each reaction. Cycling conditions were 95 °C for 10 min, followed by 35 cycles of 95 °C for 15 s, 62 °C for 20 s, 72 °C for 20 s, and a fluorescence measurement. A threshold value of 750 was set for all reactions. To confirm that no off-target amplification had occurred, a melting curve analysis was performed at the end of each qPCR program and a single melting peak was verified for each amplicon (Supplementary Figure 2). Additionally, a negative control (no template control; NTC) was included for each primer pair (Table 1 and Supplementary Figure 2). Gel electrophoresis was also performed post RT-qPCR to confirm the existence of a single amplicon (Fig. 1A and Supplementary Figure 5). Gel images were captured using a DigiDoc-It imaging system and its accompanying Doc-ItLS software. The original Fig. 1A gel image was cropped, and contrast and exposure settings for the whole image were modified using Adobe Lightroom to improve image interpretation. The original image can be found as Supplementary Figure 5.

Reference gene expression stability analysis. The expression level stability of the ten candidate RGs across the four distinct growth conditions was assessed and ranked using three commonly used software tools: BestKeeper²³, NormFinder version 0.953²⁴, and geNorm (built into qBase + [Biogazelle])^{16,25} (Figs 3 and 4;

Supplementary Tables 2 and 3). Additionally, all analyses were performed again using just candidate RGs 1, 2, 4–7, 9, and 10 to accommodate a situation where rRNA (RG3 and 8) has been depleted from the sample (Fig. 3, Supplementary Figures 3 and 4, and Supplementary Table 4).

BestKeeper is a Microsoft Excel based tool that calculates the standard deviation and coefficient of variance of the C_T value for each candidate RG, and creates an index based on the geometric mean of the C_T values, which it uses to facilitate pair-wise comparisons against and between RGs. A Pearson correlation coefficient (r-value) is calculated for each RG, with higher r-values representing greater stability (max value of 1)²³. The technical triplicate C_T values for each biological replicate were averaged prior to entry into BestKeeper and all four growth conditions were pooled and analyzed together. Candidate RGs were ranked on both their C_T value standard deviation (values > 1.0 deemed unstable²³; Supplementary Table 2) and their r-value (Fig. 3A1 and A2, and Supplementary Tables 3 and 4).

Normfinder is also a Microsoft Excel based tool that utilizes a model-based approach to compare intra- and inter-group variation between candidate RGs and then generates a stability value and ranking for each candidate gene²⁴. The lower the stability value, the more stable the gene's expression. The technical triplicate C_T values for each biological replicate were averaged and samples from all four growth conditions were pooled. Values were then converted to a log-scale per Equation 2, where E is the amplicon specific PCR efficiency, prior to entry into Normfinder²⁴. Candidate RGs were ranked on their stability value (Fig. 3B1 and B2).

$$\text{Log transformed value} = E^{-C_{T\text{ avg}}} \quad (2)$$

geNorm, formerly a Microsoft Excel based tool that has been incorporated into the qBase + software package (Biogazelle), examines the pairwise variation between RGs and creates a stability M value, where lower values represent more stably expressed genes^{16,25}. The program iterates its calculations wherein it eliminates the gene with the highest M value and then recalculates the stability of the remaining genes. Thus, geNorm creates a ranking of RG stability (Fig. 3C1 and C2). A geNorm M value lower than 1 is generally considered stable for heterogeneous growth conditions²². geNorm can also determine the minimum number of RGs required for optimal RT-qPCR normalization. Various normalization factors (NF) are calculated by first taking the geometric mean of the two most stable RGs' C_T values and then generating additional factors by adding the next best RG until all RGs are used. A pairwise variation (V_n/V_{n+1}) V value is then calculated by comparing the benefit of going from NF_n to NF_{n+1} , where n is a number of RGs (Fig. 4 and Supplementary Figure 3). Once the V value is less than 0.15, there is no additional benefit of going from n to n + 1^{16,25}.

Validation of reference gene selection. The effect of different combinations of the top candidate RGs was examined by creating multiple different NFs and normalizing the same set of expression data using REST 2009 (Qiagen). The relative expression ratio data of PD630_RS15810 going from the HN condition to either the LN, TSB, or PHE condition was normalized by a NF_2 comprising any two of the top three RGs identified across all analyses (RG3, RG7, and RG8), in addition to a NF_3 comprising all three. Additionally, RG6 was used as an example of an unsuitable RG, while RG10 was used based on literature references for other *Rhodococcus* spp^{19,20}. 95% confidence intervals were generated by REST 2009. If 95% confidence intervals failed to overlap, they were labelled as significantly different ($p < 0.05$). As two of the top three RGs identified encode rRNAs (RG3 and RG8), which are frequently depleted when performing RNA-Seq, additional NFs comprising non-rRNA RGs (RG5, RG7, and RG9) were examined in comparison to the overall top NF combination (RG7 and RG8; Supplementary Figure 4).

Data availability. The datasets generated during the current study are either included in the manuscript or Supplementary Materials or are available from the corresponding author on reasonable request.

References

- Alvarez, H. M., Mayer, F., Fabritius, D. & Steinbuchel, A. Formation of intracytoplasmic lipid inclusions by *Rhodococcus opacus* strain PD630. *Arch Microbiol* **165**, 377–386 (1996).
- Holder, J. W. *et al.* Comparative and Functional Genomics of *Rhodococcus opacus* PD630 for Biofuels Development. *PLoS Genet* **7**, e1002219, <https://doi.org/10.1371/journal.pgen.1002219> (2011).
- Yoneda, A. *et al.* Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating *Rhodococcus opacus* PD630. *Nucleic Acids Res* **44**, 2240–2254, <https://doi.org/10.1093/nar/gkw055> (2016).
- Kurosawa, K., Laser, J. & Sinskey, A. J. Tolerance and adaptive evolution of triacylglycerol-producing *Rhodococcus opacus* to lignocellulose-derived inhibitors. *Biotechnol. Biofuels* **8**, <https://doi.org/10.1186/s13068-015-0258-3> (2015).
- Lin, S. Y. & Dence, C. W. *Methods in Lignin Chemistry* (Springer, 1992).
- Hollinshead, W. D., Henson, W. R., Abernathy, M., Moon, T. S. & Tang, Y. J. Rapid metabolic analysis of *Rhodococcus opacus* PD630 via parallel ¹³C-metabolite fingerprinting. *Biotechnol. Bioeng.* **113**, 91–100, <https://doi.org/10.1002/bit.25702> (2016).
- Kurosawa, K., Plassmeier, J., Kalinowski, J., Ruckert, C. & Sinskey, A. J. Engineering L-arabinose metabolism in triacylglycerol-producing *Rhodococcus opacus* for lignocellulosic fuel production. *Metab. Eng.* **30**, 89–95, <https://doi.org/10.1016/j.ymben.2015.04.006> (2015).
- Kurosawa, K., Wewetzer, S. J. & Sinskey, A. J. Engineering xylose metabolism in triacylglycerol-producing *Rhodococcus opacus* for lignocellulosic fuel production. *Biotechnol. Biofuels* **6**, 134, <https://doi.org/10.1186/1754-6834-6-134> (2013).
- DeLorenzo, D. M., Henson, W. R. & Moon, T. S. Development of Chemical and Metabolite Sensors for *Rhodococcus opacus* PD630. *ACS Synth. Biol.* **6**, 1973–1978, <https://doi.org/10.1021/acssynbio.7b00192> (2017).
- DeLorenzo, D. M., Rottinghaus, A. G., Henson, W. R. & Moon, T. S. Molecular toolkit for gene expression control and genome modification in *Rhodococcus opacus* PD630. *ACS Synth. Biol.* <https://doi.org/10.1021/acssynbio.7b00416> (2018).
- Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
- Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63, <https://doi.org/10.1038/nrg2484> (2009).

13. Zhu, S., Qing, T., Zheng, Y., Jin, L. & Shi, L. Advances in single-cell RNA sequencing and its applications in cancer research. *Oncotarget* **8**, 53763–53779, <https://doi.org/10.18632/oncotarget.17893> (2017).
14. Bustin, S. A. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol* **29**, 23–39 (2002).
15. Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Res.* **6**, 986–994 (1996).
16. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, research0034.0031-0011 (2002).
17. Madrigal, R. G. *et al.* Use of Serial Quantitative PCR of the vapA Gene of *Rhodococcus equi* in Feces for Early Detection of *R. equi* Pneumonia in Foals. *J Vet Intern Med* **30**, 664–670, <https://doi.org/10.1111/jvim.13828> (2016).
18. Mohn, W. W. *et al.* Gene cluster encoding cholocate catabolism in *Rhodococcus* spp. *J Bacteriol* **194**, 6712–6719, <https://doi.org/10.1128/JB.01169-12> (2012).
19. Goncalves, E. R. *et al.* Transcriptomic assessment of isozymes in the biphenyl pathway of *Rhodococcus* sp. strain RHA1. *Appl Environ Microbiol* **72**, 6183–6193, <https://doi.org/10.1128/AEM.00947-06> (2006).
20. Szokol, J. *et al.* Induction and carbon catabolite repression of phenol degradation genes in *Rhodococcus erythropolis* and *Rhodococcus jostii*. *Appl. Microbiol. Biotechnol.* **98**, 8267–8279, <https://doi.org/10.1007/s00253-014-5881-6> (2014).
21. Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622, <https://doi.org/10.1373/clinchem.2008.112797> (2009).
22. Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M. & Nguyen, M. A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines. *Methods* **50**, S1–5, <https://doi.org/10.1016/j.jymeth.2010.01.005> (2010).
23. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* **26**, 509–515 (2004).
24. Andersen, C. L., Jensen, J. L. & Orntoft, T. F. 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.* **64**, 5245–5250, <https://doi.org/10.1158/0008-5472.CAN-04-0496> (2004).
25. Hellemans, J., Mortier, G., De Paep, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* **8**, R19, <https://doi.org/10.1186/gb-2007-8-2-r19> (2007).
26. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq 2. *Genome Biol* **15**, 550, <https://doi.org/10.1186/s13059-014-0550-8> (2014).
27. Kurosawa, K., Boccazzi, P., de Almeida, N. M. & Sinskey, A. J. High-cell-density batch fermentation of *Rhodococcus opacus* PD630 using a high glucose concentration for triacylglycerol production. *J. Biotechnol.* **147**, 212–218, <https://doi.org/10.1016/j.jbiotec.2010.04.003> (2010).
28. Chen, Y. *et al.* Integrated omics study delineates the dynamics of lipid droplets in *Rhodococcus opacus* PD630. *Nucleic Acids Res* **42**, 1052–1064, <https://doi.org/10.1093/nar/gkt932> (2014).
29. Goswami, L., Tejas Nambodiri, M. M., Vinoth Kumar, R., Pakshirajan, K. & Pugazhenthii, G. Biodiesel production potential of oleaginous *Rhodococcus opacus* grown on biomass gasification wastewater. *Renewable Energy* **105**, 400–406, <https://doi.org/10.1016/j.renene.2016.12.044> (2017).
30. Bender, D. A. *Amino Acid Metabolism*. 3rd edn, (Wiley-Blackwell, 2012).
31. Edwards, K. J. & Saunders, N. A. Real-time PCR used to measure stress-induced changes in the expression of the genes of the alginate pathway of *Pseudomonas aeruginosa*. *J Appl Microbiol* **91**, 29–37 (2001).
32. Pinto, F. *et al.* Improving a *Synechocystis*-based photoautotrophic chassis through systematic genome mapping and validation of neutral sites. *DNA Res.* **22**, 425–437, <https://doi.org/10.1093/dnares/dsv024> (2015).
33. O'Neil, D., Glowatz, H. & Schlumpberger, M. Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Curr Protoc Mol Biol* Chapter 4, Unit 4.19, <https://doi.org/10.1002/0471142727.mb0419s103> (2013).

Acknowledgements

This work was supported by U.S. Department of Energy (DE-SC0012705 and DE-SC0018324 to TSM). DMD is the recipient of an NSF Graduate Research Fellowship, DGS-1143954.

Author Contributions

D.M.D. performed all experiments. D.M.D. and T.S.M. analyzed data and wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-24486-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018