SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

Growth phase proteomics of the heterotrophic marine bacterium *Ruegeria pomeroyi*

Dasha Krayushkina[1], Emma Timmins-Schiffman[2], Jessica Faux[3], Damon H. May[2], Michael Riffle[4], H. Rodger Harvey[3] & Brook L. Nunn[2*]

The heterotrophic marine bacterium, *Ruegeria pomeroyi*, was experimentally cultured under environmentally realistic carbon conditions and with a tracer-level addition of ^{13}C-labeled leucine to track bacterial protein biosynthesis through growth phases. A combination of methods allowed observation of real-time bacterial protein production to understand metabolic priorities through the different growth phases. Over 2000 proteins were identified in each experimental culture from exponential and stationary growth phases. Within two hours of the ^{13}C-labeled leucine addition, *R. pomeroyi* significantly assimilated the newly encountered substrate into new proteins. This dataset provides a fundamental baseline for understanding growth phase differences in molecular physiology of a cosmopolitan marine bacterium.

Background & Summary

Ecosystems are sustained by biogeochemical cycling of major elements essential to organism survival. Principally driven by microbes, elemental cycling pathways depend on processes including the uptake and incorporation of inorganic nutrients, consumption by heterotrophic consumers, and degradation and remineralization of organic material. Proteins catalyze and mediate these processes essential for organism-environment interaction, making them windows into ecosystem health and function[1].

Protein synthesis and consumption essential to many marine biogeochemical processes are difficult to fully explore due to the limited ability to examine proteins *in situ*. To minimize this limitation, the rapidly advancing field of proteomics is providing insights into many processes that have not previously been examined[2–4]. Bacterial proteins are often masked by the more abundant inputs of eukaryotic marine organisms, making it technically difficult to detect the full suite of active bacterial processes in mixed environmental samples[5–10]. Datasets that provide a foundation to better examine ecosystem cycling, degradation, and preservation of proteins that emphasize bacteria's role in these processes are necessary to improve current understanding of biogeochemical cycling. Understanding diverse bacterial functional roles in the environment will require surveys of natural communities coupled with detailed, in-depth analyses of cultured species. Through high resolution mass spectrometry-based proteomics, we present the proteome of *Ruegeria pomeryoi* through cell cycle phases in realistically low carbon oceanic conditions. Further, we expand the environmentalist's molecular toolbox of methods by demonstrating how proteomic profiling of *in situ* ^{13}C incorporation into newly synthesized proteins can inform us about the timing of different cellular metabolic processes.

Ruegeria pomeroyi is a marine heterotrophic bacterium that, in part due to its generalist nature, has become the foundation of many lab manipulations to understand bacteria-environment interactions[11,12]. *R. pomeroyi's* sequenced genome makes it an ideal species for detailed investigations, such as proteomics. Fifty percent of the *R. pomeroyi* genome is considered sensitive to changing transcript abundance in response to abiotic environmental change[11]. Although proteomic profiles and cellular physiology are known to shift throughout bacterial culture experiments due to depletion in local nutrients, space, and oxygen[13,14], the underlying molecular physiological response is poorly understood through the progression of different growth phases.

[1]University of Washington, Department of Psychiatry and Behavioral Sciences, Seattle, WA, 98195, USA. [2]University of Washington, Department of Genome Sciences, Seattle, WA, 98195, USA. [3]Old Dominion University, Department of Ocean, Earth & Atmospheric Sciences, Norfolk, VA, 23529, USA. [4]University of Washington, Department of Biochemistry, Seattle, WA, 98195, USA. *email: brookh@uw.edu

Growth Phase	Protocol 1	Protocol 2	Dataset Identifier
Early Exponential	DDA proteomics		PXD008661[27]
Late Exponential	DDA proteomics	^{13}C-leucine	PXD008661[27]
Early Stationary	DDA proteomics		PXD008661[27]
Late Stationary	DDA proteomics		PXD008661[27]

Table 1. Description of samples used in analysis. "DDA proteomics" refers to data dependent acquisition proteomics, or whole proteome profiling.

It is essential to understand molecular level metabolic alterations to controlled, realistic growth conditions through the growth phases if we are to truly characterize how species respond to *in situ* environments and perturbations. Here, we thoroughly characterize the proteome of *Ruegeria pomeroyi*, a member of the Roseobacter clade, using a mass spectrometry technology capable of sequencing relatively low abundance proteins to better understand the baseline physiology of a heterotrophic marine bacterium across growth phases (Table 1). We employed whole proteome profiling, or data dependent acquisition proteomics, to uncover complete pathways and how they contribute to different growth phases (Fig. 1). In addition, *R. pomeroyi in situ* protein synthesis was tracked using trace amounts of ^{13}C-labeled leucine.

Methods

Bacterial culture and growth. *R. pomeroyi* DSS-3 was maintained on a marine basal media (28 salinity) with vitamin supplements and 10 mM glucose as the sole carbon source. Cultures were kept in continual culture, transferring every 48–72 hours once glucose concentrations were depleted to <0.1 mM.

Experimental cultures were sampled at 8-hour intervals for 48 hours to determine growth dynamics. Each experimental growth curve was established for *R. pomeroyi* by spectrophotometric analysis at an optical density of 600 nm (OD_{600}) and calibrated using 5 mL sub-samples preserved in 2% formalin and filtered onto black 0.2 µm polycarbonate filters for cell counts. Fluoroshield with DAPI histology mounting medium (Sigma-Aldrich, St. Louis, MO) was used to stain cells for direct counting[15]. Cells were visualized and counted with an Olympus BX50 microscope (Shinjuku, Tokyo) at 1250x magnification.

For proteomics sampling across growth phases, three replicate experimental cultures in 2 L flasks (and a 1 L media blank) were grown in marine basal media with 0.1 mM glucose. Three separate 2 L replicates were also grown using stable isotope probing to track assimilation of a dissolved free amino acid into new proteins at early exponential and late exponential growth phases. $^{13}C_6$-labeled leucine (^{13}C-Leu, Sigma-Aldrich) substrate was added 23 hours after inoculation, (616 nM) and the late exponential culture harvested 2 hours later to capture protein synthesis during exponential growth. Based on measured cellular growth rates and the high and low estimates for carbon content per cell in marine bacteria[16], this avoided ^{13}C-Leu depletion prior to the next sample time point (36 hrs) (Figshare File 1[17]). To capture protein synthesis during late exponential growth, a second addition of ^{13}C-Leu was added at 36 hours and the culture harvested 2 hours later to track ^{13}C-labeled proteins that remained plus newly synthesized ^{13}C-proteins.

Sampling procedures for proteomics. Triplicate bacterial cultures were sampled for proteins at 25 hours, 38 hours, 44 hours, and 66 hours, corresponding to early exponential, late exponential, early stationary, and late stationary growth phases, respectively. To track *in situ* newly synthesized proteins, triplicate cultures with the ^{13}C-leucine label addition were sampled at 25 hours and 38 hours. Samples were taken from each culture flask for data dependent acquisition (DDA) proteomic analysis with the volumes adjusted to collect similar protein amounts over cellular growth. A 10% (w/v; 4 °C) solution of trichloroacetic acid (TCA) was added to each sample to rapidly kill the bacteria and precipitate the proteins. Samples were refrigerated at 4 °C for 1 hour prior to filtration through a 0.2 µm polycarbonate filter and frozen at −80 °C until extraction.

Protein digestion. Cells were isolated by placing the filters in 1.5 mL tubes with 300 µL of 2 M urea. Samples were placed on ice and shaken on a bead beater 5 times (1 minute each) to shake the cells off the filters. The filters were removed and the cells were lysed using a microtip sonicator (Branson, Danbury, CT) in 5 rounds, 15 seconds for each sonication, placing on ice between sonication events.

Samples were digested following [18]. Briefly, proteins were reduced with 2.5 µl of 200 mM tris(2-carboxyethyl) phosphine for 1 hour at 37 °C. Reduced proteins were then alkylated with 20 µl of 200 mM iodoacetamide for 1 hour in the dark. Ammonium bicarbonate was added to dilute the urea, followed by an addition of methanol to increase the solubility of membranes. Trypsin was administered to samples at an enzyme-to-protein ratio of 1:50. Samples were vortexed and incubated overnight at 37 °C. Liquid was evaporated in a speed vac and dried peptides were reconstituted in 5% acentonitrile (ACN) + 0.1% trifluoroacetic acid.

A macro-spin C18 column (NestGroup, Southborough, MA) was used to desalt digested peptide samples following the manufacturer's instructions. Desalted peptides were dried via speedvac and 0.5 µg was brought up to a standard volume with 5% ACN with 0.1% formic acid (FA) for analysis by tandem mass spectrometry (MS/MS).

LC-MS/MS: Liquid chromatography and tandem mass spectrometry. Tandem mass spectrometry (LC-MS/MS) was performed on a Q-Exactive (QE) mass spectrometer (Thermo Fisher). Peptides were chromatographically separated on a 20 cm long, 75 µm id fused silica capillary column packed with C18 particles (Magic C18AQ, 100 Å, 5 µm; Bruker, Billerica, MA) preceded by a pre-column (2 cm length, 100 µm; Magic C18AQ, 200 Å, 5 µm, Bruker). Peptides were eluted with an acidified (FA, 0.1% w/v) water-ACN gradient (5–35% ACN)

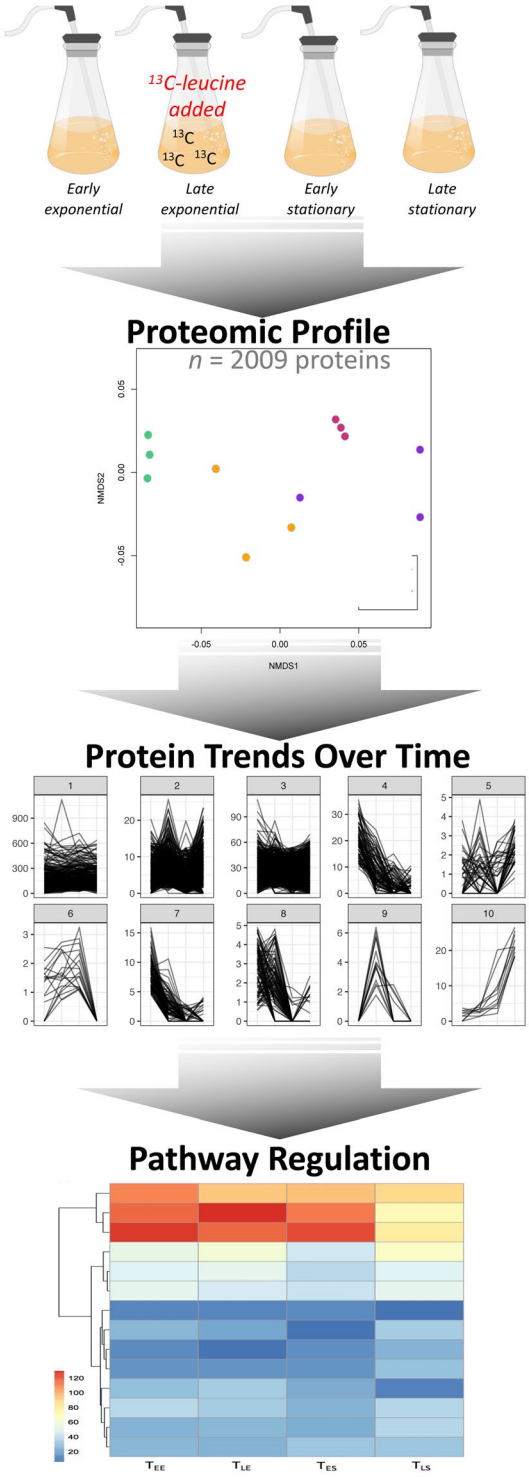

Fig. 1 Schematic of experimental design and applications for data use. Proteomes were sequenced from four growth phases of *R. pomeroyi*. These data can later be used to discover trends of protein abundance over time and how different molecular pathways are regulated throughout bacterial growth.

over 90 minutes using a nanoAcquity Ultra Performance LC (Waters; Milford, MA) at 0.25–0.3 µl/min in line with the QE. QE settings included a 30.0 s dynamic exclusion, full MS resolution of 35000, full MS AGC target of 5e5, full MS maximum IT of 100 ms, full MS scan range of 400–1600 *m/z*, MS/MS resolution of 17500, MS/MS AGC target of 5e4, MS/MS maximum IT of 50 ms, MS/MS loop count of 20, MS/MS MSX count of 1, MS/MS isolation window of 2 *m/z*, and MS/MS NCE of 25.0.

Raw files were searched against the *R. pomeroyi* proteome (http://www.uniprot.org/proteomes/UP000001023, downloaded July 11, 2013) with added common laboratory contaminants (http://www.thegpm.org/crap/) using

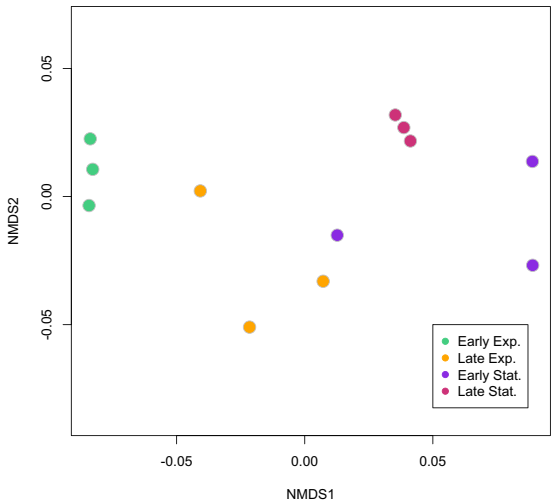

Fig. 2 Nonmetric multidimensional scaling plot of the growth experiment (*R. pomeroyi* incubated without ^{13}C) proteomics data. Each point represents the NSAF values across the entire proteome for a sample and points fall closer together if they have more similar proteomic profiles. Color indicates growth phase: Early Exponential = green, Late Exponential = orange, Early Stationary = purple, Late Stationary = pink.

Comet v. 2016.01 rev. 2[18,19]. The final protein database contained 4260 predicted *R. pomeroyi* protein sequences and 4531 sequences total. All Comet settings can be found in the comet.params file (Figshare File 2[17]) and included a concatenated decoy search, peptide mass tolerance of 2.1 amu, fully trypsin digested peptides, 2 allowed missed cleavages, variable modification of 15.9949 on methionine with a maximum of 5 per peptide, fragment bin tolerance of 1.0005 and fragment bin offset of 0.4. The Trans-Proteomic Pipeline (Peptide and ProteinProphet) was run on all files with a probability cut-off of 0.9 (corresponding to a false discovery rate of about 0.0035) for peptide and protein inferences[20,21]. Abacus[22] was used to calculate normalized spectral abundance factor (NSAF) values and find consensus protein inferences across all pep.xml and prot.xml files (Figshare File 3[17]). Stringent cut-offs were chosen for Abacus parameters to ensure inclusion of high confidence proteins in the dataset. Parameters for Abacus included 0.99 probability for the minimum PeptideProphet score of a best peptide match for a protein; 0.50 probability for the minimum PeptideProphet score for a peptide to be considered by Abacus; 0.90 probability for a protein group ProteinProphet score. The Abacus output for proteins with at least 2 unique spectral counts across replicates was used for downstream analyses (Figshare File 4[17]). The resulting dataset contained 2009 proteins, or 47% of the predicted *R. pomeroyi* proteome, which is within the range of proteome coverage for other bacterial proteomes[23–25]. All proteomics data are available via the ProteomeXchange Consortium via the PRIDE partner repository[26], accession number PXD008661[27].

^{13}C-leucine samples were similarly searched, except the following setting was changed in the comet.params file: variable_mod02 = 6.0 L 0 3–1 0 0, allowing the searching algorithm to identify peptides with leucine with or without the modification of +6, indicative of the 6- ^{13}carbons that were included in the leucine tracer experiment. TPP and Abacus were deployed as described above (Figshare File 5[7]).

Label-free protein quantification and statistical analysis. Peptide spectral counts per protein were used to estimate relative abundance of proteins for each sample in the unlabeled experiment. This was performed under the accepted assumption that the more abundant a protein is within a sample, the more peptides it will generate upon digestion, thus allowing for more spectra to be detected from those peptides[28]. Contaminant proteins were removed from the data prior to analysis.

A nonmetric multidimensional scaling (NMDS) analysis was performed on the NSAF data for the growth curve experiment using a Bray-Curtis dissimilarity matrix with log(x + 1)-transformed data in the vegan package[29] in R (https://www.R-project.org/). NMDS was used to quantify and visualize the pairwise dissimilarity in protein composition between samples of each time point.

Data Records

Figshare File 1[17]: Calculations used to derive the amount of ^{13}C-labeled leucine incorporated into cells. File is in .docx format.

Figshare File 4[17]: Proteomics data from the growth experiment (the proteomics without ^{13}C label). The first column contains the protein identifier (Uniprot accession), followed by the protein length, then the total number of unique peptides detected for that protein across all biological and technical replicates ('No. Unique Peptides Across All Replicates). The subsequent columns contain the total spectral counts and Normalized Spectral Abundance Factor values for each mass spectrometry analysis, (e.g., Early Exp 1, the first technical replicate for early exponential stage). Table is in .xlsx format.

Figshare File 5[17]: Proteomics data from the ^{13}C-labeled leucine experiment. The data are arranged similarly to Figshare File 4, except the last 4 columns are indicators of label incorporation into the protein. An asterisk in the column indicates that a peptide with a ^{13}C label was detected for that protein. Table is in .xlsx format.

Figshare File 6[17]: Raw proteomics mass spectrometry files record. Files were deposited in the online repository PRIDE under the ProteomExchange with identifier PXD008661[27]. The first column of the table contains the file name, the second column is the file format, third is how the file was produced, and fourth is the growth stage the data are derived from. Table is in .xlsx format.

Technical Validation

Technical and biological replication (reproducibility on the mass spectrometer and within time points) were assessed using NMDS (Fig. 2). There was generally clustering among time point replicates, with the exception of one of the early stationary phase points (purple) that clustered closer to late exponential phase samples than other early stationary phase samples. Throughout the mass spectrometry experiment, quality control standards were run between sets of samples to ensure reproducibility of the mass spectrometry data.

Code availability

Figshare File 2[17]: Parameters file used for Comet searches of mass spectrometry data. File is in .docx format.

Figshare File 3[17]: Parameters file used for Abacus analysis of Comet search results. File is in .docx format.

Received: 7 August 2019; Accepted: 7 November 2019;
Published online: 03 December 2019

References

1. Hawley, A. K., Brewer, H. M., Norbeck, A. D., Pasa-Tolic, L. & Hallam, S. J. Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *PNAS* **111**, 11395–11400 (2014).
2. Armengaud, J., Hartmann, E. M. & Bland, C. Proteogenomics for environmental microbiology. *Proteomics* **13**, 18–19 (2013).
3. Nunn, B. L. *et al*. Diatom proteomics reveals unique acclimation strategies to mitigate Fe limitation. *PLoS ONE* **8**, e75653 (2013).
4. Bryson, S. *et al*. Proteomic stable isotope probing reveals taxonomically distinct patterns in amino acid assimilation by coastal marine bacterioplankton. *mSystems* **1**, e00027–15 (2016).
5. Verberkmoes, N. C., Denef, V. J., Hettich, R. L. & Banfield, J. F. Systems biology: functional analysis of natural microbial consortia using community proteomics. *Nature Reviews Microbiology* **7**, 196–205 (2009).
6. Ram, R. J. *et al*. Community proteomics of a natural microbial biofilm. *Science* **308**, 1815–1920 (2005).
7. Bertrand, E. M., Saito, M. A., Jeon, Y. J. & Neilan, B. A. Vitamin B12 biosynthesis gene diversity in the Ross Sea: the identification of a new group of putative polar B12 biosynthesizers. *Environmental Microbiology* **13**, 1285–1298 (2011).
8. Moore, E. K., Nunn, B. L., Goodlett, D. R. & Harvey, H. R. Identifying and tracking proteins through the marine water column: insights into the inputs and preservation mechanisms of protein in sediments. *Geochimica et Cosmochimica Acta* **83**, 324–359 (2012).
9. Moore, E. K., Nunn, B. L., Faux, J. F., Goodlett, D. R. & Harvey, H. R. Evaluation of electrophoretic protein extraction and database-driven protein identification from marine sediments. *Limnology and Oceanography*. *Methods* **10**, 353–366 (2012).
10. Becher, D., Bernhardt, J., Fuchs, S. & Riedel, K. Metaproteomics to unravel major microbial players in leaf litter and soil environments: challenges and perspectives. *Proteomics* **13**, 2895–2909 (2013).
11. Christie-Oleza, J. A., Fernandez, B., Nogales, B., Bosch, R. & Armengaud, J. Proteomic insights into the lifestyle of an environmentally relevant marine bacterium. *The ISME Journal* **6**, 124–135 (2012).
12. Durham, B. P. *et al*. Recognition cascade and metabolite transfer in a marine bacteria-phytoplankton model system. *Environmental Microbiology* **19**, 3500–3513 (2017).
13. Amsler, C. D., Cho, M. & Matsumara, P. Multiple factors underlying the maximum motility of *Escherichia coli* as cultures enter post-exponential growth. *Journal of Bacteriology* **175**, 6238–6244 (1993).
14. Arnold, R. J., Karty, J. A., Ellington, A. D. & Reilly, J. P. Monitoring the growth of a bacteria culture by MALDI-MS of whole cells. *Analytical Chemistry* **71**, 1990–1996 (1999).
15. Porter, K. G. & Feig, Y. S. The use of DAPI for identifying and counting aquatic microflora. *Limnology and Oceanography* **25**, 943–948 (1980).
16. Vrede, K., Heldal, M., Norland, S. & Bratbak, G. Elemental composition (C, N, P) and cell volume of exponentially growing and nutrient-limited bacterioplankton. *Applied and Environmental Microbiology* **68**, 2965–2971 (2002).
17. Krayushkina, D. *et al*. Growth phase proteomics of the heterotrophic marine bacterium *Ruegeria pomeroyi*. *figshare*, https://doi.org/10.6084/m9.figshare.c.4675310 (2019).
18. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **13**, 22–24 (2012).
19. Eng, J. K. *et al*. A deeper look into comet – implementation and features. *Journal of the American Society of Mass Spectrometry* **11**, 1865–1874 (2015).
20. Pedrioli, P. G. A. Trans-Proteomic Pipeline: A Pipeline for Proteomic Analysis. In: Hubbard, S. & Jones, A. (eds) *Proteome Bioinformatics*. *Methods in Molecular Biology*™ *(Methods and Protocols)* vol 604 (Humana Press, 2010).
21. Deutsch, E. W. *et al*. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics Clin. Appl* **9**, 745–754 (2015).
22. Fermin, D., Basrur, V., Yocum, A. K. & Nesvizhskii, A. I. Abacus: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. *Proteomics* **11**, 1340–1345 (2011).
23. Nunn, B. L., Slattery, K. V., Cameron, K. A., Timmins-Schiffman, E. & Junge, K. Proteomics of *Colwellia psychrerythraea* at subzero temperatures – a life with limited movement, flexible membrantes and vital DNA repair. *Environmental Microbiology* **7**, 2319–2335 (2015).
24. Lasaosa, M. *et al*. A 2D reverse-phase x ion-pair reversed-phase HPLC-MALDI TOF/TOF-MS approach for shotgun proteome analysis. *Analytical and Bioanalytical Chemistry* **393**, 1245–1256 (2008).
25. Lipton, M. S. *et al*. Global analysis of the *Deinococcus radiodurans* proteome by using accurate mass tags. *PNAS* **17**, 11049–11054 (*2002*).
26. Vizcaíno, J. A. *et al*. update of the PRIDE database and related tools. *Nucleic Acids Research* **44**, D447–D456 (2016).
27. Timmins-Schiffman, E. & Nunn, B.L. *Ruegeria pomeroyi* growth curve. *PRIDE*, https://identifiers.org/pride.project:PXD008661 (2019).
28. Liu, H., Sadygov, R. G. & Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical Chemistry* **76**, 4193–4201 (2004).
29. Oksanen, J. *et al*. vegan: community ecology package. *R package version 2.4-1*, https://CRAN.R-project.org/package=vegan (2016).

Acknowledgements
This work was funded by NSF Chemical Oceanography grants (#1233014 and #1633939) to BLN and to HRH (#1233589). We would like to thank Mary Ann Moran for providing the *Ruegeria pomeroyi* used for the culture in this experiment. Michael J. MacCoss shared his lab and work space with us for the completion of this work. Ohad Manor created the script use for running BLAST and Maitreya Dunham gave counsel on handling of the ^{13}C-leucine data. Jake Portnova provided assistance with analysis in R. Julia Robbins created the erlenmeyer flask image in Figure 1. Thank you, as always, to TAN and IJE for keeping the science going.

Author contributions
Dasha Krayushkina – D.K. was responsible for data analysis, interpretation, and drafting of the manuscript. Emma Timmins-Schiffman – E.T.S. was equally responsible with DK for data analysis and interpretation, and drafting of the manuscript. Jessica Faux – J.F. assisted with experimental design and carried out the bacterial cultures and sampling. Damon H. May – D.H.M. assisted with data analysis. Michael Riffle – M.R. assisted with data analysis. H. Rodger Harvey – H.R.H. was principally responsible for experimental design and oversight. Brook L. Nunn – B.L.N. assisted with experimental design, was responsible for collection of mass spectrometry data, and assisted with data analysis.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to B.L.N.

Reprints and permissions information is available at www.nature.com/reprints.

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 Creative Commons Public Domain Dedication waiver http://creativecommons.org/publicdomain/zero/1.0/ applies to the metadata files associated with this article.

© The Author(s) 2019

© 2019. This work is published under http://creativecommons.org/licenses/by/4.0/(the “License”). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.