UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

Insights into the light-dark physiology of a purple sulfur bacterium and its relationship with its intimate sulfate reducing partner

Permalink

https://escholarship.org/uc/item/49x4p852

ISBN

9798186384274

Author

Castillo Villaseñor, Braulio Saúl

Publication Date

2026-03-26

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Santa Barbara

Insights into the light-dark physiology of a purple sulfur bacterium and its relationship with its intimate sulfate reducing partner

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Ecology, Evolution, and Marine Biology

by

Braulio Saúl Castillo Villaseñor

Committee in charge:

Professor Elizabeth Wilbanks, Chair

Professor Morgan Raven

Professor Alyson Santoro

June 2026

The dissertation of Braulio Saúl Castillo Villaseñor is approved.

_____

Morgan Raven

_____

Alyson Santoro

_____

Elizabeth Wilbanks, Committee Chair

March 2026

Insights into the light-dark physiology of a purple sulfur bacterium and its relationship with its intimate sulfate reducing partner

Copyright © 2026

by

Braulio Saúl Castillo Villaseñor

ACKNOWLEDGEMENTS

First of all, I need to thank the people who made this experience possible. Lizzy, you went out on a limb for me way back in 2017. You believed I could be a scientist, and afforded me this great opportunity to do some very interesting work that I am **so** excited to publish with you after this. I had a lot to learn back then (and I still really do), but I can definitely say the growth has been tremendous for me in the years since. I learned so much about thinking like a scientist from you: writing, experiments, field work, and just the great big ol' world of microbial diversity. It has been such a rich experience, and very integral to my 20s. Thank you.

Next up, my committee. Alyson and Morgan, I am very grateful to have met you when I started this journey when I first came to Santa Barbara in 2018. I still remember how you emphasized setting up my lab bench for success when we were doing ^{33}S measurements Morgan, and it's something I carried throughout my benchwork as a scientist. Alyson, being in your seminars and being a student in your classroom was also very valuable. I'm happy to have been afforded that opportunity with you, otherwise I might've never attempted to learn about something as cool as stable isotope ecology. By being in my committee, you both have seen my mistakes and missteps throughout this process, I am thankful you were there for my quals and helping me change the shape of my dissertation such that it communicated my science better. Thank you for your guidance.

My collaborators. Thank you to Shady Amin, Michael Ochensenkühn, Salah Abdelrazig and Lisa Coe. Together you helped me put together a great metabolomics dataset which was integral to the third chapter of this dissertation. Manuel Kleiner and Olivia Mathieson, thank you for helping me generate our proteomic dataset. I didn't even get that far with this dissertation, but it's really cool data.

My lab. Past and present. Taruna you were very helpful in helping me feel oriented as a young person in the lab. Also, your organization and bioinformatics skills got me out of so many a bind as I was starting. Thank you for your patience. Tori, we've just got to take things ~one step at a time~. Thank you for being a bright presence whenever I got to see you and making field work brighter and memorable, even at 1AM. Hugo, you're so smart and the best post-doctoral mentor I could have asked for when starting off graduate school. Your presence was very calming, especially as a first year graduate student trying to figure out what the Sippewissett salt marsh even looked like. Emily Junkins! We had so many memorable adventures, and I'm so happy you were there for the latter half of my graduate school experience. Thank you for being a great science mentor and friend. You kept the experience very down to earth. Finally, Anna Warsaw. Chapters two, three, and four would not have been possible without you. You helped me get this RNA through the finish line. Truly thank you. I can't believe it took six months of working together before we realized we both should be movie friends. Thank you for becoming my friend and colleague.

My family! My parents. I wouldn't be here today without both of you. Madre, you installed a deep work ethic in me and the drive to just keep going. All those border crossings… because you believed it would give me a future. I hope I made you proud with this work I'm putting out. Padre, I've always appreciated your calmer presence. I think your passion for science fiction might have been what started me on this science route I've taken. I always wanted to learn more about this world, and I definitely learned a lot about one very specific aspect of it in this little project I put together. Mi Tio Juan. I definitely wouldn't be

here today without you. Thank you. I'll miss you very much. You were always hoping to bring happiness and entertainment to my life, I see and appreciate that now. Mi Tia Juana, no one will ever outmatch you in surprises. I love your free spirit, and all the times you took it on yourself to be my second mom during my childhood. Thank you for still caring for me deeply. And finally Jaquelin, you were my first real friend who taught me the joy of laughter. I'm glad we're still friends even 14 years later. My whole family is the reason I got to where I am, and I am this person you see today. Thank you. I dedicate this dissertation to all of you.

My whole slew of friends.

College friends, I am grateful we all still talk. You gave me a second home in this country as I was still figuring a lot of things out. Jane Wei, Janet Wang, Joyce Cao, Jerry Jia, Liz Yam, Michelle Ma, Lauren Hsu, Kathy Chiang, Vicky Wang, Christine Lee, Kathy Xu, Kevin Shih, Odile Maurelli, Crystal Shan, Iris Cai, and Erin Tou. I'm glad we all still talk and have met throughout the years. Danielle Rutkowski, Rachel Giorgio, and Anita Minniefield. We met through biology, and will always be my close friends. Arlinda Shehu, Sara Hwong, and Clara Sungmin Lee thank you for being my little college family. Mitchell Poon, we did a lot of growing up together. Thank you. Joseph Simpson, my trips to Los Angeles to see you honestly helped me recharge during graduate school. They were such a light. I'm glad you also moved to California from Ithaca! Finally, but not least, Dawit Bekele, you very much deeply impacted my academic attitude and helped me believe in myself when I was very lost and intimidated within academia. You showed me so much of the world as a first-time adult. Thank you for all the trips and new experiences!

Grad school friends.

The first part of my journey. The whole Tuolumne crew! You were my home base. I could not have made it through without all of you. You guys were my rock. I'm glad we laughed and lived so much life together. James Leo Loving-Lichtenstein, Taom Sakal, Marisa Morse, Kelly Thomasson, Nicola Paul, Noe Castañeda, Bridget Vincent and Chen Xing. The wavelength we cultivated kept me going! My marine microbiology friends! Justine Albers, Natalie Dornan, Nathalie Eegholm, Jeemin Rhim, Jules Middleton, Ellie Arrington, Jon Tarn, Jose Valera, and Na Liu. I'm grateful we created a community together.

Lance Arevalo, you were an integral part to the first 6 years of my journey. I'm truly grateful that we were there for each other, I think we mutually learned a lot. Thank you for helping me through my time as a blossoming scientist, and being patient as I was learning how to be an adult.

The second part of this journey. My friends from Oregon! You gave me a second breath towards the end of this journey, when I was feeling a new loneliness. Thank you! Mackenzie Kawahara, Sunni Patton, Kareem Shehab, and Dexter Davis. I'm truly glad we became a crew that partook in so many adventures the last two years. Let's keep the adventures going.

Akhim Alexis, thank you for being there for me this last year. I'm glad we've gotten to learn so much from each other. We always have a good time and I'm glad we can make each other laugh without fail! You're a great ol' meepster and I hope I can support you back as you write your own dissertation and novel. We can do this!

And finally, Aaron Fai. Without our daily phone calls I'm not sure I could have made it out fully sane. Thanks for always bringing a smile to my face even when the going got tough. I deeply appreciate you. You're truly awesome, thank you for being such a rock.

On to the next phase of life!

VITA OF BRAULIO SAUL CASTILLO VILLASENOR
March 2026

EDUCATION

Bachelor of Arts in Biological Sciences, Cornell University, May 2018

Doctor of Philosophy in Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, March 2026

PROFESSIONAL EMPLOYMENT

2019-2025: Teaching Assistant, Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara- Cell Physiology (EEMB157), Principles of Evolution (EEMB131), Methods in Aquatic Community Ecology Laboratory (EEMB142AL), Marine Microbiology (EEMB144), Laboratory in Vertebrate Biology, Ecology, and Evolution (EEMB113L), Introductory Genetics (EEMB129)

PUBLICATIONS

Diel activity and pangenomic diversity of Candidatus "Thiohalocapsa roseobacca", a novel lineage of multicellular, millimeter-sized purple sulfur bacteria. BS Castillo Villasenor *et al*. *in prep for mBIO*

Photoheterotrophy in marine sulfate reducing bacteria? BS Castillo Villasenor *et al*. *in prep for ISME J*

AWARDS

Carl Storm Underrepresented Minority Fellowship| Summer 2025

Graduate Opportunity Fellowship| 2021-2022 and 2020-2021 Academic Year

Schmidt Family Accelerator Award| Summer 2021

Worster Award Fellowship| Summer 2021

CONFERENCES

GRC Applied and Environmental Microbiology| July 2025
"Physiological patterns underlying the metabolic interactions of a cryptic sulfur cycling bacterial consortium"

American Society of Microbiology| June 2025
"Metabolic resource sharing in a cryptic sulfur cycling bacterial consortium"

Association for the Sciences of Limnology and Oceanography| April 2025
“Physiological fluctuations underlying the metabolic interactions of a cryptic sulfur cycling bacterial consortium”

International Symposium on Microbial Ecology| August 2024
“Physiological patterns underlying the metabolic interactions of a cryptic sulfur cycling bacterial consortium”

Southern California ASM Meeting| October 2021
“Microscale spatial architecture of ‘pink berry’ microbial aggregates and their symbiotic single-cell interactions using BONCAT-FISH fluorescence imaging”

Ocean Sciences Meeting| February 2020
“Visualization of the microscale spatial architecture of ‘pink berry’ microbial aggregates and their symbiotic single-cell interactions using BONCAT and CARD-FISH fluorescence imaging”

SACNAS- The National Diversity in STEM Conference| October 2019
“Being your authentic self in graduate school, Community, Advocacy, and Resilience: Excelling Together (CARE)”

SERVICE AND ENGAGEMENT

SACNAS Graduate Chapter- Executive Board Member and Financial Advisor| October 2018-June 2024

Summer Institute for Mathematics and Science- Mentor| August 2020 and August 2021

Mentor in Microbial Culturing Research to Imanol Ulloa| September 2022-June 2023

Mentor in Microbial Genomics Research to Fox Bravo| January 2021-June 2022

ABSTRACT

Insights into the light-dark physiology of a purple sulfur bacterium and its relationship with its intimate sulfate reducing partner

by

Braulio Saúl Castillo Villaseñor

All microbes live in community, and have the potential to engage in both positive and negative interactions with each other. The complexity of naturally occurring, environmental communities, furthermore, makes the dissection of these interactions difficult and hard to interpret. Despite these interactions being difficult to detect, leveraging a multifaceted approach that employs microscale methods can aid in the decryption of these cryptic interactions. My interest in exploring microbial community interactions drove me to the pink berries of the Sippewissett salt marsh, microbial aggregates comprised of two main sulfur cycling symbionts: a purple sulfur bacteria (PB-PSB1) and a sulfate reducing bacteria (PB-SRB1). This naturally occurring community is a tractable, low-diversity system that provides a model microbial system to enable the study of bacterial metabolic interactions. As pink berry bacteria live in tightly associated, reproducible communities across space and time, we aimed to explore their physiology and use this knowledge to determine other potential instances of syntrophy between PB-PSB1 and PB-SRB1.

In the first chapter we generated 41 high quality bins from 3 individually PacBio HiFi sequenced pink berries. This contemporary database was used as a reference for all our

subsequent analyses, and allowed us to compare pink berry community composition between 2011 and 2021 as well as look at genomic variation with time in PB-PSB1. We further explored and characterized the taxonomic novelty generated by this reference bin database such that we generated 1 novel order, 10 novel genera, and 20 novel species. In the second chapter we generated a diel metatranscriptome from a lab incubation of pink berries in order to track community response to light-dark exposure. We tracked gene expression patterns in PB-PSB1 and PB-SRB1 in response to their natural light-dark cycle, and used this to characterize carbon dynamics in PB-PSB1 and its potential to be imported into PB-SRB1 for use in sulfate reduction. We found signs for anabolic growth during the day and subsequent breaking down of stored carbon at night. PB-SRB1, in turn, seemed to be engaging in greater sulfate reduction and respiratory activity during the day, as it could be stimulated by PB-SRB1 photosynthate like succinate, shikimate or other 2-hydroxyacid dehydrogenases.

For the third chapter, we carried out a molybdate inhibition experiment in order to uncouple the syntrophic relationship between PB-PSB1 and PB-SRB1. We generated a transcriptome as well as an exo- and endometabolome in order to further characterize potential phorotsynthate and carbon use in PB-PSB1 as well as identify other potential candidates for shared photosynthate between PB-PSB1 and PB-SRB1. We found some potential markers for autotrophy (eicosapentaenoic acid and 1,2-dihydroxyheptadec-16-yn-4-one,), as well as evidence for glycogen oxidation to polyhydroxyalkanoates at night by PB-PSB1. PB-SRB1 in turn, may be using organosulfonates exuded by PB-SRB1 in order to engage in dissimilatory sulfate reduction during the day. Finally, due to the dynamic biogeochemical nature of the marsh environment we explored light-dark patterns of field-sampled aggregates in order to generate insight into the potential for oxic stress in our

consortia. We found similar expression of PB-SRB1's oxidoreductases, transporters, and cytochromes between our field and lab incubated samples, potentially indicating a lack of oxygen stress that may be due to shielding by PB-PSB1. We did, however, see signs of potential oxic stress in PB-PSB1 in the form of various dehydrogenases and peroxidases. In PB-PSB1 we observed daytime upregulation of genes involved in cofactor synthesis as well as of genes involved in carbon fixation, potentially indicating a harnessing of photosynthetic energy for growth and function.

The whole of this dissertation explores the metabolic partnership between two microorganisms, giving insight into purple sulfur bacteria physiology and potential for photosynthate sharing in a naturally occurring model microbial consortium.

Table of Contents

Introduction *1*

Chapter I: Generation of a high-quality bin database and exploration of genome variation in Thiohalocapsa PB-PSB1 ..... *7*

Abstract 7

Introduction 8

Methods 13

Results and Discussion 26

Conclusions 41

References 42

Chapter II: Daytime carbon fixation and anabolic activity by Thiohalocapsa PB-PSB1 may produce small carbon compounds used by Desulfofustis PB-SRB1 ..... *49*

Abstract 49

Introduction 51

Methods 56

Results and Discussion 61

Conclusions 83

References 86

Chapter III: Chemical insights into autotrophic activity in Thiohalocapsa PB-PSB1 and potential use of fixed organosulfonates by Desulfofustis PB-SRB1 ..... *93*

Abstract **93**

Introduction **94**

Methods **97**

Results and Discussion **105**

Conclusions **125**

References **128**

Chapter IV: Similarities in Thiohalocapsa PB-PSB1 and Desulfofustis PB-SRB1 metabolic capabilities in both lab and environmental conditions despite oxygen exposure ***133***

Abstract **133**

Introduction **135**

Methods **137**

Results and Discussion **139**

Conclusions **149**

References **151**

Appendix ***154***

Chapter I Supplemental Information **154**

Chapter II Supplemental Information **173**

Chapter III Supplemental Information **214**

Chapter IV Supplemental Information **220**

Introduction

All organisms live in communities, whether this be a eukaryotic host and its microbiome, or a group of bacteria living in tandem together. Just like the larger scaled eukaryotes we are all familiar with, the type of interactions microorganisms can engage in range from positive to negative. The nature of the interactions microbes can engage in is quite varied, ranging from sharing nutrients to establish a stable community, to releasing metabolites to result in the exact opposite scenario. Furthermore, as the scale of these interactions is much smaller than ours, discerning the occurrence and nature of these interactions can prove to be difficult. With this dissertation I employed a suite of methods in order to generate multiple scales of evidence aimed at elucidating the microbial physiology of a model microbial consortium, the pink berries of the Sippewissett salt marsh. I hoped to generate a window into this consortium's dynamics that would help me determine whether I could find instances of shared compounds and interactions that lead to the stable community we see forming across space and time (Wilbanks et al. 2014, 2022; Doré et al. 2024).

How microbial communities come to stably form is a varied puzzle which takes various lines of evidence to decipher. Eukaryotic organisms, for example, are known to select for specific microbiomes via a variety of methods such as through the host immune system, host behavior itself, or physiological regulation of the host environment (Wilde et al. 2024). Bacterial co-existence furthermore, can be the product of contemporaneous competition where the resulting community has undergone multiple modes of active competition. This can include storing of nutrients to deprive competitors, blocking colonization sites and increasing dispersal, and production of antimicrobial toxins (Hibbing et al. 2010). Diatoms,

for example, have been shown to select for bacterial partners via the use of secondary metabolites (Shibl et al. 2020). These bacterial interactions, moreover, can be cryptic as they take place over short distances and transient timescales, avoiding detection by bulk methods measuring net rates of compound flux. In order to elucidate these microscale interactions between microorganisms, therefore, we need to employ multifaceted, multi-scaled approaches.

Such interactions, furthermore, are able to facilitate cryptic biogeochemical cycles, which have important global impacts. Cryptic cycling of sulfur is particularly notable, as this can lead to greater carbon remineralization rates than otherwise estimated in areas like coastal shelf sediments (Jørgensen 1982), and has been thought to be possible in oxygen minimum zones (Crowe et al. 2018). Demonstrating these relationships can be difficult and can be facilitated by the integrated use of multiple, microscale tools. Cryptic cycling of carbon, for example, has been demonstrated in organisms such as ANME archaea in the sulfate methane transition zone, where methane produced from organic matter oxidation was oxidized back to carbon dioxide by sulfate reduction (Beulig et al. 2019). Such a relationship could not have been explained by just measuring net methane and sulfate fluxes, and was instead discerned by the joint use of radiotracers, isotope budgets, and quantitative PCR.

Studying microscale bacterial interactions, therefore, can help us understand microbial community formation and cooperation, but also can also help elucidate interactions that have wider scale geochemical impacts. The study of these interactions in the environment can prove difficult, however, which is why I pursued the study of a tractable, naturally occurring microbial consortium that can be experimented with in a lab environment. I hoped to elucidate and examine cryptic host-symbiont interactions in bacterial communities by

studying the pink berries of the Sippewissett salt marsh. This bacterial community is consistently composed mainly of a host like purple sulfur bacterium (PB-PSB1) and a partner sulfate reducing bacterium (PB-SRB1), as well as various other heterotrophs like *Bacteroidia* (Wilbanks et al. 2014). The two main symbionts have been shown to participate in cryptic cycling of sulfide and sulfate via stable isotope fractionation and nanoSIMS measurements. The work of this dissertation is meant to elucidate other potential cryptic interactions taking place in this bacterial community and glean insights into community assembly and microscale biogeochemical interactions within this one model consortium.

PB-SRB1, for example, has been shown to take sulfate from PB-PSB1's sulfide oxidation and use it for sulfate reduction. Other cryptic interactions, therefore, may be taking place in this consortium, such as the sharing of photosynthate from the dominant PB-PSB1 to PB-SRB1 as an electron donor for sulfate reduction. SRB, however, are typically known to degrade small carbon compounds like short chain fatty acids and TCA cycle intermediates (i.e. lactate, acetate, malate, fumarate) (Widdel and Bak 1992). While the degradation from complex photosynthate to smaller carbon compound could be enabled by another community member, it is also possible that PB-SRB1 could be cryptically, but directly degrading PB-PSB1's photosynthate.

In order to understand the physiological landscape behind these cryptic interactions I first generated, via PacBio HiFi long read sequencing of three individual pink berry aggregates, a contemporary, high-quality reference database of bins to represent my pink berry communities (Chapter 1). With this analysis I compared pink berry community composition between 2011 and 2021, genomic variation in PB-PSB1 between 2011 and 2021, as well as characterized the newfound novelty from my reference bin database. We followed this by

exploring gene expression patterns in PB-PSB1 and PB-SRB1 in response to their natural light-dark cycle in the lab environment (Chapter 2). In PB-PSB1 we explored diel patterns in carbon catabolism, storage, and general anabolic growth in order to understand its photosynthetic capabilities. In PB-SRB1 we explored the consequent potential for gene expression indication of small carbon compound import and use for sulfate reduction. We followed this with a molybdate inhibition experiment in order to disrupt PB-SRB1 activity and uncouple the syntrophic relationship between PB-PSB1 and PB-SRB1 (Chapter 3). We explored potential indicators of PB-PSB1 photosynthetic activity in the endo- and exometabolome as well as its use of carbon storage compunds, as well as explored this chemical space for PB-SRB1 sulfate reduction electron donor candidates. Finally, we sampled pink berry material directly in the environment, presumably capturing the effects of oxygen stress on the consortium (Chapter 4). We explored the diel patterns in PB-PSB1 for similarities to our lab incubations as well as for signs of oxygen stress, as this could affect PB-PSB1 activity and consequently availability of shared photosynthate for PB-SRB1.

References

Beulig, F., H. Røy, S. E. McGlynn, and B. B. Jørgensen. 2019. “Cryptic CH4 Cycling in the Sulfate–Methane Transition of Marine Sediments Apparently Mediated by ANME-1 Archaea.” *The ISME Journal* 13 (2): 250–62. https://doi.org/10.1038/s41396-018-0273-z.

Crowe, Sean A., Raymond P. Cox, CarriAyne Jones, et al. 2018. “Decrypting the Sulfur Cycle in Oceanic Oxygen Minimum Zones.” *The ISME Journal* 12 (9): 2322–29. https://doi.org/10.1038/s41396-018-0149-2.

Doré, Hugo, Amy R. Eisenberg, Emily N. Junkins, et al. 2024. “Targeted Hypermutation of Putative Antigen Sensors in Multicellular Bacteria.” *Proceedings of the National Academy of Sciences* 121 (9): e2316469121. https://doi.org/10.1073/pnas.2316469121.

Hibbing, Michael E., Clay Fuqua, Matthew R. Parsek, and S. Brook Peterson. 2010. “Bacterial Competition: Surviving and Thriving in the Microbial Jungle.” *Nature Reviews Microbiology* 8 (1): 15–25. https://doi.org/10.1038/nrmicro2259.

Jørgensen, Bo Barker. 1982. “Mineralization of Organic Matter in the Sea Bed—the Role of Sulphate Reduction.” *Nature* 296 (5858): 643–45. https://doi.org/10.1038/296643a0.

Salman, Verena, Tingting Yang, Tom Berben, Frieder Klein, Esther Angert, and Andreas Teske. 2015. “Calcite-Accumulating Large Sulfur Bacteria of the Genus *Achromatium* in Sippewissett Salt Marsh.” *The ISME Journal* 9 (11): 2503–14. https://doi.org/10.1038/ismej.2015.62.

Shibl, Ahmed A., Ashley Isaac, Michael A. Ochsenkühn, et al. 2020. “Diatom Modulation of Select Bacteria through Use of Two Unique Secondary Metabolites.” *Proceedings of the National Academy of Sciences* 117 (44): 27445–55. https://doi.org/10.1073/pnas.2012088117.

Widdel, Friedrich, and Friedhelm Bak. 1992. “Gram-Negative Mesophilic Sulfate-Reducing Bacteria.” In *The Prokaryotes*, edited by Albert Balows, Hans G. Trüper, Martin Dworkin, Wim Harder, and Karl-Heinz Schleifer. Springer New York. https://doi.org/10.1007/978-1-4757-2191-1_21.

Wilbanks, Elizabeth G., Hugo Doré, Meredith H. Ashby, Cheryl Heiner, Richard J. Roberts, and Jonathan A. Eisen. 2022. “Metagenomic Methylation Patterns Resolve Bacterial Genomes of Unusual Size and Structural Complexity.” *The ISME Journal* 16 (8): 1921–31. https://doi.org/10.1038/s41396-022-01242-7.

Wilbanks, Elizabeth G., Ulrike Jaekel, Verena Salman, et al. 2014. “Microscale Sulfur Cycling in the Phototrophic Pink Berry Consortia of the S Ippewissett S Alt M Arsh.” *Environmental Microbiology* 16 (11): 3398–415. https://doi.org/10.1111/1462-2920.12388.

Wilde, Jacob, Emma Slack, and Kevin R. Foster. 2024. “Host Control of the Microbiome: Mechanisms, Evolution, and Disease.” *Science* 385 (6706): eadi3338. https://doi.org/10.1126/science.adi3338.

Chapter I: Generation of a high-quality bin database and exploration of genome variation in *Thiohalocapsa* PB-PSB1

Braulio S. Castillo Villaseñor, Emily N. Junkins, and Elizabeth G. Wilbanks

Abstract

Our representations of microbial genomes allow us determine the functional potential for the processes our organisms of interest are engaging in. The technology and manner in which we build these representations are constantly improving, and in this chapter I use PacBio HiFi sequencing in order to build a contemporary bin for the main photosynthetic host PB-PSB1 of a bacterial consortium, the pink berries of the Sippewissett salt marsh. I additionally generated updated bins for a slew of other new members generating 41 new bins total. We found that PB-PSB1 and PB-SRB1 consistently dominate the community across our sampling scheme of 10 years. The other members of our database, furthermore, represents newfound novelty in the manner of 1 novel order, 10 novel genera, and 20 novel species. Finally we delved into the genome of PB-PSB1 and characterized both its functional and structural architecture. PB-PSB1's core genome was enriched in functions related to central metabolism while its flexible genome was enriched in defense related functions. We also examined a few examples of consistent insertions and deletions between 2011 and 2021.

Introduction

Sequenced microbial genomes are the basis off which we as microbial ecologists can start interrogating the functional potential of our genomes of interest. Both the cost and quality of genomes we are able to generate has vastly improved, especially since the early 2000s (Sboner et al. 2011), and much of this is due to the progress in sequencing technology available. In order to probe the physiology of our bacterial pink berry consortium, we first needed to generate updated, high-quality bins that can serve as an accurate baseline informing the physiological potential of consortium members like PB-PSB1 and PB-SRB1.

The Genomic Standards Consortium developed a standard for high quality metagenome-assembled genomes where completion is greater than 90%, contamination is less than 5%, and assembly is in few fragments such that rRNA genes (23S, 16S, and 5S) and at least 18 tRNAs are present (The Genome Standards Consortium et al. 2017). Incomplete and contaminated MAGs may arise from variation or error in any step in their generation, from sequencing to assembly. While an incomplete bin could be missing DNA encoding for genes key to a particular metabolism, a complete bin would give a more accurate and holistic understanding of the functional, metabolic capacity of the bin's source organism (Chen et al. 2020). Short-read, Illumina data is highly accurate (Stoler and Nekrutenko 2021), and improvements have been made possible by new technologies like long-read sequencing which is able to span repetitive regions in genomes, generate longer contigs, and thus reduce contamination and improve completion, leading to an improvement in overall bin quality (Espinosa et al. 2024). Furthermore, it is possible to undertake a hybrid approach and combine short-read HiSeq with long-read PacBio approaches to yield better assembled, high quality bins (Jin et al. 2022).

A single MAG, however, is unlikely to accurately represent the diversity seen in a natural population. A study comparing *E. coli* isolates to a metagenomic dataset, for example, determined that their ~95% complete MAGs reflected only 77% of the core genome and 50% of the variable genome (Meziti et al. 2021). Furthermore, when comparing genes present in their MAGs as opposed to their isolates, the authors saw an average of 4.4% of sequences attributed to binning errors in contrast to the checkM average estimate of 2.63%. Even "completed" MAGs therefore, have a lot of room for error in their "representation" of the organism of interest. This highlights the needs for accurate, well assembled bins that are a comprehensive reference, in order to capture the diverse physiological landscape of the focal organism. In fact, a study with mangrove sediments that utilized the hybrid HiSeq-PacBio discussed above, was able to recover double the number of high quality MAGs as well as more novel MAGs than when using short reads alone (Z.-F. Zhang et al. 2023). This better represented the diversity of members within their sampled community, as well as improved their understanding of the functional potential within mangrove sediments.

Good, high-quality bins can also help with ecological and evolutionary questions. By utilizing single-nucleotide variants it is possible to characterize strain-level variation in populations, as has been used to determine host-microbiome specificity with geographically distinct populations of humans (Truong et al. 2017). Further, by being able to probe nucleotide-level evolutionary changes it is possible to study evolutionary processes like sweeps of variant gains or losses (Garud et al. 2019).

Beyond implications of the bin itself, in order to assign taxonomy we depend on reference databases like the Genome Taxonomy Database. This, however, is inherently limited by the quality of the composite references themselves (Parks et al. 2022a). By

generating well assembled references it is possible to further improve community databases as well as paint a more accurate genomic picture of an organism, hence, we see biome specific reference genome catalogs being created by various research groups (Gurbich et al. 2023). Ultimately, to further validate these metagenomic assemblies, being able to leverage 'omic data on multiple scales (transcriptomics, metabolomics), can corroborate the functional hypotheses generated by metagenomic data and lead to the creation of better physiological pathway maps (Mallick et al. 2017).

The pink berries of the Sippewissett salt marsh have been found in three geographic locations: Little Sippewissett salt marsh, Great Sippewissett salt marsh, and Penzance Point. In 2011, 16S rRNA clone library Sanger sequencing data was generated from Little Sippewissett salt marsh berries collected in 2007 and 2010, as well as from Penzance Point berries collected in 2011 (Wilbanks et al. 2014). PB-PSB1 and PB-SRB1 were shown to make up 60-70% of the community, and the remaining composition was mostly diverse *Bacteroides*, diatom chloroplasts, *Alphaproteobacteria*, and *Cyanobacteria* (Figure 1). This dataset, however, was generated via an 8F primer with high sequencing bias against *Alphaproteobacteria*.

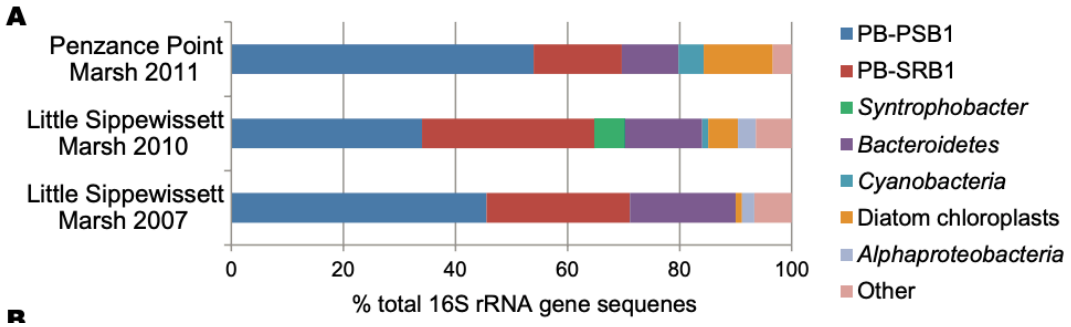


Figure 1. Figure 2A from (Wilbanks et al. 2014). Pink berry member relative abundance from clone library generated 16S rRNA sequences: 2011 Penzance Point (89 clones), 2010 Little Sippewissett (94 clones), and 2007 Little Sippewissett (90 clones).

Roche 454 Titanium and Illumina HiSeq co-assembled data was used for metagenomic analysis to produce bins for PB-PSB1 (20-25X coverage) and PB-SRB1 (15-20X coverage). These bins were highly fragmented, however, such that PB-PSB1 was in 1386 scaffolds and PB-SRB1 was in 422 scaffolds. Their sizes were also much smaller than what we have determined from our current reference bins (PB-PSB1 at 6.7 Mb instead of the current 7.9 Mb and PB-SRB1 at 3.7 Mb instead of the current 4.6 Mb). Our current, NCBI-deposited reference genomes for PB-PSB1 and PB-SRB1 were generated via PacBio RSII in 2011 and refined with manual curation. PB-SRB1 was recovered in 71 contigs and PB-PSB1 was closed to a single scaffold upon manual curation (Wilbanks et al. 2022). This PacBio RSII sequencing, furthermore, when complemented by measures like tetranucleotide frequency or methylation profiles yielded the following bin clusters: PB-PSB1, PB-SRB1, *Micavibrionaceae*, *Oceanicaulis*, and *Rhodobacteraceae* (Figure 2). *Micavibrionaceae*, *Oceanicaulis*, and *Rhodobacteraceae* had not been previously detected because of primer bias in our clone library data.

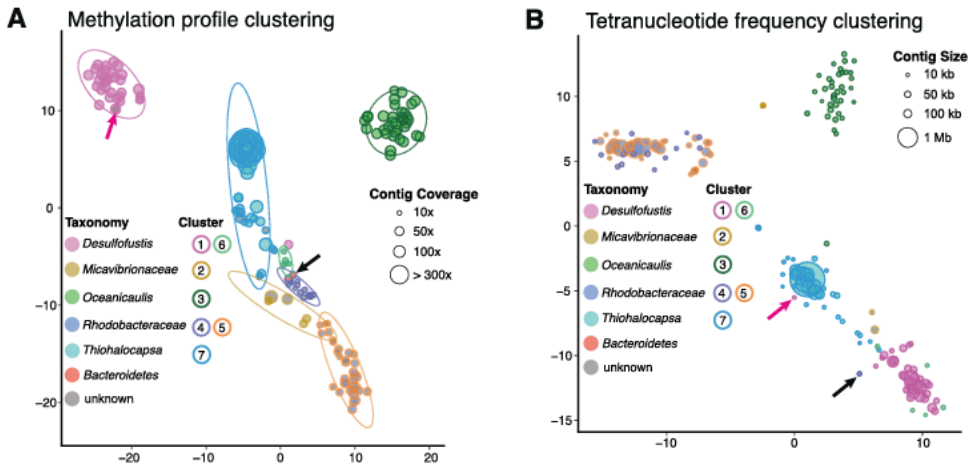


Figure 2. Figure 3 from (Wilbanks et al. 2022). Contig similarity visualized via t-distributed stochastic neighbor embedding of (A) methylation profiles or (B) tetranucleotide frequencies. Point size scaled by either (A) contig coverage or (B) contig length.

For this project we carried out improved PacBio HiFi sequencing, ten years after the original PacBio RSII sequencing that was done in order to generate our initial 2010 bins as well as our 2011 NCBI reference genomes for PB-PSB1 and PB-SRB1. We hoped to generate higher quality bins that would serve as a new reference for future studies (Chapters 2 and 3), as well as be indicative of genomic variation with time. We were able to generate novel bins which offered insight into other heterotrophic pink berry community members. Additionally, we observed multiple instances of insertions and deletions between 2011 and 2021 that were present in two of our PacBio HiFi samples. Our high-quality sequencing of multiple samples also allowed us to examine inter-genome variability and more fully characterize the functional pangenomic landscape available to PB-PSB1. We explored the functional enrichment of the core as opposed to the flexible fraction generated from three PB-PSB1 bins.

Methods

Berry sequencing and Bin Generation

Three pink berry aggregates were individually extracted for long-read DNA using methods described previously (Doré et al. 2024), prepared as PacBio HiFi libraries, then sequenced at the Joint Genome Institute (JGI) via the PacBio SEQUELIIe platform. The three pink berries were denoted sample001, sample002 and sample024.

JGI implemented their standard pipeline for assembly and binning of data (J workflow, Figure 3). All three sequencing runs were assembled with flye (flye v2.9, [--meta -t 32 --pacbio-hifi]; Kolmogorov et al. 2019), using the default min_overlap parameter of sample N90 which is ~3.5kb for our samples. The set J assemblies were polished with racon (racon v1.4.13 [-u - t 36], Vaser et al. 2017), reads were mapped to assembly contigs with minimap2 (minimap2 v2.17 [-t 16 -a -x map-pb], Li 2018), and BBtools was used to determine coverage as well as filter reads (BBtools v38.79, Bushnell 2014). metabat2 was used for binning (metabat2 v2.15, Kang et al. 2019), gtdbtk was used for lineage classification (gtdb-tk v1.3, GTDB release 95, Chaumeil et al. 2020), and checkm was used for determining bin contamination and completion (checkm v1.1.3, Parks et al. 2015).

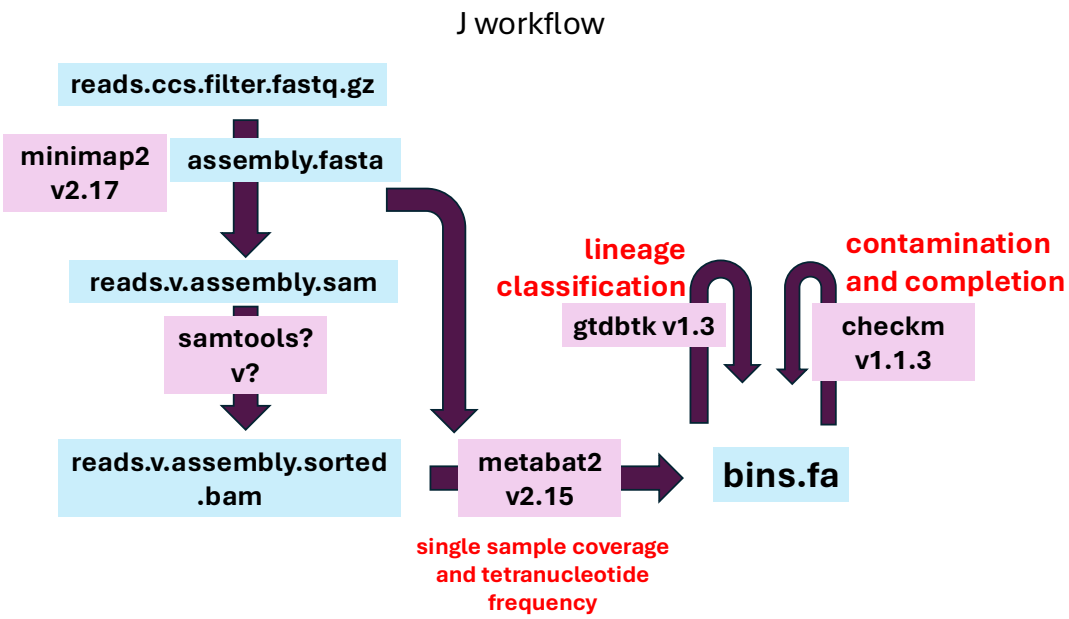


Figure 3. J workflow-Overview of JGI's standard assembly and binning process for our HiFi sequencing

The JGI pipeline used metabat2 for binning, implementing only single sample coverage and tetranucleotide frequency. We attempted two additional workflows in order to try to improve binning and assembly. We first tried to improve binning by providing coverage from all three of our samples and seeing the impact of richer covariance on bin quality (E workflow, Figure 4).

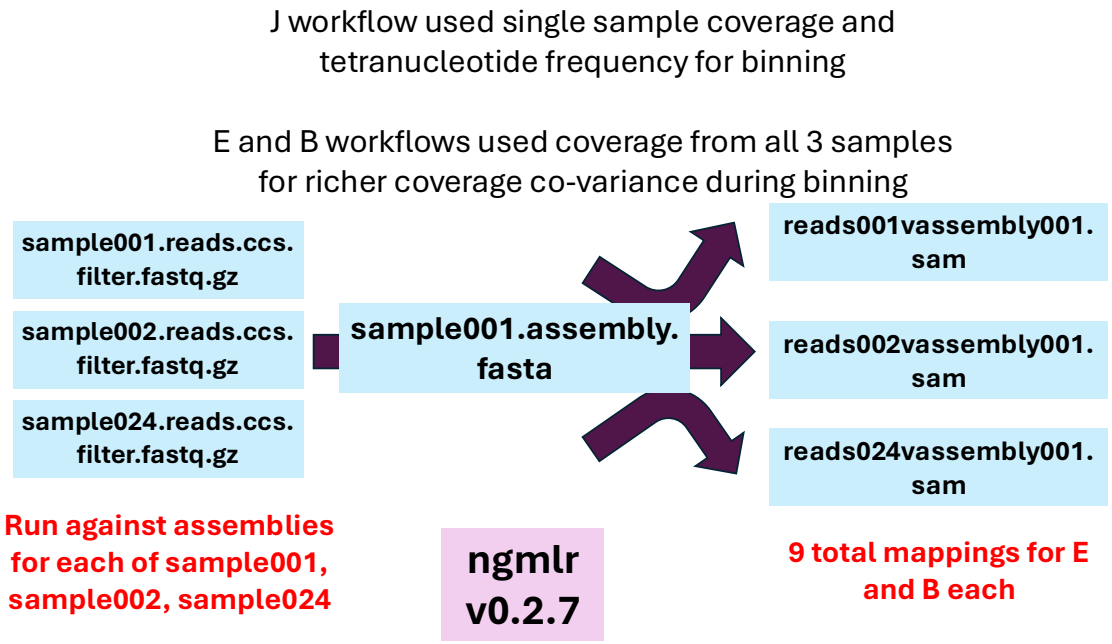


Figure 4. Covariance of from mapping reads for each sample to one assembly, sample specific mappings were provided to metbat2 in order to improve binning in E and B workflows

We also tried to improve assembly of higher coverage organisms in our samples by using a longer overlap parameter during assembly, in order to hopefully disambiguate and streamline assembly graph formation in repeat rich regions. For this workflow we also provided sample covariance from all 3 samples as in workflow E (B workflow, Figure 5). Thus, while all three sequencing runs were assembled with flye (flye v2.9, [--meta -t 32 --pacbio-hifi]), workflows J and E used the default min_overlap, whereas for workflow B we changed the min_overlap parameter to 7kb (flye v2.9-b1768).

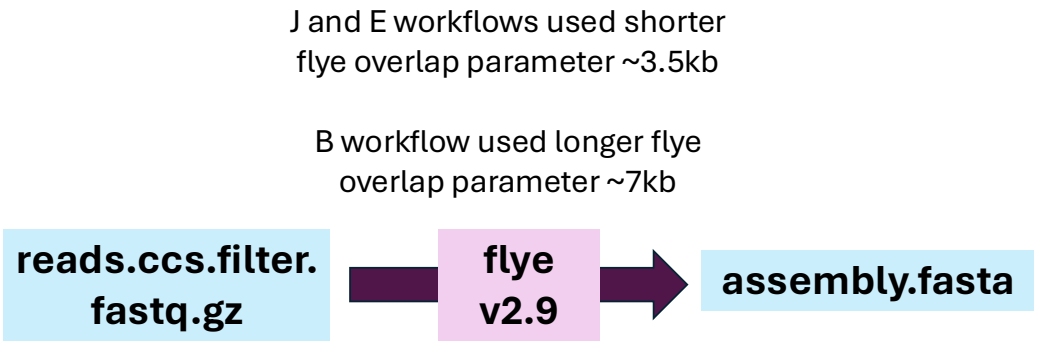


Figure 5. For workflow B we increased the min_overlap parameter during assembly in order to improve assembly graphs

Set E and B assemblies both underwent the same, following workflow (Figure 6). ngmlr was used for mapping such that reads from sample001, sample002, and sample024 were each mapped to the assembly from sample001, sample002, and sample024 in order to yield 9 total mappings (ngmlr v0.2.7, Sedlazeck et al. 2018). samtools was used to generate the appropriate bam, sorted.bam, and sorted.bam.bai files (samtools v1.6, Li et al. 2009). metabat2's jgi_summarize_bam_contig_depths was then used to look at the covariance between each set of reads mapping to an individual assembly[how mapping of reads from sample001, sample002 and sample024 to sample00X assembly covary] (metabat2 v2.12.1, Kang et al. 2019). metabat2 then used these files to generate bins from each assembly. Following this, checkm was used to obtain quality and coverage for bins (checkm v1.2.2, Parks et al. 2015), and gtdbtk was used for taxonomic classification (gtdb-tk v2.1.1, GTDB release 207, Chaumeil et al. 2020).

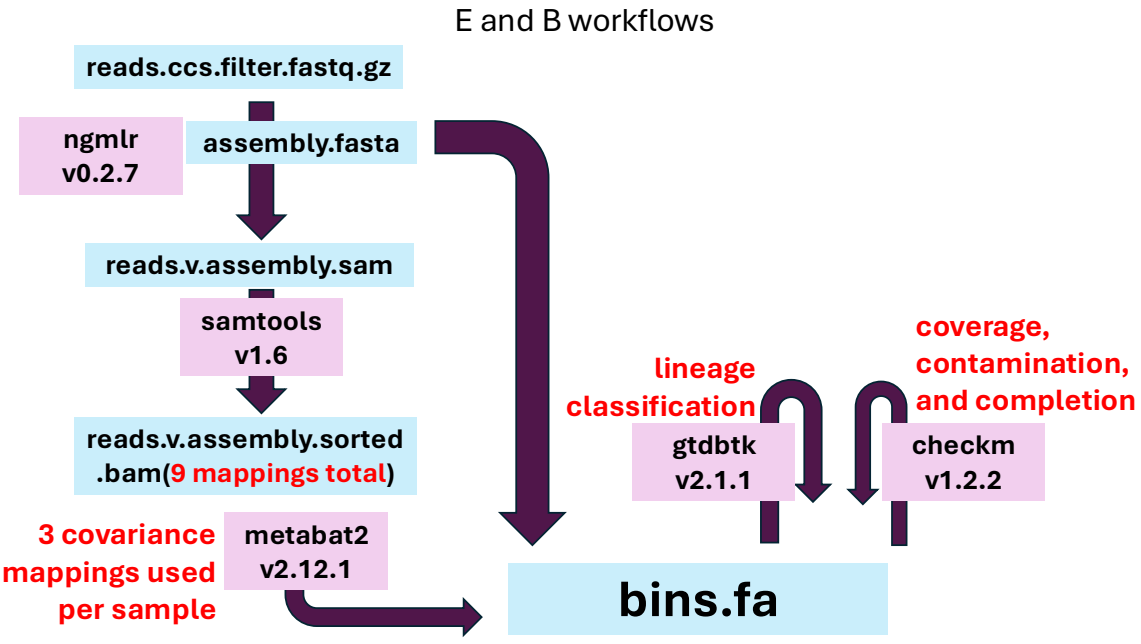


Figure 6. Overview of E and B workflows. Both E and B workflows implement sample covariance in order to improve binning.

Pangenome Generation

We aimed to use the high quality bins obtained from our HiFi sequencing in order to generate a cross-sample genomic representation of major community members. We called this a "pangenome" where content between the three samples as well as that unique to each sample was compiled together into one "genes.fna" pangenome (Figure 7). To start this, dRep was used in order to identify clusters of closely related bins across our 3 samples and 3 workflows (dRep v3.4.5, Olm et al. 2017). Specifically, we used parameters like number of scaffolds, longest scaffold, coverage, contamination, and completion in order to determine which organism across the J, E, or B workflows best represents the organism for each sample (i.e. for PB-PSB1 organism one bin in sample001, sample002 and sample024 each). We chose major community members that consistently show up across our samples and at a higher relative abundance like PB-PSB1, PB-SRB1, *Oceanicaulis*, *Rhodobacteraceae*, and *Micavibrio* as well as other heterotrophs with high quality bins in all three samples: *Lewinella*, and UBA10426 (*Salibacteraceae*). For PB-PSB1 and PB-SRB1 we also included the 2010 reference genomes: PB-PSB1 reference (CP050890.1) and PB-SRB1 reference (JAEQMT000000000.1). The genbanks used for annotation and location information were obtained from annotations generated by NCBI's prokaryotic genome annotation pipeline (PGAP v6.9, Tatusova et al. 2016). This was then provided to panaroo in order to generate a representative pangenome across our 3 samples (panaroo v1.5.0, Tonkin-Hill et al. 2020).

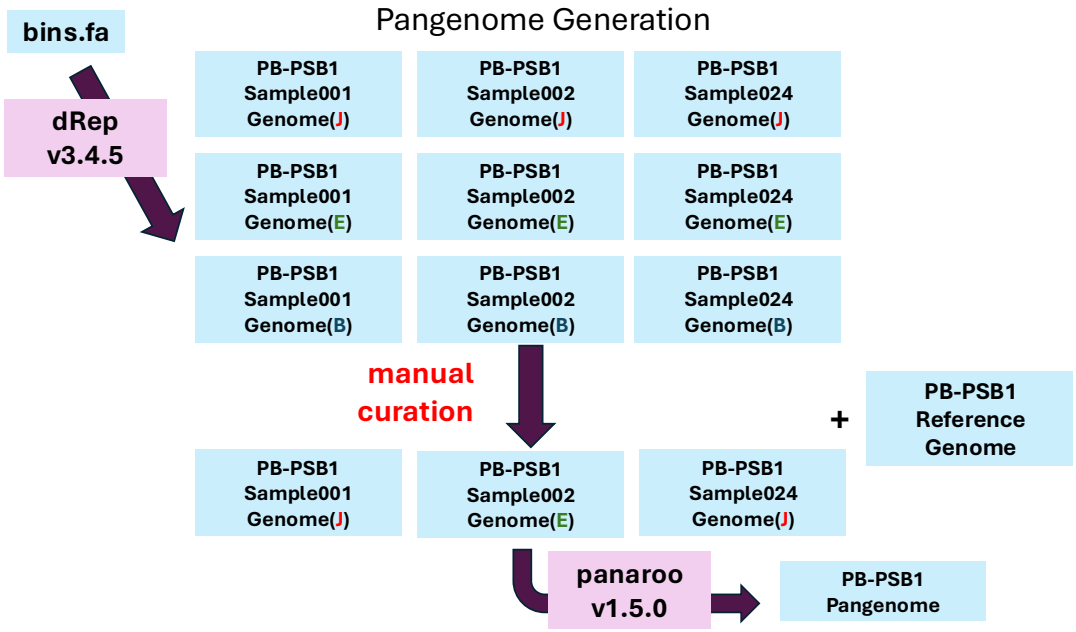


Figure 7. Overview for pangenome generation of major pink berry community members

Metatranscriptome Count Database Generation

We aimed to use this comprehensive genomic representation of coding space across our major community members to generate a reference database that can be used for comprehensive and accurate metatranscriptome counting. Since our community was also composed of more organisms than the 7 in our pangenomes, we were able to use the clustered dRep data to further determine the single best representative bin between our three samples for the rest of our bins. This, along with our pangenomes, became our dereplicated representative bin database (Figure 8). Only the best quality bin (low number of contigs, low contamination, high completion) was chosen from one sample. These bins were annotated via IMG's annotation pipeline. Overall, our reference database consists of 7 pangenome files which are complemented by a further 34 bins generated by either the J, E, or B method and of which only one representative was chosen (41 bins total).

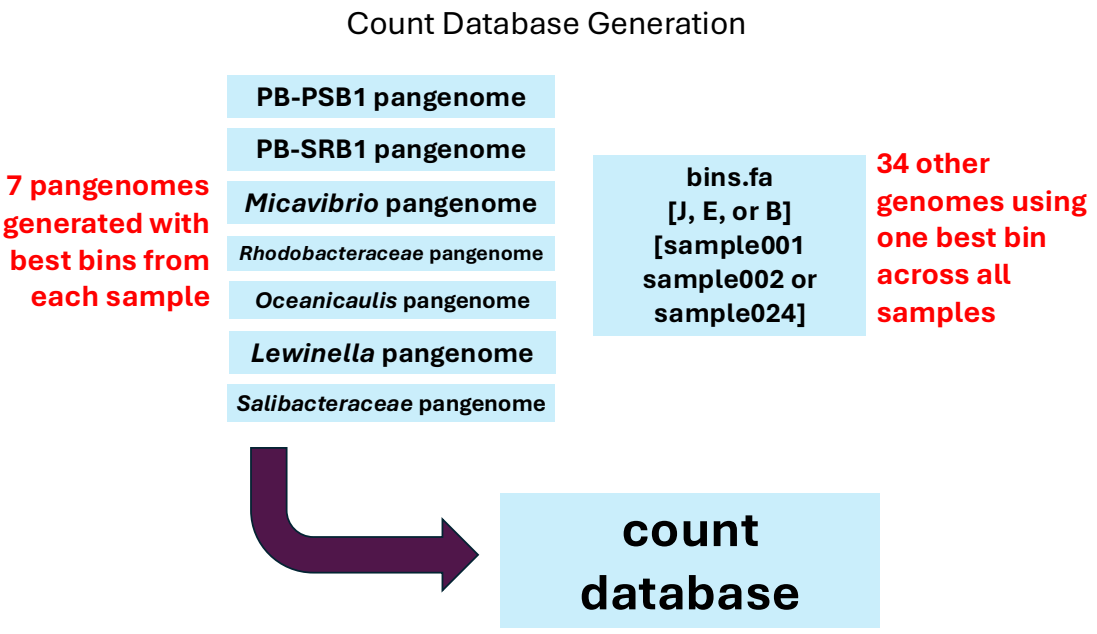


Figure 8. Overview of metatranscriptome count reference database generation with pangenomes for major organisms as well as bins for the rest of community members

This database may contain multiple copies of the same gene, or shared gene content across bins. In order to eliminate ambiguity for counting algorithms we collapsed repeats unto themselves such that if the repeat belonged to the same source bin the repeat was kept. If the repeat crossed multiple bins, however, the sequence was dropped as any mapping to this could have been erroneously assigned to the wrong bin (Figure 9). The 41 bins were thus run through cd-hit-est in order to group sequences into a 100% similarity threshold, -c 1.0 [sequence identity threshold] -G 0 [use local sequence identity] -aS 0.95 [alignment coverage for the shorter sequence, alignment must cover 95% of sequence] -d 0 [take fasta defline and stop at first space], allowing us to find instances of repeat sequences (cd-hit-est v4.8.1, Fu et al. 2012). This 41-bin repeatless database was used to unambiguously count reads during later metatranscriptome work [note: this file, out.db.rm.mxdsp, is in genes.fna format, is not separated by bin, is in nucleotide space, and does not contain an unbinned component].

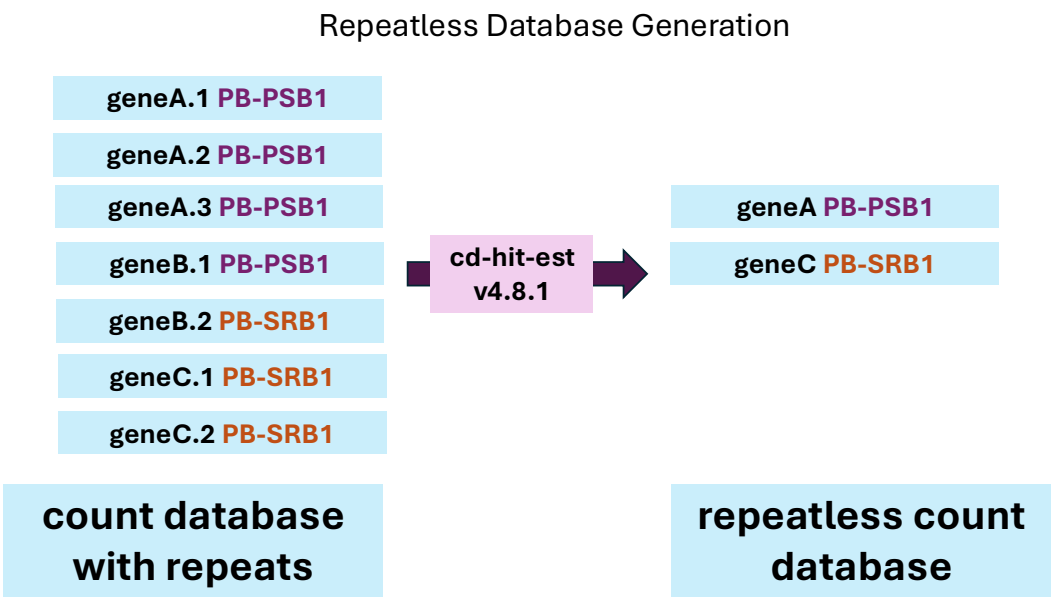
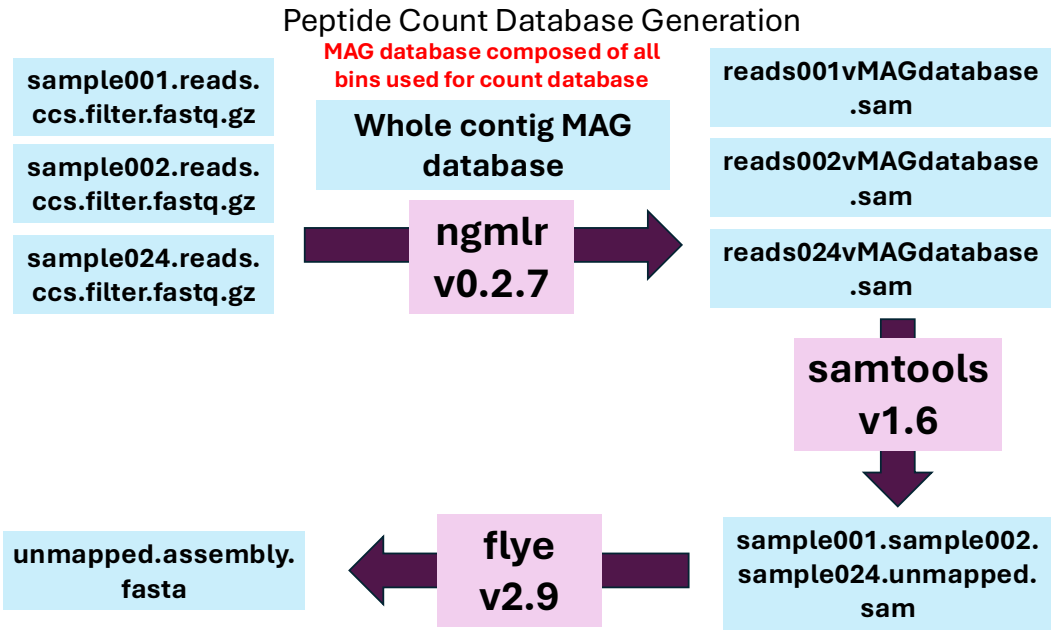


Figure 9. Illustration of our strategy for eliminating repeats across all the bins used for our metetranscriptome count database.

Peptide Mapping Database Generation

To gather a greater amount of peptide data for our metaproteome dataset we further needed to produce an unbinned file to generate a database for peptide mapping. This essentially was produced by taking reads across all our 3 samples that did not map to any of our bins used to generate the repeatless count database (Figure 10). Firstly, sample001, sample002, and sample024 PacBio reads were separately mapped back to a whole contig MAG database with ngmlr (for pangenomes each of the composite bins were used, ngmlr v0.2.7). The unmapped reads from each of these three bam's were pulled with samtools, then concatenated together (samtools v1.6). These unmapped reads were assembled with flye (min_overlap of 3kb, the default, flye v2.9.6-b1802), and prodigal was used to call open reading frames and translate (prodigal v2.6.3, Hyatt et al. 2010). They were then concatenated and denoted as unbinned. Following this, the gene entries in the RNA database file (out.db.rm.mxdsp), were separated back into bins and translated with emboss transeq

(EMBOSS: The European Molecular Biology Open Software Suite v6.5.7.0, Rice et al. 2000). These bins and the unbinned file were processed with cd-hit to eliminate cross species repeats, -c 1.0 [sequence identity threshold] -G 0 [use local sequence identity] -aS 0.95 [alignment coverage for the shorter sequence, alignment must cover 95% of sequence] -d 0 [take fasta defline and stop at first space] (cd-hit v4.8.1), split back by bins and the one unbinned file, then used for peptide mapping.



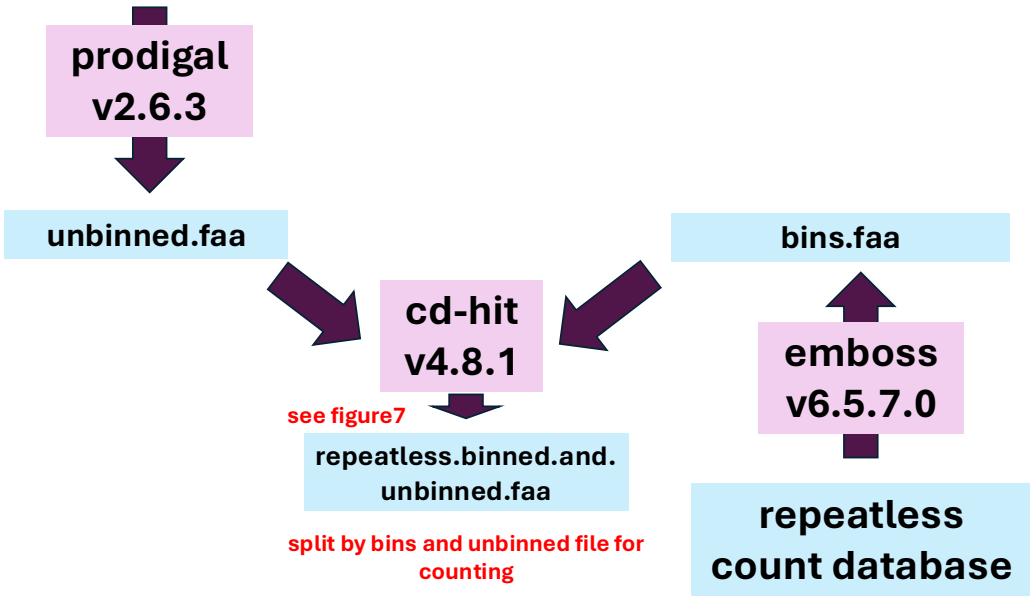


Figure 10. Overview of our strategy for generating an .faa file for reads that did not map to any of our bins. This unbinned file as well as peptide translation of our bins was used as a database for peptide mapping.

SingleM and GTDB

We needed to generate relative abundance estimates for both our short-read Illumina as well as our long-read sequencing runs. We used Sandpiper generated relative abundance for short-read runs (Sandpiper v1.1.9, Woodcroft et al. 2025) which relies on release 226 of the Genome Taxonomy Database (Parks et al. 2022b). For our long read sequencing we generated relative abundance values with singleM (SingleM v0.20.3, Woodcroft et al. 2025) which also relies on release 226 of the Genome Taxonomy Database. They were then both summarized with SingleM (v0.20.3).

While we used release 207 of the Genome Taxonomy Database for taxonomy assignment of the pink berry bins in our database, our comparison to uncultured representatives was done with bins online was also done under release 226 (Parks et al. 2022b).

Pangenome Generation for Functional Enrichment Analysis

We further generated a pangenome via anvi'o in order to visualize flexible and core genome content across our samples and analyze for functional enrichment in the core as opposed to the flexible genome (anvi'o v8, Eren et al. 2020). In order to compare the differences between flexible and core genome content in PB-PSB1 we took the anvi'o annotations for the PB-PSB1 reference genome, sample001, and sample002. We then analyzed the genes by annotation type (COG Category, COG function, KOfam, or KEGG class); and calculated how many genes were annotated out of total gene instances (we counted each member of each gene cluster across our three genomes). This was labelled as annotation coverage. Gene instances account for the presence of each member of a gene cluster across genomes, and within a genome itself; while gene clusters collapse all these instances into one (Figure 11).

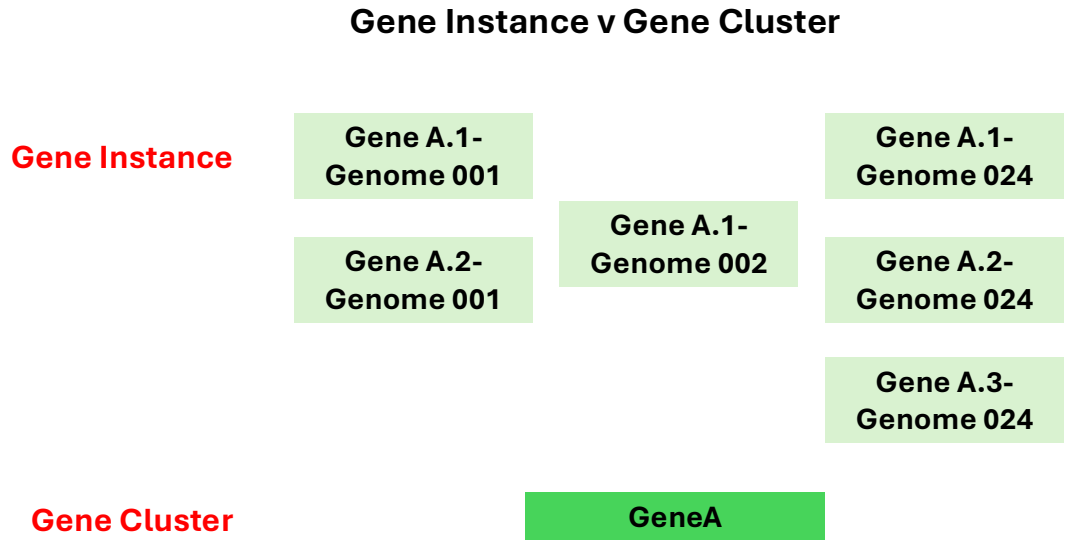


Figure 11. Gene instances are what are used as counts in our core genome v. flexible genome functional enrichment. A gene instance counts each member of a gene across our 3 samples, whereas a gene cluster collapses all these instances for a gene into one.

We continued analyzing by annotation type (ie COG category), and calculated gene proportion annotated by functional category (ie Defense Mechanisms). We took genes annotated for a specific functional category and divided this by total genes belonging to that annotation type. This proportion was then used to construct contingency tables to perform Fisher's exact test, then error corrected with a false discovery rate threshold of 0.05 to determine significantly enriched genes in either the core or flexible genome splits of PB-PSB1 (Figure 12).

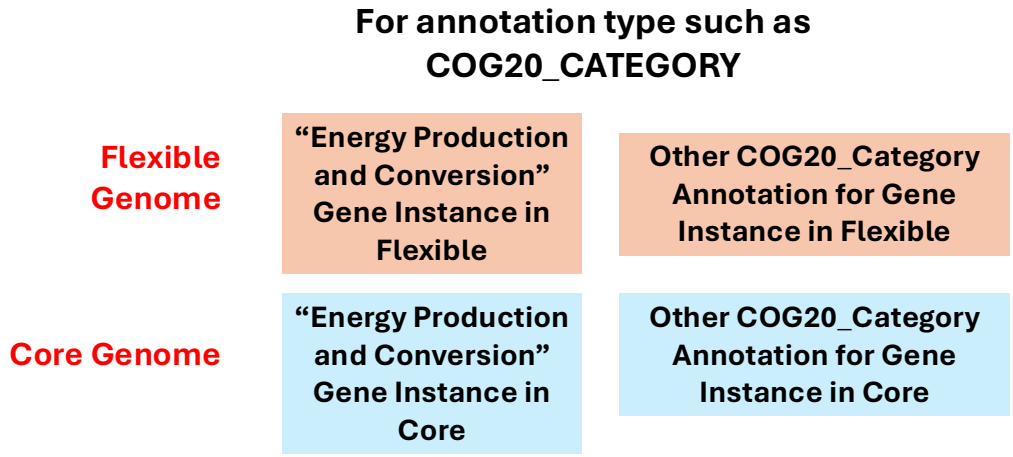


Figure 12. Overview of contingency table for fisher's exact test for core genome v. flexible genome functional enrichment. Gene instances for a specific annotation are compared for membership within the flexible genome relative to everything else with an annotated gene instance in the flexible genome. That same specific annotation is compared for membership in the core genome relative to everything else annotated in the core genome.

The total genes annotated for a specific category may not add up to total gene instances with an annotation, as some genes may have cross-category annotations.

Determination of Gene Gain/Loss in PB-PSB1 over Time

Coding sequences from the 2010 PB-PSB1 reference genome (CP050890.1) were pulled from anvi'o to make a blast database (blast v2.16.0, Sayers et al. 2025). The sample001 and sample002 PB-PSB1 bins were then blastp'd against the reference, -evalue 1e-5 -max_target_seqs 1, and manipulated with R in order to determine which sequences were

present in the reference but gone in sample001 and sample002 (R v4.4.2). To do this, genes were considered present in both the reference and sample00x if identity and query coverage were both over 95%. The remaining reference genes were then considered “reference genes absent in sample00x”. These were compiled for sample001 and sample002 and were the list of genes we explored to see if they truly had disappeared from our 2021 samples (sample001 and sample002). To do this, genomes were visualized in Geneious (Geneious v2025.2.2), and alignment was performed with the mauve plugin.

Results and Discussion

Sequencing and Bin Quality Changes with Time and Technology

We have carried out sequencing efforts of pink berry aggregates over a time period spanning both ends of a decade (2010, 2011, and 2021). We see a marked improvement with time enabled by long read technology such that we see a greater generation of sequencing data with less input biomass (Table 1).

Bins of just PB-PSB1 and PB-SRB1 were generated in 2010 from these earliest sequencing runs, and were highly fragmented and numbering in the hundreds of scaffolds (Supplemental Table 1-2). These bins were improved by both Australian (Parks et al. 2018), and European groups in larger studies that re-assembled and re-binned existing sequencing datasets (Mitchell et al. 2018). The bins generated from their reworking our 2010 PB-PSB1 and PB-SRB1 bins were much less fragmented (1386 to 690 scaffolds for PB-PSB1, stayed at a similar number for PB-SRB1). Despite this, while our 2010 PB-PSB1 bin was 6.7Mb big, the Parks *et al*. bins were much smaller (3.8Mb and 5.4Mb), as was the Mitchell *et al*. bin (5.9Mb). While this deviation was less evident in PB-SRB1 we still saw a similar parallel such that our 2010 PB-SRB1 bin was largest at 3.7Mb big, Parks *et al*.'s were smaller than our reference (2.6Mb and 3.3Mb), and Mitchell *et al*.'s was just slightly smaller than our 2010 reference (3.6Mb). A later study from our group in 2022 integrated PacBio RSII technology and manual curation to vastly improve our bins such that they were greater in genome size (PB-PSB1 was 7.9Mb and PB-SRB1 was 4.7Mb), had fewer scaffolds (PB-PSB went from 1386 scaffolds to 1, PB-SRB1 went from 422 scaffolds to 71), and displayed higher completion. In this current study we used PacBio HiFi sequencing without any manual curation and we were able to consistently produce bins for both PB-PSB1 and PB-SRB1 that

had a low scaffold count (4-26 scaffolds for PB-PSB1 and 2-9 scaffolds for PB-SRB1), and high completion.

Our community, furthermore, consists of various other *Alphaproteobacteria* as well as *Bacteroides* members. Within the *Alphaproteobacteria* we saw *Oceanicaulis alexandrii*, a *Micavibrio* and a *Rhodobacteraceae*. We saw similar quality trajectories for each of these genera such that *Oceanicaulis alexandrii* was binned into a single contig with PacBio HiFi, *Micavibrio* similarly also collapsed into bins of much smaller scaffold size (1-2), and the *Rhodobacteraceae* saw an order of magnitude scaffold reduction between Illumina and PacBio (both RSII/HiFi) albeit HiFi resulted in less contigs than RSII as well as higher completion. Within the *Bacteroides*, we saw a *Salibacteraceae* (UBA10426) whose scaffold number we were able to reduce from about 100 to just 2 scaffolds at about 100-300X coverage, as well as a *Lewinella* which we got to 10-15 scaffolds with much lower coverage (20X). We also got a lower scaffold bin for *Cyclobacteriaceae* (UBA10427) (300 to 9 scaffolds), as well as for a *Bacteroidales* (UBA10428) (600 to 90 scaffolds). Our sampling thus got higher quality bins for both our main community members as well as various other *Alphaproteobacteria* and *Bacteroides* members of the community.

date sampled	Sample name	sample description	Citation	Platform	Bases (Gigabases)	Read N50	Library layou	NCBI BioProject	NCBI Biosample	SRA run	SRA Experiment	Notes
2010-07-01	pinkB1	10 pooled berries	Wilbanks et al. 2014	Roche 454 Titanium GS FLX+	0.09402	350	Single	PRJNA214436	SAMN02303577	SRR948504	SRX332170	454 and Hiseq data co-assembled for MAGs in Wilbanks et al 2014
2010-07-01	pinkB2	10 pooled berries	Wilbanks et al. 2014	Illumina HiSeq 2500	0.5256	100	Paired	PRJNA214436	SAMN02303578	SRR948505	SRX332174	454 and Hiseq data co-assembled for MAGs in Wilbanks et al 2014
2010-07-01	pinkB3	10 pooled berries	Wilbanks et al. 2014	Illumina MiSeq	5.49	250	Paired	PRJNA214436	SAMN02303579	SRR948829	SRX332175	Not assembled for Wilbanks et al 2014; this data was assembled and binned by: Parks DH et al. Nat Biotechnol, 2018 Nov;36(10):996-1004; AND Mitchell AL et al., Nucleic Acids Res, 2018 Jan 4;46(D1):D726-D735
2011-07-01	s01_2011	pool of 13 aggregates from 2 - 4 mm in diameter	Wilbanks et al. 2022	PacBio RS II	4.72 (raw) 0.58 (filtered)	6.1 kb	Single	PRJNA684324	SAMN17054265	SRR13296979, SRR13296980, SRR13296981, SRR13296982, SRR13296983, SRR13296984, SRR13296988, SRR13296989, SRR13296990, SRR13296991, SRR13296992, SRR13297009, SRR13297010, SRR13297016, SRR13297017, SRR13297018, SRR14255432	SRX9725939, SRX9725938, SRX9725937, SRX9725936, SRX9725935, SRX9725934, SRX9725930, SRX9725929, SRX9725928, SRX9725927, SRX9725926, SRX9725909, SRX9725908, SRX9725902, SRX9725901, SRX9725900, SRX10617869	
2011-07-01	berry9_2011	one large berry ~9 mm in diameter	Wilbanks et al. 2022	PacBio RS II	25.4 (raw) 7.14 (filtered)	2.9 kb	Single	PRJNA684324	SAMN17053571	SRR13296985, SRR13296986, SRR13296987, SRR13296993, SRR13296994, SRR13296995, SRR13296996, SRR13296997, SRR13296998, SRR13296999, SRR13297000, SRR13297001, SRR13297002, SRR13297003, SRR13297004, SRR13297005, SRR13297006, SRR13297007, SRR13297008, SRR13297011, SRR13297014, SRR13297015, SRR13297019, SRR13297020, SRR14255433, SRR14255434	SRX9725933, SRX9725932, SRX9725931, SRX9725925, SRX9725924, SRX9725923, SRX9725922, SRX9725921, SRX9725920, SRX9725919, SRX9725918, SRX9725917, SRX9725916, SRX9725915, SRX9725914, SRX9725913, SRX9725912, SRX9725911, SRX9725910, SRX9725907, SRX9725904, SRX9725903, SRX9725899, SRX9725898, SRX10617868, SRX10617867	
2011-07-01	berry9_2011+ s01_2011	in silico combination of all pacbio reads from berry9_2011 and s01_2011	Wilbanks et al. 2022	PacBio RS II	7.72 (filtered) 0.58 (error-corrected)	6.9 kb	Single	PRJNA684324	SAMN17228542	*see above*	*see above*	This was co-assembled and used for downstream pacbio analysis in Wilbanks et al2022
2011-07-01	berry9_2011	one large berry ~9 mm in diameter	Wilbanks et al. 2022	Illumina HiSeq 2500	1.19	150	Paired	PRJNA684324	SAMN17053571	SRR13297008	SRX9725910	
2011-07-01	s02_2011	pool of 10 aggregates from 2 - 4 mm in diameter	Wilbanks et al. 2022	Illumina HiSeq 2500	35.72	150	Paired	PRJNA684324	SAMN17054266	SRR13297012	SRX9725906	
2011-07-01	s02_2011	pool of 10 aggregates from 2 - 4 mm in diameter	Wilbanks et al. 2022	Illumina MiSeq 2500	6.86	250	Paired	PRJNA684324	SAMN17054266	SRR13297013	SRX9725905	
2021-07-30	LS01_2021_001	1 pink berry aggregate	this study	PacBio HiFi	15.78 (filtered) 15.94 (raw)		Single					
2021-07-30	LS01_2021_002	1 pink berry aggregate	this study	PacBio HiFi	16.68 (filtered) 16.85 (raw)		Single					
2021-07-30	LS01_2021_024	1 pink berry aggregate	this study	PacBio HiFi	13.17 (filtered) 13.29 (raw)		Single					

Table 1. Sequencing stats for our various short-read and long-read sequencing runs from berries in 2010, 2011 and 2021.

Short-Read Illumina Data Demonstrates Little Variation in Community Composition over Time, Long-Read Sequencing Allowed for Capturing much more Taxonomic Depth

While previous estimates from 16S rRNA clone library Sanger sequencing data showed PB-PSB1 and PB-SRB1 to make up 60-70% of the community, we recently regenerated estimates with previous Illumina sequencing data from 2010 and 2011. PB-PSB1 was consistently present at about 50% relative abundance and PB-SRB1 was between 7.5-9.5%. The next most abundant community members consistently seem to be the *Rhodobacteraceae* and *Micavibrio*, with *Oceanicaulis alexandrii* showing some variation. Various heterotrophs make up the rest of the community to varying degrees, although we consistently detect *Salibacteraceae*, *Winogradskyella*, and *Salinispira*.

While our short-read estimates were mostly consistent with each other, our long-read samples showed more variation (Figure 13). While the 2011 long-read runs varied in the abundance attributed to community members, like a greater abundance of *Rhodobacteraceae* via PacBio RSII, the moleculo sequencing missed the presence of *Alphaproteobacteria* like *Micavibrio* and *Oceanicaulis alexandrii*. Interestingly our 2021 HiFi sequencing captured a much greater taxonomic depth than any of our other samples, yet it consistently displayed the same taxa at the greatest abundance, albeit with variation in actual abundance values (PB-PSB1, PB-SRB1, *Rhodobacteraceae*, *Salibacteraceae*, *Micavibrio*, *Cyclobacteriaceae*, *Oceanicaulis alexandrii*, *Luminiphilus*, *Winogradskyella*, and *Salinispira* make up the majority of all HiFi samples). PB-PSB1, for example, displayed the lowest relative abundance out of any of our samples in hifi_001, at a relative abundance of 16.87%, yet it was always at the highest abundance compared to any other community member.

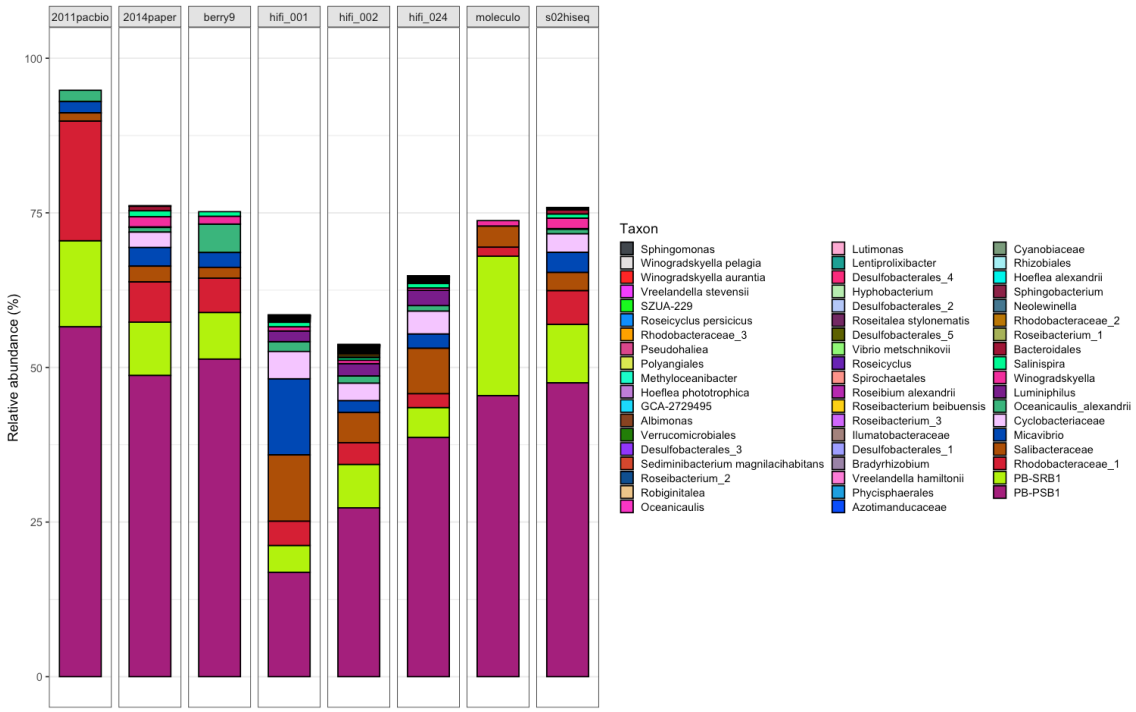


Figure 13. Community relative abundance in berries sampled in 2010, 2011, and 2021. “2014paper”, “berry9”, and “s02hiseq” were sequenced with Illumina short-read technology in 2010 and 2011. “2011pacbio” was sequenced in 2011 with PacBioRSII technology and “moleculo” in 2011 with moleculo. “hifi_001”, “hifi_002”, and “hifi_024” were all sequenced in 2021 with PacBio HiFi technology. RA values were generated by singleM. Remaining values correspond to “unassigned”.

Novel Order, Genera, and Species added to the Genome Taxonomy Database by this Study

Our PacBio HiFi sequencing generated 41 distinct, high-quality bins. Only 8 of these bins had a confident species level assignment from the Genome Taxonomy Database (PB-SRB1, UBA10428 sp003508755-*Bacteroidales*, UBA10426 sp003499525-*Salibacteraceae*, UBA10427 sp003500135-*Cyclobacteriaceae*, *Micavibrio*, *Rhodobacteraceae*, *Oceanicaulis alexandrii*, and PB-PSB1, GTDBr226). Two bins belonged to a novel order, albeit we did not determine whether they belonged to the same novel order, but we did determine the ANI between the two bins to be below 80%. Further, 11 bins represented novel genera. Two of these 11 bins were assigned to family SZUA-146 with an ANI of 79.2-79.4%. Conservatively we thus estimate this study to have generated 1 novel order and 10 novel genera. The

remaining 20 bins represent novel species, such that even if a bin had the same species designation, their ANI was well below 95% and thus we considered these separate species (Figure 14A).

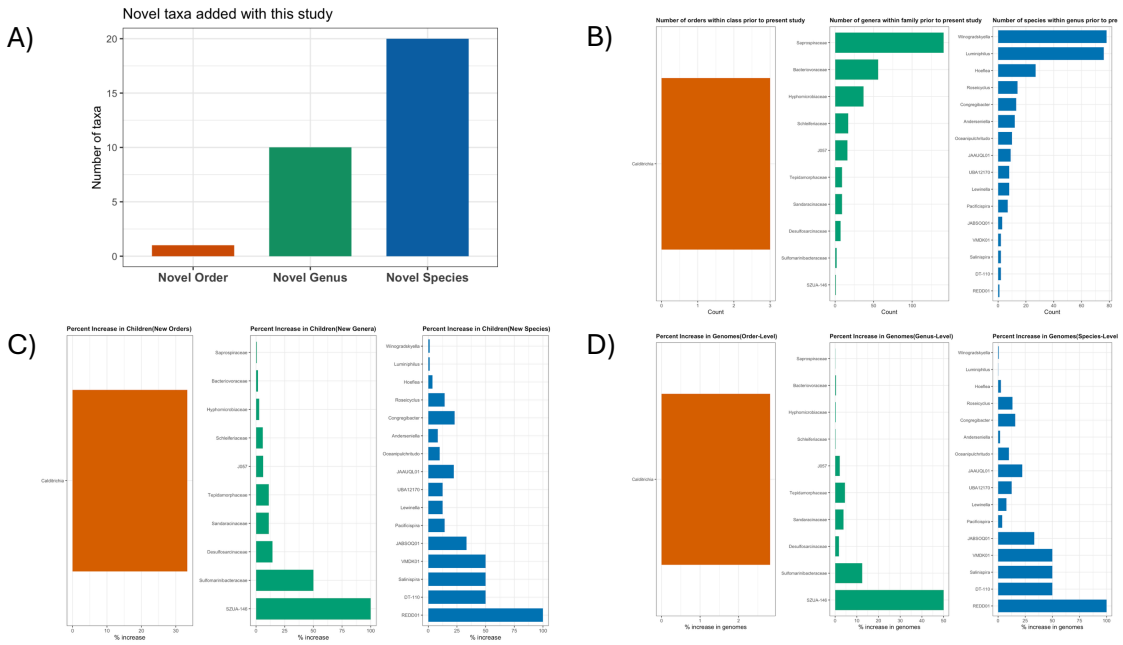


Figure 14. Comparison of our novel bin taxonomy to that in the online Genome Taxonomy Database (Parks et al. 2018). (A) Number of taxa added to this study representing novel orders, genera, or species. (B) The number of distinct children already present within a parent class in the Genome Taxonomy Database. (C) Percent increase in children within a parent class due to the bins generated by our study. (D) Percent increase in genomes within a parent class due to the bins generated by our study.

To contextualize the value of these new bins, we examined the extent to which these novel MAGs increased both richness (e.g. the number of genera within a family) and sampling (e.g. the total number of genomes available in for the family) at that taxonomic level.

The families with the most member genera (children) were *Saprospiraceae*, *Bacteriovoraceae* and *Hyphomicrobiaceae*, while the least members corresponded to *Desulfosarcinaceae*, *Sulfomarinibacteraceae*, and SZUA-146. The genera with the greatest amount of member species were *Winogradskyella* and *Luminiphilus*, whereas the least rich

consisted of JABSOQ01, VMDK01, *Salinispira*, DT-110, and REDD01. Generally the genera and species with the least amount of member "children" were the most impacted by our study and saw the greatest percent increase when measured by either percent increase of "children" as well as genomes deposited. The most impacted genera were *Sulfomarinibacteraceae* and SZUA-146. The most impacted species were JABSOQ01, VMDK01, *Salinispira*, DT-110, and REDD01 (1).

We further investigated the genome information and data for bins in genera and species with no cultured isolates (Supplemental Tables 10-20). We aimed to explore the taxonomic levels where our binning added the most amount of novelty. We found that other species within the SZUA-146 genus came from marine sources like the Atlantic and Pacific oceans or marine adjacent sources like sand from Port Phillip Bay. Similarly, all these bins fall between 3-4Mb and were decently complete (80-90%).

Three other bins were classified as novel species under JABSOQ01. The two smaller, lower quality bins were from activated sludge and biofloc rearing sediment, whereas the biggest and most similarly complete bin to ours (~95%) originated from station ALOHA in the Pacific Ocean.

Two other bins were classified under the VMDK01 genus. The less complete one originated from activated sludge in Hong Kong and was bigger than ours, while the more complete one originated from coastal seater off Georgia and was smaller than our bin.

DT-110 also spanned just two species, were decently complete (80-90%), and were both smaller than our bin (5Mb v ~7Mb). One of these bins originated from Pacific Ocean water from the North Pacific gyre, whereas the other originated from a Chinese ballast tank.

REDD01 has only one other species classified under this genus, and displays a smaller genome size (2.7Mb v 4Mb), and originates from a Canadian alkaline salt lake. Largely the uncultured bins most closely related to ours are of marine or marine-adjacent origin. Often, our bins allow for somewhat better reconstructed genomes and improve our bin quality for these underrepresented genera and species.

PB-PSB1 Pangenome

We compared PB-PSB1 bins from sample001, sample002, sample024 as well as the 2011 PacBio reference genome (CP050890, Wilbanks et al. 2022) (Figure 14A). The "core genome" was assigned to clusters that spanned all four bins, whereas the remainder was denoted as "flexible genome". The PB-PSB1 core genome consisted of 4882 total gene clusters, whereas the flexible genome consisted of 933 gene clusters. We observed that a sizable portion of gene content was shared across our reference, sample001 and sample002 but was missing in sample024. This might in fact be a binning issue such that a missing contig actually should be in our sample024 bin and this portion of the flexible genome would thus instead belong in the core genome. Furthermore, sample024 consisted of the least amount of sequencing data and the corresponding PB-PSB1 bin had the most amount of scaffolds and was smallest in size. We thus thought it to be possible that the PB-PSB1 sample024 bin actually had a misbinned contig and was not an accurate presentation of PB-PSB1. We therefore generated a further pangenome without our sample024 bin. We compared just sample001, sample002, and the 2011 PacBio reference genome (Figure 14B).

This new pangenome resulted in 5282 core gene clusters, and 397 flexible gene clusters.

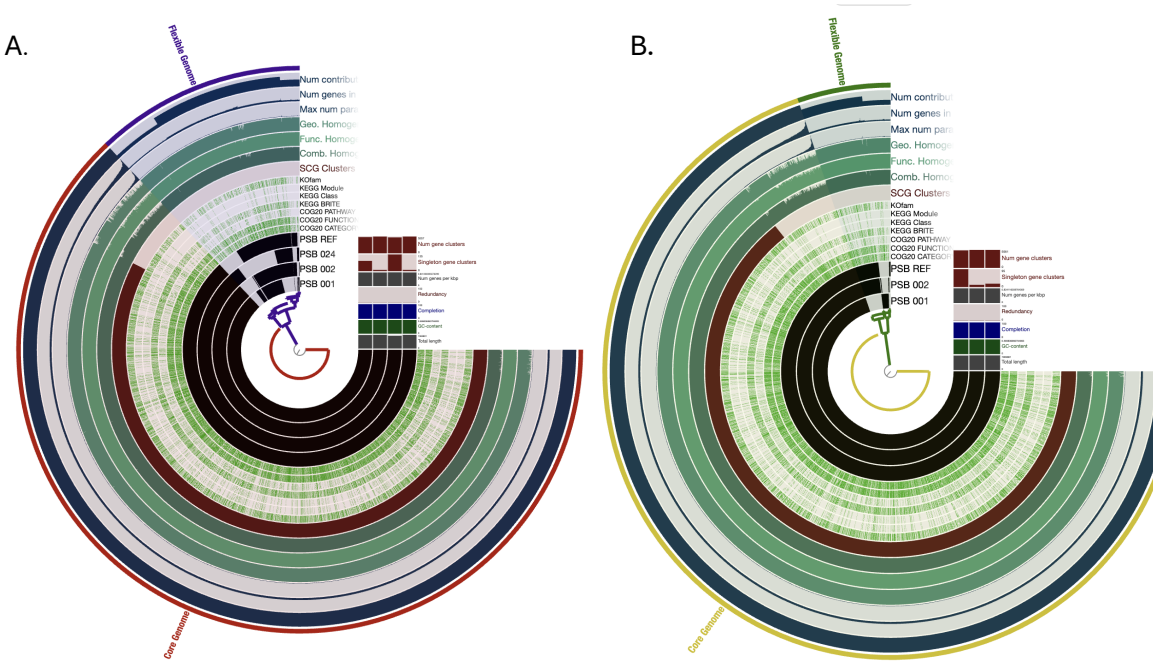


Figure 14. (A)Visualization of PB-PSB1 pangenome generated via anvi'o. Composite bins are from sample001, sample002, sample024, and the 2011 PacBio reference genome. (B) Composite bins are from sample001, sample002, and the 2011 PacBio reference genome.

PB-PSB1's Flexible Genome is Functionally Enriched for Defense Mechanisms whereas its Core Genome is Enriched for Central Metabolic Capacities

We aimed to investigate the differences between core and flexible genome content in PB-PSB1, comparing the reference versus our sample001 and sample002 bins. We found general agreement between the annotation coverage for both gene instances and gene clusters across the different types of annotations approached by anvio: COG category, COG function, KEGG KOfam, and KEGG class (Table 4). We saw the greatest annotation coverage for COG category, where we saw ~67% annotation coverage in the core genome, and ~47% in the flexible. This was followed by KEGG KOfam annotations where we saw ~45% annotation coverage in the core genome, and ~27% in the flexible genome. We saw the least

annotation coverage by KEGG class, such that the core was roughly ~8% annotated and the flexible was at ~4%.

	COG Category	COG Function	KEGG Kofam	KEGG Class
Gene Instance Core	12713/19096	12713/19096	8249/19096	1509/19096
Gene Instance Flexible	338/698	338/698	201/698	27/698
Gene Cluster Core	3612/5282	3612/5282	2430/5282	496/5282
Gene Cluster Flexible	183/397	183/397	106/397	15/397

Table 4. Number of genes annotated by a specific annotation type across our different annotation schemes in the core and flexible genome. This is done for both gene instances and gene clusters (see methods).

We then compared the gene instances that were significantly enriched in either the core or flexible fractions (Supplemental Table 21). We found defense mechanisms to be enriched in the flexible genome (COG category), as well as pathways for lysine and methane metabolism in addition to gene sets for drug resistance (KEGG class). The core genome was enriched for the transposons and prophages, coenzyme transport and metabolism, nucleotide transport and metabolism, intracellular trafficking, and cell motility (COG category). Counts and proportions annotated were similar across gene instances and gene clusters for significantly enriched COG categories (Figure 15). While raw counts are higher in all instances of the core genome, once accounting for the proportion of each of these genomes that was annotated the flexible genome can be visualized to have a higher proportion of defense mechanisms (Figure 15CD).

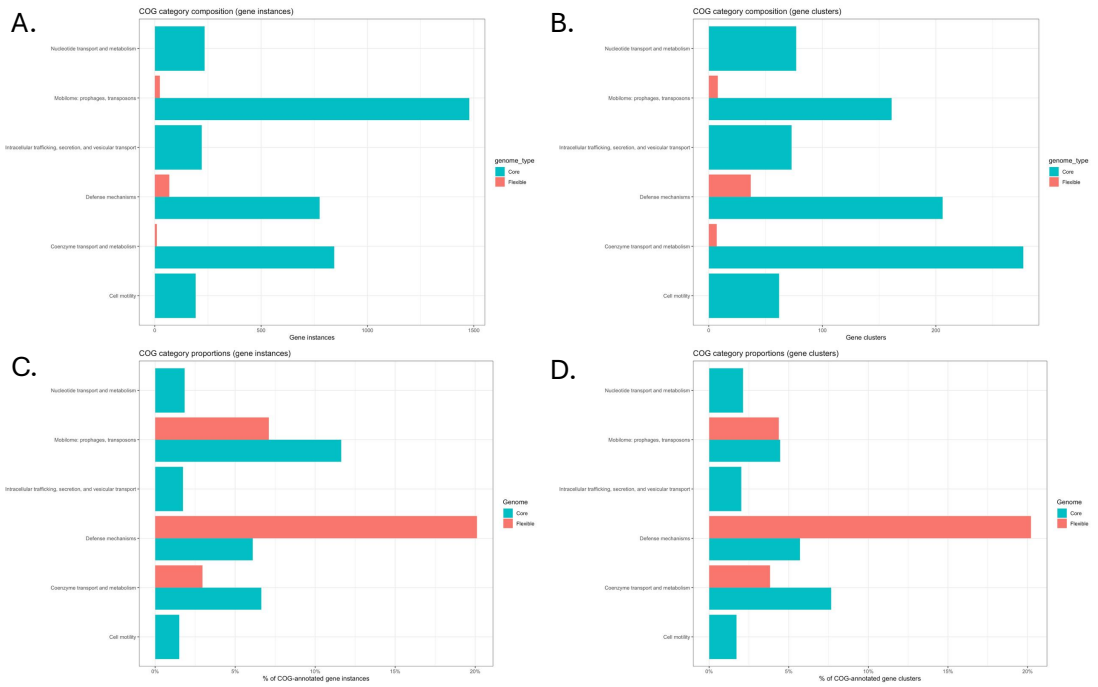


Figure 15. COG categories enriched in either the core or flexible genomes. Proportions refer to number annotated in a specific COG category over all genes annotated to a COG category. (A) Gene instance counts. (B) Gene cluster counts. (C) Gene instance proportions. (D) Gene cluster proportions.

Within our COG category enrichment, we saw an enrichment for defense mechanisms in the flexible genome, which when dissected by function we specifically saw an enrichment for various CRISPR associated proteins. Proteins 1,3-6 of the type IIIB/C or CMR complex (Haft et al. 2005, 45), which is the main RNA cleaving complex (J. Zhang et al. 2012), were absent in the core genome. We also saw proteins 3 and 6 of the type IIIA/D CSM complex, which is another type of effector complex (You et al. 2019). We saw an enrichment for *cas* proteins like *cas2* which is associated with spacer integration; as well as *cas10* which is associated with crRNA and target binding, target DNA cleavage, and signal transduction (Makarova et al. 2025). Typically, the core genome is comprised of genes contributing to basic biology and maintenance, whereas the flexible genome can confer selective advantage by aiding in adaptation to a new niche (Medini et al. 2005).

Such novel niches may be the result of newfound stresses like antibiotic exposure or viral attack. Studies have found niche-specific enrichment for genes associated with antibiotic resistance, phage defense, and stress tolerance; all in gene cassettes whose movement was facilitated by integrons (Ghaly et al. 2023). This niche adaptation can thus expand the flexible genome to help promote survival against new stresses.

In a study with *Pseudomonas aeruginosa*, furthermore, the flexible genome portions were found to be enriched for metal resistance genes, however this was thought to synergistically co-select for antibiotic resistance genes as the flexible portion seemed to have an elevated fraction dedicated to resistance (García-Rivera et al. 2024). Plasmids, moreover, were enriched in antibiotic resistance genes and in the same area as integrons, highlighting the machinery that may help such genomic structures spread.

The core genome, in contrast, is thought to be comprised of functions necessary for central metabolism and basic function as opposed to niche specific adaptation. In *Pseudomonas aeruginosa*, for example, the core genome was shown to display respiratory, coenzyme, nucleotide, lipid, carbohydrate and amino acid metabolism genes (Valot et al. 2015). This coincides with our core genome's enrichment of genes involved in coenzyme and nucleotide metabolism. Intracellular trafficking and cell motility are also needed for basic function, so their enrichment is sensible.

While our COG20 category was enriched for transposons in the core genome, our KEGGkofam analysis was enriched for putative transposons in the core genome but transposons in the flexible genome. While studies of *Pseudomonas aeruginosa* exhibit an enrichment of mobile genetic elements in the flexible genome, with integrons being found only in the flexible genome (Ozer et al. 2014), 10% of our genome is composed of

transposases and could justify our differing values for enrichment. This highlights the dynamic nature of genomes enabled by genomic rearrangement machinery like transposons and integrons. A simple presence/absence determination in "X" genomes, furthermore, may miss variation that can be captured by transposon copy number. While this warrants further analysis, we proceeded to look at specific examples of genomic rearrangement within our samples as facilitated by transposases and integron integrases.

Genomic Variation between PB-PSB1 Genomes Sampled in 2010 and 2021

In order to look at specific examples of the dynamic nature of PB-PSB1's genome we compared the content of our reference genome sampled in 2011 to our new bins generated in 2021 with PacBio HiFi. We looked for genes present in 2011, that were absent in our "modern" 2021 bins. We then aligned these two regions to look for visual evidence of insertion or deletion into our 2021 bins.

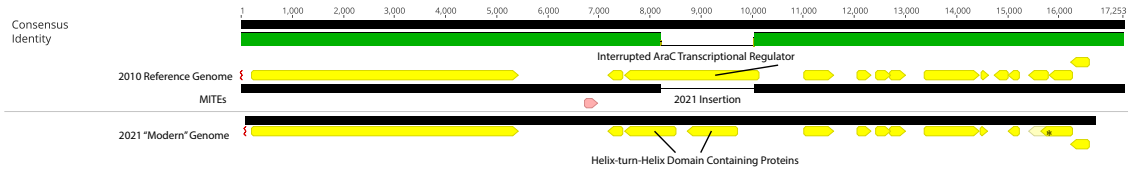


Figure 16. Geneious alignment between 2011 reference genome and 2021 "modern" genome of PB-PSB1, displaying an insertion into 2021.

In one such region we found evidence for a small insertion (Figure 16), such that a helix-turn-helix domain protein was inserted into two of the 2021 genomes, interrupting a portion of AraC. Some of AraC was lost by this insertion and thus was picked up by our workflow to find portions of the reference genome missing in 2021. The interrupted gene as well as the insertion are both annotated as helix-turn-helix domain containing proteins.

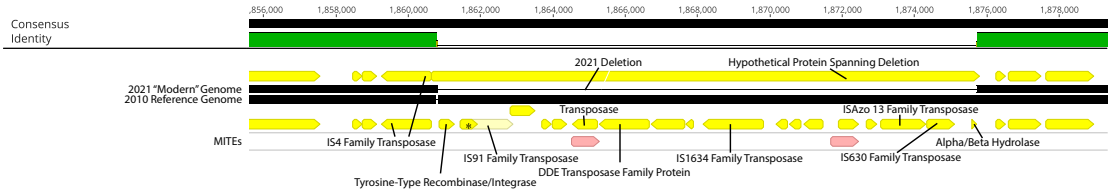


Figure 17. Mauve alignment between 2011 reference genome and 2021 "modern" genome of PB-PSB1, displaying a deletion into 2021.

Furthermore, we were also able to detect deletions (Figure 17). A genomic span of our reference genome preceded by an IS4 family transposase was deleted in two of our modern 2021 genomes. This 15kb deleted region from our reference genome contained various transposases, and in our modern 2021 genome is spanned by a hypothetical protein.

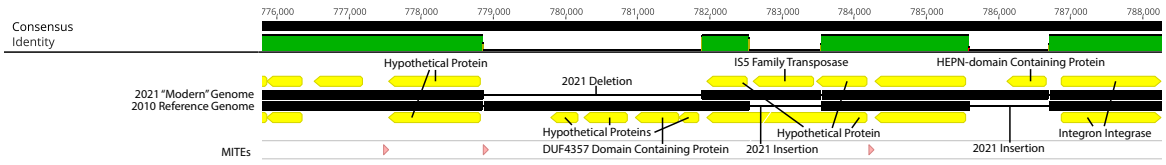


Figure 18. Mauve alignment between 2011 reference genome and 2021 "modern" genome of PB-PSB1, displaying one deletion and two insertions into 2021

Our genome is highly dynamic and also displayed evidence of both deletions and insertions taking place in close proximity (Figure 18). This genomic span in PB-PSB1 displays one deletion and two insertion between our 2011 to 2021 genomes. The deleted portion is preceded by a hypothetical protein that is flanked by 2 MITE regions and contains hypothetical proteins flanking a DUF4357 domain containing protein. The insertion allowed

the addition of an IS5 family transposase and a HEPN domain containing protein to the PB-PSB1 genome. The second insertion is followed by an integron integrase.

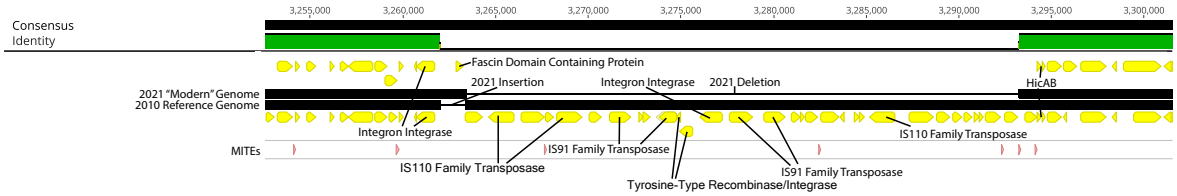


Figure 19. Mauve alignment between 2011 reference genome and 2021 "modern" genome of PB-PSB1, displaying one small insertion and a bigger deletion into 2021.

Finally, we saw variability in the size of our deletions as is evidenced by a 30kb deletion region we investigated (Figure 19). This genomic span exhibits a slightly smaller than 30kb deletion from our reference to our 2021 genome, as well as a small insertion of fascin-domain containing protein. The deleted portion exhibited a wide variety of transposases (IS91 and IS110) as well as integrases. This area of genomic rearrangement is spanned by and integron integrase and both components of the type II toxin-antitoxin system HicAB.

Between 2011 and 2021 we saw evidence of both insertions and deletions, and both of these showed variability in the size of genomic region affected. These areas tended to be repeat rich as well as populated by transposases, integrases, and/or MITEs. This affords the potential for the genomic rearrangement we saw evidenced throughout our genomes between 2011 and 2021.

Conclusions

To better represent our pink berry community, in this chapter we generated 41 contemporaneous (2021), high quality bins via long-read PacBio HiFi sequencing of three individual pink berry aggregates. This new sequencing allowed us to generate updated estimates of community member relative abundances such that we revealed much greater taxonomic depth than our earlier 2011 sequencing. We found, however, that our communities are still consistently dominated by the same two organisms, PB-PSB1 and PB-SRB1 even ten years later. We took bins without a confident species level designation and compared them to the Genome Taxonomy Database in order to explore the novelty generated from our dataset. Conservatively, we produced members of 1 novel order, 10 novel genera, and 20 novel species.

We further delved into our bin for the main member of the consortium, PB-PSB1, and generated a pangenome to capture the coding space of this organism across our three aggregates. In terms of function, we found the core genome of PB-PSB1 to be enriched in functions related to central metabolism while the flexible genome was enriched in functions related to defense. Structurally, we were able to detect variation over time such that we found multiple instances of insertions and deletions between 2011 and 2021. This helped us generate evidence illustrating the dynamic nature of the genome of PB-PSB1 as well as characterize its functional content using contemporary, really high-quality bins.

References

Bushnell, Brian. 2014. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Released. https://bbmap.org.

Chaumeil, Pierre-Alain, Aaron J. Mussig, Philip Hugenholtz, and Donovan H. Parks. 2020. "GTDB-Tk: A Toolkit to Classify Genomes with the Genome Taxonomy Database." *Bioinformatics* 36 (6): 1925–27. https://doi.org/10.1093/bioinformatics/btz848.

Chen, Lin-Xing, Karthik Anantharaman, Alon Shaiber, A. Murat Eren, and Jillian F. Banfield. 2020. "Accurate and Complete Genomes from Metagenomes." *Genome Research* 30 (3): 315–33. https://doi.org/10.1101/gr.258640.119.

Doré, Hugo, Amy R. Eisenberg, Emily N. Junkins, et al. 2024. "Targeted Hypermutation of Putative Antigen Sensors in Multicellular Bacteria." *Proceedings of the National Academy of Sciences* 121 (9): e2316469121. https://doi.org/10.1073/pnas.2316469121.

Eren, A. Murat, Evan Kiefl, Alon Shaiber, et al. 2020. "Community-Led, Integrated, Reproducible Multi-Omics with Anvi'o." *Nature Microbiology* 6 (1): 3–6. https://doi.org/10.1038/s41564-020-00834-3.

Espinosa, Elena, Rocio Bautista, Rafael Larrosa, and Oscar Plata. 2024. "Advancements in Long-Read Genome Sequencing Technologies and Algorithms." *Genomics* 116 (3): 110842. https://doi.org/10.1016/j.ygeno.2024.110842.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. 2012. "CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data." *Bioinformatics* 28 (23): 3150–52. https://doi.org/10.1093/bioinformatics/bts565.

García-Rivera, Celia, Carmen Molina-Pardines, José M. Haro-Moreno, Mónica Parra Grande, Juan Carlos Rodríguez, and Mario López-Pérez. 2024. "Genomic Analysis of

Antimicrobial Resistance in Pseudomonas Aeruginosa from a 'One Health' Perspective." *Microorganisms* 12 (9): 1770. https://doi.org/10.3390/microorganisms12091770.

Garud, Nandita R., Benjamin H. Good, Oskar Hallatschek, and Katherine S. Pollard. 2019. "Evolutionary Dynamics of Bacteria in the Gut Microbiome within and across Hosts." *PLOS Biology* 17 (1): e3000102. https://doi.org/10.1371/journal.pbio.3000102.

Ghaly, Timothy M., Vaheesan Rajabal, Anahit Penesyan, et al. 2023. "Functional Enrichment of Integrons: Facilitators of Antimicrobial Resistance and Niche Adaptation." *iScience* 26 (11): 108301. https://doi.org/10.1016/j.isci.2023.108301.

Gurbich, Tatiana A., Alexandre Almeida, Martin Beracochea, et al. 2023. "MGnify Genomes: A Resource for Biome-Specific Microbial Genome Catalogues." *Journal of Molecular Biology* 435 (14): 168016. https://doi.org/10.1016/j.jmb.2023.168016.

Haft, Daniel H., Jeremy Selengut, Emmanuel F. Mongodin, and Karen E. Nelson. 2005. "A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes." *PLoS Computational Biology* 1 (6): e60. https://doi.org/10.1371/journal.pcbi.0010060.

Hyatt, Doug, Gwo-Liang Chen, Philip F. LoCascio, Miriam L. Land, Frank W. Larimer, and Loren J. Hauser. 2010. "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification." *BMC Bioinformatics* 11 (1): 119. https://doi.org/10.1186/1471-2105-11-119.

Jin, Hao, Lijun You, Feiyan Zhao, et al. 2022. "Hybrid, Ultra-Deep Metagenomic Sequencing Enables Genomic and Functional Characterization of Low-Abundance Species in the Human Gut Microbiome." *Gut Microbes* 14 (1): 2021790. https://doi.org/10.1080/19490976.2021.2021790.

Kang, Dongwan D., Feng Li, Edward Kirton, et al. 2019. “MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies.” *PeerJ* 7 (July): e7359. https://doi.org/10.7717/peerj.7359.

Kolmogorov, Mikhail, Jeffrey Yuan, Yu Lin, and Pavel A. Pevzner. 2019. “Assembly of Long, Error-Prone Reads Using Repeat Graphs.” *Nature Biotechnology* 37 (5): 540–46. https://doi.org/10.1038/s41587-019-0072-8.

Li, Heng. 2018. “Minimap2: Pairwise Alignment for Nucleotide Sequences.” *Bioinformatics* 34 (18): 3094–100. https://doi.org/10.1093/bioinformatics/bty191.

Li, Heng, Bob Handsaker, Alec Wysoker, et al. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Makarova, Kira S., Sergey A. Shmakov, Yuri I. Wolf, et al. 2025. “An Updated Evolutionary Classification of CRISPR–Cas Systems Including Rare Variants.” *Nature Microbiology* 10 (12): 3346–61. https://doi.org/10.1038/s41564-025-02180-8.

Mallick, Himel, Siyuan Ma, Eric A. Franzosa, Tommi Vatanen, Xochitl C. Morgan, and Curtis Huttenhower. 2017. “Experimental Design and Quantitative Analysis of Microbial Community Multiomics.” *Genome Biology* 18 (1): 228. https://doi.org/10.1186/s13059-017-1359-z.

Medini, Duccio, Claudio Donati, Hervé Tettelin, Vega Masignani, and Rino Rappuoli. 2005. “The Microbial Pan-Genome.” *Current Opinion in Genetics & Development* 15 (6): 589–94. https://doi.org/10.1016/j.gde.2005.09.006.

Meziti, Alexandra, Luis M. Rodriguez-R, Janet K. Hatt, Angela Peña-Gonzalez, Karen Levy, and Konstantinos T. Konstantinidis. 2021. “The Reliability of Metagenome-

Assembled Genomes (MAGs) in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived from the Same Fecal Sample." *Applied and Environmental Microbiology* 87 (6): e02593-20. https://doi.org/10.1128/AEM.02593-20.

Mitchell, Alex L., Maxim Scheremetjew, Hubert Denise, et al. 2018. "EBI Metagenomics in 2017: Enriching the Analysis of Microbial Communities, from Sequence Reads to Assemblies." *Nucleic Acids Research* 46 (D1): D726–35. https://doi.org/10.1093/nar/gkx967.

Olm, Matthew R., Christopher T. Brown, Brandon Brooks, and Jillian F. Banfield. 2017. "dRep: A Tool for Fast and Accurate Genomic Comparisons That Enables Improved Genome Recovery from Metagenomes through de-Replication." *The ISME Journal* 11 (12): 2864–68. https://doi.org/10.1038/ismej.2017.126.

Ozer, Egon A., Jonathan P. Allen, and Alan R. Hauser. 2014. "Characterization of the Core and Accessory Genomes of Pseudomonas Aeruginosa Using Bioinformatic Tools Spine and AGEnt." *BMC Genomics* 15 (1): 737. https://doi.org/10.1186/1471-2164-15-737.

Parks, Donovan H., Maria Chuvochina, Christian Rinke, Aaron J. Mussig, Pierre-Alain Chaumeil, and Philip Hugenholtz. 2022a. "GTDB: An Ongoing Census of Bacterial and Archaeal Diversity through a Phylogenetically Consistent, Rank Normalized and Complete Genome-Based Taxonomy." *Nucleic Acids Research* 50 (D1): D785–94. https://doi.org/10.1093/nar/gkab776.

Parks, Donovan H., Maria Chuvochina, Christian Rinke, Aaron J. Mussig, Pierre-Alain Chaumeil, and Philip Hugenholtz. 2022b. "GTDB: An Ongoing Census of Bacterial and Archaeal Diversity through a Phylogenetically Consistent, Rank Normalized and Complete Genome-Based Taxonomy." *Nucleic Acids Research* 50 (D1): D785–94. https://doi.org/10.1093/nar/gkab776.

Parks, Donovan H., Maria Chuvochina, David W. Waite, et al. 2018. “A Standardized Bacterial Taxonomy Based on Genome Phylogeny Substantially Revises the Tree of Life.” *Nature Biotechnology* 36 (10): 996–1004. https://doi.org/10.1038/nbt.4229.

Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. 2015. “CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes.” *Genome Research* 25 (7): 1043–55. https://doi.org/10.1101/gr.186072.114.

Rice, Peter, Ian Longden, and Alan Bleasby. 2000. “EMBOSS: The European Molecular Biology Open Software Suite.” *Trends in Genetics* 16 (6): 276–77.

Sayers, Eric W., Jeffrey Beck, Evan E. Bolton, et al. 2025. “Database Resources of the National Center for Biotechnology Information in 2025.” *Nucleic Acids Research* 53 (D1): D20–29. https://doi.org/10.1093/nar/gkae979.

Sboner, Andrea, Xinmeng Mu, Dov Greenbaum, Raymond K. Auerbach, and Mark B. Gerstein. 2011. “The Real Cost of Sequencing: Higher than You Think!” *Genome Biology* 12 (8): 125. https://doi.org/10.1186/gb-2011-12-8-125.

Sedlazeck, Fritz J., Philipp Rescheneder, Moritz Smolka, et al. 2018. “Accurate Detection of Complex Structural Variations Using Single-Molecule Sequencing.” *Nature Methods* 15 (6): 461–68. https://doi.org/10.1038/s41592-018-0001-7.

Stoler, Nicholas, and Anton Nekrutenko. 2021. “Sequencing Error Profiles of Illumina Sequencing Instruments.” *NAR Genomics and Bioinformatics* 3 (1): lqab019. https://doi.org/10.1093/nargab/lqab019.

Tatusova, Tatiana, Michael DiCuccio, Azat Badretdin, et al. 2016. “NCBI Prokaryotic Genome Annotation Pipeline.” *Nucleic Acids Research* 44 (14): 6614–24. https://doi.org/10.1093/nar/gkw569.

The Genome Standards Consortium, Robert M. Bowers, Nikos C. Kyrpides, et al. 2017. “Minimum Information about a Single Amplified Genome (MISAG) and a Metagenome-Assembled Genome (MIMAG) of Bacteria and Archaea.” *Nature Biotechnology* 35 (8): 725–31. https://doi.org/10.1038/nbt.3893.

Tonkin-Hill, Gerry, Neil MacAlasdair, Christopher Ruis, et al. 2020. “Producing Polished Prokaryotic Pangenomes with the Panaroo Pipeline.” *Genome Biology* 21 (1): 180. https://doi.org/10.1186/s13059-020-02090-4.

Truong, Duy Tin, Adrian Tett, Edoardo Pasolli, Curtis Huttenhower, and Nicola Segata. 2017. “Microbial Strain-Level Population Structure and Genetic Diversity from Metagenomes.” *Genome Research* 27 (4): 626–38. https://doi.org/10.1101/gr.216242.116.

Valot, Benoît, Christophe Guyeux, Julien Yves Rolland, Kamel Mazouzi, Xavier Bertrand, and Didier Hocquet. 2015. “What It Takes to Be a Pseudomonas Aeruginosa? The Core Genome of the Opportunistic Pathogen Updated.” *PLOS ONE* 10 (5): e0126468. https://doi.org/10.1371/journal.pone.0126468.

Vaser, Robert, Ivan Sović, Niranjan Nagarajan, and Mile Šikić. 2017. “Fast and Accurate de Novo Genome Assembly from Long Uncorrected Reads.” *Genome Research* 27 (5): 737–46. https://doi.org/10.1101/gr.214270.116.

Wilbanks, Elizabeth G., Hugo Doré, Meredith H. Ashby, Cheryl Heiner, Richard J. Roberts, and Jonathan A. Eisen. 2022. “Metagenomic Methylation Patterns Resolve Bacterial

Genomes of Unusual Size and Structural Complexity." *The ISME Journal* 16 (8): 1921–31. https://doi.org/10.1038/s41396-022-01242-7.

Wilbanks, Elizabeth G., Ulrike Jaekel, Verena Salman, et al. 2014. "Microscale Sulfur Cycling in the Phototrophic Pink Berry Consortia of the S Ippewissett S Alt M Arsh." *Environmental Microbiology* 16 (11): 3398–415. https://doi.org/10.1111/1462-2920.12388.

Woodcroft, Ben J., Samuel T. N. Aroney, Rossen Zhao, et al. 2025. "Comprehensive Taxonomic Identification of Microbial Species in Metagenomic Data Using SingleM and Sandpiper." *Nature Biotechnology*, ahead of print, July 16. https://doi.org/10.1038/s41587-025-02738-1.

You, Lilan, Jun Ma, Jiuyu Wang, et al. 2019. "Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-Transcriptional Interference." *Cell* 176 (1–2): 239-253.e16. https://doi.org/10.1016/j.cell.2018.10.052.

Zhang, Jing, Christophe Rouillon, Melina Kerou, et al. 2012. "Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity." *Molecular Cell* 45 (3): 303–13. https://doi.org/10.1016/j.molcel.2011.12.013.

Zhang, Zhi-Feng, Li-Rui Liu, Yue-Ping Pan, Jie Pan, and Meng Li. 2023. "Long-Read Assembled Metagenomic Approaches Improve Our Understanding on Metabolic Potentials of Microbial Community in Mangrove Sediments." *Microbiome* 11 (1): 188. https://doi.org/10.1186/s40168-023-01630-x.

Chapter II: Daytime carbon fixation and anabolic activity by *Thiohalocapsa* PB-PSB1 may produce small carbon compounds used by *Desulfofustis* PB-SRB1

Braulio S. Castillo Villaseñor, Anna E. Warsaw, Emily N. Junkins, and Elizabeth G. Wilbanks

Abstract

Microbes living in communities like biofilms are able to engage in tightly coupled interactions which are hard to detect. Tackling this in communities that are uncultured, furthermore, necessitates probing that may at times be indirect. We look at microbial interactions between PB-PSB1 and PB-SRB1 by tracking changes and patterns due to light-dark exposure in the pink berry biofilm over its diel cycle. By exposing communities to light we hoped to gain insight into physiological processes that lead to PB-PSB1 photosynthate, and then track light-dark patterns that could indicate its use in PB-SRB1. We found daytime expression of genes demonstrating anabolic growth and carbon catabolism during the day in PB-PSB1. At night we found that PB-PSB1 may experiencing nitrogen limitation in response to carbon fixation, and may also be breaking down stores of glycogen to generate energy. We found that PB-SRB1 may be reacting to daytime photosynthate from PB-PSB1, as it may engage in greater daytime sulfate reduction as well as electron transport. We identified some

potential compounds it may be importing and oxidizing during the day, including products of photorespiration like glycolate.

Introduction

Microbial biofilms can be a predominant lifestyle across various ecosystems (Stoodley et al. 2002). This macroenvironment allows for microheterogeneity where a wide variety of interactions can take place, enabling various forms of cooperation and communication (West et al. 2007). Specifically, this can allow for the sharing of metabolites between microbes, which can itself take many forms (D'Souza et al. 2018). Upon sharing metabolites in a community, it may be beneficial to undergo metabolic specialization, as it has been modelled to lead to greater pathway efficiency within syntrophic systems as compared to monocultures (Harvey et al. 2014). Within our own biofilm we thus aimed to see if some kind of metabolic specialization took place between its two main members, such that there would be a characteristic flux in specific metabolic intermediates of carbon, specifically photosynthate, between partners PB-PSB1 and PB-SRB1.

Whereas their natural setting within tidal pools in the Sippewissett salt marsh exposes aggregates to variable oxygen, temperature, and light across a diel cycle, we decided to control for temperature and oxygen via a controlled incubation where the biofilm was exposed to a variable light regime. As PB-PSB1 is an anoxygenic photosynthesizer we expected the "host" PB-PSB1 to exhibit a direct response to light and thus exhibit distinct diel patterns. We then aimed to track whether its partner PB-SRB1 correspondingly exhibits diel patterns and whether they are direct responses to the variable light regime or an indirect response to light influenced by PB-PSB1 exudates. These distinct physiological patterns could then be a display of metabolic specialization where specific resources are shared by PB-PSB1 to PB-SRB1. Specifically, PB-PSB1 could be sharing photosynthetic carbon, as

has been shown in cases like the diatom *Pseudonitzchia multiseries* providing organic carbon to *Sulfitobacter* (Amin et al. 2015).

Potential for Carbon Fixation in PB-PSB1 and its effects on PB-SRB1

PSB are known to photosynthesize anoxygenically and engage in carbon fixation (Pfennig 1977). Purple sulfur bacteria like *Allochromatium vinosum*, furthermore, have been found to possess *cso* genes encoding carboxysomes that compartmentalize and facilitate carbon fixation (Weissgerber et al. 2011). Carboxysomes have yet to be visualized in *A. vinosum*, leaving their actual role in PSB carbon fixation unresolved (Weissgerber et al. 2011). Pink berries themselves, however, have been shown to photosynthesize and fix carbon with light (Wilbanks et al. 2014), thus they could potentially be upregulating the machinery encoding such complexes and their activation during the day.

While during the day PB-PSB1 may be photosynthesizing and fixing carbon, it is possible that PB-PSB1 could be releasing this photosynthate at night via respiratory and fermentative processes. In other mats with green sulfur bacteria and cyanobacteria, photosynthate is made during the day and released at night by fermentation, producing mixed organic acids and possibly hydrogen (Burow et al. 2012, 2013). In similar mats also including SRB most of the carbon fixed during the day by cyanobacteria was found to be used up at night by SRB and *Chloroflexi* (Lee et al. 2014). Furthermore, this depletion of stored carbon could be linked to production of polyhydroxyalkanoates (PHAs), as the purple layers of microbial mats from the Great Sippewissett have been found accumulate PHAs at night, and deplete them during the day (Rothermich et al. 2000). This nighttime accumulation of PHAs is thought to be due to the breakdown of stored carbon. Similarly, studies of *Allochromatium vinosum* found that at night, glucose is respired, producing

polyhydroxyalkanoates and ATP, and reducing stored elemental sulfur to sulfide. During the day, the produced polyhydroxyalkanoate is converted to structural cell material (van Gemerden 1968a, 1968b, 1968c). PB-PSB1 stored carbon could thus be broken down at night, leading to polyhydroxyalkanoate and ATP production, and reduction of stored elemental sulfur to sulfide.

In terms of light entrained SRB activity, a study in an Israeli mat found higher sulfate reduction during the day, and activity in the sulfate reducing layer also seemed to be enhanced by ethanol which was thought to be fermentatively produced by cyanobacteria (Fründ and Cohen 1992). A study in a North Carolina mat measured SRB nitrogen fixation by monitoring nitrogenase activity via the acetylene reduction assay under molybdate inhibited and non-inhibited conditions (Steppe and Paerl 2002). Average nighttime nitrogenase activity potentially attributable to SRB went down at night by an average of 31.1%. While inhibitor studies can introduce artefacts, these findings suggested SRB-driven nitrogen fixation displayed diel fluctuations driven by the availability of reduced carbon compounds. Light entrained PB-SRB1 activity, therefore, may be intimately linked to PB-PSB1's photosynthetic activity making reduced carbon available to the consortium.

Potential for Nitrogen Fixation in PB-PSB1

PSB have been known to fix nitrogen for some time (Lindstrom et al. 1950). Pink berries, when exposed to a light-dark cycle have themselves been shown to incorporate $^{15}N_2$ via NanoSIMS (Wilbanks et al. 2010). From this data alone it is not possible to determine whether PB-PSB1 nitrogen fixation would be light or dark entrained, however, both remain a distinct possibility. In terms of the natural environment, experiments in a microbial mat in a Bahamian lagoon showed that anoxygenic phototrophs would fix nitrogen as they

photosynthesize, as photosynthetic energy from this process fueled both carbon and nitrogen fixation (Pinckney and Paerl 1997). Culture experiments, on the other hand, show cyanobacteria to be capable of both light and dark nitrogen fixation, where *Trichodesmium* fixes nitrogen with daylight (Capone et al. 1997), and *Crocosphaera watsonii* fixes nitrogen at nighttime (Dron et al. 2012). The sensitivity of nitrogenase to oxygen, however, would require some kind of temporal or spatial segregation to occur in order to allow nitrogen fixation concurrent with oxic conditions. Our incubations, however, were made anoxic and thus nitrogenase activity would not reflect influence of oxygen, but may respond to a light or temporal cue.

Schema for Diel Activity in PB-PSB1 and PB-SRB1

By first understanding the diel activity of PB-PSB1 we may subsequently understand the potential for resource sharing between PB-PSB1 and PB-SRB1. During the daytime PB-PSB1 may be engaging in anoxygenic photoautotrophy where it is oxidizing sulfide, and using ATP and reducing equivalents from this process to fix and store carbon. At night PB-PSB1 could then be heterotrophically oxidizing this stored carbon to produce ATP and polyhydroxyalkanoates, reducing stored sulfur to sulfide. Nitrogen fixation on the other hand, could be happening during either day or nighttime. PB-SRB1 may correspondingly react to photosynthate from PB-PSB1 and display some signs of importing and using carbon from PB-PSB1.

We characterized the diel transcriptomes of PB-PSB1 and PB-SRB1 and the corresponding transcriptional response to light-dark changes. We characterized the physiological response of PB-PSB1 denoting potential for daytime growth as well as

nighttime carbon and nitrogen transformations. We then examined whether PB-SRB1 could be reacting to PB-PSB1 photosynthate by consequently importing and oxidizing it.

Methods

Lab Diel Incubation

Berry samples were collected from site LS01 (41°34′33.01′N, 70°38′21.24′W) at 9AM on 22 July 2023 during mid-tide. Marsh water was collected separately and 0.02u filtered. A portion of this water was used to wash berries from marsh sediment, and the rest was used for incubations and sparged with nitrogen gas to render incubations anoxic. Twenty berries were incubated per serum vial (on same day of collection), and amended to a concentration of 21mM ^{12}C sodium bicarbonate, followed by 2.09mM neutralized sodium sulfide. Incubation set up started at 10AM, with berry sorting into vials done by 3:20PM, and the final anoxic incubation set up done at 7:30PM. Samples were then put on an artificial light-dark cycle under an incandescent light bulb [light:6AM-8PM; dark:8PM-6AM] with the first sample being collected at 7:42PM. Samples were destructively sampled in triplicate by freezing in liquid nitrogen at the following times: predusk (7:30PM), postdusk (8:30PM), peak dark (1AM), predawn (5:30AM), postdawn (6:30AM), and peak light (1PM) for three days; then sampled at peak light and peak dark for two additional days (Figure 1).

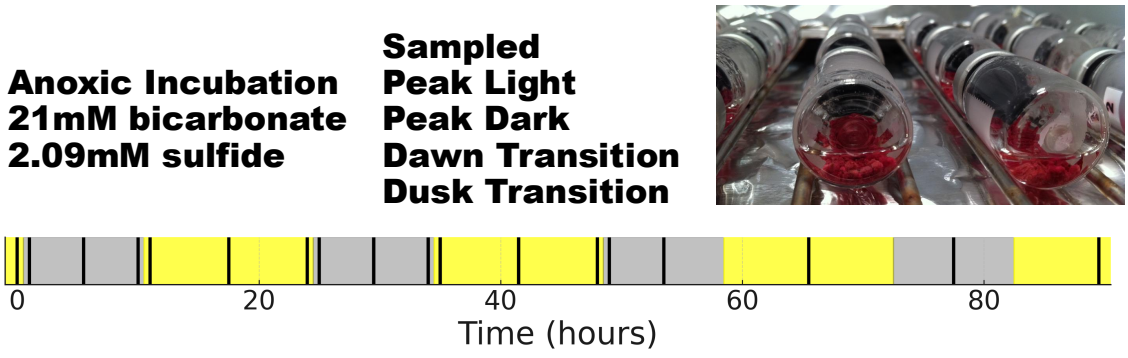


Figure 1. Sampling schematic for pink berries on a diel cycle over five days. Berries were samples at dusk (7:30PM and 8:30PM) and dawn (5:30AM and 6:30AM) transitions as well as peak dark (1AM) and peak light (1PM) times.

RNA Extraction and Sequencing

Berries were extracted with Zymo R2040 kit. Diel extractions totaled 22 time points in triplicate. Each sample consisted of 5 berries extracted together. Samples were treated with DNAse I in 2 hour increments (Zymo DNAse I set, E1011), varying in length from 2 to 6 hours total DNAse I incubation. For QC we quantified RNA with Qubit (at least 200 total ng of RNA, up to 1525 ng in one instance), verified lack of DNA contamination with PCR (if DNA present, sample underwent an additional 2 hour incubation with DNAse I), and RNA quality was checked with the Tape Station (RIN score of at least 6, often around 7-8). If RNA quantity fell below these limits, then an additional 5-berry extraction was pooled to increase concentration.

Note: While Tape Station can be typically used to check for large DNA remaining in sample, it is still possible for DNA to be "hiding" in RNA peak. The best way to check for contamination is to use PCR to amplify contaminating DNA. If there are no bands then sample is not DNA contaminated, and this can be sent for sequencing.

Samples were sequenced by the Joint Genome Institute under grant 508543. This was sequenced 2x151 using the Illumina NovaSeqX platform, and BBDuk was used to remove various contaminants, remove adapter reads, and right quality trim reads. Reads that mapped via BBMap to human, cat, dog and mouse references as well as common microbial contaminants were removed; ribosomal reads were also removed (BBtools v39.03, Bushnell 2014). The final filtered fastq was used for metatranscriptome analysis.

RNA Differential Expression Analysis

The repeatless 41-bin database in the genes.fna style (out.db.rm.mxdsp from Chapter 1), was used as an index for pseudoalignment of sequenced reads with kallisto for counting

(kallisto v0.48.0, Bray et al. 2016). Counts were then separated by bin, filtered for genes whose total count across samples was 10 or higher, and differential expression analysis was run with DESeq2, followed by shrinkage analysis via "apeglm" (DESeq2 v1.46.0, Love et al. 2014). The p-adjusted, or q-value, below 0.05 was then used to determine significance. MA plots were generated with plotMA from DESeq2. Bar, volcano, and PCA plots were generated with ggplot2 (q-value less than 0.05 and $\log_2$ (fold change) greater than |1| was used for bar and volcano plots; ggplot2 v3.5.2, Wickham 2009). PCoA plots were also generated via ggplot2, and ordinations were based on Bray-Curtis dissimilarity. Heatmaps filtered by significance were generated via pheatmap (50 most significant genes and another based on the 50 most expressed of the significant genes; pheatmap v1.0.13, Kolde, n.d.), and ClusterGVis was used to generate clustered line maps and heatmaps (clustered by TCSeq; v0.1.4, Zhang et al. 2026). plotly was used for interactive volcano plots and heatmaps (plotly v4.11.0, Sievert, n.d.).

ComplexUpset was used in order to generate upset plots for exploration of gene membership across differing DESeq2 significance comparisons (ComplexUpset v.1.3.3, Lex et al. 2014). To make "specialized" interactive heatmaps, genes that were significant across different conditions of interest were compiled into one list then clustered via ClusterGVis. This matrix was then exported to plotly and used to explore diel patterns across genes of interest.

Functional COG/KEGG Enrichment in PB-PSB1 and PB-SRB1 Clusters

We determined patterns of expression between DESeq2-generated significantly expressed genes from our multiple light-dark comparisons and clustered relative to each other via

ClusterGVis. We wanted to see how these light-dark patterns compared to the rest of the genome of origin so we carried out a cluster-wise functional enrichment.

To generate COG and KEGG annotations the PGAP-annotated genbanks used for our panaroo-generated pangenomes in chapter 1 for PB-PSB1 and PB-SRB1 were imported into anvi'o (anvi'o v8, Eren et al. 2020). These were then annotated via anvi-run-ncbi-cogs and anvi-run-kegg-kofams. These functions were then exported back out of anvi'o and merged with the panaroo pangenome ID's generated in chapter 1 for our repeatless count database. This allowed us to match COG and KEGG annotations back to the genes composing our ClusterGVis determined clusters of DESeq2 significant genes.

For our functional analysis we analyzed by four different annotation types (COG Category, COG function, KOfam, or KEGG class). We defined the "foreground" as the number of genes for a function annotated within a specific annotation type, i.e. COG category, in the current cluster of interest, and the "background" was the number of genes for that same function annotated outside of that cluster (Figure 2).

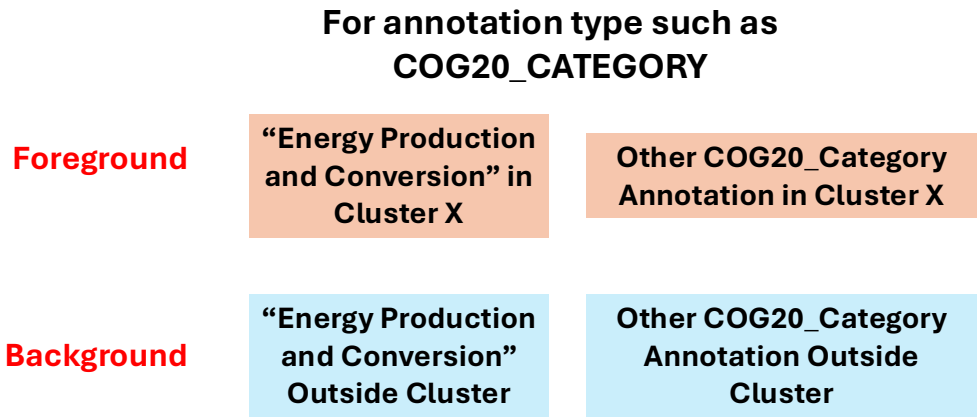


Figure 2. Overview of contingency table for fisher's exact test for cluster functional enrichment. A specific annotation is compared for membership within a cluster relative to everything else annotated in that cluster (foreground). That same specific annotation is compared for membership outside of the cluster of interest relative to everything else annotated outside the cluster (background).

This proportion of foreground to background was then used to construct contingency tables to perform Fisher's exact test, then error corrected with a false discovery rate threshold of 0.05 to determine significantly enriched genes in either the current cluster of interest in either PB-PSB1 or PB-SRB1. The total genes annotated for a specific category may not add up to total gene instances with an annotation, as some genes may have cross-category annotations.

No genes were determined to be significantly functionally enriched after multiple testing correction. This was likely because the size of cluster membership was so small relative to the rest of the genome, meaning that our multiple testing correction led to nothing resulting as significant. For exploration we thus plotted points that were nominally significant, meaning they just had an unadjusted p-value less than 0.05.

Results and Discussion

PB-PSB1 displayed rapid transcriptional response to light-dark transitions while PB-SRB1's lagging change with light may be a response to photosynthetic exudate

To determine light-dark changes in pink berry physiology potentially linked to photosynthetic exudate we first explored diel patterns in physiology of the main pink berry consortium members: PB-PSB1 and PB-SRB1. To start this, we explored the metatranscriptome space across our 65 lab diel incubation samples, which we found to be dominated by transcripts from PB-PSB1, reaching transcripts per million counts at 10^6, which was in turn 2 orders of magnitude higher than PB-SRB1 at 10^4 (Supplemental Figure A). The next two most dominant bins were UBA10427 (*Cyclobacteriaceae*) and UBA12170 (*Bacteroidales*). For comparison, our metagenomic coverage from the 2021 PacBio HiFi sequencing was at a much different abundance where PB-PSB1 was at 300-500x, PB-SRB1 at 100-150x, UBA10427 at 60-90x, and UBA12170 at 10x.

Our sampling of pink berries allowed us to observe potential patterns of PB-PSB1 expression at light-dark transitions as we sampled berries before and after dawn/dusk. We could further observe for a sustained light/dark response within a 24-hour cycle as we also sampled at peak light and peak dark conditions. For PB-PSB1, we found that 50.19% (ie 3344 genes significant out of 6662 genes analyzed [count 10 or above across samples]) of analyzed genes were significantly different between all light and all dark samples (Figure 3). PB-PSB1 transcripts, furthermore, demonstrated rapid light-dark differentiation, wherein they consistently occupied a significantly different space between the light-dark transitions at dawn and at dusk (Figure 4), that was sustained through peak lighting conditions. This is relatively fast, as sampling for light-dark transitions took place in as little as an hour.

Moreover, as we continued this sampling scheme over three days, with two additional days of "peak" sampling, we could observe whether this light-dark transition was sustained over time. While we see a consistent light-dark shift with time, we also observed both light and dark points migrating jointly, potentially as a response to some abiotic or biotic incubation change. To illustrate the change produced in transcripts throughout this incubation, our samples were determined to be significantly different between the first half and the latter half of the experiment (Supplemental Figure B).

PB-PSB1 seems to display a sweeping response to light exposure as about 50% of analyzed genes were significantly different due to light-dark differences. As these changes occur quite reproducibly within 30 minutes of a light change, this may suggest some direct sensing of light. This does not, however, exclude the possibility of some entrained rhythm causing this light change, as the light change was set to coincide with local sunup and sundown timing.

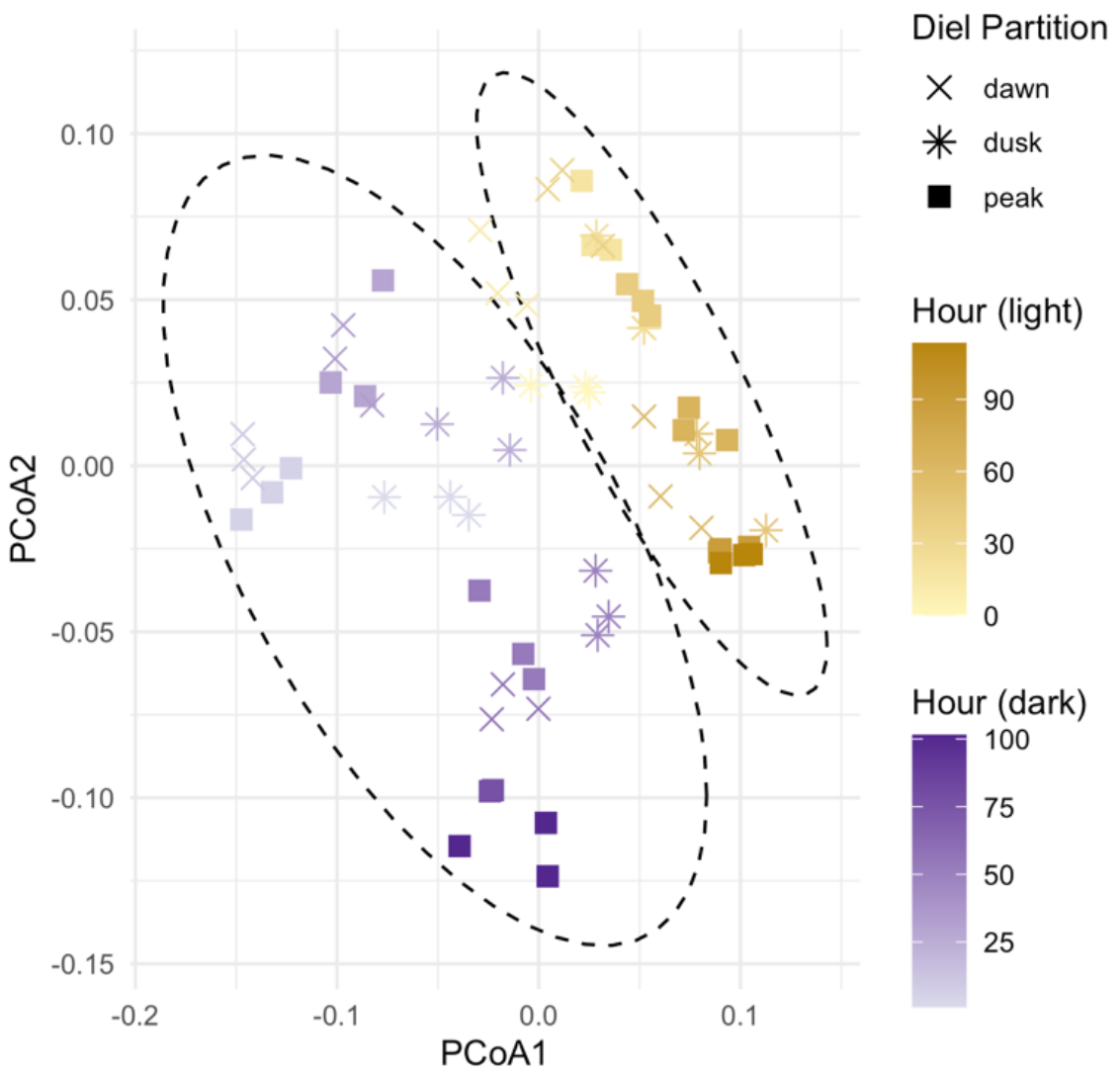


Figure 3. Differences in PB-PSB1 transcripts between light-dark conditions over time. PCoA for PB-PSB1 using Bray-Curtis distance where 95% confidence ellipses are based on incubation light v. dark conditions, and corresponding light-dark condition PERMANOVA results are F=20.704, R^2=0.24735, p= 0.001. Gold tones correspond to light time points, and purple tones correspond to dark time points. Time points taken before and after dawn are denoted by "X", time points taken before and after dusk are denoted by "*", and time points taken at either peak light or peak dark are indicated by squares.

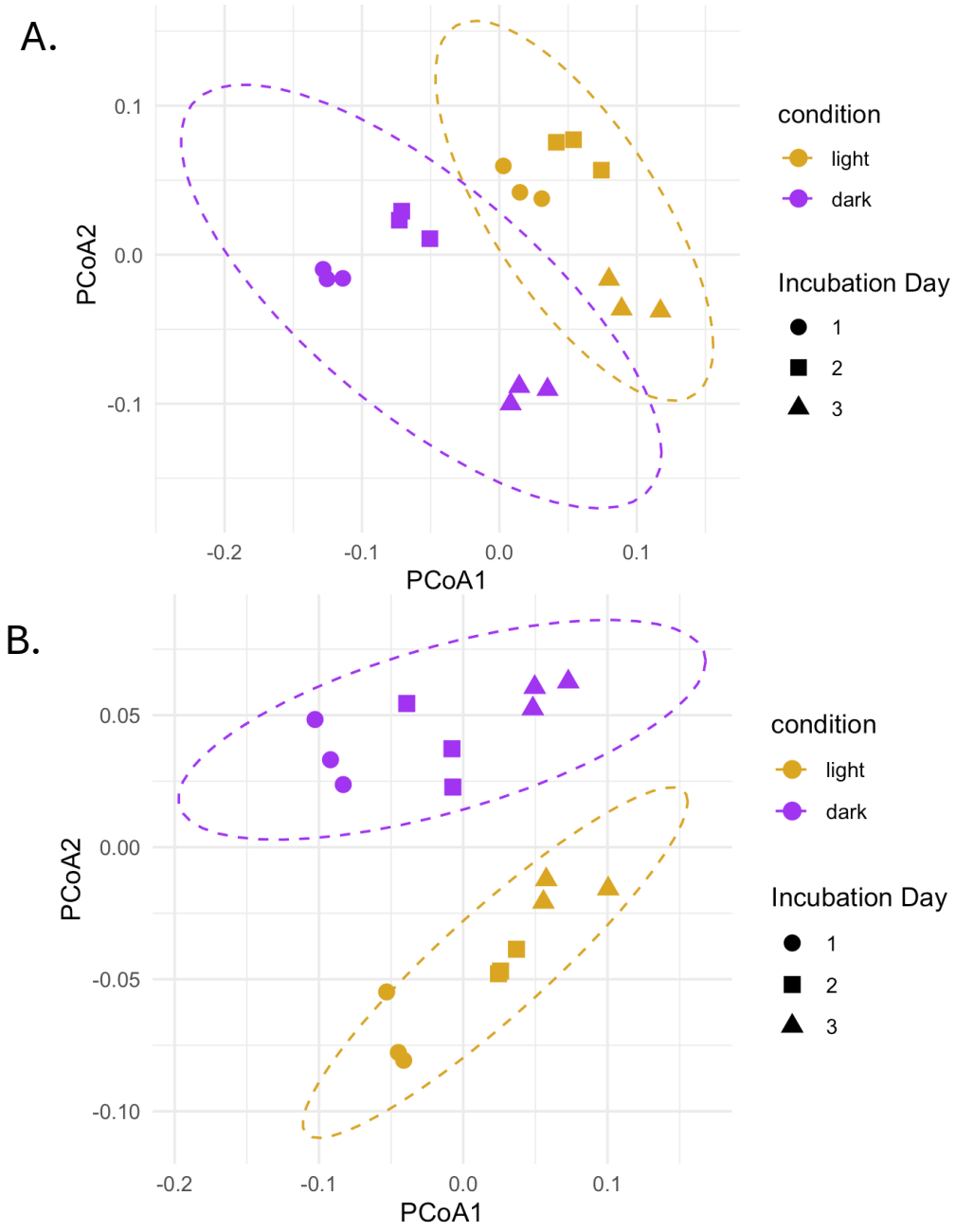


Figure 4. Differences in PB-PSB1 transcripts between light-dark transitions at **(A)** dawn and **(B)** dusk shown as a PCoA ordination using Bray-Curtis distance. Ellipses represent 95% confidence, based on incubation light v. dark conditions. Light-dark condition PERMANOVA results are F=7.4578, R^2=0.31792, p= 0.001 **(panel A)**, and F=4.4518, R^2=0.21767, p= 0.002 **(panel B).** Gold tones correspond to light time points, and purple tones correspond to dark time points.

PB-SRB1's light-dark differentiation between samples, on the other hand, was more noticeable with the experiment's progression (Figure 5). For the first half of our incubation,

we see no significant light-dark differentiation in PB-SRB1's transcripts (Figure 6A), however, we do see a significant light-dark differentiation within the second half of our experiment (Figure 6B). This difference could potentially be a reaction of PB-SRB1 to either something being exuded by other community members like PB-PSB1 with incubation progression, or a reaction to abiotic experimental incubation conditions. Additionally, there may be a greater response to light within a day, as PB-SRB1 transcripts are significantly differently between light-dark conditions at peak sampling points (noon/1AM), and could be a reaction to a within day exudate from another community member (Supplemental Figure C). For PB-SRB1, 6.57% (ie 249 genes significant out of 3789 genes analyzed [count 10 or above across samples]) of analyzed genes were significantly different between all light and all dark samples. This transcriptional response to light exposure in PB-SRB1 was much more modest than PB-PSB1's, potentially indicating a more targeted response in expression. While PB-SRB1 could be directly responding to light, it did not display differences as immediately with light like PB-PSB1, and instead displayed significant differences in expression at peak light/dark sampling. This apparent lag in response to light exposure, both within a day and with incubation progression, as well as the magnitude of PB-SRB1's light exposure response led us to believe PB-SRB1 may be sensing some specific shift in photosynthetic exudates produced via another community member's detection of light, namely PB-PSB1.

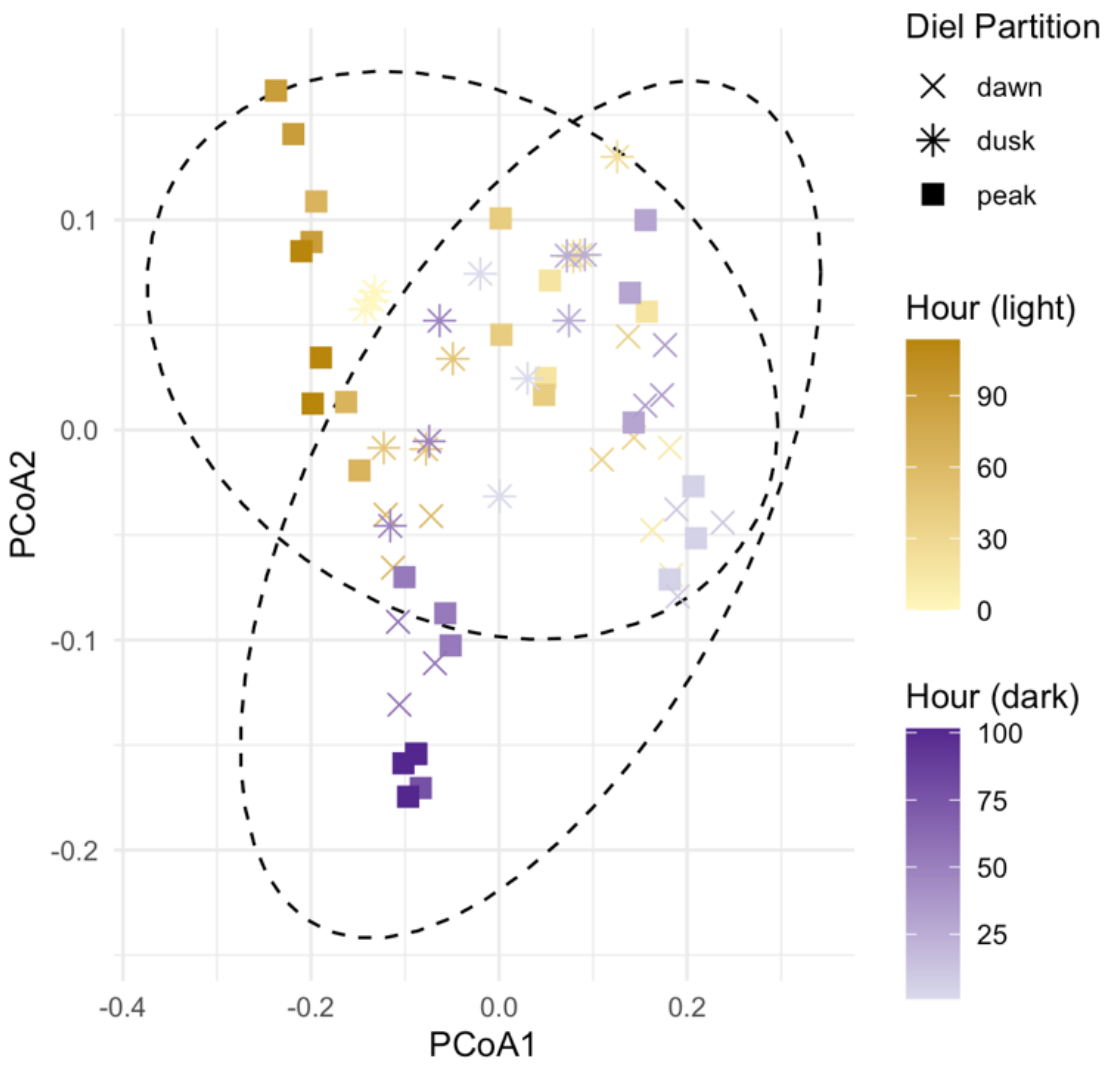


Figure 5. Differences in PB-SRB1 transcripts between light-dark conditions over time. PCoA for PB-SRB1 using Bray-Curtis distance where 95% confidence ellipses are based on incubation light v. dark conditions, and corresponding light-dark condition PERMANOVA results are F=3.7814, R^2=0.05662, p= 0.003. Gold tones correspond to light time points, and purple tones correspond to dark time points. Time points taken before and after dawn are denoted by "X", time points taken before and after dusk are denoted by "*", and time points taken at either peak light or peak dark are indicated by squares.

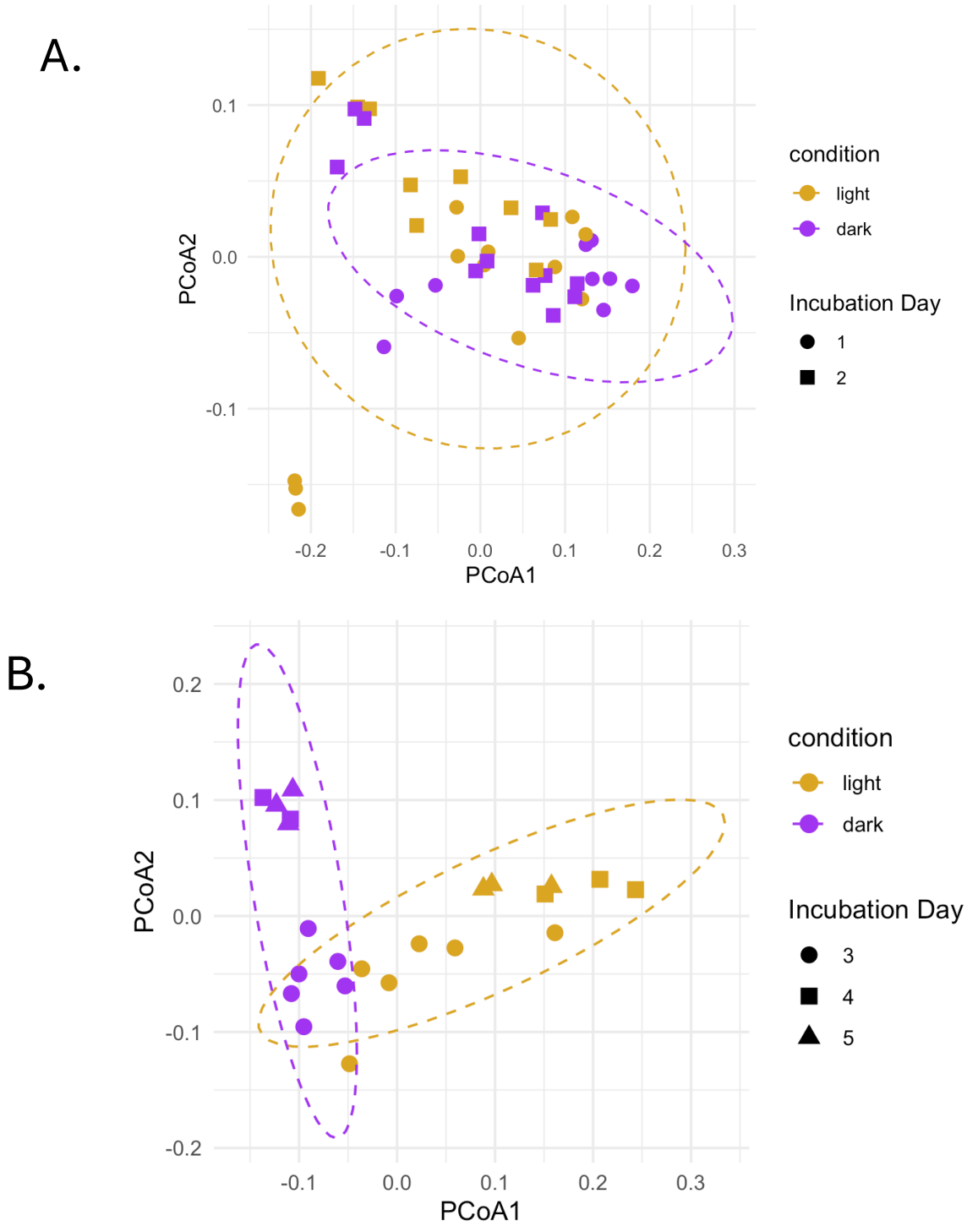


Figure 6. The *Desulfofustis* PB-SRB1 genome-wide transcriptional response to light was largely overlapping during early time points **(A)**, but exhibited pronounced differences later in the incubation **(B).** Shown in the panels are a PCoA ordination based on Bray-Curtis distance between PB-SRB1 gene expression levels on early incubation (days 1-2, **panel** A), and late incubation days (days 3-5, **panel B).** 95% confidence ellipses are based on incubation light v. dark conditions. Corresponding light-dark condition PERMANOVA results are F=1.6865, R^2=0.04046, p= 0.071 **(panel A)**, and F=6.9038, R^2=0.24742, p= 0.001 **(panel B)**.

Diel Cycling by PB-PSB1 and PB-SRB1

Due to the sustained, sharp light-dark transitions we saw at dawn and dusk in PB-PSB1, we wanted to further explore the functional content of genes responding to changes in light-dark conditions. To do this, significantly different genes between the various light-dark conditions that we compared were clustered ($p_{adj} < 0.05$, and $\log_2$ (fold change) >1), leading to distinct groupings. PB-PSB1 genes with significant differences between light and dark conditions formed four clusters with distinct diel patterns of gene expression (Figure 7). One of our clusters displayed sharp light-dark transitions, such that its genes were upregulated in the daytime and downregulated in the nighttime (PB-PSB1 cluster 4). Another cluster displayed the exact opposite pattern, such that while it also displayed crisp light-dark transitions, its genes were downregulated in the daytime, but upregulated at night (PB-PSB1 cluster 3). Furthermore, a third cluster displayed a similar pattern of daytime downregulation and nighttime upregulation, but its transitions were more delayed and less crisp as genes at these intermediate time points displayed more intermediate values of expression (PB-PSB1 cluster 2). Our final cluster showed a much less coherent pattern that eventually seems to develop into nighttime upregulation and daytime downregulation, but was somehow interrupted towards the beginning of our incubation (PB-PSB1 cluster 1).

Furthermore, While PB-SRB1's light-dark transitions were less evident than those of PB-PSB1, we still saw evidence of light-dark differentiation in its transcripts. We thus also aimed to explore potential patterns of diel cycling in PB-SRB1. After similar clustering to PB-PSB1 where we compared different light-dark conditions and pulled genes where (p_{adj}< 0.05, and $\log_2$ (fold change) > 1), PB-SRB1 only displayed 2 clusters (Figure 8). Overall, the light-dark transitions for PB-SRB1's transcripts were much less sharp, and lagged behind the

light-dark transition time. Of these clusters, one had transcripts that were roughly upregulated at night and downregulated during the daytime (PB-SRB1 cluster 2). Our other cluster was downregulated at night and upregulated during the day (PB-SRB1 cluster 1).

We conducted a functional enrichment analysis of genes enriched within each of our clusters. All the enrichments were nominal (unadjusted p-value < 0.05), however, as after multiple testing correction all p-values were not significant. This exploration, however, allowed us to glean functions that were particular to either daytime or nighttime conditions.

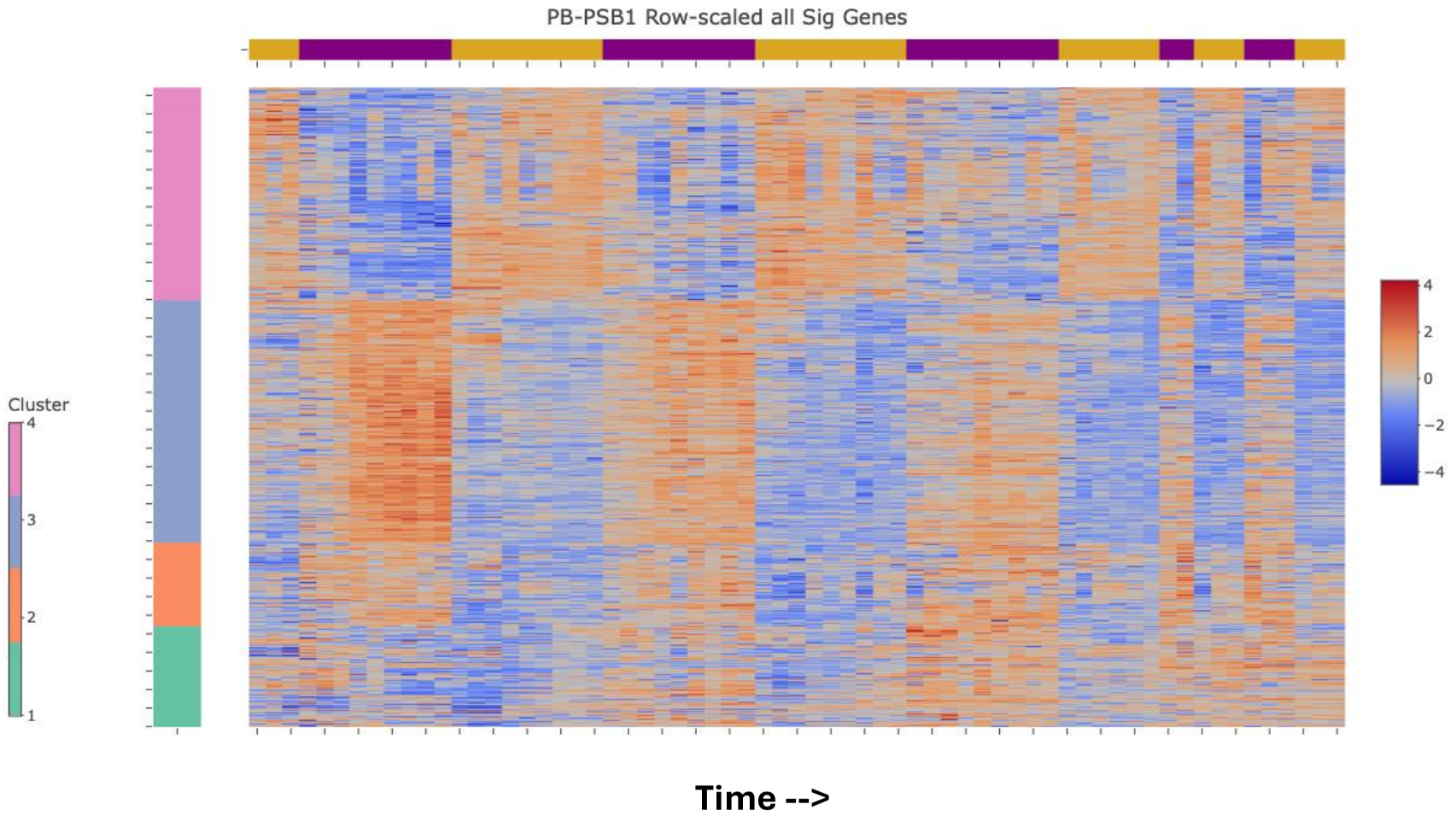


Figure 7. PB-PSB1 genes differentially expressed between all light and dark condition comparisons that were performed (Supplemental Table 1), shrinkage adjusted and row-scaled ($p_{adj} < 0.05$, $\log_2$ fold-change > 1). Clustered with ClusterGVis and visualized with plot_ly. From left to right we see experiment progression with time, gold bars correspond to light time points and purple bars to dark time points. For first three days we sampled light/dark transitions, as well as peak light and peak dark conditions. Last two days were just peak time points.

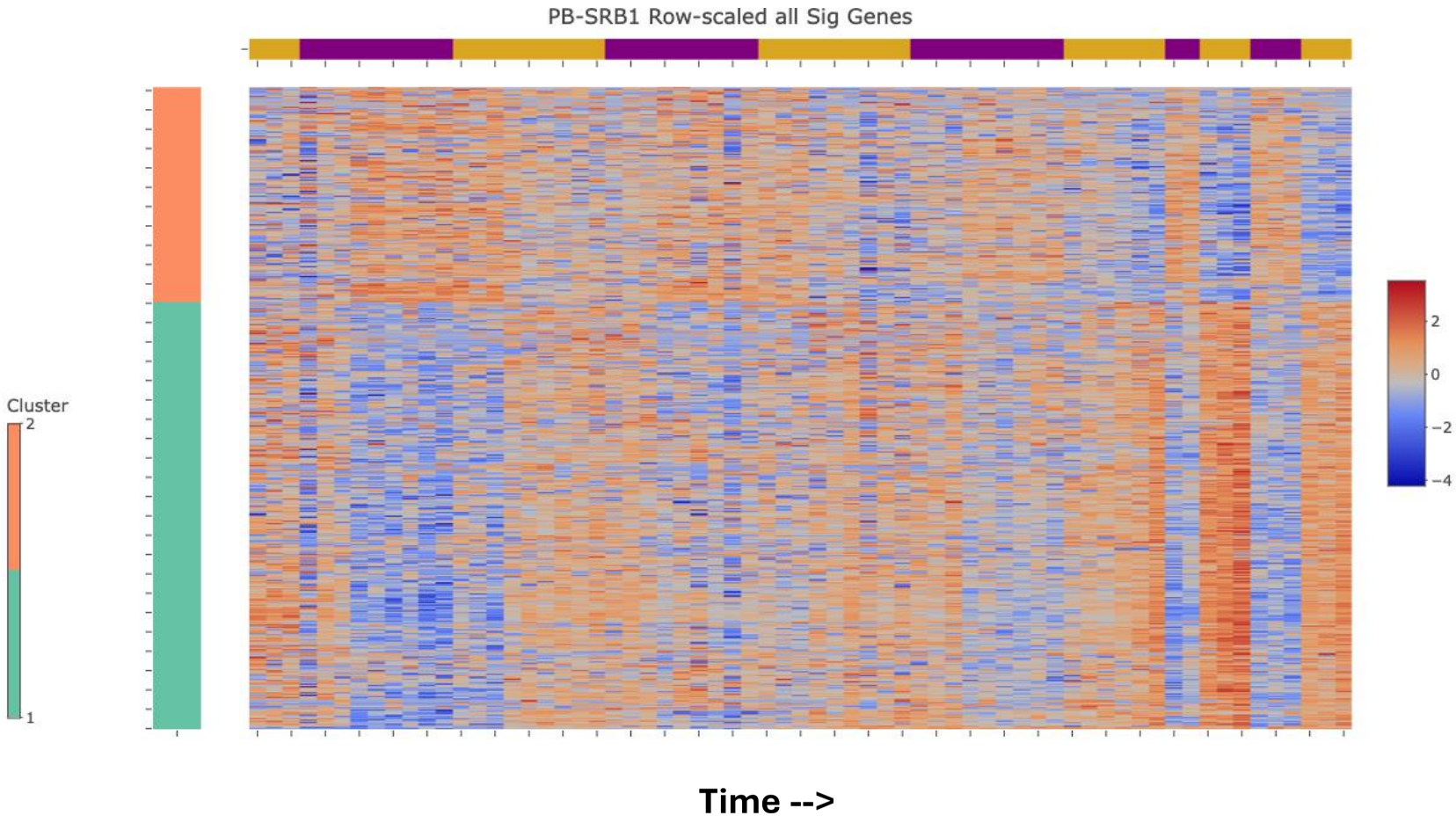


Figure 8. PB-SRB1 genes differentially expressed between all light and dark condition comparisons that were performed (Supplemental Table 2), shrinkage adjusted and row-scaled ($p_{adj} < 0.05$, $\log_2$ fold-change > 1). Clustered with ClusterGVis and visualized with plot_ly. From left to right we see experiment progression with time, gold bars correspond to light time points and purple bars to dark time points. For first three days we sampled light/dark transitions, as well as peak light and peak dark conditions. Last two days were just peak time points.

Daytime Upregulation of genes in PB-PSB1 may indicate an increase in growth rate and anabolic activity

When we explored genes that were upregulated in our daytime cluster we found evidence suggesting increased growth rate and anabolic during the day. We found genes, for example, involved in translation such as various ribosomal proteins and their subunits (PSB__G5092), maturation factors and ribosome binding ATPases (PSB__G1569). Correspondingly we also found tRNA associated genes that enable transfer of various moieties (PSB__G1703), including glycosyl and methyl in tRNA (PSB__G1543 and PSB__G2865). We found genes involved in cell division such as those enabling DNA replication like the replication initiator and DNA polymerase (PSB__G952 and PSB__G1527). Correspondingly, the cell would need to engage in membrane and peptidoglycan synthesis. Accordingly, we found a SPOR domain containing protein which would be involved in septal formation (PSB__G6786) alongside glycosyltransferases (PSB__G6706) and lipopolysaccharide transporters (PSB__G6748). In order to support this translation and eventual cell division various enzymes would need the support of appropriate vitamins and cofactors. We correspondingly found daytime upregulation of genes involved in biosynthetic pathways for thiamin (PSB__G1552), riboflavin (PSB__G6836), heme (PSB__G912), biotin (PSB__G1061), and vitamin B12 (PSB__G1653) (Figure 9).

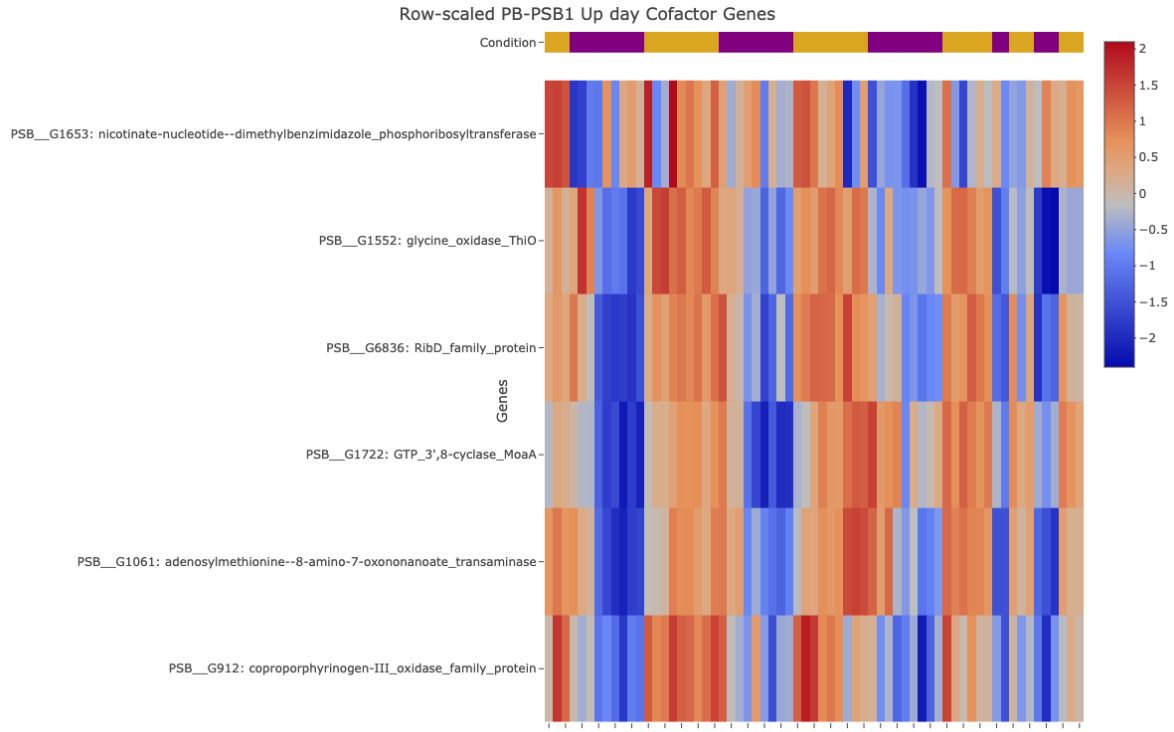


Figure 9. Various genes associated with cofactor synthesis in PB-PSB1 that were significantly different in some form of light and dark comparison. Shrinkage adjusted, row-scaled, and visualized with plot_ly. From left to right we see experiment progression with time, gold bars correspond to light time points and purple bars to dark time points.

Furthermore, while we did not see daytime upregulation of genes involved in carbon fixation, we did see daytime upregulation of genes involved in carbon catabolism. Genes indicating carbon flow through glycolysis (6-phosphofructokinase, PSB__G1114), for example, may suggest routing of fixed carbon resulting in production of glycolytic intermediates that can be geared towards anabolism. For example, generation of glycolysis intermediates such as glucose 6-phosphate, can result in production of ribose 5-phosphate via the pentose phosphate pathway, thus resulting in a precursor of nucleotide synthesis. Furthermore, the glycolytic end product, pyruvate, could be routed via the TCA cycle as many of its intermediates also serve as precursors for anabolic production, such as alpha-ketoglutarate and its conversion to glutamate and glutamine or proline. A gene we did find differentially expressed in our cluster, PHB depolymerase (PSB__G428), would indicate

consumption of the cell's stored carbohydrates during the day, presenting another source of carbon that is able to flow into anabolic production. Furthermore, we also found upregulation of genes in this cluster that could directly suggest this carbon rerouting, like acetyl-coA carboxylase (PSB__G2198), which shifts carbon towards fatty acid biosynthesis, thus providing a direct link between central metabolism and anabolism.

We further saw an enrichment for both glycine oxidase and glycolate dehydrogenase (PSB__G1552 and PSB__G1545), which both form glyoxylate and metabolize photorespiration byproducts. If glycolate is forming due to photorespiration, then its subsequent transformation to glyoxylate offers the cell another entry point for carbon into central metabolism. If it is operated on enzymes like those in the glyoxylate shunt, then the cell can generate other central intermediates that could be further shunted into anabolic pathways.

Together the flow of carbon into central metabolic pathways like glycolysis and the TCA cycle paints the picture of a PB-PSB1 cell that may be harnessing photosynthetic energy in order to grow and divide. The cell would actively be routing carbon towards anabolic production, and synthetizing cofactors and vitamins during the day.

Nighttime upregulation of genes in PB-PSB1 may indicate generation of new fixed nitrogen as well as new bacteriochlorophyll

We then explored genes that showed nighttime upregulation. We saw evidence for nighttime ammonium generation as well as nitrogenase activation. We found actual ammonium transporters (PSB__G1322) as well as genes that would result in ammonium generation like guanine deaminase (PSB__G1058). We also saw multiple transglutaminases, such that peptide bonds would be forming between amino acids (PSB__G2228,

PSB__G3068). Interestingly, the upregulation of NADP-dependent isocitrate dehydrogenase during nighttime (PSB__G1208) catalyzes the formation of alpha-ketoglutarate. Alpha-ketoglutarate is the skeleton upon which glutamine and glutamate synthase operate in order to incorporate ammonium into metabolic pathways. Apart from these genes involved in ammonium generation and nitrogen mobilization we additionally observed nighttime upregulation of some nitrogenase and nitrogen transport genes. Of these genes, *nifA* has been shown to be a *nif* activator (Buchanan-Wollaston et al. 1981)(PSB__G731), and *nifX* has been shown to help with the biosynthesis of the FeMo cofactor for nitrogenase (Shah et al. 1999) (PSB__G4132).

Furthermore, we found that at the start of the dark period, or right at dusk, upregulation of *ptsN* (PSB__G3581), would precede many of the above-mentioned genes (Figure 10). *ptsN* is a component of the PTS^{Ntr} phosphotransferase system along with *manX*, that together respond to nitrogen and carbon limitation as has been studied in *Rhizobium leguminosarum* bv. *viciae* 3841 (Sánchez-Cañizares et al. 2020). Together they work in a way such that when carbon is limiting (and nitrogen is rich), ManX stimulates the TCA cycle and carbon oxidation while PtsN stimulates potassium uptake. Under nitrogen-limiting conditions (and carbon is rich), PtsN stimulates amino acid uptake and nitrogen metabolism, while the TCA cycle and carbon oxidation are decreased and production of polymers for carbon storage is increased. The upregulation of *ptsN* is not well understood, but it is part of the *rpoN* operon and it has been shown that sigma e, an indicator of cell envelope stress, promotes transcription of all proteins encoded by the *rpoN* operon (Sikora et al. 2024). *rpoN* itself encodes for sigma-54 and is involved in transcribing genes for assimilation of ammonia and glutamate under N limitation in Salmonella strain SK416 and SK419 (Hirschman et al.

1985). Within PB-PSB1's genome, furthermore, we also see these two genes located together in one operon.

This potential nighttime response of increased nitrogen fixation could be a response to a skewed carbon: nitrogen ratio, as cells have a carbon: nitrogen ratio at which they can optimally operate (Vrede et al. 2002). PB-PSB1 has been shown to fix carbon during the daytime (Wilbanks et al. 2014), leading to a bigger C:N ratio that would be decreased with nighttime nitrogen fixation in order to restore intracellular nutrient balance. This initial uptick we see of *ptsN* expression, therefore, could potentially be a response to stress induced by nitrogen limitation and may be related to the subsequent upregulation we see of *nifA* and other nitrogen processing genes.

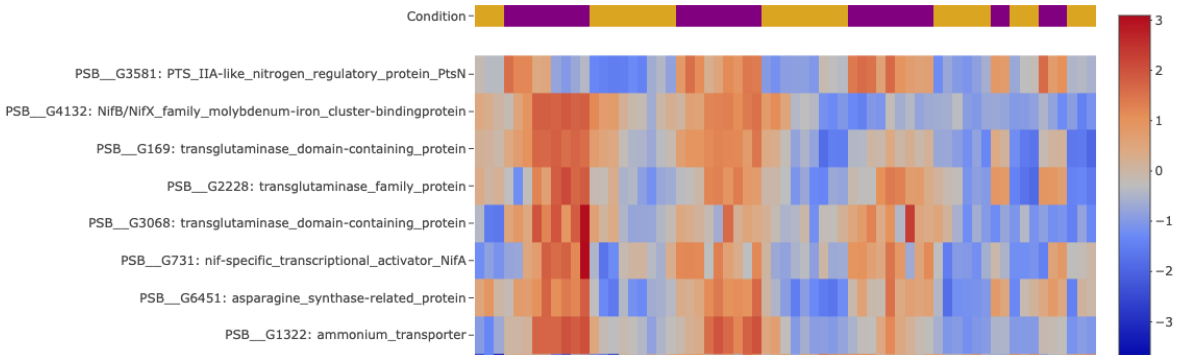


Figure 10. Nitrogenase and nitrogen transport genes in PB-PSB1 that were significantly different in some form of light and dark comparison. Shrinkage adjusted, row-scaled, and visualized with plot_ly. From left to right we see experiment progression with time, gold bars correspond to light time points and purple bars to dark time points.

In this night cluster we further saw genes involved in bacteriochlorophyll synthesis and transport. Moreover, we observed genes like geranylgeranyl diphosphate reductase (PSB__G5163) and phytoene desaturase (PSB__G1093) which are involved in bacteriochlorophyll and carotenoid synthesis. We also saw genes encoding bacteriochlorophyll delivery family transporters (PSB__G961). Cultures of PSB like

Thiocapsa roseopersicina, have been grown under a light-oxic/dark-anoxic regime and been shown to generate bacteriochlorophyll at night due to oxygen sensitivity (De Wit and Van Gemerden 1990). While our incubations were anoxic there may still be some sources of oxygen. Cyanobacteria in our consortia may have generate small daytime pockets of oxygen, or there may be some sort of circadian response programmed to regenerate bacteriochlorophyll at night.

PB-PSB1 may be engaging in nighttime breakdown of glycogen and using consequent electrons to respire stored elemental sulfur

Further inspection of nighttime upregulated genes indicated potential signs of glycogen breakdown and movement. We further saw, for example, an upregulation of glycogen phosphorylase (PSB__G5402) and various glycosyltransferase genes at night (PSB__G1589, PSB__G6749, and PSB__G807) (Figure 11). This may be an indication of carbon breakdown, specifically daytime photosynthate. The PB-PSB1 called may be oxidizing stored carbon in order to produce ATP.

We interestingly saw nighttime upregulation of all three subunits of complex I of the electron transport chain *nuoABC*, (PSB__G4088, PSB__G3553, PSB__G2638), allowing transfer of electrons from NADH to the quinone pool. Furthermore, we also saw nighttime upregulation of *napH*, encoding for the ferredoxin subunit of quinol dehydrogenase (PSB__G2055). While this is typically associated with transferring electrons to NapAB nitrate reductase in organisms like *E. coli* (Brondijk et al. 2002), this could be a reasonable means for our organism to transfer subsequently electrons to a protein in the periplasmic space after transfer of electrons to the quinone pool by *nuoABC*. The cell could then use these electrons from stored carbon for respiratory reduction of stored elemental sulfur.

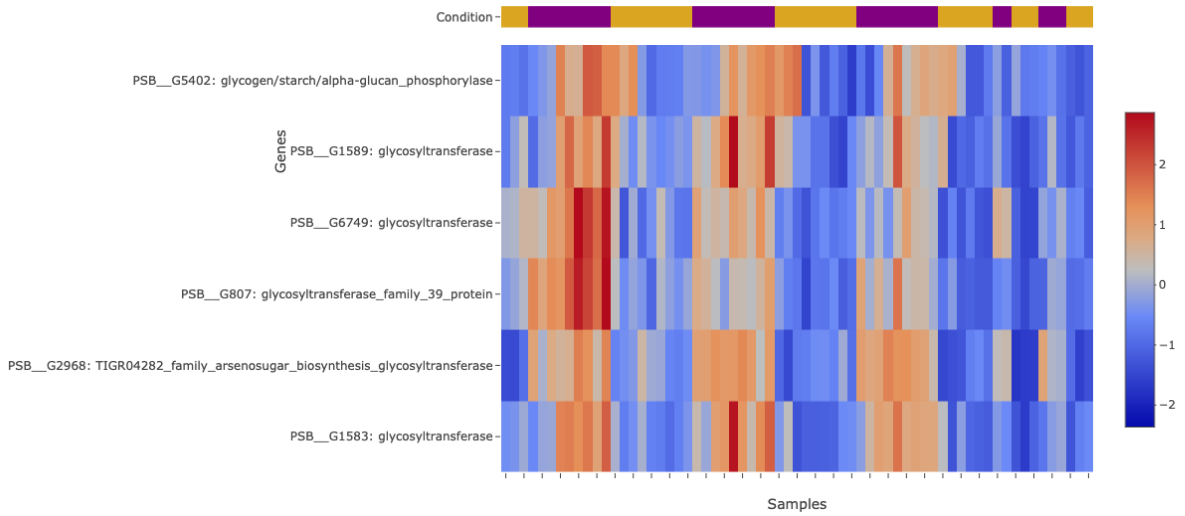


Figure 11. Carbon breakdown and movement genes in PB-PSB1 that were significantly different in some form of light and dark comparison. Shrinkage adjusted, row-scaled, and visualized with plot_ly. From left to right we see experiment progression with time, gold bars correspond to light time points and purple bars to dark time points.

Daytime upregulation of genes in PB-SRB1 may indicate increased daytime sulfate reduction stimulated by PB-PSB1 photosynthate

We explored daytime upregulation of genes by PB-SRB1. We observed daytime upregulation of various cytochrome oxidases (SRB__G4333, SRB__G2281, SRB_G405) as well as peroxiredoxin (SRB__G2293). Studies across sulfate reducers like *Desulfovibrio* have demonstrated the employment of cytochrome *c* oxidase and cytochrome *bd* oxygen reductase in oxygen defense (Dolla et al. 2006). Thus, these genes in PB-SRB1 may all indicate oxygen stress, potentially from micro-pockets of oxygen generated from cyanobacteria in the community that could be engaging in photosynthesis.

Furthermore, we found PB-SRB1 to express genes for respiratory complexes like various subunits of NADH-quinone oxidoreductase (SRB__G1095, SRB__G1284, SRB__G927) and ATPase (SRB__G497). We also saw upregulation of expression for other electron transport

complexes like RnfA (SRB__G1449 and SRB__G2345) and RsxE (SRB__G2221). Together these electron mobilizing complexes may suggest increased electron transport during the day, suggesting stimulation of catabolic activity in PB-SRB1.

We earlier saw in PB-PSB1 daytime upregulation of glycine oxidase and glycolate dehydrogenase (PSB__G1552 and PSB__G1545), which both produce glyoxylate from glycine and glycolate, respectively. These are products of photorespiration and may suggest either daytime photoassimilation or consumption of endogenously produced glycolate. The production and potential availability of this compound in the consortium is of interest as PB-SRB1's closest cultured relative, *Desulfofustis glycolicus*, was isolated via growth on glycolate (Friedrich et al. 1996). We thus further pursued the potential use of PB-PSB1 produced photosynthate by PB-SRB1.

***Upregulation of* dsrD**

Across all our PB-SRB1 samples, we saw an upregulation of *dsrD* within lighted conditions (SRB__G3326). *dsrD* has been shown to be an allosteric activator of *dsrAB*, which codes for dissimilatory sulfite reductase (*dsrAB*) (Ferreira et al. 2022), the final step in sulfate reduction where sulfite is reduced to sulfide (Figure 12). Since this was upregulated during the daytime we decided to investigate the possibility that this could be the result of PB-SRB1 being stimulated by photosynthate from PB-PSB1. This would need photosynthate to be imported via PB-SRB1 transporters followed by oxidation of the photosynthate via dehydrogenases. Sulfate reducing bacteria are capable of using a wide variety of compounds as electron donors, such as short and long-chain fatty acids as well as aromatic compounds (Muyzer and Stams 2008). Complementing this daytime of oxidation of photosynthate, we also nighttime upregulation of a glycogen debranching protein which would indicate usage of

stored carbon once the cell has potentially oxidized photosynthate during the day (SRB__G205). Potential greater daytime sulfate reduction as well as use of respiratory complexes made us posit stimulation from PB-PSB1 photosynthate. Our focus thus shifted to mining for daytime upregulation of corresponding dehydrogenases and transporters in PB-SRB1's genome, as this could potentially indicate import of photosynthate from PB-PSB1 and subsequent oxidation of photosynthate.

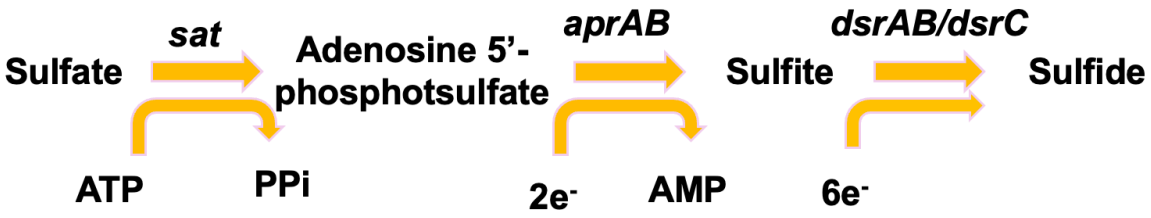


Figure 12. Sulfate reduction pathway, including relevant genes and intermediates

More dehydrogenases and transporters are significantly differentially expressed during the day than at night by PB-SRB1

We found more dehydrogenases upregulated by PB-SRB1 during the day than at night (only 3 at night, 14 during the day) (Figure 13 and Supplemental Figure D). Furthermore, we found four times as many transporters upregulated by PB-SRB1 during the day than at night (10 at night, 37 during the day) (Figure 14 and Supplemental Figure E). Within these "nighttime" transporters, few showed any distinct nighttime transport pattern. This seems to point to the transport of a wider variety of substrates during daytime conditions. This daytime preference may be pointing use of photosynthate. Specifically, we found dehydrogenases for succinate, shikimate, and 2-hydroxyacid to be upregulated during the day and when we looked for potential corresponding transporters in the daytime we saw upregulation of major

facilitator superfamily (MFS) and tripartite ATP-independent periplasmic (TRAP) transporter proteins (Figure 14). MFS transporters are a wide-ranging super family, and could potentially encompass metabolites generated by PB-PSB1 as they are transporters of amino acid derivatives, sugars, and pyruvate derivatives (Reddy et al. 2012). TRAP transporters, furthermore are known for transport of dicarboxylic acids (Janausch et al. 2002). It is thus possible for succinate, shikimate, or some 2-hydroxyacid to be imported into PB-SRB1 cells during daytime photosynthetic conditions. Interestingly, of the photorespiration byproducts we mentioned earlier glyoxylate is a 2-oxoacid.

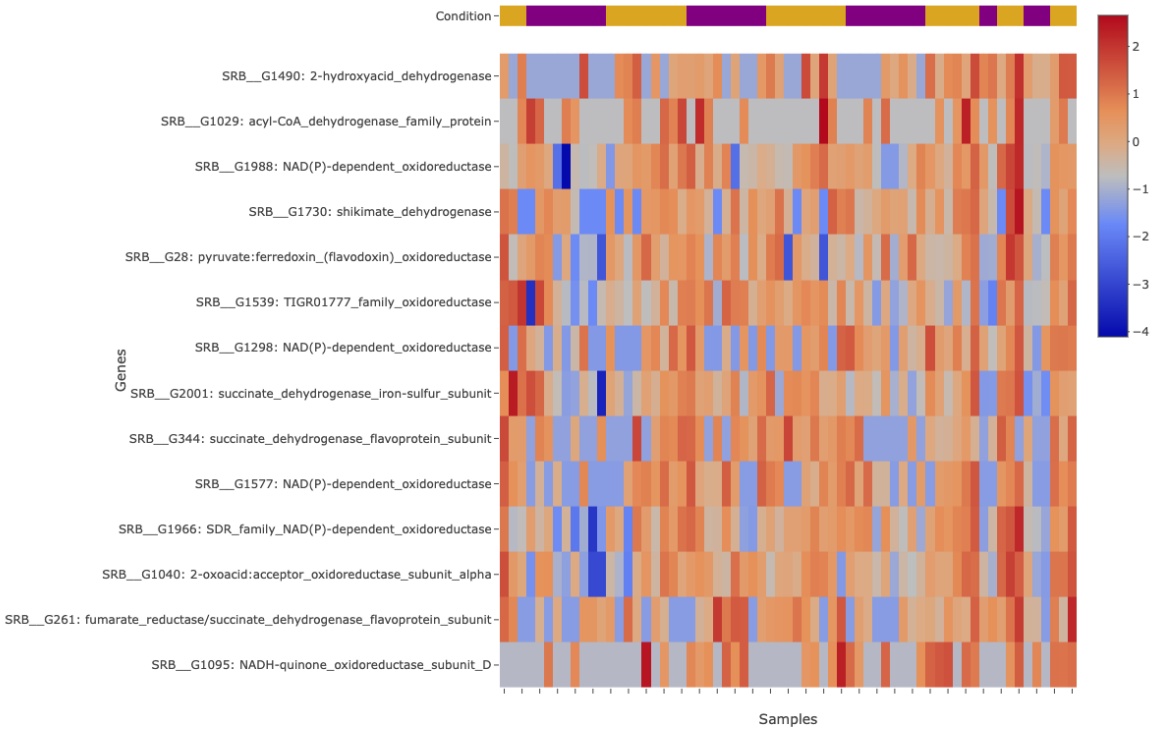


Figure 13. Dehydrogenase genes in PB-SRB1 that were significantly different in some form of light-dark comparison. Shrinkage adjusted, row-scaled, and visualized with plot_ly. From left to right we see experiment progression with time, gold bars correspond to light time points and purple bars to dark time points.

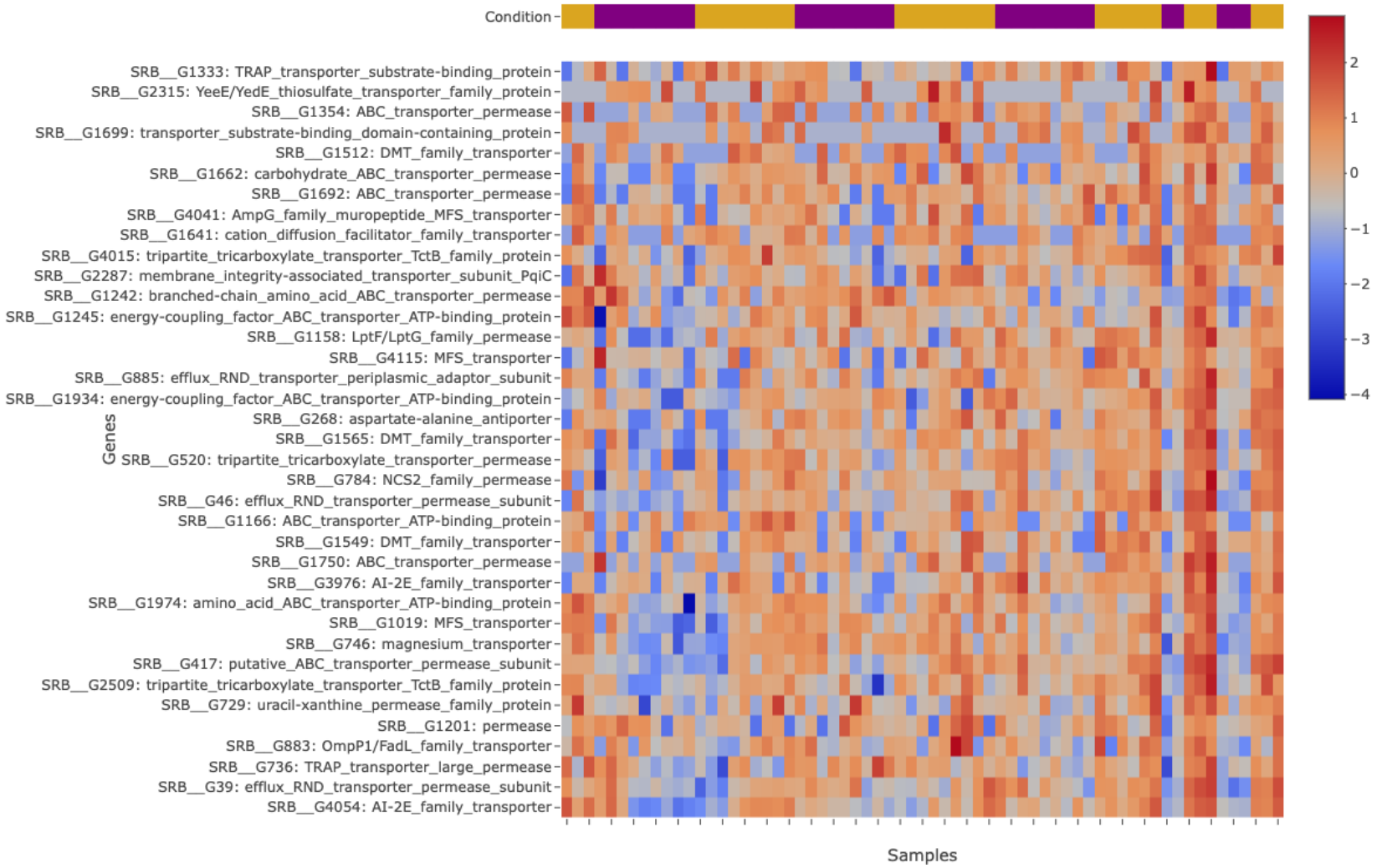


Figure 14. Transporter genes in PB-SRB1 that were significantly different in some form of light-dark comparison. Shrinkage adjusted, row-scaled, and visualized with plot_ly. From left to right we see experiment progression with time, gold bars correspond to light time points and purple bars to dark time points.

Conclusions

We hoped to understand the potential for sharing of photosynthate between PB-PSB1 and PB-SRB1. In order to do this we examined changes due to light exposure in order to see if this would help us glean information about PB-PSB1 photosynthate, if we also saw changes in PB-SRB1 this could potentially give us insight into its subsequent use by PB-SRB1. To do this we explored the gene expression patterns of the two main members of the pink berry consortium, PB-PSB1 and PB-SRB1, over their diel cycle in a controlled lab environment. *Thiohalocapsa* PB-PSB1 exhibited sweeping, genome-wide changes in transcription over a diel cycle, with rapid, reproducible responses to light-dark transitions, suggesting either circadian and/or direct sensing of light. *Desulfofustis* PB-SRB1, on the other hand, differentially regulated genes over diel cycles with distinct clusters of light-responsive and dark-responsive genes. Due to the timing and magnitude of this transcriptional response, which both lagged several hours behind the light dark shift and was more enhanced at later incubation timepoints, we conclude that SRB1 is most likely sensing diel shifts in PSB1's photosynthetic exudates.

When we examined differentially upregulated clusters in our organisms we found that genes upregulated in the light in *Thiohalocapsa* PB-PSB1 suggest an overall increase in growth rate and anabolic activity, including genes involved in translation (ribosomal proteins, tRNA modification/translation efficiency), cell division and envelope biogenesis (DNA replication, membrane biosynthesis, peptidoglycan synthesis/remodeling), and vitamin/cofactor biosynthesis (thiamine, riboflavin,, heme, biotin, vitamin B12 synthesis). Furthermore, while carbon fixation genes in PB-PSB1 were not differentially expressed with light, we observed daytime upregulation indicative of carbon catabolism (6-

phosphofructokinase in glycolysis, PHB depolymerase for consumption of stored polyhydroxyalkanoates), connections between central carbon metabolism and anabolism (acetyl-CoA carboxylase as conduit to fatty acid biosynthesis), and metabolism of photorespiration by-products (glycolate/glyoxylate).

In terms of nighttime, in PB-PSB1 we also found a cluster of genes that exhibited nighttime upregulation and daytime downregulation. Within this cluster we saw an upregulation of genes associated with nitrogenase activation as well as multiple genes that generated ammonium. This may be a reaction to daytime carbon fixation in pink berries as shown by previous work. The cell is then relatively more limited in nitrogen and would have to regenerate nitrogen for intracellular nutrient balance. For PB-PSB1's nighttime upregulated cluster we also observed genes that would be involved in breaking down carbon stores like glycogen (glycogen phosphorylase), as well as glycosyl transferases that would be moving this carbon. This may be indicating a breakdown of daytime fixed carbon, as the cell could oxidize stored carbon to produce ATP. Corresponding with the nighttime upregulation in genes associated with glycogen breakdown, we observed a suite of genes which suggest increased respiratory activity. We conclude this may represent a pathway for respiratory reduction of stored elemental sulfur in PB-PSB1. In this same nighttime upregulated cluster, furthermore, we saw an upregulation of genes involved in bacteriochlorophyll synthesis and transport. Older cultures of PSB like *Thiocapsa roseopersicina*, have been grown under a light-oxic/dark-anoxic regime and been shown to generate bacteriochlorophyll at night due to oxygen sensitivity (De Wit and Van Gemerden 1990). While our incubations were anoxic there may still be some sources of oxygen. Cyanobacteria in our consortia may have

generate small daytime pockets of oxygen, or there may be some sort of circadian response programmed to regenerate bacteriochlorophyll at night.

Interestingly, the products of photorespiration – e.g. glycolate – may represent a shared metabolite in the consortia. Daytime upregulation in glycolate oxidase by the phototroph PB-PSB1 suggests either photoassimilation or consumption of endogenously produced glycolate. This evidence is intriguing as *Desulfofustis* PB-SRB1's closest cultured relative, *Desulfofustis glycolicus*, was isolated as a glycolate oxidizing sulfate reducer. Finally, differentially transcribed genes in *Desulfofustis* PB-SRB1 suggest enhanced rates of sulfate reduction during the day stimulated by the daytime availability of photosynthate. These conclusions are based on the daytime upregulation of *dsrD*, an activator of dissimilatory sulfite reductase, the final step in sulfate reduction, and daytime upregulation in numerous transporters and dehydrogenases (succinate, shikimate and 2-hydroxyacid dehydrogenases), as well as upregulation of multiple genes for complexes involved in electron transport.

References

Amin, S. A., L. R. Hmelo, H. M. Van Tol, et al. 2015. “Interaction and Signalling between a Cosmopolitan Phytoplankton and Associated Bacteria.” *Nature* 522 (7554): 98–101. https://doi.org/10.1038/nature14488.

Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. “Near-Optimal Probabilistic RNA-Seq Quantification.” *Nature Biotechnology* 34 (5): 525–27. https://doi.org/10.1038/nbt.3519.

Brondijk, T. H. C., D. Fiegen, D. J. Richardson, and J. A. Cole. 2002. “Roles of NapF, NapG and NapH, Subunits of the Escherichia Coli Periplasmic Nitrate Reductase, in Ubiquinol Oxidation: Roles of NapF, G and H in Ubiquinol Oxidation in E. Coli.” *Molecular Microbiology* 44 (1): 245–55. https://doi.org/10.1046/j.1365-2958.2002.02875.x.

Buchanan-Wollaston, V., M. C. Cannon, J. L. Beynon, and F. C. Cannon. 1981. *Role of the nifA Gene Product in the Regulation of Nif Expression in Klebsiella Pneumoniae*. 294.

Burow, Luke C., Dagmar Woebken, Brad M. Bebout, et al. 2012. “Hydrogen Production in Photosynthetic Microbial Mats in the Elkhorn Slough Estuary, Monterey Bay.” *The ISME Journal* 6 (4): 863–74. https://doi.org/10.1038/ismej.2011.142.

Burow, Luke C., Dagmar Woebken, Ian PG Marshall, et al. 2013. “Anoxic Carbon Flux in Photosynthetic Microbial Mats as Revealed by Metatranscriptomics.” *The ISME Journal* 7 (4): 817–29. https://doi.org/10.1038/ismej.2012.150.

Bushnell, Brian. 2014. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Released. https://bbmap.org.

Capone, Douglas G., Jonathan P. Zehr, Hans W. Paerl, Birgitta Bergman, and Edward J. Carpenter. 1997. "*Trichodesmium* , a Globally Significant Marine Cyanobacterium." *Science* 276 (5316): 1221–29. https://doi.org/10.1126/science.276.5316.1221.

De Wit, Rutger, and Hans Van Gemerden. 1990. "Growth and Metabolism of the Purple Sulfur Bacterium Thiocapsa Roseopersicina under Combined Light/Dark and Oxic/Anoxic Regimens." *Archives of Microbiology* 154 (5): 459–64. https://doi.org/10.1007/BF00245228.

Dolla, Alain, Marjorie Fournier, and Zorah Dermoun. 2006. "Oxygen Defense in Sulfate-Reducing Bacteria." *Journal of Biotechnology* 126 (1): 87–100. https://doi.org/10.1016/j.jbiotec.2006.03.041.

Dron, Anthony, Sophie Rabouille, Pascal Claquin, Bertrand Le Roy, Amélie Talec, and Antoine Sciandra. 2012. "Light-Dark (12:12) Cycle of Carbon and Nitrogen Metabolism in Crocosphaera Watsonii WH8501: Relation to the Cell Cycle: Daily C and N Metabolism in C. Watsonii WH8501." *Environmental Microbiology* 14 (4): 967–81. https://doi.org/10.1111/j.1462-2920.2011.02675.x.

D'Souza, Glen, Shraddha Shitut, Daniel Preussger, Ghada Yousif, Silvio Waschina, and Christian Kost. 2018. "Ecology and Evolution of Metabolic Cross-Feeding Interactions in Bacteria." *Natural Product Reports* 35 (5): 455–88. https://doi.org/10.1039/C8NP00009C.

Eren, A. Murat, Evan Kiefl, Alon Shaiber, et al. 2020. "Community-Led, Integrated, Reproducible Multi-Omics with Anvi'o." *Nature Microbiology* 6 (1): 3–6. https://doi.org/10.1038/s41564-020-00834-3.

Ferreira, Delfim, Ana C. C. Barbosa, Gonçalo P. Oliveira, Teresa Catarino, Sofia S. Venceslau, and Inês A. C. Pereira. 2022. "The DsrD Functional Marker Protein Is an Allosteric Activator of the DsrAB Dissimilatory Sulfite Reductase." *Proceedings of the*

National Academy of Sciences 119 (4): e2118880119. https://doi.org/10.1073/pnas.2118880119.

Friedrich, M., N. Springer, W. Ludwig, and B. Schink. 1996. "Phylogenetic Positions of Desulfofustis Glycolicus Gen. Nov., Sp. Nov. and Syntrophobotulus Glycolicus Gen. Nov., Sp. Nov., Two New Strict Anaerobes Growing with Glycolic Acid." *International Journal of Systematic Bacteriology* 46 (4): 1065–69. https://doi.org/10.1099/00207713-46-4-1065.

Fründ, Claudia, and Yehuda Cohen. 1992. "Diurnal Cycles of Sulfate Reduction under Oxic Conditions in Cyanobacterial Mats." *Applied and Environmental Microbiology* 58 (1): 70–77. https://doi.org/10.1128/aem.58.1.70-77.1992.

Gemerden, H. van. 1968a. "Growth Measurements of *Chromatium* Cultures." *Archiv Für Mikrobiologie* 64 (2): 103–10. https://doi.org/10.1007/BF00406968.

Gemerden, H. van. 1968b. "On the ATP Generation by *Chromatium* in Darkness." *Archiv Für Mikrobiologie* 64 (2): 118–24. https://doi.org/10.1007/BF00406970.

Gemerden, H. van. 1968c. "Utilization of Reducing Power in Growing Cultures of *Chromatium*." *Archiv Für Mikrobiologie* 64 (2): 111–17. https://doi.org/10.1007/BF00406969.

Harvey, Emily, Jeffrey Heys, and Tomáš Gedeon. 2014. "Quantifying the Effects of the Division of Labor in Metabolic Pathways." *Journal of Theoretical Biology* 360 (November): 222–42. https://doi.org/10.1016/j.jtbi.2014.07.011.

Hirschman, J., P. K. Wong, K. Sei, J. Keener, and S. Kustu. 1985. "Products of Nitrogen Regulatory Genes ntrA and ntrC of Enteric Bacteria Activate glnA Transcription in Vitro: Evidence That the ntrA Product Is a Sigma Factor." *Proceedings of the National Academy of Sciences* 82 (22): 7525–29. https://doi.org/10.1073/pnas.82.22.7525.

Janausch, I. G., E. Zientz, Q. H. Tran, A. Kröger, and G. Unden. 2002. “C4-Dicarboxylate Carriers and Sensors in Bacteria.” *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1553 (1–2): 39–56. https://doi.org/10.1016/S0005-2728(01)00233-X.

Kolde, Raivo. n.d. *Pheatmap: Pretty Heatmaps*. V. R package version 1.0.13. https://github.com/raivokolde/pheatmap.

Lee, Jackson Z., Luke C. Burow, Dagmar Woebken, et al. 2014. “Fermentation Couples Chloroflexi and Sulfate-Reducing Bacteria to Cyanobacteria in Hypersaline Microbial Mats.” *Frontiers in Microbiology* 5. https://doi.org/10.3389/fmicb.2014.00061.

Lex, Alexander, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. 2014. “UpSet: Visualization of Intersecting Sets.” *IEEE Transactions on Visualization and Computer Graphics* 20 (12): 1983–92. https://doi.org/10.1109/TVCG.2014.2346248.

Lindstrom, ES, SR Tove, and PW Wilson. 1950. “Nitrogen Fixation by the Green and Purple Sulfur Bacteria.” *Science* 112 (August): 197–98.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

Muyzer, Gerard, and Alfons J. M. Stams. 2008. “The Ecology and Biotechnology of Sulphate-Reducing Bacteria.” *Nature Reviews Microbiology* 6 (6): 441–54. https://doi.org/10.1038/nrmicro1892.

Pfennig, N. 1977. “PHOTOTROPHIC GREEN AND PURPLE BACTERIA: A COMPARATIVE, SYSTEMATIC SURVEY.” *Annual Review of Microbiology* 31 (1): 275–90. https://doi.org/10.1146/annurev.mi.31.100177.001423.

Pinckney, J. L., and H. W. Paerl. 1997. “Anoxygenic Photosynthesis and Nitrogen Fixation by a Microbial Mat Community in a Bahamian Hypersaline Lagoon.” *Applied and Environmental Microbiology* 63 (2): 420–26. https://doi.org/10.1128/aem.63.2.420-426.1997.

Reddy, Vamsee S., Maksim A. Shlykov, Rostislav Castillo, Eric I. Sun, and Milton H. Saier. 2012. “The Major Facilitator Superfamily (MFS) Revisited.” *The FEBS Journal* 279 (11): 2022–35. https://doi.org/10.1111/j.1742-4658.2012.08588.x.

Rothermich, Mary M., Ricardo Guerrero, Robert W. Lenz, and Steve Goodwin. 2000. “Characterization, Seasonal Occurrence, and Diel Fluctuation of Poly(Hydroxyalkanoate) in Photosynthetic Microbial Mats.” *Applied and Environmental Microbiology* 66 (10): 4279–91. https://doi.org/10.1128/AEM.66.10.4279-4291.2000.

Sánchez-Cañizares, Carmen, Jürgen Prell, Francesco Pini, et al. 2020. “Global Control of Bacterial Nitrogen and Carbon Metabolism by a PTS^{Ntr} -Regulated Switch.” *Proceedings of the National Academy of Sciences* 117 (19): 10234–45. https://doi.org/10.1073/pnas.1917471117.

Shah, Vinod K., Priya Rangaraj, Ranjini Chatterjee, et al. 1999. “Requirement of NifX and Other *Nif* Proteins for In Vitro Biosynthesis of the Iron-Molybdenum Cofactor of Nitrogenase.” *Journal of Bacteriology* 181 (9): 2797–801. https://doi.org/10.1128/JB.181.9.2797-2801.1999.

Sievert, Carson. n.d. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny*. Chapman and Hall/CRC. https://plotly-r.com.

Sikora, Florian, Lara Veronika Perko Budja, Olja Milojevic, Amelia Ziemniewicz, Przemyslaw Dudys, and Boris Görke. 2024. “Multiple Regulatory Inputs Including Cell

Envelope Stress Orchestrate Expression of the *Escherichia Coli rpoN* Operon." *Molecular Microbiology* 122 (1): 11–28. https://doi.org/10.1111/mmi.15280.

Steppe, Tf, and Hw Paerl. 2002. "Potential N2 Fixation by Sulfate-Reducing Bacteria in a Marine Intertidal Microbial Mat." *Aquatic Microbial Ecology* 28: 1–12. https://doi.org/10.3354/ame028001.

Stoodley, P., K. Sauer, D. G. Davies, and J. W. Costerton. 2002. "Biofilms as Complex Differentiated Communities." *Annual Review of Microbiology* 56 (1): 187–209. https://doi.org/10.1146/annurev.micro.56.012302.160705.

Vrede, Katarina, Mikal Heldal, Svein Norland, and Gunnar Bratbak. 2002. "Elemental Composition (C, N, P) and Cell Volume of Exponentially Growing and Nutrient-Limited Bacterioplankton." *Applied and Environmental Microbiology* 68 (6): 2965–71. https://doi.org/10.1128/AEM.68.6.2965-2971.2002.

Weissgerber, Thomas, Renate Zigann, David Bruce, et al. 2011. "Complete Genome Sequence of *Allochromatium Vinosum* DSM 180(T)." *Standards in Genomic Sciences* 5 (3): 311–30. https://doi.org/10.4056/sigs.2335270.

West, Stuart A., Stephen P. Diggle, Angus Buckling, Andy Gardner, and Ashleigh S. Griffin. 2007. "The Social Lives of Microbes." *Annual Review of Ecology, Evolution, and Systematics* 38 (1): 53–77. https://doi.org/10.1146/annurev.ecolsys.38.091206.095740.

Wickham, Hadley. 2009. *Ggplot2: Elegant Graphics for Data Analysis*. Springer New York. https://doi.org/10.1007/978-0-387-98141-3.

Wilbanks, Elizabeth G., Parris T. Humphrey, Ulrike Jaekel, Cristina Moraru, and Victoria J. Orphan. 2010. *Eco-Physiology of Macroscopic Pink and Green Bacterial Consortia of the Mighty Little Sippewissett Salt Marsh*.

Wilbanks, Elizabeth G., Ulrike Jaekel, Verena Salman, et al. 2014. “Microscale Sulfur Cycling in the Phototrophic Pink Berry Consortia of the S Ippewissett S Alt M Arsh.” *Environmental Microbiology* 16 (11): 3398–415. https://doi.org/10.1111/1462-2920.12388.

Zhang, Jun, Hongyuan Li, Wenjun Tao, and Jun Zhou. 2026. “1 ClusterGVis: An Advanced Visualization and Clustering Tool for 2 Gene Expression Analysis.” *Genomics, Proteomics, and Bioinformatics*, ahead of print, January 22. https://doi.org/https://doi.org/10.1093/gpbjnl/qzag005.

Chapter III: Chemical insights into autotrophic activity in *Thiohalocapsa* PB-PSB1 and potential use of fixed organosulfonates by *Desulfofustis* PB-SRB1

Braulio S. Castillo Villaseñor, Anna E. Warsaw, Emily N. Junkins, Michael A. Ochsenkühn, Salah Abdelrazig, Shady A. Amin, and Elizabeth G. Wilbanks

Abstract

Tightly coupled interactions between microbes engaged in intimate symbioses are hard to dissect. We hoped to probe the syntrophic relationship between the main members of the pink berry consortium, PB-PSB1 and PB-SRB1, by intentionally uncoupling their relationship via exposure of sulfate reduction inhibitor. Via use of molybdate and prolonged light or dark exposure we hoped to glean insight into the chemical identity of photosynthate produced by PB-PSB1. While molybdate specifically exhibited PB-SRB1 activity, we found the main difference in pink berry metabolomic space to be due to prolonged light or dark exposure. By delving into light-dark differences in this chemical space we were able to find some potential markers for autotrophy (eicosapentaenoic acid, 1,2-dihydroxyheptadec-16-yn-4-one, cis-7-hexadecenoic), as well as indicators of incomplete glycogen oxidation to polyhydroxyxalkanoates by PB-PSB1. When we looked at differences due to molybdate we found that PB-SRB1 may be oxidizing a compound with an aromatic moiety that might not originate from PB-PSB1, however, in terms of photosynthate it may be catabolizing organosulfonates exuded by PB-PSB1 in order to produce its energy.

Introduction

PB-PSB1 and PB-SRB1 are consistently found together across our sites in Great Sippewissett, Little Sippewissett, and Penzance Point (Wilbanks et al. 2014, 2022; Doré et al. 2024). As shown in chapter one, the genomes of these organisms have also remained highly similar within the span of our ten years of sampling. This partnership thus seems to be highly specific, and is reminiscent of microbiome recruitment in marine diatoms (Shibl et al. 2020). PB-PSB1's photosynthate, specifically, could be shared with PB-SRB1 in addition to sulfur, and be used as an electron donor for PB-SRB1's sulfate reduction. While we started to examine potential sharing of carbon between our two main community members, PB-PSB1 and PB-SRB1, by analyzing pink berry transcripts over a "normal" diel cycle in chapter two, we decided to further probe for shared photosynthate in chapter three.

We first aimed to identify potential shared carbon photosynthate by collecting chemical evidence resulting from prolonged light or dark exposure via metabolite collection. By looking at light generated carbon compounds we would potentially be able to identify photosynthate, while dark generated compounds would give us insight into dark carbon use. Second, to further dissect this partnership, we aimed to specifically inhibit PB-SRB1 and thus potentially uncouple their activity. Molybdate is a known sulfate reduction inhibitor (Oremland and Capone 1988), therefore by incubating pink berries with molybdate we thought it should be possible to disrupt PB-SRB1's energy generation and disturb PB-PSB1's and PB-SRB1's known sulfur sharing activity. Under this disturbed regime, potentially shared photosynthate from PB-PSB1 that would natively be consumed by PB-SRB1 could be accumulating under molybdate inhibited conditions (unless some other community member consumes this). If we see a depletion of a compound under molybdate inhibited conditions,

furthermore, then the identified compound itself may instead be a product of PB-SRB1 metabolism and no longer produced once PB-SRB1 is inactive. Since the partnership between these two organisms is so specific across space and time, we believe that this accumulation or depletion would be specific to PB-SRB1 and thus help us identify shared photosynthate.

Interestingly, many studies exploring sulfate reducer inhibition by molybdate have specifically targeted anaerobic digestors or salt marsh sediments (Banat et al. 1983; Isa and Anderson 2005; Ranade 1999), focusing on characterizing the community's metabolic output due to a disturbed metabolic regime. While these have used a variety of analytical techniques, none of these studies, to our knowledge, have produced a metatranscriptome that provides an 'omic insight to the nature of molybdate inhibition. Our study could thus produce novel insight into "normal" sulfate reducer physiology and its subsequent inhibition due to molybdate.

We performed two sets of experiments. In the first experiment we pre-incubated for 2.5 days a set of 600 berries with ^{13}C bicarbonate to be fixed by PB-PSB1, and then subsetted this into individual incubations where a group was exposed to molybdate and another group exposed to no molybdate. Each of these subsets was then exposed to continuous light conditions or continuous dark conditions for 48 hours. We collected RNA, endometabolites, and exometabolites. This was then followed by a second experiment where berries were not pre-incubated and instead immediately individually incubated and exposed to ^{12}C bicarbonate for carbon fixation. This was similarly subsetted to molybdate and non-molybdate exposed conditions, as well as continuous light and continuous dark conditions, albeit this incubation was only for 24 hours. We did not collect RNA for these ^{12}C incubated samples.

In order to understand this disturbed relationship, we characterized the metatranscriptomic space in order to verify that our molybdate inhibition indeed specifically affected the gene expression of PB-SRB1. We then explored whether there was a significant difference in our metabolomic space due to light exposure. We explored the endometabolome and characterized dark carbon transformations in the pink berry biofilm, informing what kind of transformation may be generating ATP for community members. We were also able to determine some possible compounds as markers for autotrophic activity in the pink berry biofilm. Our molybdate inhibition experiments identified PB-SRB1 as potentially being able to oxidize some sort of compound with an aromatic moiety, as well as generate energy via dissimilatory catabolism of organosulfonates.

Methods

^{13}C Lab Continuous Light/Dark Incubation

All glassware was baked in a muffle furnace at 450C for 4-5 hours then cooled overnight. Tubing and plastic was 10% acid washed. Berries were collected from site LS01 (41°34′33.01″N, 70°38′21.24″W) at noon on 16 August 2022 during high tide. Water was collected separately for both washing of berries from sediment and for incubations themselves. Water was 0.02u filtered and used to wash sediment off berries, and then 2 sets of about 600 berries were pre-incubated for ~two and a half days under an artificial light-dark cycle in ^{13}C bicarbonate (collected on a Tuesday morning and separated into smaller incubations on Friday morning). For the batch pre-incubation, berries were placed in 0.02u filtered marsh water and amended with 1.5 mM neutralized sodium sulfide as well as 21.17 mM 99 at% sodium bicarbonate. Upon termination of the pre-incubation a subset of berries were frozen in liquid nitrogen for endometabolite and RNA extraction. The media was filtered in series for exometabolites, first in an Oasis HLB (polar) column followed by a BondElut C18 (nonpolar) column. This was first primed with LCMS grade methanol, followed by DI water, the filtrate, and a final wash in DI. After this, berries were moved into an artificial continuous light or continuous dark regime.

For individual incubations (20 berries/vial), berries were partitioned into the following four experiments (Figure 1): continuous light with no molybdate, continuous light with molybdate, continuous dark with no molybdate, and continuous dark with molybdate. Berries were incubated on either continuous light or dark at the beginning of their natural light-dark cycle (6AM or 7PM respectively, thus first 12 hours of incubation were “natural”). Time point zero consisted of subsampling of the batch filtrate, then individual incubations were

subsequently sampled at 12 hours, 24 hours, and 48 hours. All sampling was done in triplicate, leading to 7 light time points (1 zero time point, 3 molybdate (12, 24 and 48h), and 3 no molybdate (12, 24 and 48h)) and 7 dark time points (1 zero time point, 3 molybdate (12, 24 and 48h), and 3 no molybdate (12, 24 and 48h)). The 20 berries corresponding to each incubation were frozen in liquid nitrogen (RNA and endometabolites), and the media was filtered via Oasis HLB and BondElut C18 columns like the batch filtrate (exometabolites). For individual incubations 35.29 mM ^{13}C sodium bicarbonate, 1.5 mM neutralized sodium sulfide, and 10 mM sodium molybdate was used.

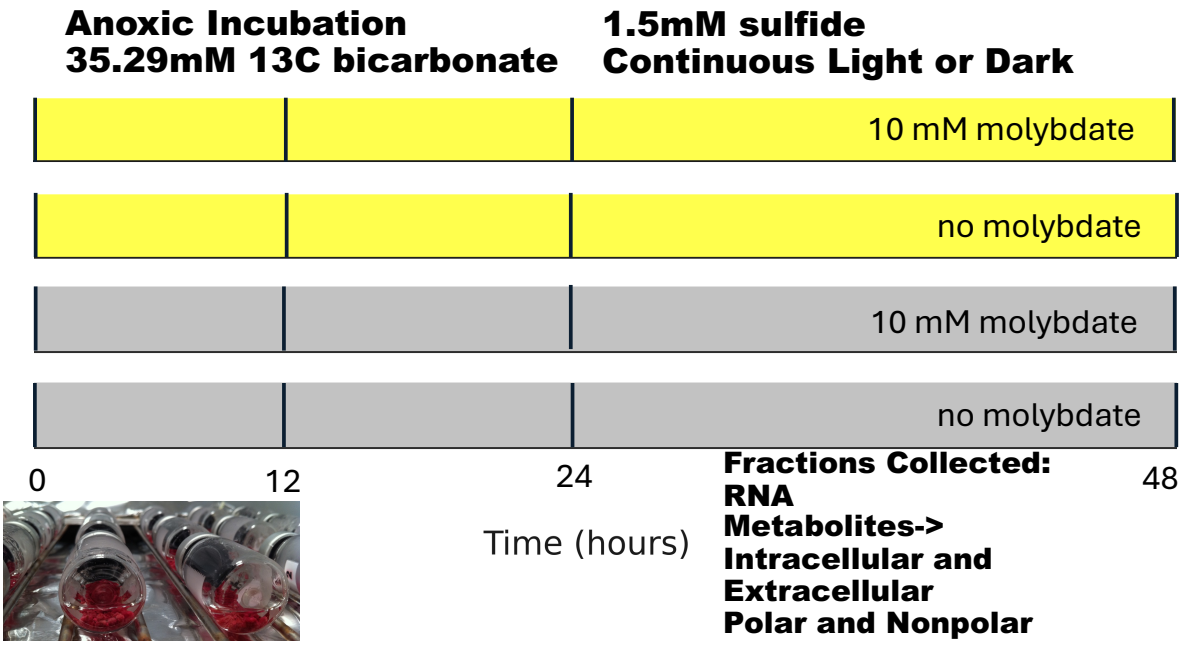


Figure 1. Overview of our continuous lighting incubation conducted with ^{13}C bicarbonate. Samples were collected at 0, 12, 24, and 48 hours; with or without molybdate; continuous light or continuous dark.

^{12}C Lab Continuous Light/Dark Incubation

A separate experiment was carried out with ^{12}C sodium bicarbonate. Berries were collected from site LS01 (41°34′33.01′N, 70°38′21.24′W) around ~11AM on 24 August 2022 during low tide. Water was collected separately for both washing of berries from sediment

and for incubations themselves. Water was 0.02u filtered and used to wash sediment off 280 berries which were then separated into individual vials for incubation (20 berries/vial). Gassing of vials to render anoxic started at 2:50PM, with our last incubation gassed by 4:19PM. These were then amended with sulfide, bicarbonate, and/or molybdate and our incubations started at 5PM (Figure 2).

While the same four type of incubations were carried out (continuous light or dark in molybdate or no molybdate), there was no pre-incubation and only endometabolites were sampled at incubation start (0 hours, no exometabolites). This was then followed by one instance of endometabolite and exometabolite sampling at 24 hours (time point 1). Tissue was collected at both points for RNA, however this was not sequenced. Tissue that could still be sequenced exists for light at 24 hours, dark at 0 hours, and dark at 24 hours. Incubations were started simultaneously so berries were off their regular light-dark cycle, with both incubations running from 5PM to about 6PM the next day. Thus "light" berries were exposed to ~24 hours on top of their regular 12 hours of light on day of sampling to total 36 hours of continuous light. "Dark" berries were exposed to 24 hours of dark after their regular 12 hours of light and thus the first 12 "dark" hours were within their regular rhythm and more akin to our ^{13}C experiment. For individual incubations 10 mM ^{12}C sodium bicarbonate, 1.5 mM neutralized sodium sulfide, and 10 mM sodium molybdate was used.

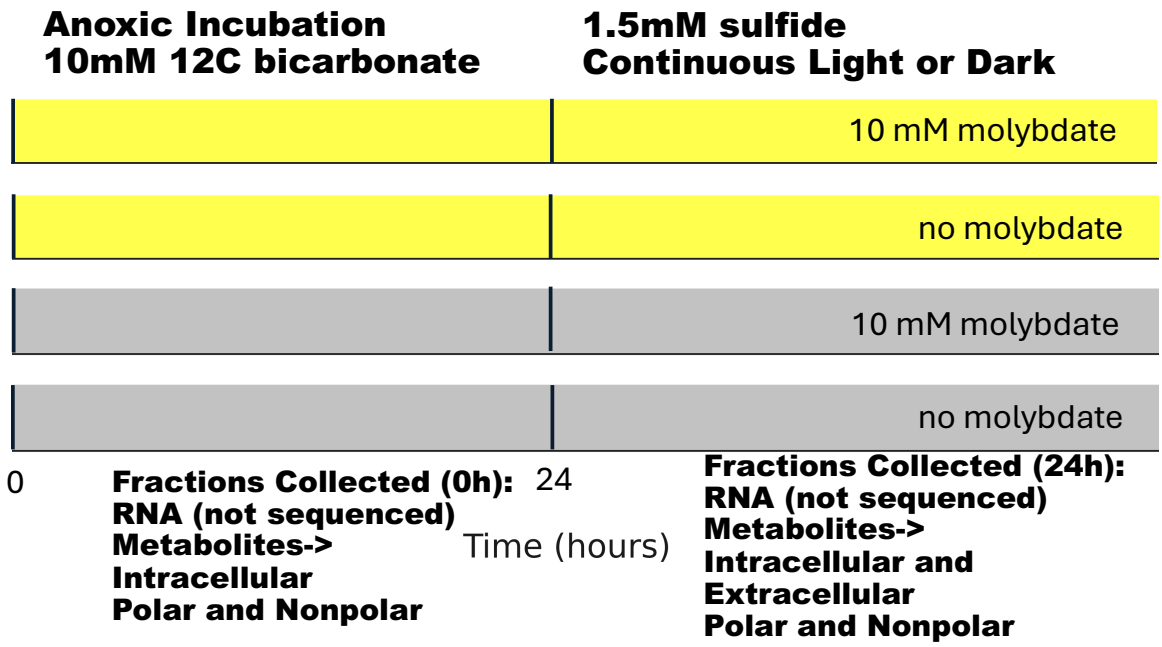


Figure 2. Overview of our continuous lighting incubation conducted with ^{12}C bicarbonate. Samples were collected at 0 and 24 hours; with or without molybdate; continuous light or continuous dark.

RNA Extraction and Sequencing

Berries were extracted with Zymo R2040 kit. Extractions were distributed across 14 time points sampled in triplicate, leading to 42 total extractions. Each sample consisted of 5 berries extracted together. Samples were treated with DNAse I in 2 hour spurts (Zymo DNAse I set, E1011), varying in length from 2 to 6 hours total DNAse I incubation. For QC we quantified RNA with Qubit (least was 275 total ng of RNA, up to 1300 ng in one instance), verified lack of DNA contamination with PCR (if DNA present, sample underwent an additional 2 hour incubation with DNAse I), and RNA quality was checked with the Tape Station (RIN score of at least 6, often around 7-8). If RNA quantity fell below these limits, then an additional 5-berry extraction was pooled to increase concentration.

Note: While Tape Station can be typically used to check for large DNA remaining in sample, it is still possible for DNA to be “hiding” in RNA peak. The best way to check for

contamination is to use PCR to amplify contaminating DNA. If there are no bands then sample is not DNA contaminated, and this can be sent for sequencing.

Samples were sequenced by the Joint Genome Institute under grant 508949. This was sequenced 2x151 using the Illumina NovaSeqX platform, and BBDuk was used to remove various contaminants, remove adapter reads, and right quality trim reads. Reads that mapped via BBMap to human, cat, dog and mouse references as well as common microbial contaminants were removed; ribosomal reads were also removed (BBtools v39.19, Bushnell 2014). The final filtered fastq was used for metatranscriptome analysis.

RNA Differential Expression Analysis

The repeatless 41-bin database in the genes.fna style (out.db.rm.mxdsp from Chapter 1), was used as an index for pseudoalignment of sequenced reads with kallisto for counting (kallisto v0.48.0, Bray et al. 2016). Counts were then separated by bin, filtered for genes whose total count across samples was 10 or higher, and differential expression analysis was run with DESeq2, followed by shrinkage analysis via "apeglm" (DESeq2 v1.46.0, Love et al. 2014). The p-adjusted, or q-value, below 0.05 was then used to determine significance. MA plots were generated with plotMA from DESeq2. Bar, volcano, and PCA plots were generated with ggplot2 (q-value less than 0.05 and $\log_2$ (fold change) greater than |1| was used for bar and volcano plots; ggplot2 v3.5.2, Wickham 2009). PCoA plots were also generated via ggplot, and ordinations were based on Bray-Curtis dissimilarity. Heatmaps filtered by significance were generated via pheatmap (50 most significant genes and another based on the 50 most expressed of the significant genes; pheatmap v1.0.13, Kolde, n.d.), and ClusterGVis was used to generate clustered line maps and heatmaps (clustered by TCSeq;

v0.1.4, Zhang et al. 2026). plot_ly was used for interactive volcano plots and heatmaps (plotly v4.11.0, Sievert, n.d.).

ComplexUpset was used in order to generate upset plots for exploration of gene membership across differing DESeq2 significance comparisons (ComplexUpset v.1.3.3, Lex et al. 2014). To make "specialized" interactive heatmaps, genes that were significant across different conditions of interest were compiled into one list then clustered via ClusterGVis. This matrix was then exported to plot_ly and used to explore differential patterns across genes of interest (molybdate v no molybdate).

Metabolite extraction

For endometabolite extraction, five berries were treated as one replicate and centrifuged together in 70% methanol in milliQ water. Extracted supernatant was dried down in a 96-well plate with nitrogen gas. This was followed by a 50% methanol-50% chloroform centrifugation where supernatant from the same sample was also dried down in a separate plate.

For exometabolite extraction, filtrate from columns was first eluted in 2% ammonium hydroxide in methanol, followed by 2% formic acid in methanol into the same well. This elution was done for both C18 and HLB columns, with a well for each. This was done into a 96-well plate and dried down with nitrogen gas. Finally, each column was additionally used for an ethyl acetate elution into a glass test tube and dried down with nitrogen gas. (one well for C18 with basic and acidic methanol elution, one well for HLB with basic and acidic methanol elution, one test tube for C18 ethyl acetate elution, one test tube for HLB ethyl acetate elution).

LC-MS Analysis

Metabolites were shipped to NYU Abu Dhabi and processed by the lab of Shady Amin. Chromatography was performed using Vanquish UHPLC system (Thermo Fisher Scientific, USA) on a BEH Z-HILIC column (Waters, USA) and a BEH Z-HILIC guard column. Mobile phases: (A) 20 mM ammonium carbonate, (B) acetonitrile. Total analysis time was 24 min. Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific, USA) was used in simultaneous ESI+ and ESI- modes for full LC-MS profiling and to generate data dependent MS/MS (ddMS/MS) accurate mass spectra for identification (most intense peaks, TopN).

LC-MS Data Processing

The raw LC–MS datasets of the ^{13}C labelled and unlabelled samples in the study were processed with Compound Discoverer (Thermo Fisher Scientific, USA) for untargeted Isotope-Assisted Metabolomics (IAM) peak-picking, peak alignment, peak deconvolution and univariate analysis. Metabolite identification was performed by Compound Discoverer metabolite search engine linked to different online and local spectral database. Metabolite identification was performed by matching the accurate masses and the MS/MS fragmentation spectra of the detected peaks in the samples with an in-house high-resolution orbitrap spectral database of standards (mzVault, Thermo Fisher Scientific, USA) and an online orbitrap fragmentation database of authentic standards (mzCloud, Thermo Fisher Scientific,USA). The confidence in metabolite identification was assigned as level 1-4 (L1-4) based on the recommendation by Chemical Analysis Working Group, Metabolomics Standards Initiative (MSI).

Multivariate analysis (MVA) was performed using Simca P18 (Sartorius Stedim Data Analytics, Sweden). The analytical performance of the LC-MS run was assessed using the pooled QC approach using unsupervised principal component analysis (PCA). PCA was also used to give an unbiased overview for any possible trends, groupings and outliers within the sample datasets. Partial least squares - discriminant analysis (PLS-DA) and orthogonal partial least squares - discriminant analysis (OPLS-DA) were used for modelling the differences between samples. Only the polar, methanol-based fractions for both endo and exometabolites were analyzed.

Results and Discussion

Continuous lighting and molybdate experiments lead to a light response in PB-PSB1 and a molybdate response in PB-SRB1

We perturbed normal pink berry conditions by exposing berries to continuous lighting as well as molybdate. We examined how the transcriptomic space in this perturbed and off "its normal diel cycle" experiment compared to both our lab-incubated and environmentally sampled diel experiments. Transcripts belonging to PB-PSB1 mostly split cleanly by experiment type as well as light-dark conditions (Supplemental Figure A). When looking at the continuous lighting molybdate and no-molybdate experiments, however, there is little overlap between light and dark samples suggesting separation due to light exposure. When comparing samples' exposure to molybdate, furthermore, there is a lot of overlap between molybdate and no-molybdate samples. This potentially suggests no effect from molybdate and greater segregation due to lighting.

PB-SRB1, furthermore, also split along experiment type albeit its light-dark separation was much less distinct in lab incubated samples (Supplemental Figure B). We see further differences, however, when examining the continuous lighting PB-SRB1 molybdate and no-molybdate samples such that while samples don't really segregate by lighting they do seem to be affected by the molybdate treatment.

PB-PSB1 significantly reacts to light in the no-molybdate experiment whereas PB-SRB1 does not

Knowing that there might be some separation in pink berry transcripts due to light in our continuous experiments, we compared transcript counts between the no-molybdate experiment continuous light and continuous dark treatments. We found that for PB-PSB1

25.67% of analyzed genes (1692 of 6591 genes) were significantly different between light and dark conditions (Figure 3A). We saw a greater spread in the PB-PSB1 dark samples whereas the light samples were much more tightly grouped together. PB-SRB1, on the other hand, exhibited no genes analyzed that were significantly different between light-dark conditions (0 of 3206 genes, Figure 3B). Both light and dark samples for PB-SRB1 exhibited tight, overlapping grouping.

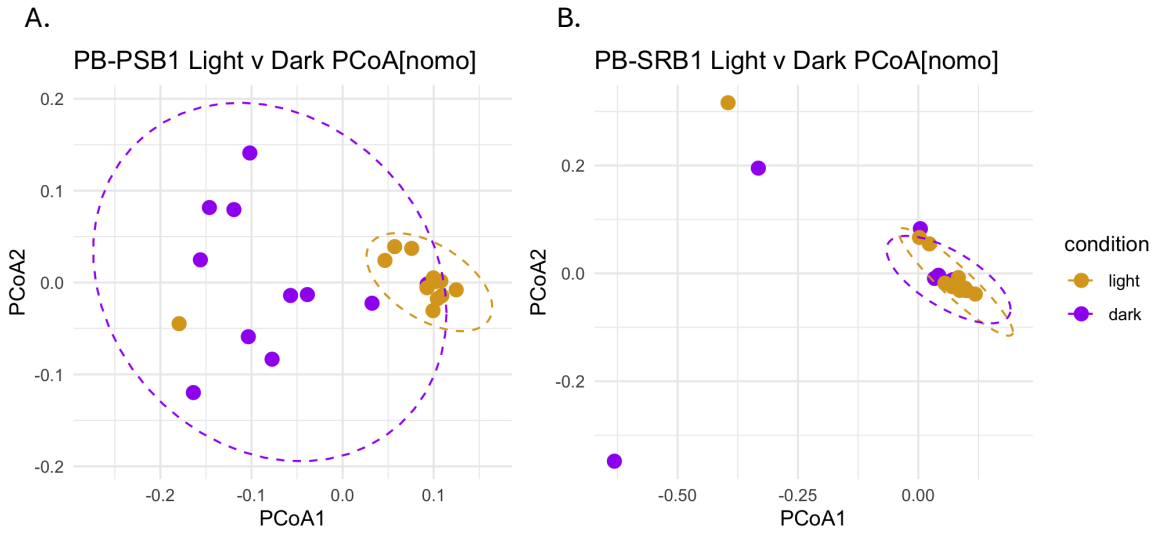


Figure 3. (A) Differences in PB-PSB1 transcripts between continuous light and continuous dark samples within native conditions. PCoA for PB-PSB1 using Bray-Curtis distance where 95% confidence ellipses are based on incubation light v. dark conditions, and corresponding light-dark condition PERMANOVA results are $F=8.042$, $R^2=0.27691$, $p= 0.001$. (B) Differences in PB-SRB1 transcripts between continuous light and continuous dark samples incubated without molybdate. PCoA for PB-SRB1 using Bray-Curtis distance where 95% confidence ellipses are based on incubation light v. dark conditions, and corresponding light-dark condition PERMANOVA results are $F=1.0511$, $R^2=0.04767$, $p= 0.314$. Gold tones correspond to light time points, and purple tones correspond to dark time points.

PB-SRB1 significantly reacts to incubation with molybdate whereas PB-PSB1 does not

We further wanted to explore whether pink berry community members were significantly reacting to molybdate exposure, and thus compared transcript counts between the no-molybdate and molybdate treatments. We found that for PB-PSB1, only 0.43% of genes showed a significant difference between treatments (29 of 6629 genes, Figure 4A),

with only a phosphatase PAP2 family protein showing an adjusted p-value < 0.05 and absolute $\log_2$ (fold change) > 1 greater in the no-molybdate condition. PB-SRB1, in contrast, had many more analyzed genes displaying a significant difference, 16.36% (543 of 3319 genes, Figure 4B).

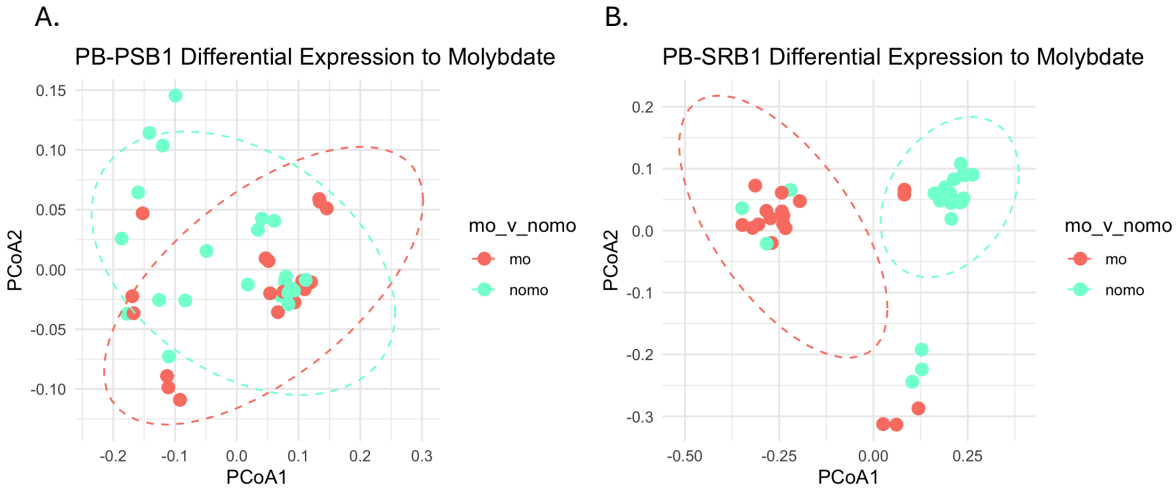


Figure 4. (A) Differences in PB-PSB1 transcripts between molybdate incubated and non-molybdate incubated samples. PCoA for PB-PSB1 using Bray-Curtis distance where 95% confidence ellipses are based on incubation molybdate v. no-molybdate conditions, and corresponding molybdate-no-molybdate condition PERMANOVA results are F=1.5227, R^2=0.03758, p= 0.154. (B) Differences in PB-SRB1 transcripts between molybdate incubated and non-molybdate incubated samples. PCoA for PB-SRB1 using Bray-Curtis distance where 95% confidence ellipses are based on incubation molybdate v. no-molybdate conditions, and corresponding molybdate-no-molybdate condition PERMANOVA results are F=7.5031, R^2=0.16135, p= 0.001. Salmon tones correspond to molybdate time points, and teal tones correspond to no-molybdate time points.

After filtering this list to those statistically significant genes with at least an average two fold change in abundance (abs ($\log_2$ (fold change) > 1 and $p_{adj} < 0.05$), we identified 498 genes – 97% of which were found to have decreased expression due to our molybdate treatment (thus only 14 genes increased in expression with molybdate treatment). In order to determine whether the $\log_2$ (fold change) towards no-molybdate conditions is actually an increase in expression or just due to less expression under molybdate exposure we examined organism specific TPM across samples and found molybdate treatment to cause a global

decrease in transcription for sulfate reducing bacteria (Supplemental Figure C). Specifically, most bins seemed to show no change in TPM distribution due to molybdate exposure, except for PB-SRB1 and a *Desulfosarcinaceae* bin, such that molybdate incubated samples were consistently at the lower end of the TPM distribution. We tested whether the median TPM between our native and molybdate exposed treatments was significantly different for these two bins and found our molybdate exposed bins to exhibit a significantly lower TPM (Figure 5).

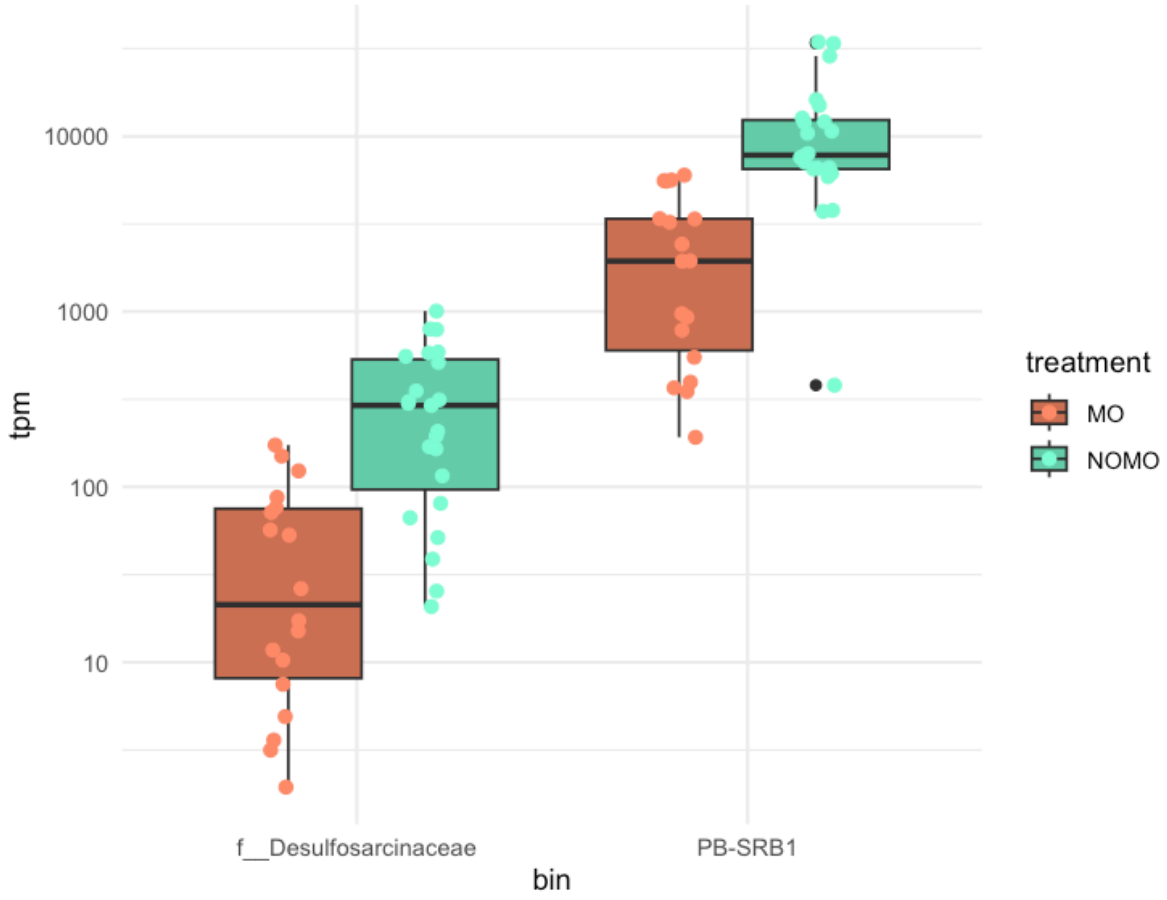


Figure 5. Lower expression in pink berry sulfate reducers PB-SRB1 and *Desulfosarcinaceae* due to molybdate exposure. Wilcoxon rank sum test was used to compare medians and were both found to be significant (PB-SRB1 pvalue=0.0001, *Desulfosarcinaceae* pvalue=0.0001). Teal indicates no molybdate exposure, salmon indicates molybdate exposure.

Genes that were significantly differentially expressed with a negative $\log_2$ (fold change) under molybdate incubated conditions, therefore, may in fact not be downregulated but rather be the result of lower overall expression due to molybdate inhibition of sulfate reduction as the cell is no longer able to participate in electron transport due to stoppage of sulfate reduction. Whether genes with a positive $\log_2$ (fold change) on the other hand, are potentially a result of actual upregulation would need more rigorous examination as the statistical software we used assumes most genes are not being differentially expressed (Love et al. 2014).

Ligh exposure causes widespread changes in the pink berry metabolome

To complement our gene-level analysis we also examined the cellular endometabolome as well as the excreted exometabolome from our incubations (1100 features for endometabolome and 847 features for exometabolome, with an assigned annotation- Table 1).

	exometabolome	endometabolotome
automated annotation assigned to feature (used for ordinations)	847	1100
ambiguous annotation for feature ("similar to")	1835	1432
no annotation assigned to feature (blank)	3492	3711
Total features	6174	6243
metabolite with some method of identification: m/z, RT, MS/MS (used for heatmaps and statistics)	136	267

Table 1. Green: number of features with an automated annotation assigned (used for sample ordinations), an ambiguous annotation assigned, or no annotation assigned totaling to number seen in "Total features". Blue: number of metabolites compared between samples for $\log_2$ (fold change), determined based on annotation and LC-MS identification (m/z, RT, MS/MS used to assign confidence; these metabolites were used for heatmaps and $\log_2$ abundance comparisons).

We hoped to examine the differences between samples in order to determine the degree to which our light/dark exposure treatment affected the overall pink berry endo- and exometabolomes. Since we also used two isotopes, ^{12}C and ^{13}C, which underwent no pre-incubation and a pre-incubation, respectively, we determined whether time before individual sub-incubation affected the overall metabolome space. Our ^{13}C-stable isotope assisted metabolomic approach was meant to be supplemented by our ^{12}C incubations in case of potential complications, however, both were integrated into our analysis. Our ^{13}C incubations had a 2.5 day pre-incubation before continuous lighting exposure whereas ^{12}C incubations were immediately exposed to continuous lighting for incubation. This could potentially explain the greater drift we see in our ^{12}C samples with ordination, which eventually seem to stabilize at the point where most of our ^{13}C samples cluster together (Supplemental Figure D). While we only sampled our ^{12}C incubations at 0 and 24 hours we additionally sampled our ^{13}C incubations at 12 and 48 hours.

When accounting for all ^{13}C labelled samples, we found a significant separation due to lighting exposure in both the endometabolome and the exometabolome (Figure 6). Of note, the ^{13}C batch pre-incubations' exometabolite time point zero consisted of filtering a far greater amount of water than the individual incubations through a column (500mL v 30mL), which led to outliers. These outliers were removed for ordinations in this section in order to help parse out patterns in our individual incubations.

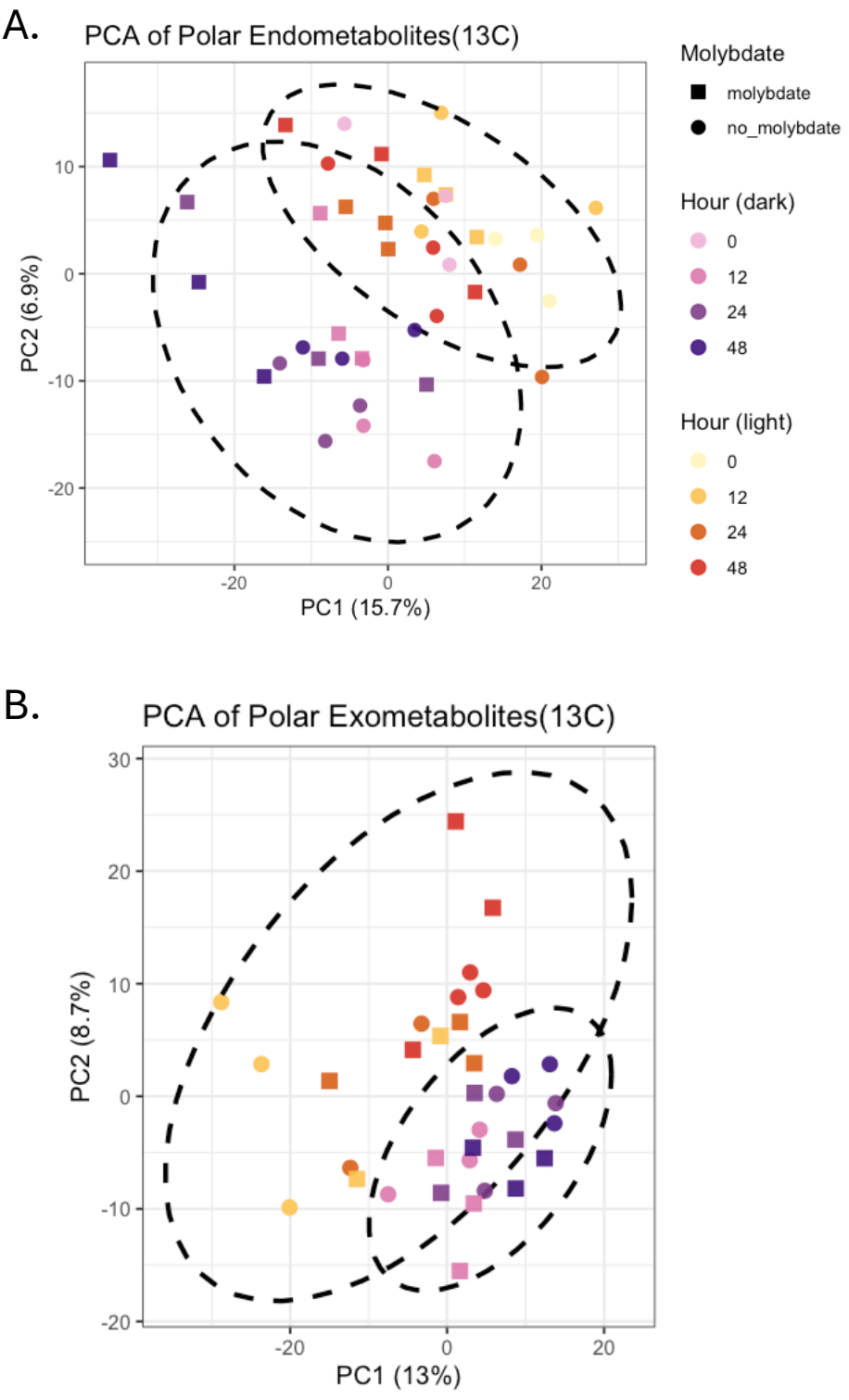
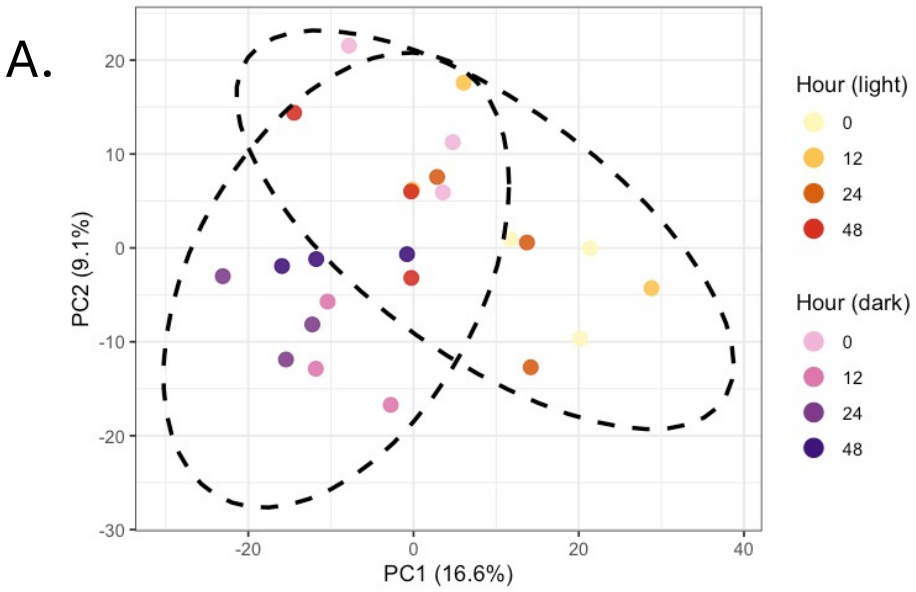


Figure 6. Full breadth of ^{13}C incubated metabolite samples are affected by light-dark exposure in (A) the endometabolome and (B) the exometabolome (outliers were removed for the exometabolome). PCAs using Euclidean distance where 95% confidence ellipses are based on incubation light v. dark conditions, and

corresponding light-dark condition PERMANOVA results are (A) F=3.7257, R^2=0.08521, p=0.001. (B) F=3.0425, R^2=0.08682, p=0.001. Red tones correspond to continuous light exposure and purple tones correspond to continuous dark exposure. Stronger shades indicate longer incubation time.

We further examined just native condition samples and still observed significant differences due to light exposure (Figure 7). This change in chemical space due to lighting may potentially be driven by photosynthate produced by the dominant photosynthetic organism in the pink berry, PB-PSB1.



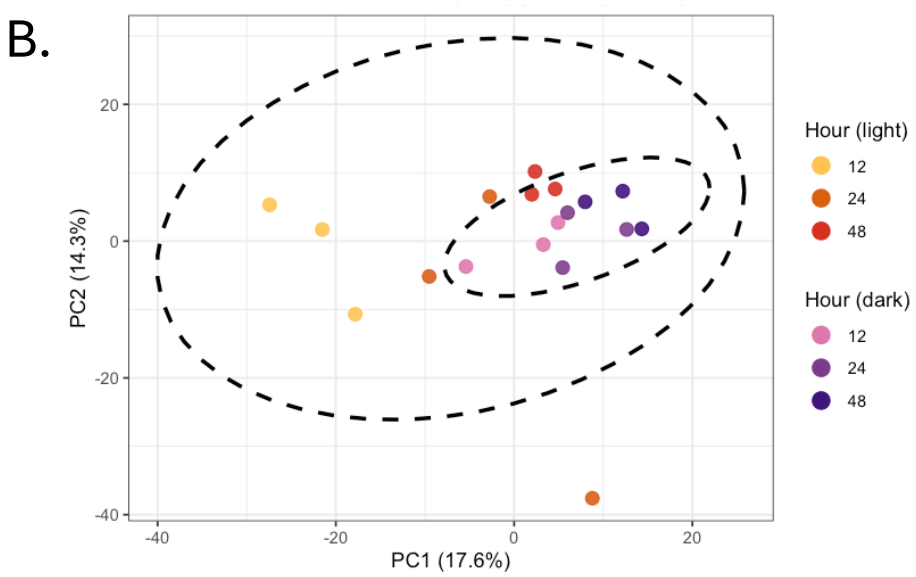


Figure 7. Full breadth of ^{13}C incubated metabolite samples without molybdate exposure for (A) the endometabolome and (B) the exometabolome. PCAs using Euclidean distance where 95% confidence ellipses are based on incubation light v. dark conditions, and corresponding light-dark condition PERMANOVA results are (A) F=2.531, R^2=0.10318, p=0.001. (B) F=2.1724, R^2=0.11955, p=0.001. Red tones correspond to continuous light exposure and purple tones correspond to continuous dark exposure. Stronger shades indicate longer incubation time.

cis-7-hexadecenoic acid , eicosapentaenoic acid, and 1,2-dihydroxyheptadec-16-yn-4-one may be biomarkers for PB-PSB1 autotrophic activity

We hoped to determine what metabolites may be involved in these shifts due to light in the overall ordination space (Figures 6 and 7), as they could denote patterns in photosynthate that may be produced in PB-PSB1 and potentially consumed by PB-SRB1. In order to focus on which metabolites to analyze from the large global patterns we observed, we further looked for significant $\log_2$ (fold change) between light-dark exposure in metabolites from the same time point-isotope combination. We found two exometabolites to be significantly more abundant in light conditions than in dark conditions: eicosapentaenoic acid and 1,2-dihydroxyheptadec-16-yn-4-one (Supplemental Figure E). Earlier time points for these

metabolites had a higher abundance under light conditions as the ^{13}C time points zero and one showed strong $\log_2$ (fold change), however only time point one was determined to be significant, likely due to time point zero only having one replicate. Eicosapentaenoic acid and 1,2-dihydroxyheptadec-16-yn-4-one, therefore, seem to display greater accumulation at the beginning of light exposed incubations than in the dark (Figure 8).

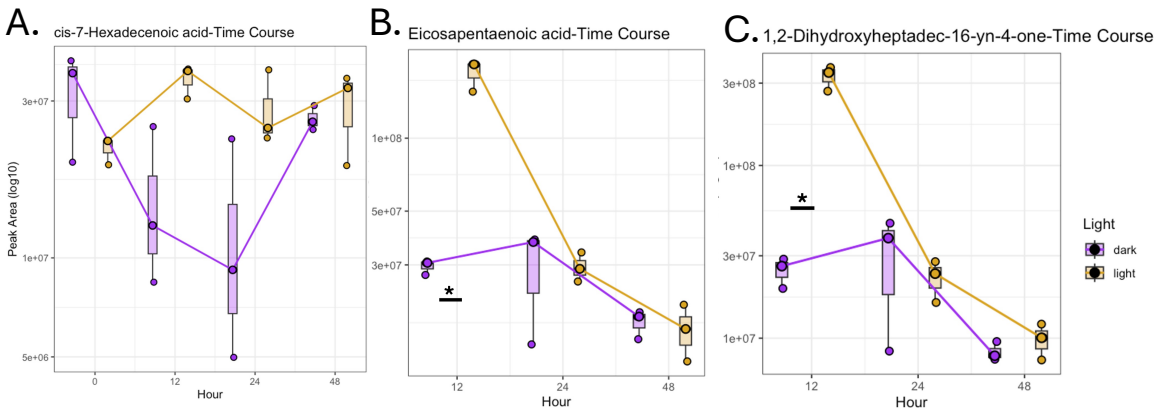


Figure 8. Metabolite peak area for metabolites potentially marking autotrophy high in some light conditions. Time course in ^{13}C experiment across light/dark exposure conditions. Light conditions denoted by gold and dark by purple. Circles denote no molybdate exposure and are connected by solid lines. Asterisks denote a significant $\log_2$ (fold change) between light-dark samples. Metabolites are (A) cis-7-hexadecenenoic acid, (B) eicosapentaenoic acid, (C) 1,2-dihydroxyheptadec-16-yn-4-one.

Eicosapentaenoic acid was consistently higher in our earlier time points across both our ^{12}C and ^{13}C experiments, then subsequently underwent a sustained decrease in our ^{13}C time course. This pattern was mostly similar for 1,2-dihydroxyheptadec-16-yn-4-one from our observed $\log_2$ (fold change) intensities such that we also observed a sustained decrease in our no molybdate light treatment. Furthermore, while not significant, we observed higher amounts of cis-7-hexadecenoic acid in light conditions than dark conditions throughout several of our time course samples.

Studies with labelled bicarbonate have shown ^{13}C incorporation into fatty acids and has been linked to chemolithoautotrophic activity (Knief et al. 2003a). Specifically, studies with *Thiobacillus* have resulted in strong incorporation of label into cis-7-hexadecenoic acid. The higher amounts of cis-7-hexadecenoic acid that we are seeing in several of our light than in our dark conditions may similarly be indicative autotrophic activity within the pink berry, potentially PB-PSB1.

While eicosapentaenoic acid has itself been thought to be a biomarker for diatoms, it may actually be produced by bacteria as a study has shown this to be produced by *Gammaproteobacteria* isolates from intertidal flats (Freese et al. 2009). Furthermore, this study also found *Gammaproteobacteria* to be a potential source of eicosapentaenoic acid in sediment deeper than where eukaryotic diatoms were found. Eicosapentaenoic acid, in addition, has been found to compose 5% of the fatty acid pool from a *Thiohalocapsa* dominated mat (Hubas et al. 2017). PB-PSB1 therefore could potentially be a reasonable source for the eicosapentaenoic acid seen in our exometabolome.

We could not find any literature on the production of our other exometabolite, 1,2-dihydroxyheptadec-16-yn-4-one, by photosynthetic bacteria. 1,2-dihydroxyheptadec-16-yn-4-one, interestingly, is a terminal alkyne, which are widespread in cyanobacteria and marine mollusks and are made by hybrid polyketide synthase/nonribosome peptide synthases (Chai et al. 2016). While not directly linked to purple sulfur bacteria, we would like to pursue whether this could be a marker for purple sulfur bacteria and their photosynthetic activity.

Eicoapentaenoic acid, 1,2-dihydroxyheptadec-16-yn-4-one, and cis-7-hexadecnoic acid may potentially be markers for autotrophic activity. If eicoapentaenoic acid and 1,2-dihydroxyheptadec-16-yn-4-one are derived from some kind of photosynthetic pigment, their

initial higher presence in light than dark conditions may be reflecting photosynthetic activity. The subsequence decrease, however, may be the result of the continuous light exposure to which we subjected our organisms. Bacteriochlorophyll is known to photo-oxidize and degrade, as has been shown with *Cereibacter sphaeroides* (previously *Rhodobacter*), a purple non-sulfur bacterium (Tandori et al. 2005). Ultimately the similarity of these compounds to other unsaturated fatty acids which are carotenoid precursors as well as the relationships of these compounds to photosynthesis lead us to posit the potential for these compounds as markers of photosynthetic activity.

Metabolomic evidence for nighttime oxidation of glycogen to polyhydroxyalkanoates

In order to find other potential links to photosynthesis we looked at diel patterns of other reduced carbon compounds with sugar-type structure. We examined endometabolites that exhibited a significant difference between light-dark conditions at the same time point (Supplemental Figure F). We found a dark decrease in sucrose, glucosamine, and 5-hydroxymethyl-2-furaldehyde relative to light conditions alongside a dark increase in 2-hydroxybutyrate, 3-hydroxybutyrate, and 3-hydroxyvalerate relative to light conditions (Figure 9). It should be noted that we could only identify our glucosamine annotation with low confidence as we only had m/z to make this identification.

The higher presence of sucrose and glucosamine in the light than in dark conditions could indicate some kind of dark decrease of sugar type subunits. While glucosamine did now show a significant difference throughout its time course, it was identified with low confidence and could be an indicator of some other carbohydrate type compound. 5-hydroxymethyl-2-furaldehyde has been found to be a result of the thermal decomposition of glucose and fructose (Zirbes et al. 2013), as well as by the dehydration of glucose (Tao et al.

2014). While these processes seem mostly abiotic, some portion of our extraction and LC/MS workflow could have reasonably resulted in the formation of 5-hydroxymethyl-2-furaldehyde and thus explain its showing similar patterns to other carbohydrates in our endometabolome. As glucose is a component subunit of glycogen, and our transcriptome of PB-PSB1 showed nighttime upregulation of a glycogen phosphorylase which would break up glycogen into its component subunits (PSB__G5402); this dark increase in sucrose, glucosamine, and 5-hydroxymethyl-2-furaldehyde might be indicating glycogen breakdown with prolonged dark exposure.

2-hydroxybutyrate, 3-hydroxybutryate, and 3-hydoxyvalerate all displayed a higher presence in the dark than in our light conditions, and are potentially the reason we see a lower presence of sucrose and glucosamine in the dark than in the light. Daytime carbon fixation and generation of glycogen could be followed by its incomplete oxidation resulting in the production of polyhydroxyalkanoates, leading to energy generation alongside carbon storage. As hydroxybutyrate and hydroxyvalerate are components of polyhydroxyalkanoates (Higuchi-Takeuchi et al. 2016), our results could reflect their nighttime synthesis. Furthermore, these polyhydroxyalkanoate monomers were labelled in ^{13}C and thus presumably originated from PB-PSB1 photosynthate.

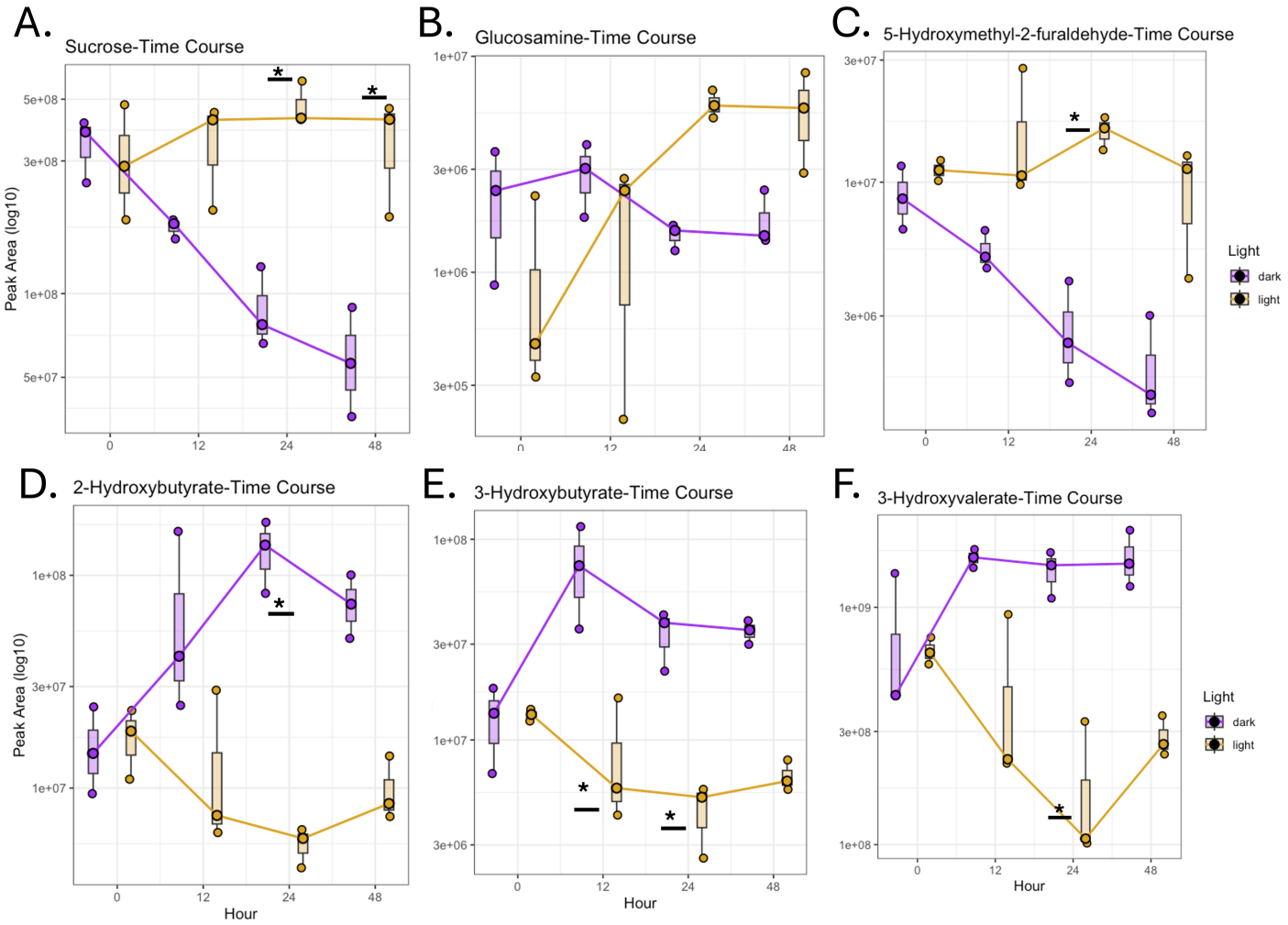


Figure 9. Glycogen decrease and corresponding PHA monomer increase in the dark. Metabolite peak area for metabolites involved in glycogen breakdown over time course in ^{13}C experiment across light/dark exposure conditions. Light conditions denoted by gold and dark by purple. Circles denote no molybdate exposure and are connected by solid lines. Asterisks denote a significant log_2 fold change between light-dark samples. Metabolites are (A) sucrose, (B) glucosamine, (C) 5-hydroxymethyl-2-furaldehyde, (D) 2-hydroxybutyrate, (E) 3-hydroxybutyrate, and (F) 3-hydroxyvalertae.

1,2,4-benzenetricarboxlic acid may be a product of PB-SRB1 metabolism

Metabolite abundance in response to light-dark exposure allowed us to glean insight into autotrophic activity as well as potential photosynthate fate in the pink berry consortium. We hoped examining differences due to molybdate exposure may help us detect specific PB-SRB1 activity. As explored earlier, the transcriptome for our sulfate reducers seems to have been affected by molybdate (Figures 4B, 5), thus any metabolites that are significantly different due to molybdate exposure sparked our interest. While no exometabolite displayed a significant difference due to molybdate exposure, we found several endometabolites to display a significant difference due to molybdate across our ^{12}C and ^{13}C experiments (Supplemental Figure G). Prolylglycine, N,N-dimethylglycine, and beta-glycerophosphate were all higher in our molybdate treated 24hour-^{12}C dark time point, however this pattern was inconsistent across our other sample comparisons. In contrast, 1,2,4-benzenetricarboxylic acid displayed a consistent $\log_2$ (fold change) in our molybdate-no molybdate comparisons.

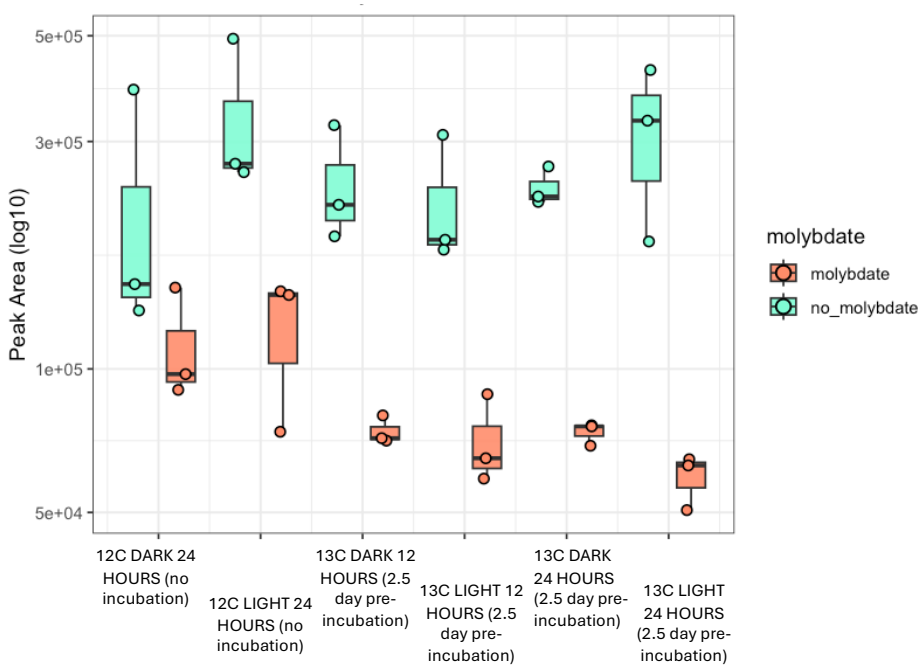
This consistent $\log_2$ (fold change) resulted in a pattern where peak area was lower for our molybdate conditions irrespective of light/dark exposure (Figure 10). Since peak areas were lower due to molybdate exposure, this could potentially indicate a need for the presence of a sulfate reducer to result in the higher presence seen in our no molybdate conditions. 1,2,4,-benzentricarboxylic acid could thus somehow be a product of PB-SRB1 metabolism. When PB-SRB1 is exposed to molybdate, the cell shutdown we potentially saw in the transcriptome would lead to the lower metabolite peak area we observed. To explore this possibility we examined genes expressed in our non-molybdate exposed conditions for genes involved in

benzene or tricarboxylic acid metabolism. We found genes encoding an ethylbenzene dehydrogenase related protein (SRB__G1266) as well as a Bug family tripartite tricarboxylate transporter (SRB_G1396) within the genome of PB-SRB1 as well as its expression in the native PB-SRB1 transcriptome (Supplemental Figures H and I).

Ethylbenzene dehydrogenase itself catalyzes the ferricenium-dependent oxidation of ethylbenzene to (S)-1-phenylethanol, first characterized from a denitrifiying *Azoarcus* strain EbN1 (Kniemeyer and Heider 2001). While the mechanism for 1,2,4-benzentricarboxylic acid production is unclear, PB-SRB1 is normally able to transcribe a gene related to ethylbenzene dehydrogenase and could thus be interacting with a benzene type compound in some way. The Bug family of tricarboxylate transporters, furthermore, can transport a wide cadre of compounds, thus we hoped to determine whether it would be possible to transport 1,2,4-benzentricarboxylic acid. While there are no reports of a tricarboxylated benzene type compound being transported by a Bug family member, TphC, which is itself a part of the Bug family of tricarboxylate transporters, is able to transport terephthalate, a dicarboxylated benzene (Gautom et al. 2021). This is an example of benzene type compound being transported by this family and could be the means by which PB-SRB1 imports a 1,2,4-benzentricarboxylic acid precursor from PB-PSB1, or exports its waste. Sulfate reducing bacteria are capable of using a wide cadre of compounds as a source of electrons, such that multiple studies show various other sulfate reducing bacteria being able to use benzene as an electron donor to power sulfate reduction (Keller et al. 2018; Phelps et al. 1998). It is certainly possible, therefore, for PB-SRB1 to be using some type of benzene compound as an electron donor, producing 1,2,4-benzentricarboxylic acid as a result. When we looked for labelling of 1,2,4-benzentricarboxylic acid, however, we found no ^{13}C label, which indicates

either lack of an origin from PB-PSB1 or turnover times that do not fall within detection for the span of our incubation.

A.



B.

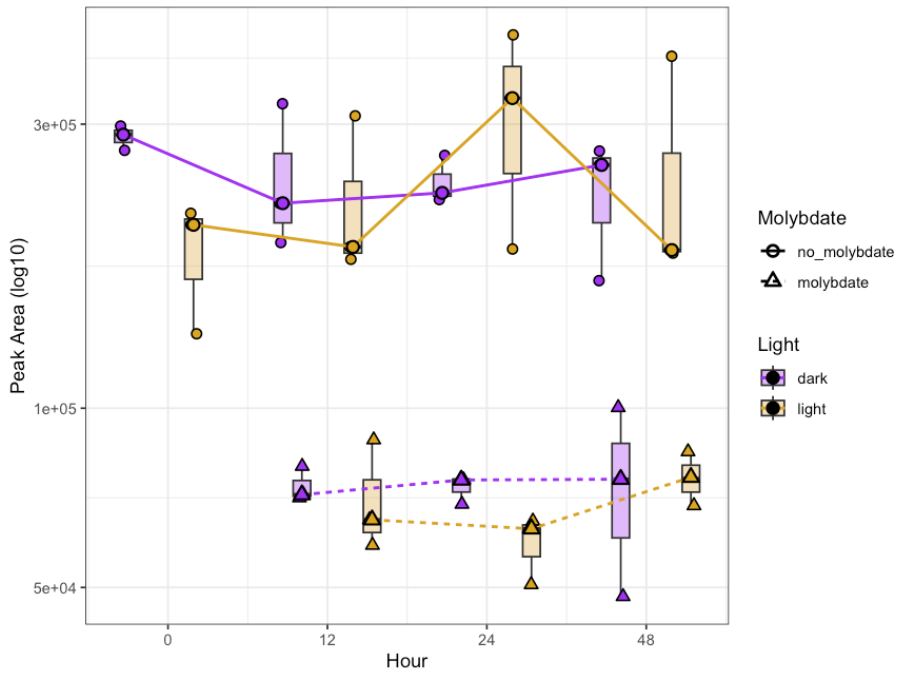


Figure 10. Higher 1,2,4-benzentricarboxylic acid peak area across all our native conditions that were never exposed to molybdate. (A) 1,2,4-benzenetricarboxylic acid log_{10}-scaled peak area comparing molybdate/no molybdate conditions for our early time point individual-incubations for both our ^{12}C and ^{13}C experiments. Molybdate exposure denoted by salmon and no molybdate by teal. (B) 1,2,4-benzentricarboxylic acid peak area for time course in ^{13}C experiment across both molybdate and light/dark exposure conditions. Light conditions denoted by gold and dark by purple. Circles denote no molybdate exposure and are connected by solid lines, triangles denote molybdate exposure and are connected by dashed lines. Asterisks denote a significant log_2 (fold change) between molybdate-no molybdate samples.

Potential for use of small organosulfonates by PB-SRB1

We looked further into other potential differences in the pink berry metabolome due to molybdate exposure. While not statistically significant, our PLS-DA analysis did determine methanesulfonic acid to be an important contributor to the chemical space. We thus examined its abundance patterns over our incubation and found a potential build up of methanesulfonic acid in the light with molybdate exposure (Figure 11). This could potentially indicate direct use of methanesulfonic acid by PB-SRB1 under native conditions, such that upon molybdate exposure methanesulfonic acid use was prevented and thus led to its buildup in our sample. There is also the possibility, however, that methanesulfonic acid may have been a photodegradation product of dissolved organic sulfur as this "build up" was not observed in the dark (Ossola et al. 2019). This would mean that PB-SRB1 typically uses some dissolved organic sulfur and, upon molybdate exposure, this substrate was not used and thus photodegraded to methanesulfonic acid with light exposure.

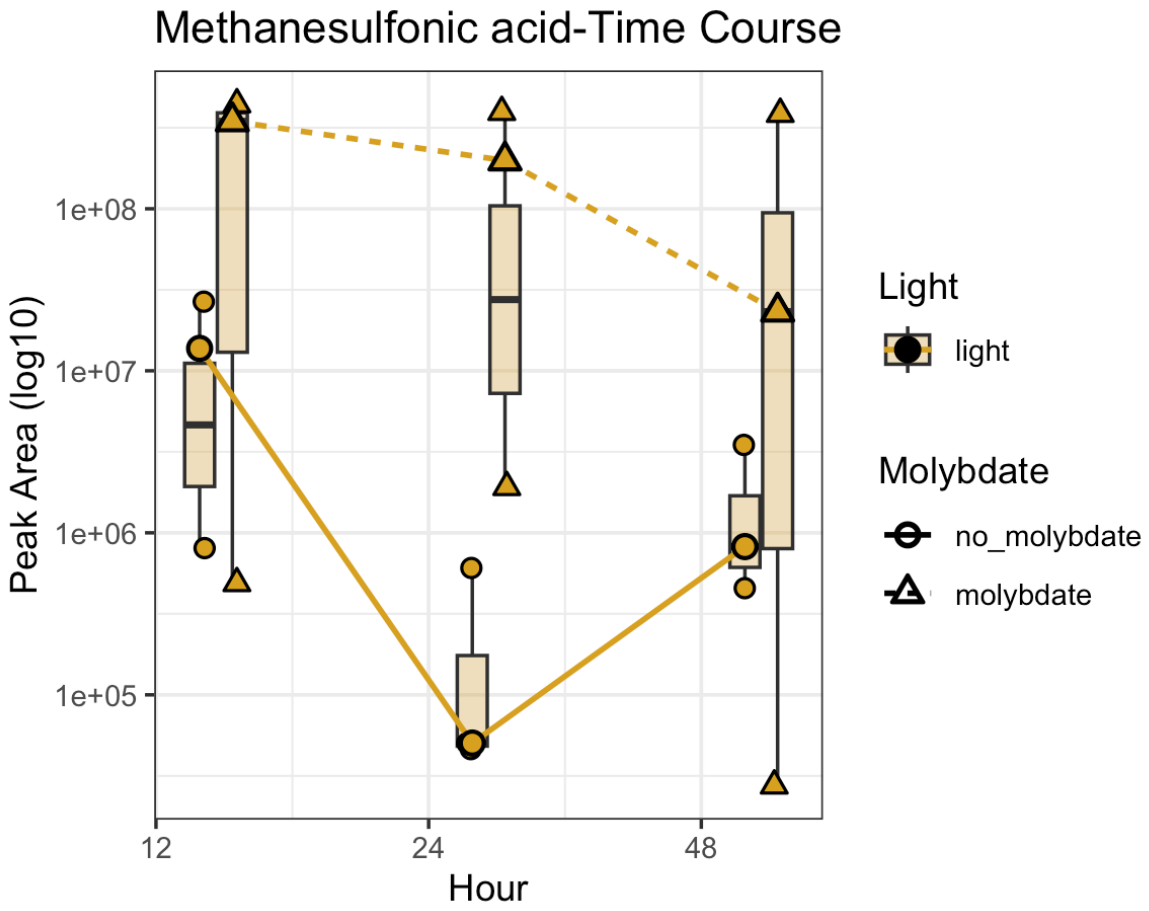


Figure 11. Methanesulfonic acid peak area under light exposed molybdate condition. Time course in ^{13}C experiment across both molybdate and light/dark exposure conditions. Light conditions denoted by gold and dark by purple. Circles denote no molybdate exposure and are connected by solid lines, triangles denote molybdate exposure and are connected by dashed lines.

Dissolved organosulfonates, furthermore, have been shown to be an important substrate for sulfate reducers in marine microbial mats (Visscher et al. 1999). We examined the potential for use dissolved organic sulfur by PB-SRB1. While the potential dissolved organic sulfur compounds that could be used are widely varied (Cook and Denger 2002), we specifically decided to look at taurine use, and looked for genes used by *Desulfovibrionaceae* (Ye et al. 2023). Taurine dependent aminotransferase and pyruvate dependent alanine dehydrogenase work together to produce sulfoacetaldehyde from taurine (Figure 12). This is

then the substrate for production of sulfite via sulfoacetaldehyde acetyltransferase. When we looked at native expression of PB-SRB1 we found expression of both an aminotransferase family protein encoding gene (SRB__G640) as well as sulfoacetaldehyde acetyltransferase (SRB__G325) in most of our samples, but we rarely found either of these genes expressed in our molybdate exposed samples. Furthermore, we found methanesulfonic acid to be labelled with ^{13}C meaning methanesulfonic acid originated from PB-PSB1 photosynthate. Together this led us to posit that PB-SRB1 may potentially be engaging in dissimilatory catabolism of organosulfonates.

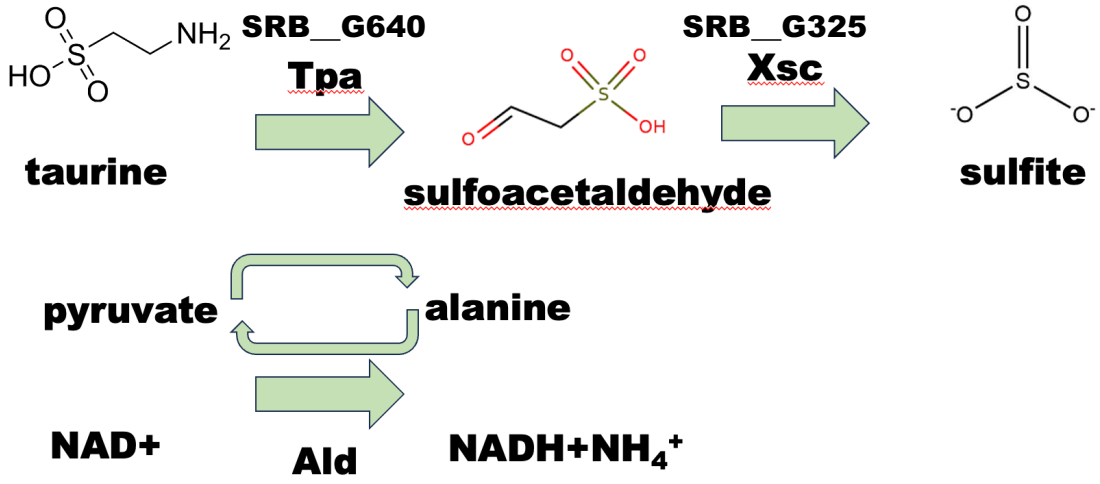


Figure 12. Taurine use in *Taurinivorans muris* strain LT0009, adapted from Ye *et al*. 2023. Tpa is taurine dependent aminotransferase. Xsc is sulfoacetaldehyde acetyltransferase. Ald is pyruvate dependent alanine dehydrogenase.

Conclusions

We hoped to probe the potential for a carbon sharing relationship between PB-PSB1 and PB-SRB1 by intentionally uncoupling their relationship via use of a sulfate reduction inhibitor. We hoped this exposure to molybdate would be specific to sulfate reducing bacteria in our samples and found that exposure to molybdate changed the transcriptomic space in PB-SRB1, such that 16.36% of its genes were significantly different due to molybdate exposure. When comparing median TPM between native and molybdate exposed conditions, we found counts in molybdate exposed samples to be significantly lower. This difference seemed not to affect PB-PSB1. This would indicate that our molybdate treatment was specific to sulfate reducers and potentially indicated shutdown of SRB cells.

We complemented our gene expression measurements with metabolomic sampling. We found that metabolites from within the pink berry biofilm as well as its exudates were significantly different between light and dark exposure conditions. As *Thiohalocapsa* PB-PSB1 is the major photosynthetic organism in the biofilm, we conclude that this light-dark difference is due to the photosynthetic capabilities of PB-PSB1 changing the overall chemical space in the pink berry biofilm. We first explored how this metabolomic space may give us insight into light-dark physiology of pink berries and identity of photosynthate. The endometabolome indicated that in the dark glycogen reserves were incompletely oxidized for maintenance energy and salvaged as polyhydroxyalkanoate inclusions. Data supporting glycogen breakdown included the dark decrease in sucrose, glucosamine, and 5-hydroxymethyl-2-furaldehyde (a chemical byproduct of glucose and fructose breakdown). We also saw a corresponding increase in polyhydroxyalkanoate monomers like 2-hydroxybutyrate, 3-hydroxybutyrate, and 3-hydroxyvalerate, indicative of nighttime

oxidation of glycogen to polyhydroxyalkanoates, such that the cell is generating ATP but still storing some of its carbon. Labelling data from polyhydroxyalkanoate monomers helped us conclude origin from PB-PSB1. We further identified two exometabolites which were significantly greater at earlier time points in our incubations when exposed to light than when exposed to dark conditions. These exometabolites were eicosapentaenoic acid, which has been found in *Gammaproteobacteria* isolates from intertidal flats, and 1,2-dihydroxyheptadec-16-yn-4-one which had a terminal alkyne end, a moiety which has been found widespread in cyanobacteria. We also found cis-7-hexadecenoic acid to be greater in some light conditions in our incubations, and this metabolite has been found to be a marker for active autotrophy in *Thiobacillus* (Knief et al. 2003b). The likeness to other unsaturated fatty acids which are precursors to carotenoids as well as these links to photosynthesis led us to believe these compounds may be markers for autotrophic activity.

We then hoped to glean insight into what carbon could be shared to PB-SRB1, by examining differences in molybdate exposed conditions. We found that *Desulfofustis* PB-SRB1 oxidizes a compound with an aromatic moiety in order to produce 1,2,4-benzentricarboxylic acid; however labelling data indicated that this compound is likely not a high-turnover metabolite produced by *Thiohalocapsa* sp. PB-PSB1. Furthermore, *Desulfofustis* PB-SRB1 likely produces sulfide from dissimilatory catabolism of organosulfonates exuded by *Thiohalocapsa* PB-PSB1. We detected a buildup of methanesulfonic acid in light-exposed molybdate conditions; this compound could either be utilized directly by PB-SRB1 under native conditions, or may result from the photodegradation of other sulfonates (taurine, DMS). We explored potential use of PB-SRB1 of a dissolved organic sulfur compound, taurine, and found native expression of two genes

needed for its use: taurine dependent aminotransferase and sulfoacetaldehyde acetyltransferase. Furthermore, methanesulfonic acid was ^{13}C labelled and thus likely originated from PB-PSB1 photosynthate.

References

Banat, I. M., D. B. Nedwell, and M. T. Balba. 1983. “Stimulation of Methanogenesis by Slurries of Saltmarsh Sediment after the Addition of Molybdate to Inhibit Sulphate-Reducing Bacteria.” *Microbiology* 129 (1): 123–29. https://doi.org/10.1099/00221287-129-1-123.

Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. “Near-Optimal Probabilistic RNA-Seq Quantification.” *Nature Biotechnology* 34 (5): 525–27. https://doi.org/10.1038/nbt.3519.

Bushnell, Brian. 2014. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Released. https://bbmap.org.

Chai, Qiu-Ye, Zhen Yang, Hou-Wen Lin, and Bing-Nan Han. 2016. “Alkynyl-Containing Peptides of Marine Origin: A Review.” *Marine Drugs* 14 (11): 216. https://doi.org/10.3390/md14110216.

Cook, Alasdair M., and Karin Denger. 2002. “Dissimilation of the C2 Sulfonates.” *Archives of Microbiology* 179 (1): 1–6. https://doi.org/10.1007/s00203-002-0497-0.

Doré, Hugo, Amy R. Eisenberg, Emily N. Junkins, et al. 2024. “Targeted Hypermutation of Putative Antigen Sensors in Multicellular Bacteria.” *Proceedings of the National Academy of Sciences* 121 (9): e2316469121. https://doi.org/10.1073/pnas.2316469121.

Freese, Elke, Heike Rütters, Jürgen Köster, Jürgen Rullkötter, and Henrik Sass. 2009. “Gammaproteobacteria as a Possible Source of Eicosapentaenoic Acid in Anoxic Intertidal Sediments.” *Microbial Ecology* 57 (3): 444–54. https://doi.org/10.1007/s00248-008-9443-2.

Gautom, Trishnamoni, Dharmendra Dheeman, Colin Levy, et al. 2021. “Structural Basis of Terephthalate Recognition by Solute Binding Protein TphC.” *Nature Communications* 12 (1): 6244. https://doi.org/10.1038/s41467-021-26508-0.

Higuchi-Takeuchi, Mieko, Kumiko Morisaki, Kiminori Toyooka, and Keiji Numata. 2016. “Synthesis of High-Molecular-Weight Polyhydroxyalkanoates by Marine Photosynthetic Purple Bacteria.” *PLOS ONE* 11 (8): e0160981. https://doi.org/10.1371/journal.pone.0160981.

Hubas, Cédric, Dominique Boeuf, Bruno Jesus, Najet Thiney, Yann Bozec, and Christian Jeanthon. 2017. “A Nanoscale Study of Carbon and Nitrogen Fluxes in Mats of Purple Sulfur Bacteria: Implications for Carbon Cycling at the Surface of Coastal Sediments.” *Frontiers in Microbiology* 8 (October): 1995. https://doi.org/10.3389/fmicb.2017.01995.

Isa, M. Hasnain, and G. K. Anderson. 2005. “Molybdate Inhibition of Sulphate Reduction in Two-Phase Anaerobic Digestion.” *Process Biochemistry* 40 (6): 2079–89. https://doi.org/10.1016/j.procbio.2004.07.025.

Keller, Andreas H., Sabine Kleinsteuber, and Carsten Vogt. 2018. “Anaerobic Benzene Mineralization by Nitrate-Reducing and Sulfate-Reducing Microbial Consortia Enriched From the Same Site: Comparison of Community Composition and Degradation Characteristics.” *Microbial Ecology* 75 (4): 941–53. https://doi.org/10.1007/s00248-017-1100-1.

Knief, Claudia, Karlheinz Altendorf, and André Lipski. 2003a. “Linking Autotrophic Activity in Environmental Samples with Specific Bacterial Taxa by Detection of^{13} C-labelled Fatty Acids.” *Environmental Microbiology* 5 (11): 1155–67. https://doi.org/10.1046/j.1462-2920.2003.00510.x.

Knief, Claudia, Karlheinz Altendorf, and André Lipski. 2003b. “Linking Autotrophic Activity in Environmental Samples with Specific Bacterial Taxa by Detection of^{13} C-labelled

Fatty Acids." *Environmental Microbiology* 5 (11): 1155–67. https://doi.org/10.1046/j.1462-2920.2003.00510.x.

Kniemeyer, Olaf, and Johann Heider. 2001. "Ethylbenzene Dehydrogenase, a Novel Hydrocarbon-Oxidizing Molybdenum/Iron-Sulfur/Heme Enzyme." *Journal of Biological Chemistry* 276 (24): 21381–86. https://doi.org/10.1074/jbc.M101679200.

Kolde, Raivo. n.d. *Pheatmap: Pretty Heatmaps*. V. R package version 1.0.13. https://github.com/raivokolde/pheatmap.

Lex, Alexander, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. 2014. "UpSet: Visualization of Intersecting Sets." *IEEE Transactions on Visualization and Computer Graphics* 20 (12): 1983–92. https://doi.org/10.1109/TVCG.2014.2346248.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

Oremland, Ronald S., and Douglas G. Capone. 1988. "Use of 'Specific' Inhibitors in Biogeochemistry and Microbial Ecology." In *Advances in Microbial Ecology*, vol. 10.

Ossola, Rachele, Julie Tolu, Baptiste Clerc, Paul R. Erickson, Lenny H. E. Winkel, and Kristopher McNeill. 2019. "Photochemical Production of Sulfate and Methanesulfonic Acid from Dissolved Organic Sulfur." *Environmental Science & Technology* 53 (22): 13191–200. https://doi.org/10.1021/acs.est.9b04721.

Phelps, Craig D., Lee J. Kerkhof, and Lily Y. Young. 1998. "Molecular Characterization of a Sulfate-Reducing Consortium Which Mineralizes Benzene." *FEMS Microbiology Ecology* 27 (3): 269–79. https://doi.org/10.1111/j.1574-6941.1998.tb00543.x.

Ranade, D. 1999. “Evaluation of the Use of Sodium Molybdate to Inhibit Sulphate Reduction during Anaerobic Digestion of Distillery Waste.” *Bioresource Technology* 68 (3): 287–91. https://doi.org/10.1016/S0960-8524(98)00149-7.

Shibl, Ahmed A., Ashley Isaac, Michael A. Ochsenkühn, et al. 2020. “Diatom Modulation of Select Bacteria through Use of Two Unique Secondary Metabolites.” *Proceedings of the National Academy of Sciences* 117 (44): 27445–55. https://doi.org/10.1073/pnas.2012088117.

Sievert, Carson. n.d. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny*. Chapman and Hall/CRC. https://plotly-r.com.

Tandori, Júlia, Zsolt Tokaji, Katalin Misurda, and Péter Maróti. 2005. “Thermodynamics of Light-induced and Thermal Degradation of Bacteriochlorins in Reaction Center Protein of Photosynthetic Bacteria.” *Photochemistry and Photobiology* 81 (6): 1518–25. https://doi.org/10.1562/2005-05-31-RA-556.

Tao, Fu-Rong, Chen Zhuang, Yue-Zhi Cui, and Jing Xu. 2014. “Dehydration of Glucose into 5-Hydroxymethylfurfural in SO3H-Functionalized Ionic Liquids.” *Chinese Chemical Letters* 25 (5): 757–61. https://doi.org/10.1016/j.cclet.2014.01.044.

Visscher, Pieter T., Rachel F. Gritzer, and Edward R. Leadbetter. 1999. “Low-Molecular-Weight Sulfonates, a Major Substrate for Sulfate Reducers in Marine Microbial Mats.” *Applied and Environmental Microbiology* 65 (8): 3272–78. https://doi.org/10.1128/AEM.65.8.3272-3278.1999.

Wickham, Hadley. 2009. *Ggplot2: Elegant Graphics for Data Analysis*. Springer New York. https://doi.org/10.1007/978-0-387-98141-3.

Wilbanks, Elizabeth G., Hugo Doré, Meredith H. Ashby, Cheryl Heiner, Richard J. Roberts, and Jonathan A. Eisen. 2022. “Metagenomic Methylation Patterns Resolve Bacterial Genomes of Unusual Size and Structural Complexity.” *The ISME Journal* 16 (8): 1921–31. https://doi.org/10.1038/s41396-022-01242-7.

Wilbanks, Elizabeth G., Ulrike Jaekel, Verena Salman, et al. 2014. “Microscale Sulfur Cycling in the Phototrophic Pink Berry Consortia of the S Ippewissett S Alt M Arsh.” *Environmental Microbiology* 16 (11): 3398–415. https://doi.org/10.1111/1462-2920.12388.

Ye, Huimin, Sabrina Borusak, Claudia Eberl, et al. 2023. “Ecophysiology and Interactions of a Taurine-Respiring Bacterium in the Mouse Gut.” *Nature Communications* 14 (1): 5533. https://doi.org/10.1038/s41467-023-41008-z.

Zhang, Jun, Hongyuan Li, Wenjun Tao, and Jun Zhou. 2026. “1 ClusterGVis: An Advanced Visualization and Clustering Tool for 2 Gene Expression Analysis.” *Genomics, Proteomics, and Bioinformatics*, ahead of print, January 22. https://doi.org/https://doi.org/10.1093/gpbjnl/qzag005.

Zirbes, Lara, Bach Kim Nguyen, Dirk C. De Graaf, et al. 2013. “Hydroxymethylfurfural: A Possible Emergent Cause of Honey Bee Mortality?” *Journal of Agricultural and Food Chemistry* 61 (49): 11865–70. https://doi.org/10.1021/jf403280n.

Chapter IV: Similarities in *Thiohalocapsa* PB-PSB1 and *Desulfofustis* PB-SRB1 metabolic capabilities in both lab and environmental conditions despite oxygen exposure

Braulio S. Castillo Villaseñor, Anna E. Warsaw, Emily N. Junkins, and Elizabeth G. Wilbanks

Abstract

Tightly coupled interactions are hard to dissect, particularly so in naturally occurring microbial communities. Our previous studies have looked at syntrophs within the pink berry aggregate, and examined differences in interactions due to either light or molybdate exposure. In this chapter we hoped to dissect the coupled interactions between PB-PSB1 and PB-SRB1 by probing differences in physiology due to oxygen exposure. To do this we sampled pink berry aggregates directly in the field as daytime sampled aggregates are actively exposed to oxygen, whereas, nighttime exposed berries would encounter anoxic conditions such as those in our lab incubations (chapters two and three). By examining the effects of oxygen exposure in the pink berry biofilm we hoped to determine whether PB-PSB1 engaged in chemoautotrophic activity and thus potentially changed the identity of reduced carbon available to PB-SRB1. We found PB-PSB1 to engage in many similar processes to our non-oxygen exposed lab incubations, such that it engaged in daytime vitamin synthesis and nighttime bacteriochlorophyll synthesis. We also found multiple genes

indicating daytime carbon fixation in PB-PSB1. When examining effects due to oxygen exposure we found some signs of oxygen stress in PB-PSB1 that were not found in our lab incubations, however, PB-SRB1displayed similar signs of oxygen stress in both field and lab samples. While we have not been able to determine gene expression denoting chemoautotrophic growth in PB-PSB1 and potential differences in reduced carbon available to PB-SRB1, we found similar expression of 2-oxoacid dehydrogenases in both our lab and field samples such that PB-SRB1 may be oxidizing small carbon compounds despite oxygen exposure.

Introduction

Over the last two chapters we characterized various aspects of the potential for a carbon sharing partnership between PB-PSB1 and PB-SRB1 in the pink berry consortium. These experiments, however, were intentionally rendered anoxic in order to isolate the effects of oxygen on the physiology of both PB-PSB1 and PB-SRB1. Purple sulfur bacteria are in fact metabolically versatile and, when exposed to oxygen, may also be capable of chemolithoautotrophic growth. The corresponding carbon fixation in organisms like *Thiocapsa roseopersicina* is thought to take place via the reductive pentose phosphate pathway (Wit and Gemerden 1987a). The occurrence of chemolithoautrophic growth, therefore, is capable of influencing the nature and timing of the reduced organic carbon available to the cell.

We thus also wanted to measure the potential effects of oxygen exposure to pink berry member physiology. In order to understand this effect it is necessary to characterize the nature of the salt marsh pond environment our organisms are exposed to. Studies of oxygen and sulfide profiles in the Sippewissett salt marsh have found daytime oxic, but low sulfide conditions at the sediment-water interface. At night, however, the sediment seems to be anoxic and sulfidic (Salman et al. 2015). Older culture studies of *Thiocapsa roseopersicina* have captured similar conditions, such that *T. roseopersicina* was exposed to an oxic light and a dark anoxic regime (De Wit and Van Gemerden 1990a). Under this regime glycogen production went up during the day and down night, mirroring an opposite pattern in bacteriochlorophyll *a* such that bcl *a* went up at night and down during the day. Sulfide saw and immediate drawdown with daytime, and gradually built up again in the dark. Elemental

sulfur, in contrast, underwent an immediate uptick at the beginning of the light period and then slowly started decreasing until it reached a state of constant concentration in the dark.

Typically, if in a fully anoxic environment, such as that in our lab incubations from chapter two, purple sulfur bacteria can photoautotrophically oxidize sulfide in order to fix and store carbon during the day (Witt and Gemerden 1990). In the environment, however, due to the low levels of sulfide at this sediment-water interface, PB-PSB1 would be sulfide and electron donor limited. In the nighttime, furthermore, stored carbon could be heterotrophically oxidized to produce ATP and stored sulfur may get reduced to sulfide (De Wit and Van Gemerden 1990a; Wit and Gemerden 1987b). As this depends on how much carbon has been stored, the cell would be ATP and electron acceptor limited. If oxygen is present, however, the PSB cells can also chemoautotrophically oxidize stored sulfur, and respire oxygen (Nelson et al. 1986; Overmann and Pfennig 1992).

We hoped to understand the potential effect of oxygen on chemoautotrophic growth of PB-PSB1, the major member of the pink berry microbial consortium. We aimed to discern the effects of exposing PB-PSB1 to oxygen would have on other community members as this would potentially also affect the nature and timing of reduced carbon available to other community members. We expected daytime conditions to expose the consortium to oxygen stress and potentially elicit signs of chemoautotrophy. We then hoped to examine whether PB-SRB1 exhibited some corresponding change in its physiology. In order to do this we collected berries sampled directly from the environment and examined significantly differently expressed genes between light-dark samples and compared effects due to light exposure, then we looked for any indication of a response to potential oxygen exposure in either PB-PSB1 or PB-SRB1.

Methods

Environmental Peak Light/Dark Sampling

Berries were sampled from site LS01 (41°34′33.01′N, 70°38′21.24′W) to correspond with peak light and peak dark times (1AM and 1PM) on 28 July 2023 and preserved in RNAlater. Sampling in each case took between 30-45 minutes such that 100mL of berries were repeatedly washed in marsh water to clean out debris. This was then placed in 500mL of RNAlater, and stored over two nights at 4°C to allow RNAlater to penetrate tissue. After the two days supernatant was removed, berries were sorted into groups of 20, then moved to -80°C for long term storage.

RNA Extraction, Sequencing, and Differential Expression Analysis

Berries were extracted with Zymo R2040 kit. Environmental sample extractions consisted of 10 replicates at 2 time points. Each sample consisted of 5 berries extracted together. Samples were treated with DNAse I in 2 hour increments (Zymo DNAse I set, E1011), varying in length from 2 to 6 hours total DNAse I incubation. For QC we quantified RNA with Qubit (at least 200 total ng of RNA, up to 1525 ng in one instance), verified lack of DNA contamination with PCR (if DNA present, sample underwent an additional 2 hour incubation with DNAse I), and RNA quality was checked with the Tape Station (RIN score of at least 6, often around 7-8). If RNA quantity fell below these limits, then an additional 5-berry extraction was pooled to increase concentration.

Samples were sequenced by the Joint Genome Institute under grant 508543. This was sequenced 2x151 using the Illumina NovaSeqX platform, and BBDuk was used to remove various contaminants, remove adapter reads, and right quality trim reads. Reads that mapped via BBMap to human, cat, dog and mouse references as well as common microbial

contaminants were removed; ribosomal reads were also removed (BBtools v39.03, Bushnell 2014). The final filtered fastq was used for metatranscriptome analysis. RNA differential expression analysis was carried out following the same protocol as in chapter 2.

Results and Discussion

Field-sampled transcripts in both PB-PSB1 and PB-SRB1 display a significant difference due to light-dark conditions

We hoped to first determine whether field sampled aggregates still exhibited the same light dark separation we found in our lab-incubated samples. To probe this light-dark separation we analyzed our environmental samples and found a stark light-dark separation for PB-PSB1 transcripts, such that 47.19% (ie 3127 genes significant out of 6627 genes analyzed [count 10 or above across samples]) of analyzed genes were significantly different between all light and all dark samples (Figure 1). Our environmental sampling, however, was only done at peak light/dark conditions and thus done at a much less fine-scale than our diel incubation, so we were not able to capture light-dark transitions at dawn or at dusk. For PB-SRB1, moreover we found similarly clear light-dark separation, and observed that 11.22% (ie 347 genes significant out of 3092 genes analyzed [count 10 or above across samples]) of analyzed genes were significantly different between all light and all dark samples (Figure 2). Whether this response is a direct response to light or some entrained rhythm would need more detailed sampling that captures the light-dark transition seen in the marsh environment.

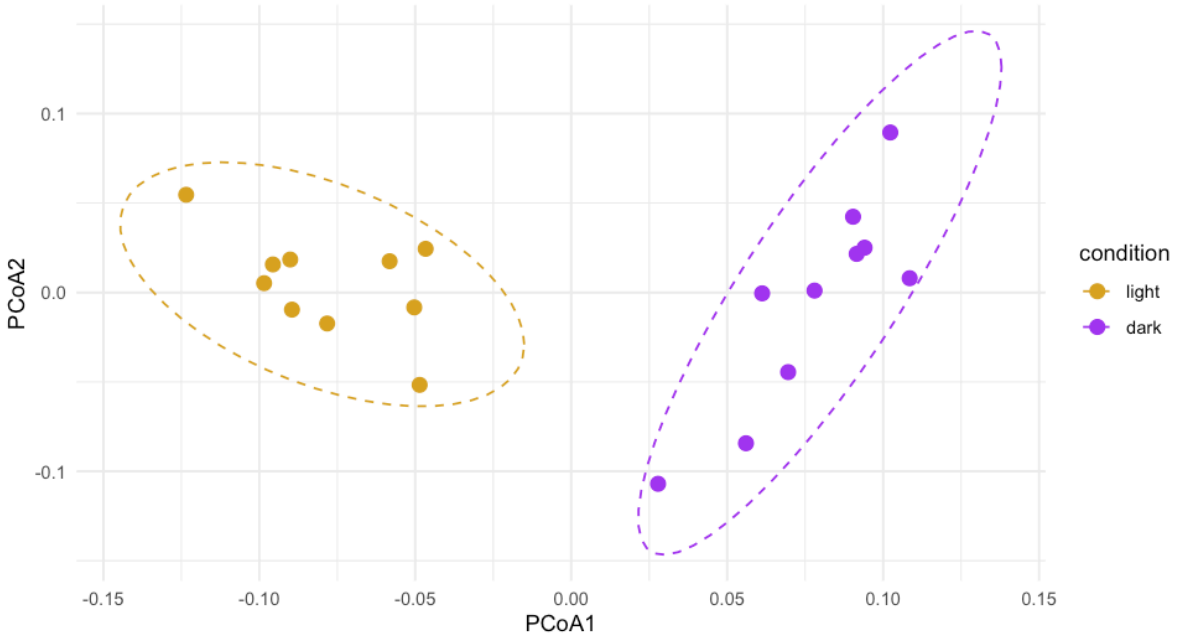


Figure 1. Differences in PB-PSB1 transcripts between light-dark conditions in environmentally sampled berries. PCoA for PB-PSB1 using Bray-Curtis distance where 95% confidence ellipses are based on incubation light v. dark conditions, and corresponding light-dark condition PERMANOVA results are F=17.712, R^2=0.49597, p= 0.001. Gold tones correspond to light time points, and purple tones correspond to dark time points.

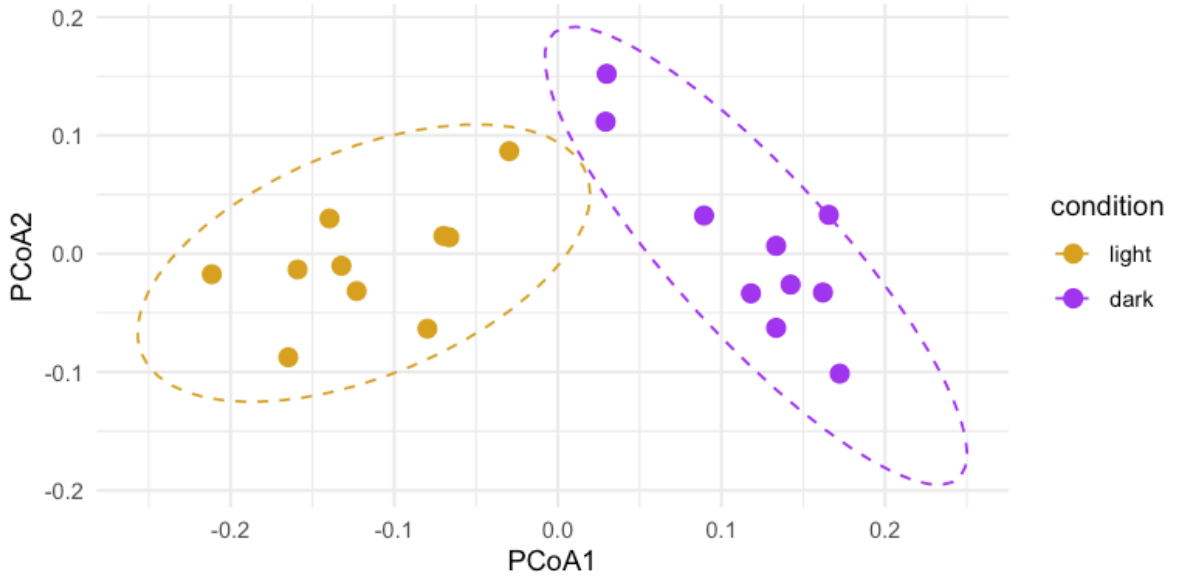


Figure 2. Differences in PB-SRB1 transcripts between light-dark conditions in environmentally sampled berries. PCoA for PB-SRB1 using Bray-Curtis distance where 95% confidence ellipses are based on incubation light v. dark conditions, and corresponding light-dark condition PERMANOVA results are F=7.3127, R^2=0.28889, p= 0.001. Gold tones correspond to light time points, and purple tones correspond to dark time points.

When we further compared the transcripts from our field samples to our lab samples we found samples to cluster more closely together by field versus lab sampling in both PB-PSB1 and PB-SRB1. Specifically, PB-PSB1 transcripts cleanly separated by light exposure, as well as by field versus lab sampling approach (Supplemental Figure A). While PB-SRB1 transcripts cleanly separated by field versus lab sampling approach, they only displayed cleaner separation due to lighting in the environmentally sampled time points (Supplemental Figure B). While time points from earlier on in the lab incubation showed much more overlap between lighting conditions, the PB-SRB1 transcripts from later in the lab incubation showed a more distinct light-dark separation. Together this denoted light-dark separation in both PB-PSB1 and PB-SRB1 across peak light conditions in both field sampled and lab-incubated aggregates.

Diel Cycling Exhibited by PB-PSB1

In order to further characterize the nature of the light-dark separation we saw in our environment samples we clustered significantly different genes with an $|LFC| > 1$ and $p_{adj} < 0.05$. For PB-PSB1, two clusters were upregulated with daylight and downregulated at nighttime. The third cluster exhibited the opposite pattern such that it was upregulated at night and downregulated during the day (Figure 3). We subsequently performed the same clustering as for PB-SRB1 such that we also clustered significantly different genes with an $|LFC| > 1$ and $p_{adj} < 0.05$. We found three clusters such that two clusters were significantly downregulated during the day and upregulated at night (Figure 4). The third cluster displayed the opposite pattern such that it was upregulated during the day and downregulated at night.

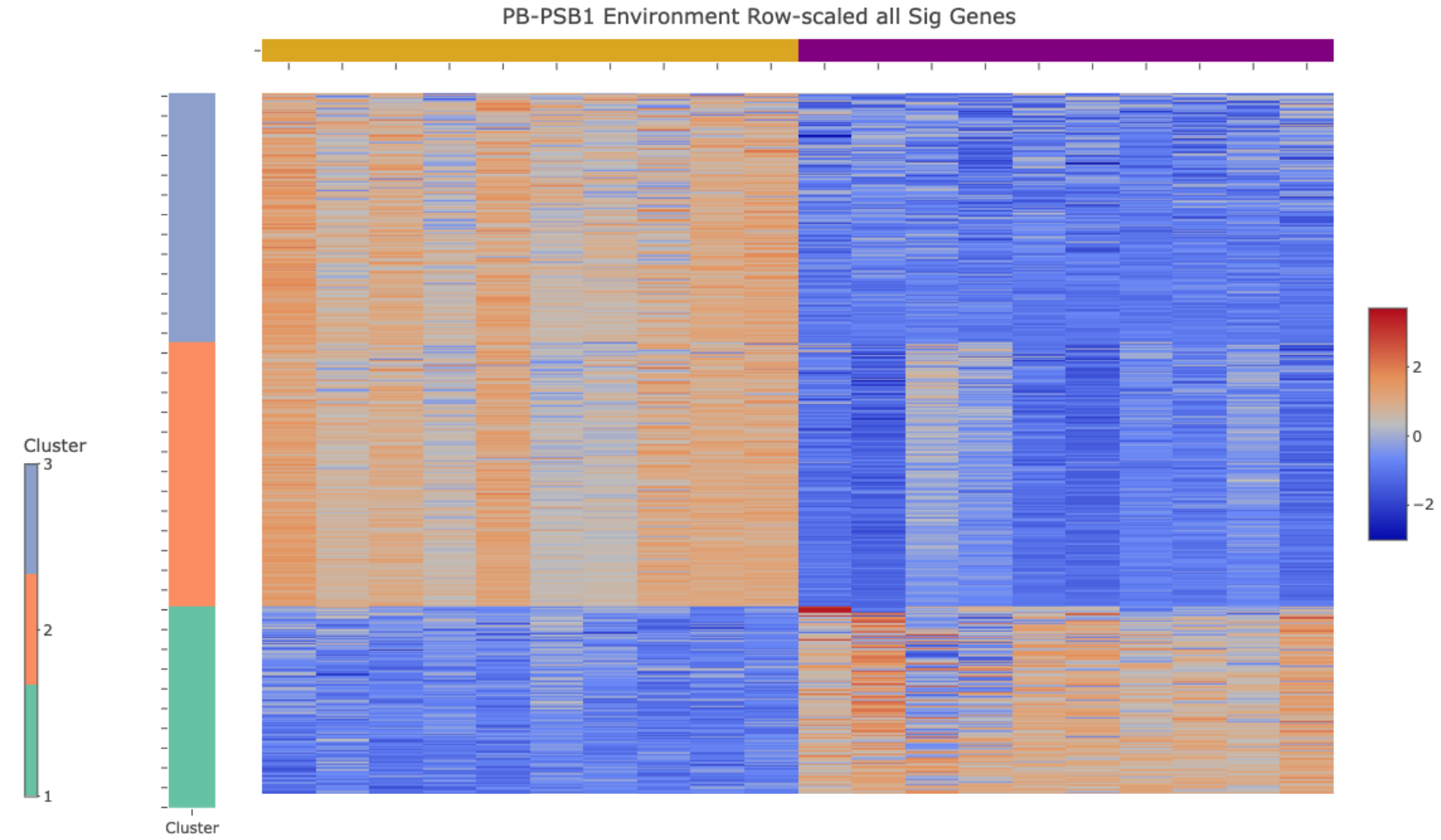


Figure 3. PB-PSB1 genes differentially expressed between all light and dark environmental samples (|LFC|>1 and p_{adj}>0.05), shrinkage adjusted and row-scaled. Clustered with ClusterGVis and visualized with plot_ly. Gold bars correspond to light time points and purple bars to dark time points. Cluster membership is described by left bar.

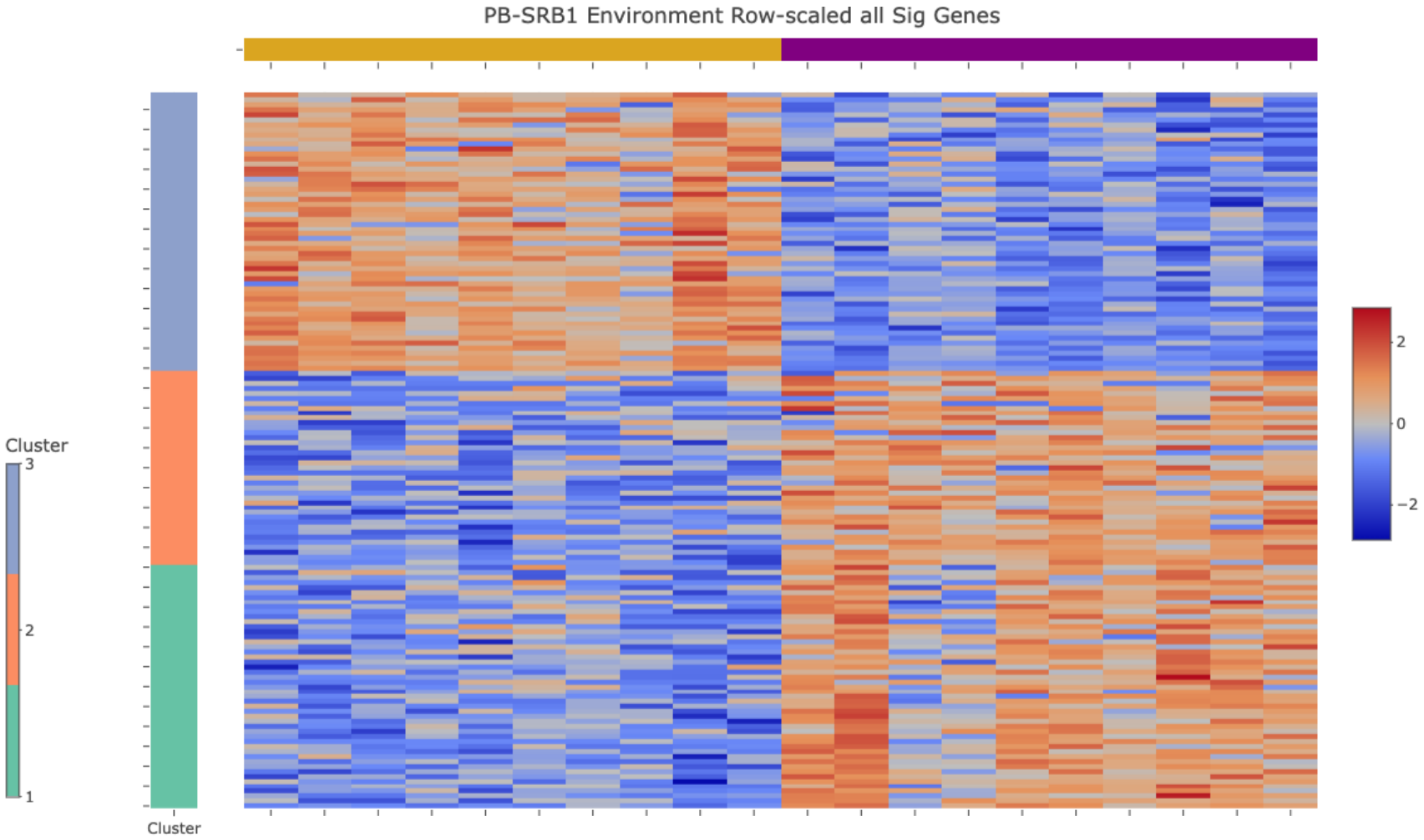


Figure 4. PB-SRB1 genes differentially expressed between all light and dark environmental samples (|LFC|>1 and padj>0.05), shrinkage adjusted and row-scaled. Clustered with ClusterGVis and visualized with plot_ly. Gold bars correspond to light time points and purple bars to dark time points. Cluster membership is described by left bar.

Daytime upregulation of cofactor synthesis and carbon fixation in PB-PSB1

We first examined the clusters of genes that exhibited daytime upregulation in PB-PSB1 and found many genes involved in cofactor and vitamin synthesis. Specifically, we observed genes involved in the synthesis of compounds such as vitamin B12, folate, biotin, and riboflavin. In terms of vitamin B12, we found upregulation in genes encoding many of the precorrin interacting proteins needed for its synthesis (PSB__G1011, PSB__G2386, PSB__G1320), as well as metal chelating proteins aiding in its synthesis (CobW, Young et al. 2021) (PSB__G1269). Within folate metabolism we found synthesis related genes like the tetrahydrofolate regenerating 5-methyltetrahydrofolate-homocysteine methyltransferase (PSB__G1663) as well as dihydroneopterin aldolase (PSB__G6682, Haußmann et al. 1998). Genes related to biotin synthesis such as dethiobiotin synthtase (PSB__G2647) and malonyl-ACP O-methyltransferase BioC (PSB__G2161), alongside genes encoding for proteins dependent on biotin for its activity like an acetyl-CoA carboxylase biotin carboxylase subunit (PSB__G447) were also upregulated.

Together, upregulation of these genes paints a similar picture as seen in our lab incubations, were upregulation of groups of genes involved in vitamin and cofactor synthesis during the day may be the result of the cell reacting to photosynthetic energy provided by the sun. Accordingly, we saw genes that are part of various carbon fixation pathways to be upregulated in our daytime conditions. In terms of the Calvin cycle, we found upregulation for the small chain of ribulose−bisphosphate carboxylase as well as ribose 5 phosphate isomerase (PSB__G4153 and PSB__G2891). In terms of the hydroxypropionate bicycle we found upregulation of 3-methylfumaryl-CoA hydratase (PSB__G6466, Zarzycki et al. 2009) and within the 3-hydroxypropionate/4-hydroxybutyrate cycle we found upregulation of

acrylyl-CoA reductase (PSB__G1815, Berg et al. 2007). While PB-PSB1 encodes for most of the genes for the hydroxypropionate bicycle, this pathway tends to be used for carbon fixation in some green sulfur bacteria (Hügler and Sievert 2011). Parts of the cycle have, additionally, sometimes been found within *Haliaceae* of the *Gammaproteobacteria* (Zarzycki et al. 2009). Within the PB-PSB1 genome, the pathway was mostly complete except for a malonyl-coA reductase. The incomplete pathway may help in assimilating trace amounts of small organic compounds like acetate, propionate, and succinate even under oxic conditions as none of the cycle's enzymes are oxygen sensitive (Zarzycki et al. 2009). Other *Chromatiales* like *Allochromatium vinosum* and *Marichromatium purparatum* are known for utilizing similar, small organic compounds (Imhoff and Trüper 1980; Nicolay et al. 1983), however we have not analyzed their genomes for encoding of the specific genes we observed in PB-PSB1. Notwithstanding, it may still be possible for PB-PSB1 to use this incomplete hydroxypropionate bicycle for assimilation of similar small carbon compounds.

Nighttime upregulation of bacteriochlorophyll synthesis genes in PB-PSB1

We subsequently examined nighttime upregulation of genes in PB-PSB1. We found nighttime upregulation in genes involved in chlorophyll synthesis such as a lycopene beta cyclase which converts lycopene to carotene (PSB__G1362, Zhao et al. 2020), potentially indicating nighttime bacteriochlorophyll synthesis. This aligns with older studies of *Thiocapsa roseopersicina*, where when the organism was grown under a light-oxic/dark-anoxic regime bacteriochlorophyll synthesis increased during the night due to bacteriochlorophyll's oxygen sensitivity (De Wit and Van Gemerden 1990b). Whether this is an actual response to oxygen exposure or instead some kind of entrained response would require further study.

The oxygen stress response in PB-PSB1 differs between field and lab samples but is similar for PB-SRB1

Having observed differences due to light-dark exposure in PB-PSB1, we next aimed to examine whether we could detect a difference in transcriptional response that could indicate a response to daytime environmental oxygen. Within PB-PSB1 We observed daytime upregulation for a hydrogenase maturation factor (PSB__G1215, HypD), as well as a proton-translocation NAD(P) transhydrogenase (PSB__G1536). Functioning hydrogenases could oxidize hydrogen and reduce oxygen, potentially relieving oxygen stress if necessary. While knowing if such a function was possible in these PB-PSB1 hydrogenases would need specific characterization, hydrogenases have been studied in *Hydrogenophaga pseudoflava* and have been shown to first scavenge oxygen before participating in their normal catalytic activity (Siebert et al. 2026). Additionally, we found an L−tyrosine peroxygenase (PSB__G2238), which would be capable of detoxifying peroxide. Furthermore, we saw enrichment for a nitrogenase-stabilizing/protective protein (PSB__G3880). Typically nitrogenase stabilizing proteins like NifW are known for interacting with the MoFe cofactor in nitrogenase in order to protect it from oxygen damage (Kim and Burgess 1996). The daytime upregulation of these hydrogenases and peroxygenases in PB-PSB1 was not seen in our lab incubated samples, and could thus be a response unique to our field collected samples and oxygen encountered therein.

We proceeded to look at daytime upregulation of genes that PB-SRB1 may be employing to deal with oxic stress. We found daytime upregulation of cytochrome *c* oxidase and cytochrome *d* ubiquinol oxidase in both our lab and field samples (SRB__G2281 and SRB__G4333). If these are indeed indicative of oxic stress then we might be observing a

similar response to oxic stress by PB-SRB1 in both lab and field incubated samples. This could potentially mean that PB-PSB1's own response to oxygen stress may be reducing enough oxygen that it protects other community members in the pink berry biofilm from oxygen exposure. We would require a wider survey in terms of number genes and their patterns of expression by PB-SRB1 in order to truly determine if PB-SRB1's response is indeed similar between lab and field conditions.

Field PB-SRB1 samples express a similar number of oxidoreductases and transporters between light and dark conditions

In our lab incubated samples from chapter two, light-dark differences resulted in our recording more annotated oxidoreductases and transporters upregulated during the day than at night by PB-SRB1. The annotations of these genes helped us infer potential electron donors for daytime PB-SRB1 sulfate reduction using compounds like 2-oxoacids. As oxygen exposure in the environment may influence the identity and production of photosynthate by PB-PSB1 we aimed to look at how PB-SRB1 expression of oxidoreductases and transporters between light-dark conditions changed in our environmental samples.

We found a similar number of oxidoreductase and transporter genes expressed between day and night sampling conditions. Within our environmental transcriptomes we found four oxidoreductase genes upregulated during the day. Among these, 2-oxoacid was an operative substrate for one (SRB__G1040 and SRB__G1652) and the others were of the SDR family type (SRB__G1966 and SRB__G568). We found two night upregulated oxidoreductases, one operated on glycerol 3-phosphate (SRB__G776) and the other operated on L-glutamate gamma-semialdehyde (SRB__G59). In terms of transporter genes we found three upregulated during the day (SRB__G3976, SRB__G1974, and SRB__G1096), and we found

five upregulated at night (SRB__G553, SRB__G1551, SRB__G1826, SRB__G922, and SRB__G1061). It is interesting to note that our lab incubations also exhibited daytime upregulation of a 2-oxoacid dehydrogenase as potential compounds used by this complex could be oxoglutrarate, pyruvate, oxobutyrate and glyoxylate. As we observed previously in chapter two, potential use of a photorespiration byproduct like glyoxylate is intriguing as PB-SRB1's closest cultured relative was isolated on glycolate oxidizing media (Friedrich et al. 1996), and its transformation to glycolate was upregulated by PB-PSB1 in our lab incubations.

While we did not observe major differences in transcription between day and night sampling for the number of genes encoding oxidoreductases and transporters, we did find genes that were similarly transcribed between environmental light and dark conditions. This struck us as interesting as environmentally sampled aggregates would have been exposed to a much richer pool of potential substrates due to their location on the marsh sediment. Similar gene expression between conditions despite this richer available pool may suggest specific substrate use by PB-SRB1, and hint at the specific association between PB-PSB1 and PB-SRB1. Potentially, our field samples could be oxidizing a 2-oxoacid similar to our lab incubated samples.

Conclusions

We hoped to determine how changes in oxygen exposure affected the syntrophic relationship between PB-PSB1 and PB-SRB1. In order to do this both PB-PSB1 and PB-SRB1 were sampled at peak lighting conditions in the field. When these samples were compared for difference between light-dark conditions we saw a significant difference such that 47.19% of analyzed genes in PB-PSB1 and 11.22% of analyzed genes in PB-SRB1 were significantly different. Whether this indicates a direct response to a light shift or some entrained rhythm would need more detailed sampling. We then honed our focus to examining light-dark differences in gene expression by PB-PSB1. We found some similarities to our lab incubated samples such that our environmental metatranscriptome of PB-PSB1 again showed a daytime upregulation of genes involved in cofactor synthesis: specifically adenosylcobalamin, folate, biotin, and riboflavin. This could again be indicative of energy being harnessed from photosynthesis. This photosynthesis, furthermore, could provide energy for carbon fixation, as we saw genes upregulated in various carbon fixation pathways: the Calvin cycle, the hydroxypropionate bicycle, and the 3-hydroxypropionate/4-hydroxybutyrate cycle. We then looked at dark expression of genes in PB-PSB1. Our environmental metatranscriptome of PB-PSB1 showed a nighttime upregulation of bacteriochlorophyll synthesis genes. This, in turn, may be a response to oxygen exposure in the environment as our daytime sampling would have been oxic, however this does not rule out some sort of entrained response as we also saw nighttime expression of bacteriochlorophyll synthesis genes in our lab-incubated samples.

Furthermore, our interest in pairing an environmental transcriptome to our lab incubations lay in determining potential differences due to oxic stress. In PB-PSB1 we saw

some potential signs of oxic stress in PB-PSB1 as we observed a daytime upregulation of genes associated with hydrogenases, a peroxygenase that can detoxify peroxides, as well as *nifW* which encodes for a stabilizing protein that protects the nitrogenase cofactor from oxidative damage. Signs for differential oxic stress between lab and field conditions were less obvious in PB-SRB1. Within our environmental transcriptomes of PB-SRB1 we saw daytime expression of a cytochrome *c* oxidase and cytochrome *d* ubiquinol oxidase, which were also expressed in our lab incubations. If these are signs of oxic stress, then we may be observing similar levels of oxic stress in PB-SRB1 in both our lab and field samples. PB-PSB1, however showed different signs of oxic stress due to our daytime field sampling than our lab incubations, which may be indicative of PB-PSB1 taking on some role protecting the community from oxygen stress.

Within PB-SRB1's environmental metatranscriptome we did not see major differences in the number of genes encoding oxidoreductases (4 day oxidoreductases, 2 night oxidoreductases) and transporters (3 day, 5 night) upregulated in either day or night conditions. Similar to our lab incubations, however, we saw daytime upregulation of 2-oxoacid oxidoreductases. The 2-oxoacids used may be compounds like oxoglutarate, pyruvate, and oxobutyrate. The change in these central metabolism intermediates may be related to the daytime stimulation of sulfate reduction we inferred from our lab incubations. In particular, it is interesting to note that glyoxylate is a 2-oxoacid, whose transformation to glycolate we saw potentially upregulated in our lab incubations and is used by PB-SRB1's closest cultured relative. Our environmental sampling, therefore, may also be indicative of PB-SRB1's use of some photorespiration byproduct from PB-PSB1 as an electron donor for its sulfate reduction.

References

Berg, Ivan A., Daniel Kockelkorn, Wolfgang Buckel, and Georg Fuchs. 2007. “A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea.” *Science* 318 (5857): 1782–86. https://doi.org/10.1126/science.1149976.

Bushnell, Brian. 2014. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Released. https://bbmap.org.

De Wit, Rutger, and Hans Van Gemerden. 1990a. “Growth and Metabolism of the Purple Sulfur Bacterium Thiocapsa Roseopersicina under Combined Light/Dark and Oxic/Anoxic Regimens.” *Archives of Microbiology* 154 (5): 459–64. https://doi.org/10.1007/BF00245228.

De Wit, Rutger, and Hans Van Gemerden. 1990b. “Growth and Metabolism of the Purple Sulfur Bacterium Thiocapsa Roseopersicina under Combined Light/Dark and Oxic/Anoxic Regimens.” *Archives of Microbiology* 154 (5): 459–64. https://doi.org/10.1007/BF00245228.

Friedrich, M., N. Springer, W. Ludwig, and B. Schink. 1996. “Phylogenetic Positions of Desulfofustis Glycolicus Gen. Nov., Sp. Nov. and Syntrophobotulus Glycolicus Gen. Nov., Sp. Nov., Two New Strict Anaerobes Growing with Glycolic Acid.” *International Journal of Systematic Bacteriology* 46 (4): 1065–69. https://doi.org/10.1099/00207713-46-4-1065.

Hügler, Michael, and Stefan M. Sievert. 2011. “Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean.” *Annual Review of Marine Science* 3 (1): 261–89. https://doi.org/10.1146/annurev-marine-120709-142712.

Imhoff, Johannes F., and Hans G. Trüper. 1980. “Chromatium Purpuratum, Sp. Nov., a New Species of the Chromatiaceae.” *Zentralblatt Für Bakteriologie: I. Abt. Originale C: Allgemeine, Angewandte Und Ökologische Mikrobiologie* 1 (1): 61–69. https://doi.org/10.1016/S0172-5564(80)80017-0.

Kim, Seonyoung, and Barbara K. Burgess. 1996. “Evidence for the Direct Interaction of the nifW Gene Product with the MoFe Protein.” *Journal of Biological Chemistry* 271 (16): 9764–70. https://doi.org/10.1074/jbc.271.16.9764.

Nelson, Douglas C., Bo Barker Jørgensen, and Niels Peter Revsbech. 1986. “Growth Pattern and Yield of a Chemoautotrophic *Beggiatoa* Sp. in Oxygen-Sulfide Microgradients.” *Applied and Environmental Microbiology* 52 (2): 225–33. https://doi.org/10.1128/aem.52.2.225-233.1986.

Nicolay, K., H. van Gemerden, K. J. Hellingwerf, W. N. Konings, and R. Kaptein. 1983. “In Vivo 31P and ^{13}C Nuclear Magnetic Resonance Studies of Acetate Metabolism in Chromatium Vinosum.” *Journal of Bacteriology* 155 (2): 634–42. https://doi.org/10.1128/jb.155.2.634-642.1983.

Overmann, J□rg, and Norbert Pfennig. 1992. “Continuous Chemotrophic Growth and Respiration of Chromatiaceae Species at Low Oxygen Concentrations.” *Archives of Microbiology* 158 (1): 59–67. https://doi.org/10.1007/BF00249067.

Salman, Verena, Tingting Yang, Tom Berben, Frieder Klein, Esther Angert, and Andreas Teske. 2015. “Calcite-Accumulating Large Sulfur Bacteria of the Genus *Achromatium* in Sippewissett Salt Marsh.” *The ISME Journal* 9 (11): 2503–14. https://doi.org/10.1038/ismej.2015.62.

Siebert, Dominik L., Simon Grieshaber, Bastian Blombach, Volker Sieber, and Ammar Al-Shameri. 2026. “Biochemical and Mechanistic Insights into O_2 Scavenging and Stability of the Soluble Hydrogenase from *Hydrogenophaga Pseudoflava*.” *ACS Omega* 11 (1): 1769–77. https://doi.org/10.1021/acsomega.5c09877.

Wit, Rutger, and Hans Gemerden. 1987a. “Chemolithotrophic Growth of the Phototrophic Sulfur Bacterium *Thiocapsa Roseopersicina*.” *FEMS Microbiology Letters* 45 (2): 117–26. https://doi.org/10.1111/j.1574-6968.1987.tb02347.x.

Wit, Rutger, and Hans Gemerden. 1987b. “Chemolithotrophic Growth of the Phototrophic Sulfur Bacterium *Thiocapsa Roseopersicina*.” *FEMS Microbiology Letters* 45 (2): 117–26. https://doi.org/10.1111/j.1574-6968.1987.tb02347.x.

Witt, Rutger, and Hans Gemerden. 1990. “Growth of the Phototrophic Purple Sulfur Bacterium *Thiocapsa Roseopersicina* under Oxic/Anoxic Regimens in the Light.” *FEMS Microbiology Letters* 73 (1): 69–76. https://doi.org/10.1111/j.1574-6968.1990.tb03926.x.

Young, Tessa R., Maria Alessandra Martini, Andrew W. Foster, et al. 2021. “Calculating Metalation in Cells Reveals CobW Acquires CoII for Vitamin B12 Biosynthesis While Related Proteins Prefer ZnII.” *Nature Communications* 12 (1): 1195. https://doi.org/10.1038/s41467-021-21479-8.

Zarzycki, Jan, Volker Brecht, Michael Müller, and Georg Fuchs. 2009. “Identifying the Missing Steps of the Autotrophic 3-Hydroxypropionate CO_2 Fixation Cycle in *Chloroflexus Aurantiacus*.” *Proceedings of the National Academy of Sciences* 106 (50): 21317–22. https://doi.org/10.1073/pnas.0908356106.

Zhao, Zilong, Zhen Liu, and Xiangzhao Mao. 2020. “Biotechnological Advances in Lycopene β-Cyclases.” *Journal of Agricultural and Food Chemistry* 68 (43): 11895–907. https://doi.org/10.1021/acs.jafc.0c04814.

Appendix

Chapter I Supplemental Information

GTDB taxonomy	Bin ID	Citation	date sampled	total_size	Compl.	Contam	scaffold N50	maximum scaffold len	number of scaffolds	Coverage	NCBI BioProject	NCBI Biosample - Genome	Genbank Assembly Accession	Sequencing type	NCBI Biosample - physical	SRA Experiment	SRA Run
d__Bacteria; p__Pseudomonadota; c__Gammaproteobacteria; o__Chromatiales; f__Chromatiaceae; g__Thiohalocapsa; s__Thiohalocapsa sp003525925	PB-PSB1-LS01-2010a	Wilbanks et al 2014	2010-07-01	6,700,000	97.06	5.63	7,400	31,000	1386	25	PRJNA215075	SAMN17298306	GCA_000495615.1	Illumina HiSeq + Roche 454	SAMN02303577, SAMN02303578	SRX332170, SRX332174	SRR948504, SRR948505
d__Bacteria; p__Pseudomonadota; c__Gammaproteobacteria; o__Chromatiales; f__Chromatiaceae; g__Thiohalocapsa; s__Thiohalocapsa sp003525925	UBA10420	Parks DH et al., "A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life.", Nat Biotechnol, 2018 Nov;36(10):996-1004	2010-07-01	3,855,426	78.14	2.08	6,128	22,301	690	3.5	PRJNA417962	SAMN08018746	GCA_003499475.1	Illumina HiSeq	SAMN02303578	SRX332174	SRR948505
d__Bacteria; p__Pseudomonadota; c__Gammaproteobacteria; o__Chromatiales; f__Chromatiaceae; g__Thiohalocapsa; s__Thiohalocapsa sp003525925	UBA10422	Parks DH et al., "A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life.", Nat Biotechnol, 2018 Nov;36(10):996-1004	2010-07-01	5,447,945	92.05	2.93	20,002	95,463	394	106	PRJNA417962	SAMN08020484	GCA_003525925.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria; p__Pseudomonadota; c__Gammaproteobacteria; o__Chromatiales; f__Chromatiaceae; g__Thiohalocapsa; s__Thiohalocapsa sp003525925	SRR948829_binner13_Refined_13	Mitchell AL et al., "EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies.", Nucleic Acids Res, 2018 Jan 4;46(D1):D726-D735	2010-07-01	5,937,886	98.2	2.93	16,828	66,285	488	190	PRJEB45569	SAMEA115292589	GCA_963966235.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria; p__Pseudomonadota; c__Gammaproteobacteria; o__Chromatiales; f__Chromatiaceae; g__Thiohalocapsa; s__Thiohalocapsa sp003525925	PB-PSB1-LS01-2011a	Wilbanks et al. 2022	2011-07-01	7,950,631	98.89	3.62	7,950,631	7,950,631	1	500	PRJNA684324	SAMN17298418	GCA_016745215.1	PacBio RSII	SAMN17228542	***see Chapter 1-Table 1***	***see Chapter 1-Table 1***
d__Bacteria; p__Pseudomonadota; c__Gammaproteobacteria; o__Chromatiales; f__Chromatiaceae; g__Thiohalocapsa; s__Thiohalocapsa sp003525925	PSB_001	this study	2021-07-30	7,855,022	98.2	6.15	732,121	1,072,372	15	283	NA	NA	NA	PAcBio HiFi	NA	NA	NA
d__Bacteria; p__Pseudomonadota; c__Gammaproteobacteria; o__Chromatiales; f__Chromatiaceae; g__Thiohalocapsa; s__Thiohalocapsa sp003525925	PSB_002	this study	2021-07-30	7,969,801	98.89	3.62	3,203,318	3,681,167	4	493.6	NA	NA	NA	PAcBio HiFi	NA	NA	NA
d__Bacteria; p__Pseudomonadota; c__Gammaproteobacteria; o__Chromatiales; f__Chromatiaceae; g__Thiohalocapsa; s__Thiohalocapsa sp003525925	PSB_024	this study	2021-07-30	7,345,922	98.55	4.89	441,368	1,037,214	26	641.6	NA	NA	NA	PAcBio HiFi	NA	NA	NA

Supplemental Table 1. Bin statistics for PB-PSB1. Bins span our own various binning attempts throughout time as well as those of other studies (Parks et al. 2018; Mitchell et al. 2018).

GTDB taxonomy	Bin ID	Citation	date sampled	total_size	Compl.	Contam	scaffold N50	maximum scaffold len	number of scaffolds	Coverage	NCBI BioProject	NCBI Biosample - Genome	Genbank Assembly Accession	Sequencing type	NCBI Biosample - physical	SRA Experiment	SRA Run
d__Bacteria; p__Desulfobacterota; c__Desulfobulbia; o__Desulfobulbales; f__Desulfocapsaceae; g__PB-SRB1; s__PB-SRB1 sp003500125	PB-SRB1-LS01-2010a	Wilbanks et al 2014	2010-07-01	3,700,000	99	1.98	15,013	63,705	422	20	PRJNA215076	SAMN15031632	GCA_000495595.3	Illumina HiSeq + Roche 454	SAMN02303577, SAMN02303578	SRX332170, SRX332174	SRR948504, SRR948505
d__Bacteria; p__Desulfobacterota; c__Desulfobulbia; o__Desulfobulbales; f__Desulfocapsaceae; g__PB-SRB1; s__PB-SRB1 sp003500125	PB-SRB1-LS01-2011a	Wilbanks et al. 2022	2011-07-01	4,686,948	99.4	1.05	385,966	930,755	71	50	PRJNA684324	SAMN17301126	GCA_016722945.1	PacBio RSII	SAMN17228542	***see Chapter 1-Table 1***	***see Chapter 1-Table 1***
d__Bacteria; p__Desulfobacterota; c__Desulfobulbia; o__Desulfobulbales; f__Desulfocapsaceae; g__PB-SRB1; s__PB-SRB1 sp003500125	UBA10421	Parks DH et al., "A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life.", Nat Biotechnol, 2018 Nov;36(10):996-1004	2010-07-01	2,637,181	75.34	0.69	6,262	38,027	460	3.7	PRJNA417962	SAMN08018745	GCA_003499505.1	Illumina HiSeq	SAMN02303578	SRX332174	SRR948505
d__Bacteria; p__Desulfobacterota; c__Desulfobulbia; o__Desulfobulbales; f__Desulfocapsaceae; g__PB-SRB1; s__PB-SRB1 sp003500125	UBA10423	Parks DH et al., "A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life.", Nat Biotechnol, 2018 Nov;36(10):996-1004	2010-07-01	3,269,683	87.18	0.32	27,645	176,297	204	25.34	PRJNA417962	SAMN08018606	GCA_003500125.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria; p__Desulfobacterota; c__Desulfobulbia; o__Desulfobulbales; f__Desulfocapsaceae; g__PB-SRB1; s__PB-SRB1 sp003500125	SRR948829_metabat2_16	Mitchell AL et al., "EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies.", Nucleic Acids Res, 2018 Jan 4;46(D1):D726-D735	2010-07-01	3,582,348	99.4	1.19	33,549	174,826	186	39	PRJEB45569	SAMEA115292590	GCA_963966225.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria; p__Desulfobacterota; c__Desulfobulbia; o__Desulfobulbales; f__Desulfocapsaceae; g__PB-SRB1; s__PB-SRB1 sp003500125	SRB_002	this study	2021-07-30	3,716,745	86.01	2.08	598,159	1,007,845	9	148.4	NA	NA	NA	PacBio HiFi	NA	NA	NA
d__Bacteria; p__Desulfobacterota; c__Desulfobulbia; o__Desulfobulbales; f__Desulfocapsaceae; g__PB-SRB1; s__PB-SRB1 sp003500125	SRB_024	this study	2021-07-30	4,180,437	99.4	0.32	4,034,111	4,034,111	2	96.71	NA	NA	NA	PacBio HiFi	NA	NA	NA

Supplemental Table 2. Bin statistics for PB-SRB1. Bins span our own various binning attempts throughout time as well as those of other studies (Parks et al. 2018; Mitchell et al. 2018).

GTDB taxonomy	Bin ID	Citation	date sampled	total_size	Compl.	Contam	scaffold N50	maximum scaffold len	number of scaffolds	Coverage	NCBI BioProject	NCBI Biosample - Genome	Genbank Assembly Accession	Sequencing type	NCBI Biosample - physical	SRA Experiment	SRA Run
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Rhodobacterales; f__Rhodobacteraceae; g__UBA10424; s__UBA10424 sp003500165	UBA10424	Parks DH et al., "A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life.", Nat Biotechnol, 2018 Nov;36(10):996-1004	2010-07-01	3,269,265	88%	0.40%	13,516	44,573	329	17	PRJNA417962	SAMN08018742	GCA_003500165.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Rhodobacterales; f__Rhodobacteraceae; g__UBA10424; s__UBA10424 sp003500165	SRR948829_maxbin2_4	Mitchell AL et al., "EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies.", Nucleic Acids Res, 2018 Jan 4;46(D1):D726-D735	2010-07-01	3,762,636	98%	0.20%	16,775	121,856	356	29	PRJEB45569	SAMEA115292585	GCA_963966265.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Rhodobacterales; f__Rhodobacteraceae; g__UBA10424; s__UBA10424 sp003500165	PB-A2_LS01-2011-b	Wilbanks et al. 2022	2011-07-01	4,802,595	82.61	3.09	202,752	782,375	49	27	PRJNA684324	SAMN17054635	GCA_016762775.1	PacBio RSII	SAMN17053571 and SAMN17054265	NA	NA
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Rhodobacterales; f__Rhodobacteraceae; g__UBA10424; s__UBA10424 sp003500165	Rhodobacteraceae_001	this study	2021-07-30	2,991,494	85.1	5.9	117,077	304,516	40	70	NA	NA	NA	PacBio HiFi	NA	NA	NA
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Rhodobacterales; f__Rhodobacteraceae; g__UBA10424; s__UBA10424 sp003500165	Rhodobacteraceae_002	this study	2021-07-30	3,552,753	96	0	201,603	443,297	23	76.48	NA	NA	NA	PacBio HiFi	NA	NA	NA
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Rhodobacterales; f__Rhodobacteraceae; g__UBA10424; s__UBA10424 sp003500165	Rhodobacteraceae_024	this study	2021-07-30	3,732,931	96.3	1.2	186,271	304,422	29	31.9	NA	NA	NA	PacBio HiFi	NA	NA	NA

Supplemental Table 3. Bin statistics for our *Rhodobacteraceae* bin. Bins span our own various binning attempts throughout time as well as those of other studies (Parks et al. 2018; Mitchell et al. 2018).

GTDB taxonomy	Bin ID	Citation	date sampled	total_size	Compl.	Contam	scaffold N50	maximum scaffold len	number of scaffolds	Coverage	NCBI BioProject	NCBI Biosample - Genome	Genbank Assembly Accession	Sequencing type	NCBI Biosample - physical	SRA Experiment	SRA Run
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Caulobacterales; f__Maricaulaceae; g__Oceanicaulis; s__Oceanicaulis alexandrii	PB-A1_LS01-2011-a	Wilbanks et al. 2022	2011-07-01	3,235,649	95.73	2.74	29,314	113,200	186	50	PRJNA684324	SAMN17054541	GCA_016762755.1	Illumina	SAMN17054266	NA	NA
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Caulobacterales; f__Maricaulaceae; g__Oceanicaulis; s__Oceanicaulis alexandrii	PB-A1_LS01-2011-b	Wilbanks et al. 2022	2011-07-01	613,200	11.59	1.18	17,900		36	64	PRJNA684324	SAMN17054628	GCA_016762865.1	PacBio RSII	SAMN17053571 and SAMN17054265	NA	NA
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Caulobacterales; f__Maricaulaceae; g__Oceanicaulis; s__Oceanicaulis alexandrii	Oceanicaulis_001	this study	2021-07-30	3,204,013	99.84	1.06	3,204,013	3,204,013	1	40	NA	NA	NA	PacBio HiFi	NA	NA	NA
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Caulobacterales; f__Maricaulaceae; g__Oceanicaulis; s__Oceanicaulis alexandrii	Oceanicaulis_002	this study	2021-07-30	3,203,874	99.84	1.06	3,203,874	3,203,874	1	28.14	NA	NA	NA	PacBio HiFi	NA	NA	NA
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Caulobacterales; f__Maricaulaceae; g__Oceanicaulis; s__Oceanicaulis alexandrii	Oceanicaulis_024	this study	2021-07-30	3,202,119	98.8	1.38	3,202,119	3,202,119	1	15.42	NA	NA	NA	PacBio HiFi	NA	NA	NA

Supplemental Table 4. Bin statistics for our *Oceanicaulis alexandrii* bin. Bins span our own various binning attempts throughout time as well as those of other studies (Parks et al. 2018; Mitchell et al. 2018).

GTDB taxonomy	Bin ID	Citation	date sampled	total_size	Compl.	Contam	scaffold N50	maximum scaffold len	number of scaffolds	Coverage	NCBI BioProject	NCBI Biosample - Genome	Genbank Assembly Accession	Sequencing type	NCBI Biosample - physical	SRA Experiment	SRA Run
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Micavibrionales; f__Micavibrionaceae; g__UBA10425; s__UBA10425 sp003499545	PB-A3_LS01-2011-a	Wilbanks et al. 2022	2011-07-01	1,753,089	84.27	0.65	32,687	142,504	90	200	PRJNA684324	SAMN17054543	GCA_016757515.1	Illumina	SAMN17054266	NA	NA
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Micavibrionales; f__Micavibrionaceae; g__UBA10425; s__UBA10425 sp003499545	UBA10425	Parks DH et al., "A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life.", Nat Biotechnol, 2018 Nov;36(10):996-1004	2010-07-01	1,414,244	78.17	1.3	11,957	30,934	147	9.88	PRJNA417962	SAMN08018741	GCA_003499545.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Micavibrionales; f__Micavibrionaceae; g__UBA10425; s__UBA10425 sp003499545	SRR948829_binner12_Refined_10	Mitchell AL et al., "EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies.", Nucleic Acids Res, 2018 Jan 4;46(D1):D726-D735	2010-07-01	1,557,853	85.27	0.87	14,221	65,963	168	13	PRJEB45569	SAMEA115292583	GCA_963966305.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Micavibrionales; f__Micavibrionaceae; g__UBA10425; s__UBA10425 sp003499545	Micavibrio_001	this study	2021-07-30	1,642,830	85.35	0.43	1,642,830	1,642,830	1	335.04	NA	NA	NA	PacBioHiFi	NA	NA	NA
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Micavibrionales; f__Micavibrionaceae; g__UBA10425; s__UBA10425 sp003499545	Micavibrio_002	this study	2021-07-30	1,652,490	85.79	0.43	1,431,799	1,431,799	2	52.9	NA	NA	NA	PacBioHiFi	NA	NA	NA
d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Micavibrionales; f__Micavibrionaceae; g__UBA10425; s__UBA10425 sp003499545	Micavibrio_024	this study	2021-07-30	1,667,557	85.79	0.43	976,910	976,910	2	43.84	NA	NA	NA	PacBioHiFi	NA	NA	NA

Supplemental Table 5. Bin statistics for our *Micavibrio* bin. Bins span our own various binning attempts throughout time as well as those of other studies (Parks et al. 2018; Mitchell et al. 2018).

GTDB taxonomy	Bin ID	Citation	date sampled	total_size	Compl.	Contam	scaffold N50	maximum scaffold len	number of scaffolds	Coverage	NCBI BioProject	NCBI Biosample - Genome	Genbank Assembly Accession	Sequencing type	NCBI Biosample - physical	SRA Experiment	SRA Run
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Saprospiraceae;g__Lewinella_A;s__Lewinella_A sp004375085	JGIbin_sample001_66	this study	2021-07-30	4,790,503	88.5	1.5	774,553	1,131,164	10	15	NA	NA	NA	PacBioHiFi	NA	NA	NA
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Saprospiraceae;g__Lewinella_A;s__Lewinella_A sp004375085	JGIbin_sample002_107	this study	2021-07-30	5,366,041	98.5	2.5	497,929	931,891	15	23.2	NA	NA	NA	PacBioHiFi	NA	NA	NA
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Saprospiraceae;g__Lewinella_A;s__Lewinella_A sp004375085	JGIbin_sample024_23	this study	2021-07-30	5,507,181	99	1.5	971,981	1,893,917	11	16.19	NA	NA	NA	PacBioHiFi	NA	NA	NA

Supplemental Table 6. Bin statistics for our *Lewinella* bin. Bins span our own various binning strategies.

GTDB taxonomy	Bin ID	Citation	date sampled	total_size	Compl.	Contam	scaffold N50	maximum scaffold len	number of scaffolds	Coverage	NCBI BioProject	NCBI Biosample - Genome	Genbank Assembly Accession	Sequencing type	NCBI Biosample - physical	SRA Experiment	SRA Run
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Salibacteraceae;g__UBA10426;s__UBA10426 sp003499525	UBA10426	Parks DH et al., "A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life.", Nat Biotechnol, 2018 Nov;36(10):996-1004	2010-07-01	3,259,411	100	1.08	74,429	220,233	97	6.49	PRJNA417962	SAMN08018744	GCA_003499525.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Salibacteraceae;g__UBA10426;s__UBA10426 sp003499525	SRR948829_binner12_Refined_1	Mitchell AL et al., "EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies.", Nucleic Acids Res, 2018 Jan 4;46(D1):D726-D735	2010-07-01	3,219,872	97.31	0	88,952	365,877	93	8	PRJEB45569	SAMEA115292591	GCA_963966335.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Salibacteraceae;g__UBA10426;s__UBA10426 sp003499525	metabat2_LS01_001_covarybin_with_cov.8	this study	2021-07-30	3,617,951	100	0	3,572,210	3,572,210	2	273	NA	NA	NA	PacBioHiFi	NA	NA	NA
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Salibacteraceae;g__UBA10426;s__UBA10426 sp003499525	JGIbin_sample002_86	this study	2021-07-30	3,169,723	98.92	0	2,564,599	2,564,599	2	131.9	NA	NA	NA	PacBioHiFi	NA	NA	NA
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Salibacteraceae;g__UBA10426;s__UBA10426 sp003499525	metabat2_LS01_024_covarybin_with_cov.9	this study	2021-07-30	3,140,993	98.92	0	2,528,348	2,528,348	2	152.5	NA	NA	NA	PacBioHiFi	NA	NA	NA

Supplemental Table 7. Bin statistics for our *Salibacteraceae* UBA10426 bin. Bins span our own various binning attempts as well as those of other studies (Parks et al. 2018; Mitchell et al. 2018).

GTDB taxonomy	Bin ID	Citation	date sampled	total_size	Compl.	Contam	scaffold N50	maximum scaffold len	number of scaffolds	Coverage	NCBI BioProject	NCBI Biosample - Genome	Genbank Assembly Accession	Sequencing type	NCBI Biosample - physical	SRA Experiment	SRA Run
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Cyclobacteriaceae;g__UBA10427;s__UBA10427 sp003500135	UBA10427	Parks DH et al., "A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life.", Nat Biotechnol, 2018 Nov;36(10):996-1004	2010-07-01	4,303,780	94.53	5.83	18,817	74,086	332	5.1	PRJNA417962	SAMN08018743	GCA_003500135.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Cyclobacteriaceae;g__UBA10427;s__UBA10427 sp003500135	SRR948829_concoct_67	Mitchell AL et al., "EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies.", Nucleic Acids Res, 2018 Jan 4;46(D1):D726-D735	2010-07-01	4,001,563	98.51	1.8	22,474	80,297	265	6	PRJEB45569	SAMEA115292592	GCA_963966275.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Cyclobacteriaceae;g__UBA10427;s__UBA10427 sp003500135	metabat2_LS01_001_covarybin_with_cov.77	this study	2021-07-30	5,513,596	99.4	2.09	1,165,414	1,260,169	9	93	NA	NA	NA	PacBioHiFi	NA	NA	NA

Supplemental Table 8. Bin statistics for our *Cyclobacteriaceae* UBA10427 bin. Bins span our own various binning attempts as well as those of other studies (Parks et al. 2018; Mitchell et al. 2018).

GTDB taxonomy	Bin ID	Citation	date sampled	total_size	Compl.	Contam	scaffold N50	maximum scaffold len	number of scaffolds	Coverage	NCBI BioProject	NCBI Biosample - Genome	Genbank Assembly Accession	Sequencing type	NCBI Biosample - physical	SRA Experiment	SRA Run
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__UBA10428;g__UBA10428;s__UBA10428 sp003508755	UBA10428	Parks DH et al., "A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life.", Nat Biotechnol, 2018 Nov;36(10):996-1004	2010-07-01	3,471,455	66.29	1.7	6,730	28,903	607	2.55	PRJNA417962	SAMN08018753	GCA_003508755.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__UBA10428;g__UBA10428;s__UBA10428 sp003508755	SRR948829_concoct_38	Mitchell AL et al., "EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies.", Nucleic Acids Res, 2018 Jan 4;46(D1):D726-D735	2010-07-01	4,057,831	88.32	3.05	9,597	41,200	553	3	PRJEB45569	SAMEA115292584	GCA_963966325.1	Illumina MiSeq	SAMN02303579	SRX332175	SRR948829
d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__UBA10428;g__UBA10428;s__UBA10428 sp003508755	JGIbin_sample024_66	this study	2021-07-30	3,870,852	92.47	3.23	53,264	156,348	89	4.45	NA	NA	NA	PacBioHiFi	NA	NA	NA

Supplemental Table 9. Bin statistics for our *Bacteroidales* UBA10428 bin. Bins span our own various binning attempts as well as those of other studies (Parks et al. 2018; Mitchell et al. 2018).

name NCBI strainID	link	completeness	contamination	genome size	country	isolation_source	biosample	genomes in species cluster
SZUA-146	GCA_003228965.1	80.89	2.46	3,083,539	Atlantic Ocean: Mid-Atlantic Ridge; Guaymas Basin; Black smoker hydrothermal vent	marine hydrothermal chimney wall TVG10	SAMN09240130	1
SS_bin_79	GCA_013002925.1	79.62	3.64	3,296,122	Australia: Middle Park Beach; Port Phillip Bay	Beach sand	SAMN14260541	1
P01_B01_bin.59	GCA_039028935.1	90.38	2.51	4,109,248	Pacific Ocean	Pacific Ocean	SAMN32766329	1
SRR17458637_binner23_1_1705322959	GCA_963922815.1	89.27	4.43	3,994,722	USA	plastisphere metagenome	SAMEA115104826	1
JGIbin_sample002_82	this study	85.38	5.13	4,091,555	USA	LS01	NA	NA
metabat2_LS01_002_emilycovarybin_with_cov.89	this study	90.96	3.2	4,533,643	USA	LS01	NA	NA

Supplemental Table 10. Metadata for SZUA-146 bins in GTDB. We found two novel members with no genus designation under family SZUA-146, there is one other SZUA-146 genus with 4 species members

name NCBI strainID	link	completeness	contamination	genome size	country	isolation_source	biosample	genomes in species cluster
G13	GCF_031851795.1	99.45	0.82	7,674,920	China:Guangxi University	seawater, biofilm of coral aquarium	SAMN37356699	1
J057	GCA_003696835.1	89.76	1.91	6,190,056	Japan: Tokyo Prefecture; Shikinejima Island; Jinata Onsen	iron-rich hot spring	SAMN10119957	1
SSF10	GCA_019695295.1	96.45	0.05	4,898,960	Singapore: Jurong East	waste water treatment plant	SAMN19791126	1
HKST-UBA03	GCA_020438265.1	82.94	2.28	7,228,483	China:Hong Kong SAR; Shatin waste water treatment plant	activated sludge from Shatin waste water treatment plant collected monthly from 2007 through 2015	SAMN14562823	2
J06643_6	GCA_039817415.1	99.45	1.09	7,101,124	Atlantic Ocean	plastic particle	SAMN32763075	1
JTS-AC171	GCA_041741415.1	97.25	0.55	5,695,849	Norway: Svalbard, Jotun geothermal spring	biofilm from active travertine-forming spring	SAMN41306452	1
Co_bin.136	GCA_030430205.1	99.45	1.09	7,238,886	South Korea	biofloc rearing seawater	SAMN36083827	1
J06614_4	GCA_038527215.1	97.54	0.27	8,099,156	Atlantic Ocean	plastic particle	SAMN32762482	2
P01_A01_bin.108	GCA_039815525.1	82.79	1.2	7,015,350	Pacific Ocean	plastic particle	SAMN32763163	1
P01_F01_bin.74	GCA_039184535.1	79.99	4.92	6,351,373	Pacific Ocean	plastic particle	SAMN32767030	1
P01_B01_bin.41	GCA_039029315.1	98.89	2.19	12,246,603	Pacific Ocean	plastic particle	SAMN32766310	1
P01_D01_bin.90	GCA_039197765.1	95.25	1.09	7,907,653	Pacific Ocean	plastic particle	SAMN32766843	2
P01_H01_bin.183	GCA_039024295.1	95.63	1.37	5,936,669	Pacific Ocean	plastic particle	SAMN32767670	1
JTS-AC418	GCA_041753595.1	98.91	1.09	6,321,338	Norway: Svalbard, Jotun geothermal spring	biofilm from active travertine-forming spring	SAMN41306706	1
PH2015_11S_54_10	GCA_004292985.1	84.68	5.14	4,540,122	USA: California; Eel River; Dobbyn Creek near Alderpoint	microbial mat material	SAMN10167100	1
J049	GCA_003696875.1	90.88	2.02	6,568,039	Japan: Tokyo Prefecture; Shikinejima Island; Jinata Onsen	iron-rich hot spring	SAMN10119949	1
JGIbin_sample002_88	this study	80.77	1.64	6,485,499	USA	LS01	NA	NA

Supplemental Table 11. Metadata for J057 bins in GTDB. Family J057 has 16 genera under it, one species was arbitrarily chosen from each genus to explore.

name NCBI strainID	link	completeness	contamination	genome size	country	isolation_source	biosample	genomes in species cluster
FeB_10	GCA_003105185.1	94.11	7.26	4,236,048	Australia: Brisbane	Bioreactor performing anaerobic oxidation of methane coupled to iron reduction	SAMN08364728	1
AMR_MDS_1532	GCA_035358815.1	97.86	3.42	4,507,917	USA	Activated sludge of municipal wastewater treatment plants	SAMN32854737	7
JGIbin_sample001_91	this study	93.49	1.76	6,535,097	USA	LS01	NA	NA

Supplemental Table 12. Metadata for FEB-10 bins in GTDB. Genus FEB-10 has 2 species members deposited, under old database release FEB-10 was denoted to the family level.

name NCBI strainID	link	completeness	contamination	genome size	country	isolation_source	biosample	genomes in species cluster
KN72	GCF_012926505.1	100	0.43	4,206,943	NA	NA	SAMN14667155	1
EAC21	GCA_002696645.1	100	3.32	4,314,418	Indian Ocean: Eastern Africa Coastal Province	marine water sample	SAMN07620025	4
SP123	GCA_002709955.1	99.78	0.43	4,232,139	Pacific Ocean: South Pacific Ocean	marine water sample	SAMN07619986	12
ih5	GCA_017792345.1	100	2.61	4,907,757	USA:New York;Manhattan	Xenic protist cell culture	SAMN14492364	5
bin_ma_oilr2.020	GCA_029212835.1	94.24	0.65	3,902,127	USA: Damariscotta River in front of the Bigelow Laboratory for Ocean Sciences	four week timepoint of microcosm treated with crude oil after inoculation with natural surface seawater; biological replicate 2	SAMN33612194	1
Padam27	GCA_033344895.1	100	0.43	4,637,961	Australia: Sydney Harbor	non-axenic Griffithsia monilis salt water culture	SAMN36651868	1
E1-EBD-17	GCA_040217575.1	100	0.43	4,265,973	Spain: Ebro Delta	hypersaline pool; non-axenic culture containing Coleofasciculus sp. DSM 104233	SAMN36814747	2
JGIbin_sample002_40	this study	76.21	5.52	3,532,729	USA	LS01	NA	NA

Supplemental Table 13. Metadata for GCA-2696645/*Pacificispira* bins in GTDB. GCA-2696645 has now been reclassified from a genus designation to species designation, under the genus *Pacificispira*, with 7 members.

name NCBI strainID	link	completeness	contamination	genome size	country	isolation_source	biosample	genomes in species cluster
DT_12	GCA_013216195.1	94.54	0.89	4,105,046	Pacific Ocean: North Pacific Gyre; Station ALOHA	ocean water from marine abyssalpelagic zone	SAMN14675700	1
HKST-UBA07	GCA_020435085.1	67.49	0.89	1,897,141	China:Hong Kong SAR; Shatin waste water treatment plant	activated sludge from Shatin waste water treatment plant collected monthly from 2007 through 2015	SAMN14562954	1
Co_bin.56	GCA_030432925.1	92.06	0.94	3,043,218	South Korea	biofloc rearing seawater	SAMN36083972	1
metabat2_LS01_001_emilycovarybin_with_cov.176	this study	96.88	0.89	3,974,293	USA	LS01	NA	NA

Supplemental Table 14. Metadata for JABSOQ01 bins in GTDB. Three species have been deposited under genus JABSOQ01.

name NCBI strainID	link	completeness	contamination	genome size	country	isolation_source	biosample	genomes in species cluster
CK1056 this is Oceanipulchritudo coccoides	GCF_010500615.1	100	0.76	3,520,043	China: Weihai	sediment from intertidal zone	SAMN14001152	1
GSL.Bin15	GCA_009908535.1	91.22	6.76	2,711,473	USA: Bridger Bay in Great Salt Lake; Utah	Microbial Mat	SAMN13763256	1
CSMAG_288	GCA_030595525.1	85.81	0.2	2,271,176	NA	cold seep	SAMN34043555	1
PD13Bin-3-4	GCA_040830445.1	96.62	1.35	3,073,241	USA: Maunalua Bay, Oahu, Hawaii	metagenome from Accessory nidamental gland in Hawaiian bobtail squid	SAMN35715522	1
PD13Bin-3-1	GCA_040830495.1	99.32	0.68	5,070,320	USA: Maunalua Bay, Oahu, Hawaii	metagenome from Accessory nidamental gland in Hawaiian bobtail squid	SAMN35715521	1
E07Bin-2-1	GCA_040830765.1	99.32	1.49	3,993,000	USA: Maunalua Bay, Oahu, Hawaii	metagenome from Accessory nidamental gland in Hawaiian bobtail squid	SAMN35715514	1
muddycreek_2019_sw_WHONDRS-S19S_0082_B_bin.38	GCA_041163965.1	67.93	1.72	1,855,021	USA: Edgewater, Maryland	river surface water	SAMN37153627	1
Mix_sand19.2022.05.G.36	GCA_041623435.1	96.11	1.38	4,406,227	China: ZhouShan	Coastal bare tidal sandflat sediment	SAMN42105115	1
21BLK6-1.bin.89	GCA_041685165.1	99.32	0.47	3,453,401	China: Barkol Lake	saline lake sediment	SAMN37277901	1
SRR6201619_bin.15_MetaBAT_v2.12.1_MAG	GCA_913056355.1	94.12	2.06	4,308,958	Puerto Rico	marine water]	SAMEA9694326	1
metabat2_LS01_024_emilycovarybin_with_cov.26	this study	86.34	2.9	3,033,125	USA	LS01	NA	NA

Supplemental Table 15. Metadata for CK1056/*Oceanipulchritudo* bins in GTDB. CK1056 has now been reclassified to *Oceanipulchritudo* and spans 10 species, the type species of the genus is *Oceanipulchritudo coccoides*.

name NCBI strainID	link	completeness	contamination	genome size	country	isolation_source	biosample	genomes in species cluster
DLM2.Bin81	GCA_007693725.1	92.05	0	2,709,830	Canada: Cariboo Plateau	alkaline salt lake	SAMN10237383	1
JGIbin_sample002_101	this study	97.73	0	4,047,951	USA	LS01	NA	NA

Supplemental Table 16. Metadata for REDD01 bins in GTDB. Genus REDD01 has only one representative species.

name NCBI strainID	link	completeness	contamination	genome size	country	isolation_source	biosample	genomes in species cluster
UBA12170	GCA_003445975.1	97.04	1.08	2,947,046	NA	sediment	SAMN08019503	1
Zod_Metabat.155	GCA_016929955.1	87.46	2.78	2,845,310	USA: Anadarko; OK; Zodletone Spring	Freshwater Sediment	SAMN17241944	1
Zod_maxbin.0163	GCA_016933745.1	72.42	1.25	2,446,505	USA: Anadarko; OK; Zodletone Spring	Freshwater Sediment	SAMN17241752	1
SulCav_PC08_3_6_38_10	GCA_020718315.1	69.34	0.39	1,785,947	Italy: Frasassi cave	biofilm from sulfidic cave	SAMN22209628	1
AUK121	GCA_021157585.1	96.24	2.15	2,930,121	Pacific Ocean: Mexico; Pescadero Basin	marine hydrothermal sediment	SAMN18353617	1
AUK077	GCA_021158405.1	96.51	2.42	2,935,165	Pacific Ocean: Mexico; Pescadero Basin	marine hydrothermal sediment	SAMN18353573	1
GH48	GCA_027858955.1	91.94	1.34	3,462,441	Canada: Nunavut	high Arctic cold saline sulfur spring anaerobic sediment	SAMN32472399	1
Mix_sand08.2021.06.G.64	GCA_041592395.1	82.76	1.72	3,010,051	China: ZhouShan	Coastal bare tidal sandflat sediment	SAMN42103584	1
JGIbin_sample001_24	this study	98.92	1.61	4,454,126	USA	LS01	NA	NA

Supplemental Table 17. Metadata for UBA12170 bins in GTDB. Genus UBA12170 spans 8 species.

name NCBI strainID	link	completeness	contamination	genome size	country	isolation_source	biosample	genomes in species cluster
RM1_1_3	GCA_012033595.1	95.44	4.37	5,746,938	South Africa: Cape Recife	Stromatolite	SAMN14376482	1
S146_16	GCA_015492915.1	86.79	4.04	4,642,359	Pacific Ocean	deep-sea hydrothermal deposit	SAMN12876737	1
MAG.91	GCA_022360055.1	97.62	4.62	7,547,564	Saudi Arabia: Abo Shoosha reef	Porites lutea and Goniastrea edwardsi coral colonies	SAMN22830512	1
Dive100_bin10.8	GCA_024280115.1	95.54	4.24	7,222,043	Indian Ocean: Longqi hydrothermal vent field; ultraslow-spreading Southwestern Indian Ridge	diffuse vent DFF1	SAMN22253236	1
Ww83	GCA_024643455.1	93.9	4.18	5,331,874	New Zealand: Waiwera River; Auckland	sediment	SAMN19290135	1
3_bin.84	GCA_030426345.1	97.72	2.59	5,115,568	South Korea	biofloc rearing seawater	SAMN36083635	1
CSMAG_1701	GCA_030637245.1	73.16	2.12	2,709,902	NA	cold seep	SAMN34049714	1
Salar_de_Ascotan_sediment_MAG_ASE13	GCA_034523095.1	91.22	1.44	5,068,466	Chile: Atacama desert	Salar de Ascotan sediment	SAMN38083308	1
JTS-AC485	GCA_041750735.1	96.43	5.31	5,893,382	Norway: Svalbard, Jotun geothermal spring	biofilm from active travertine-forming spring	SAMN41305964	1
JGIbin_sample024_58	this study	97.02	1.88	4,996,867	USA	LS01	NA	NA
JGIbin_sample024_64	this study	95.54	3.77	7,008,344	USA	LS01	NA	NA

Supplemental Table 18. Metadata for JAAUQL01 bins in GTDB. Genus JAAUQL01 spans 9 species.

name NCBI strainID	link	completeness	contamination	genome size	country	isolation_source	biosample	genomes in species cluster
DT_110	GCA_013214265.1	91.13	0.51	5,070,646	Pacific Ocean: North Pacific Gyre; Station ALOHA	ocean water from marine abyssalpelagic zone	SAMN14675798	1
BWSM_29	GCA_030148525.1	83.2	0.42	5,032,938	China:Wuxi	ballast tank	SAMN34434854	1
JGIbin_sample002_109	this study	93.69	5.81	6,894,568	USA	LS01	NA	NA

Supplemental Table 19. Metadata for DT-110 bins in GTDB. Genus DT-110 spans 2 species

name NCBI strainID	link	completeness	contamination	genome size	country	isolation_source	biosample	genomes in species cluster
28_1	GCA_008080835.1	91.03	0.95	2,385,933	USA: coastal Georgia; Sapelo Island	Coastal seawater	SAMN12255932	1
HK-STAS-BACT-72	GCA_020635435.1	77.6	0	2,889,272	China: Hong Kong Special Administrative Region	activated sludge from Hong Kong Shatin wastewater treatment plant	SAMN15687651	1
JGIbin_sample002_28	this study	84.06	5.32	2,736,413	USA	LS01	NA	NA

Supplemental Table 20. Metadata for VMDK01 bins in GTDB. Genus VMDK01 spans 2 species.

category	annotation type	core count	core proportion	flexible count	flexible proportion	odds ratio	p_value	enriched in	p_adjusted
Defense mechanisms	COG Category	775	0.061	68	0.201	3.879	1.66E-17	Flexible	3.99E-16
Coenzyme transport and metabolism	COG Category	844	0.066	10	0.03	0.429	4.99E-03	Core	3.12E-02
Intracellular trafficking, secretion, and vesicular transport	COG Category	221	0.017	0	0	0	4.75E-03	Core	3.12E-02
Nucleotide transport and metabolism	COG Category	234	0.018	0	0	0	5.20E-03	Core	3.12E-02
Cell motility	COG Category	192	0.015	0	0	0	1.06E-02	Core	4.22E-02
Mobilome: prophages, transposons	COG Category	1479	0.116	24	0.071	0.581	9.30E-03	Core	4.22E-02
CRISPR-Cas system type III CMR-effector complex subunit Cmr4, RAMP superfamily Cas7 group (Cmr4) (PDB:4WNZ)	COG Function	0	0	4	0.012	INF	4.42E-07	Flexible	4.31E-04
tRNA1(Val) A37 N6-methylase TrmN6 (TrmN6)	COG Function	0	0	4	0.012	INF	4.42E-07	Flexible	4.31E-04
Outer membrane porin OmpC/OmpF/PhoE (OmpC) (PDB:1PHO)	COG Function	3	0	4	0.012	50.619	1.45E-05	Flexible	5.67E-03
Transcriptional regulator, contains XRE-family HTH domain (HipB) (PDB:1ADR)	COG Function	3	0	4	0.012	50.619	1.45E-05	Flexible	5.67E-03
Zinc transporter ZupT (ZupT) (PDB:5TSA)	COG Function	3	0	4	0.012	50.619	1.45E-05	Flexible	5.67E-03
CRISPR/Cas system-associated protein Cas10, large subunit of type III CRISPR-Cas systems, contains HD superfamily nuclease domain (Cas10) (PDB:6IFN)	COG Function	6	0	4	0.012	25.329	8.20E-05	Flexible	2.66E-02
CRISPR-Cas system type III CSM-effector complex subunit Csm3, RAMP superfamily Cas7 group (Csm3) (PDB:6IFN)	COG Function	12	0.001	4	0.012	12.665	6.28E-04	Flexible	3.26E-02
CRISPR/Cas system-associated endoribonuclease Cas2 (Cas2) (PDB:2I0X)	COG Function	16	0.001	5	0.015	11.908	1.64E-04	Flexible	3.26E-02
Sulfite exporter TauE/SafE/YfcA and related permeases, UPF0721 family (TauE) (PUBMED:21183667;22797525)	COG Function	12	0.001	4	0.012	12.665	6.28E-04	Flexible	3.26E-02
1-phosphofructokinase or 6-phosphofructokinase II (FruK) (PDB:2ABQ)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
ADP-heptose:LPS heptosyltransferase (RfaF) (PDB:1PSW)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Alkylated DNA repair dioxygenase AlkB (AlkB) (PDB:3KHC)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Antitoxin component of the RelBE or YafQ-DinJ toxin-antitoxin module (RelB) (PDB:4ML0)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Bacillopeptidase F, M6 metalloprotease family	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
CRISPR-Cas system type III CMR-effector complex small subunit Cmr5 (Cmr5) (PDB:2OEB)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
CRISPR-Cas system type III CMR-effector complex subunit Cmr1, RAMP superfamily Cas7 group (Cmr1) (PDB:4L6U)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
CRISPR-Cas system type III CMR-effector complex subunit Cmr3, RAMP superfamily Cas5 group (Cmr3)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
CRISPR-Cas system-associated protein Csm6, contains CARF domain (Csm6) (PDB:3QYF)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
CRISPR/Cas system CMR subunit Cmr6, Cas7 group, RAMP superfamily (Cmr6) (PDB:3X1L)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Chloramphenicol O-acetyltransferase (CatA) (PDB:1CIA)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Cys-tRNA(Pro) deacylase, prolyl-tRNA editing enzyme YbaK/EbsC (EbsC) (PDB:1DBU)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Esterase/lipase (YvaK) (PDB:4DIU)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Holliday junction resolvasome RuvABC, ATP-dependent DNA helicase subunit RuvB (RuvB) (PDB:1HQC)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02

Inorganic pyrophosphatase (Ppa) (PDB:117E)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Methionine synthase I (cobalamin-dependent), methyltransferase domain (MetH1) (PDB:1LT7)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Methionine synthase I, cobalamin-binding domain (MetH2) (PDB:4DJD)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Mg2+ and Co2+ transporter CorA (CorA) (PDB:2BBH)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Na+ or H+-translocating membrane pyrophosphatase (OVP1) (PDB:4A01) (PUBMED:11342551)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Na+/glutamate symporter (GltS)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Predicted exporter protein, RND superfamily (MMPL) (PDB:5KHN)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Predicted membrane chloride channel, bestrophin family (YneE) (PDB:4RDQ)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Pyoverdine/dityrosine biosynthesis protein Dit1 (Dit1) (PDB:3E59) (PUBMED:18156253)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Restriction endonuclease Mrr (Mrr)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Superfamily I DNA and/or RNA helicase (DNA2) (PDB:2GJK)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Tetrahydrodipicolinate N-succinyltransferase (DapD) (PDB:3BXY)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Uncharacterized membrane protein (PDB:1EUT)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Uncharacterized protein, contains ferritin-like DUF455 domain	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Very-short-patch-repair endonuclease (YcjD) (PDB:3HRL)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
mRNA-degrading endonuclease YafQ (mRNA interferase), toxin component of the YafQ-DinJ toxin-antitoxin module (YafQ) (PDB:4MMG)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
tRNA G10 N-methylase Trm11 (Trm11) (PDB:5E71)	COG Function	0	0	2	0.006	INF	6.69E-04	Flexible	3.26E-02
Multidrug efflux pump subunit AcrB (AcrB) (PDB:1IWG)	COG Function	38	0.003	6	0.018	6.026	8.88E-04	Flexible	4.22E-02
REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	COG Function	286	0.022	0	0	0	9.84E-04	Core	4.57E-02
Pathway modules; Amino acid metabolism; Lysine metabolism	KEGG Class	33	0.022	6	0.222	12.71	3.61E-05	Flexible	1.23E-03
Pathway modules; Energy metabolism; Methane metabolism	KEGG Class	40	0.027	6	0.222	10.446	9.53E-05	Flexible	1.62E-03
Signature modules; Gene set; Drug resistance	KEGG Class	33	0.022	4	0.148	7.752	3.37E-03	Flexible	3.82E-02
5-methyltetrahydrofolate--homocysteine methyltransferase [EC:2.1.1.13]	KEGG KOfam	0	0	4	0.02	INF	3.11E-07	Flexible	5.48E-04
CRISPR-associated protein Cas2	KEGG KOfam	16	0.002	5	0.025	13.111	1.08E-04	Flexible	2.48E-02
putative transposase	KEGG KOfam	442	0.054	0	0	0	3.39E-05	Core	2.48E-02
transposase	KEGG KOfam	169	0.02	13	0.065	3.305	3.87E-04	Flexible	2.48E-02
1-phosphofructokinase [EC:2.7.1.56]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase [EC:2.3.1.117]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
ATP-dependent phosphofructokinase / diphosphate-dependent phosphofructokinase [EC:2.7.1.11 2.7.1.90]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
Ala-tRNA(Pro) deacylase [EC:3.1.1.-]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02

CRISPR-associated protein Cmr1	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
CRISPR-associated protein Cmr2	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
CRISPR-associated protein Cmr3	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
CRISPR-associated protein Cmr4	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
CRISPR-associated protein Cmr5	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
CRISPR-associated protein Cmr6	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
DNA-damage-inducible protein J	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
E3 ISG15--protein ligase [EC:2.3.2.-]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
HTH-type transcriptional regulator, cell division transcriptional repressor	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
K(+)-stimulated pyrophosphate-energized sodium pump [EC:7.2.3.1]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
L-tyrosine isonitrile desaturase/decarboxylase [EC:1.14.20.9 1.14.20.10]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
L-tyrosine isonitrile synthase [EC:4.1.99.24]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
MFS transporter, PAT family, beta-lactamase induction signal transducer AmpG	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
N-succinyldiaminopimelate aminotransferase [EC:2.6.1.17]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
Nif-specific regulatory protein	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
arsenate reductase (glutaredoxin) [EC:1.20.4.1]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
chloramphenicol O-acetyltransferase type A [EC:2.3.1.28]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
glutamate:Na+ symporter, ESS family	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
glyoxylase I family protein	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
holliday junction DNA helicase RuvB [EC:5.6.2.4]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
immune inhibitor A [EC:3.4.24.-]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
inorganic pyrophosphatase [EC:3.6.1.1]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
ion channel-forming bestrophin family protein	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
lipopolysaccharide heptosyltransferase I [EC:2.4.99.23]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
malonyl-CoA/methylmalonyl-CoA synthetase [EC:6.2.1.76 6.2.1.-]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
nitrogen fixation regulatory protein	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
putative acetyltransferase [EC:2.3.1.-]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
sodium/hydrogen antiporter	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
succinyl-diaminopimelate desuccinylase [EC:3.5.1.18]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
tRNA (cmo5U34)-methyltransferase [EC:2.1.1.-]	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
two-component system, OmpR family, response regulator CpxR	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
two-component system, sensor histidine kinase LadS	KEGG KOfam	0	0	2	0.01	INF	5.63E-04	Flexible	2.48E-02
adenine-specific DNA-methyltransferase [EC:2.1.1.72]	KEGG KOfam	15	0.002	4	0.02	11.135	9.10E-04	Flexible	3.91E-02

Supplemental Table 21. Significantly enriched categories in either the core or flexible genome of PB-PSB1 across different annotation types. (COG Category, COG function, KEGG Class, KEGG KOfam)

Chapter II Supplemental Information

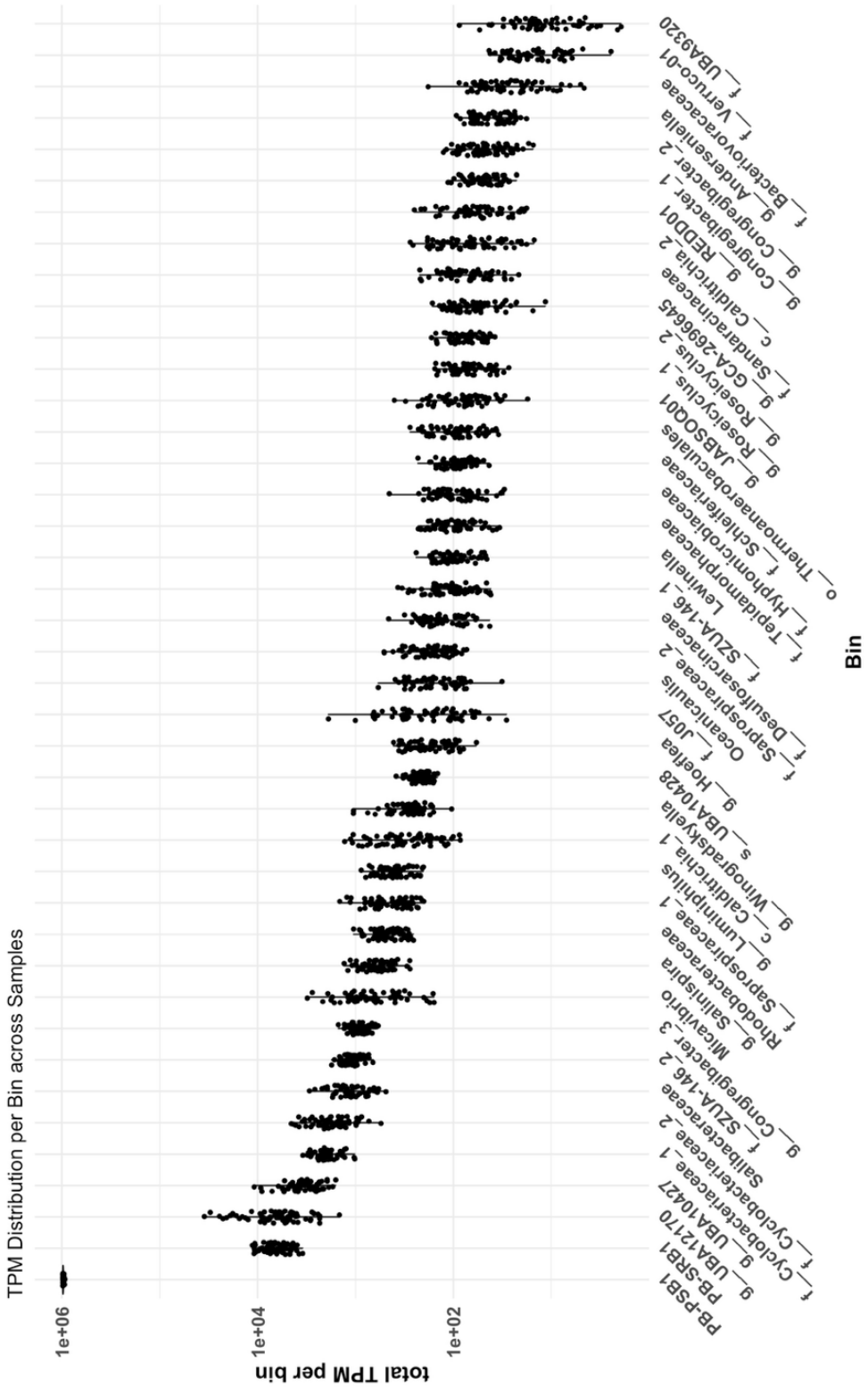


Figure 2-A. Transcriptional activity in our lab diel incubations across our pink berry bins. Visualized is total TPM per bin summed across samples, $\log_{10}$-scaled, and ordered by distribution's median in descending order.

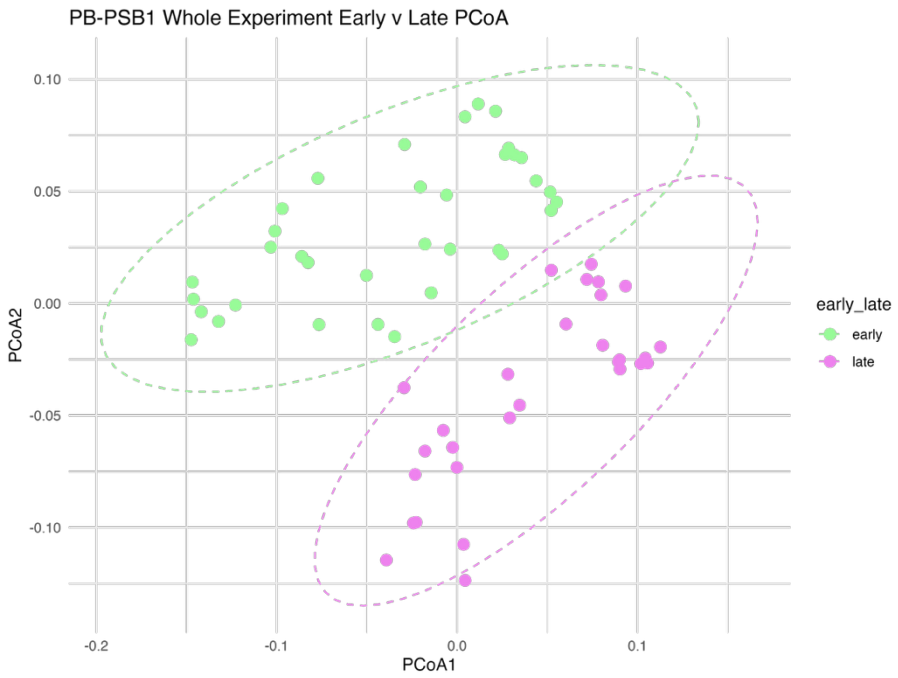


Figure 2-B. Differences in PB-PSB1 transcripts between the first 12 time points and last 10 time points of experiment (3 replicates per time point). PCoA for PB-PSB1 using Bray-Curtis distance where 95% confidence ellipses are based on incubation early v. late conditions, and corresponding early-late condition PERMANOVA results are F=15.343, R^2=0.19584, p= 0.001. Lime corresponds to beginning of experiment while pink corresponds to second half.

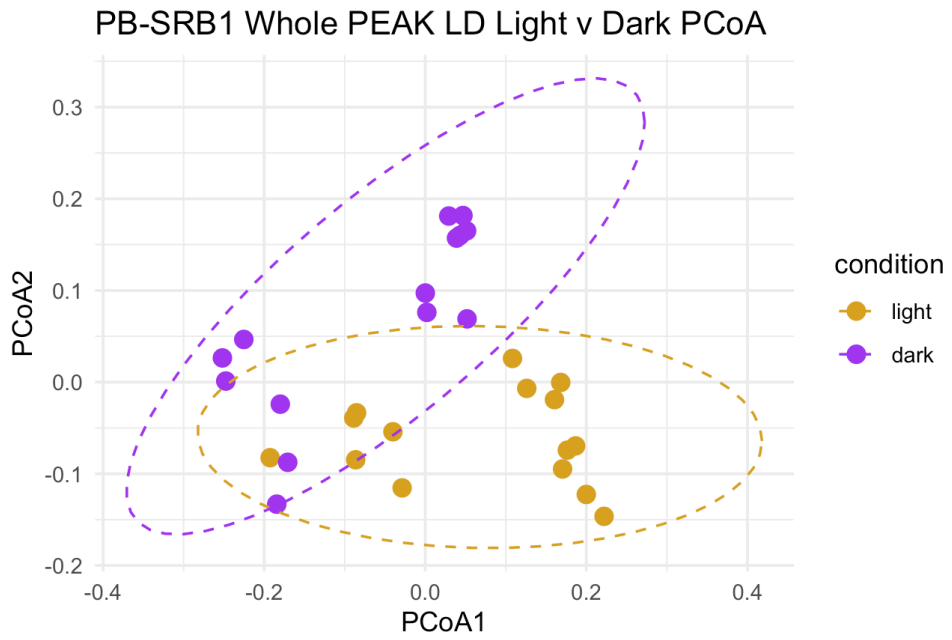


Figure 2-C. Differences in PB-SRB1 transcripts between light-dark conditions for all time points taken at peak light-dark conditions. PCoA for PB-SRB1 using Bray-Curtis distance where 95% confidence ellipses are based on incubation light v. dark conditions, and corresponding light-dark condition PERMANOVA results are $F=5.5388$, $R^2=0.17022$, $p= 0.002$. Gold tones correspond to light time points, and purple tones correspond to dark time points.

PB-PSB1 comparisons	samples compared for light v dark differences
alldawn	all pre-dawn and post-dawn time points
alldusk	all pre-dusk and post-dusk time points
dawn2	pre-dawn and post-dawn on second day of incubation
dawn3	pre-dawn and post-dawn on third day of incubation
dawn4	pre-dawn and post-dawn on fourth day of incubation
dusk1	pre-dusk and post-dusk on first day of incubation
dusk2	pre-dusk and post-dusk on second day of incubation
dusk3	pre-dusk and post-dusk on third day of incubation
peakLvpeakD	all peak light and peak dark time points sampled(noon and 1AM)
wholeexperimentLvD	all time points in experiment(peaks and transitions)

Supplemental Table 1. Different light-dark significance comparisons for PB-PSB1 run between different subsets of samples in our diel time course. This populated the significant genes seen in Figure 2-7.

PB-SRB1 comparisons	samples compared for light v dark differences
earlyallLD	all time points for the first three days of the experiment
earlypeakLD	peak time points for the first three days of the experiment
lateallLD	all time points for the second half of the experiment
latepeakLD	peak time points for the second half of the experiment
lightshift	time points artificially shifted for light conditions, such that post dusk is considered a light time point and post dawn is considered a dark time point, peaks mark the beginning of a light period
wholeexperimentLvD	all time points in experiment(peaks and transitions)
wholepeakLD	all peak light and peak dark time points sampled(noon and 1AM)

Supplemental Table 2. Different light-dark significance comparisons for PB-SRB1 run between different subsets of samples in our diel time course. This populated the significant genes seen in Figure 2-8.

Gene	product	cluster	NCBI_PGAP	COG20_FUNCTION	KOfam
PSB__G3066	transposase	1	transposase	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB__G4756	hypothetical protein	1	NA	NA	NA
PSB__G495	hypothetical protein	1	NA	NA	NA
PSB__G4735	IS5 family transposase	1	IS5 family transposase	Transposase	putative transposase
PSB__G3851	hypothetical protein	1	NA	NA	NA
PSB__G5331	hypothetical protein	1	NA	NA	NA
PSB__G4249	four helix bundle protein	1	four helix bundle protein	NA	NA
PSB__G6360	hypothetical protein	1	four helix bundle protein	NA	NA
PSB__G5728	ISAs1 family transposase	1	ISAs1 family transposase	Predicted transposase YbfD/YdcC associated with H repeats (YhhI)	NA
PSB__G4588	hypothetical protein	1	NA	NA	NA
PSB__G5171	hypothetical protein	1	NA	NA	NA
PSB__G153	efflux RND transporter permease subunit	1	multidrug efflux RND transporter permease subunit	Multidrug efflux pump subunit AcrB (AcrB) (PDB:1IWG)	multidrug efflux pump
PSB__G4024	DUF2750 domain-containing protein	1	DUF2750 domain-containing protein	NA	NA
PSB__G2768	class I SAM-dependent methyltransferase	1	class I SAM-dependent methyltransferase	2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase (UbiG) (PDB:1BHJ)	NA
PSB__G3472	HIT family protein	1	HIT family protein	Purine nucleoside phosphoramidase/Ap4A hydrolase, histidine triade (HIT) family (HinT) (PDB:1AV5) (PUBMED:20934431)	histidine triad (HIT) family protein
PSB__G6182	IS5 family transposase	1	IS5 family transposase	Transposase	putative transposase
PSB__G4371	hypothetical protein	1	NA	NA	NA
PSB__G2058	transposase	1	transposase	dTDP-4-amino-4,6-dideoxygalactose transaminase (WecE) (PDB:4PIW) (PUBMED:15271350)	NA
PSB__G6730	hypothetical protein	1	NA	NA	NA
PSB__G1424	carbohydrate ABC transporter permease	1	carbohydrate ABC transporter permease	ABC-type glycerol-3-phosphate transport system, permease component (UgpE) (PDB:1VR4)	alpha-glucoside transport system permease protein
PSB__G2025	hypothetical protein	1	NA	NA	NA
PSB__G4958	hypothetical protein	1	NA	dTDP-4-amino-4,6-dideoxygalactose transaminase (WecE) (PDB:4PIW) (PUBMED:15271350)	NA
PSB__G4455	hypothetical protein	1	NA	NA	NA
PSB__G4760	acyl carrier protein	1	acyl carrier protein (acpP)	Acyl carrier protein (AcpP) (PDB:1ACP)	acyl carrier protein
PSB__G3899	S24 family peptidase	1	S24 family peptidase	Phage repressor protein C, contains Cro/C1-type HTH and peptisase s24 domains	NA
PSB__G3530	sulfur carrier protein DsrE2	1	peroxiredoxin family protein	Peroxiredoxin family protein (YrkE) (PDB:2QS7)	uncharacterized protein
PSB__G1453	response regulator	1	response regulator	CheY-like REC (receiver) domain, includes chemotaxis protein CheY and sporulation regulator Spo0F (CheY) (PDB:6QRJ)!!!Signal transduction histidine kinase ComP (ComP) (PDB:4GT8)	two-component system, chemotaxis family, chemotaxis protein CheY
PSB__G4939	hypothetical protein	1	NA	NA	NA
PSB__G2180	protease HtpX	1	protease HtpX (htpX)	Zn-dependent protease with chaperone function (HtpX) (PDB:2YPT)	heat shock protein HtpX [EC:3.4.24.-]
PSB__G1376	hypothetical protein	1	NA	NA	NA
PSB__G535	adenylyl-sulfate reductase subunit alpha	1	adenylyl-sulfate reductase subunit alpha	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (SdhA) (PDB:1KF6)	adenylylsulfate reductase, subunit A [EC:1.8.99.2]
PSB__G3516	adenylyl-sulfate reductase subunit beta	1	adenylyl-sulfate reductase subunit beta (aprB)	NAD-dependent dihydropyrimidine dehydrogenase, PreA subunit (PreA)	adenylylsulfate reductase, subunit B [EC:1.8.99.2]
PSB__G6891	adenylyl-sulfate reductase	1	adenylyl-sulfate reductase	NA	NA
PSB__G4531	DUF6129 family protein	1	DUF6129 family protein	NA	NA
PSB__G2709	dinitrogenase iron-molybdenum cofactor biosynthesis protein	1	dinitrogenase iron-molybdenum cofactor biosynthesis protein	Predicted Fe-Mo cluster-binding protein, NifX family (NifX) (PDB:1EO1)	nitrogen fixation protein NifX
PSB__G838	nitrogenase molybdenum-iron protein alpha chain	1	nitrogenase molybdenum-iron protein alpha chain (nifD)	Nitrogenase Mo-Fe protein NifD/coenzyme F430 biosynthesis subunit CfbD (NifD/CfbD) (PDB:1FP4) (PUBMED:28225763)	nitrogenase molybdenum-iron protein alpha chain [EC:1.18.6.1]
PSB__G4633	hypothetical protein	1	NA	NA	NA
PSB__G5107	hypothetical protein	1	NA	NA	NA
PSB__G3055	GyrI-like domain-containing protein	1	GyrI-like domain-containing protein	Cyclopropanoid cyclopropyl hydrolase, GyrI-like domain (MA1133) (PDB:3B49) (PUBMED:29133784)	NA
PSB__G6810	hypothetical protein	1	NA	NA	NA
PSB__G6211	hypothetical protein	1	NA	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	NA
PSB__G878	chlorophyllide a reductase subunit Z	1	chlorophyllide a reductase subunit Z (bchZ)	Bacteriochlorophyllide reductase subunit BchZ (BchZ) (PUBMED:21886856)	3,8-divinyl chlorophyllide a/chlorophyllide a reductase subunit Z [EC:1.3.7.14 1.3.7.15]
PSB__G5312	light-harvesting antenna LH1, beta subunit	1	light-harvesting protein	NA	light-harvesting complex 1 beta chain
PSB__G5313	light-harvesting antenna LH1, beta subunit	1	light-harvesting protein	NA	light-harvesting complex 1 beta chain
PSB__G521	lipid A export permease/ATP-binding protein MsbA	1	lipid A export permease/ATP-binding protein MsbA (msbA)	ABC-type multidrug transport system, ATPase and permease component (MdlB) (PDB:2HYD)	ATP-binding cassette, subfamily B, bacterial MsbA [EC:7.5.2.6]
PSB__G2655	electron transport complex subunit E	1	electron transport complex subunit E	Na+-translocating ferredoxin:NAD+ oxidoreductase RNF, RnfE subunit (RnfE)	H+/Na+-translocating ferredoxin:NAD+ oxidoreductase subunit E
PSB__G2742	electron transport complex subunit RsxG	1	electron transport complex subunit RsxG (rsxG)	Na+-translocating ferredoxin:NAD+ oxidoreductase RNF, RnfG subunit (RnfG)	H+/Na+-translocating ferredoxin:NAD+ oxidoreductase subunit G
PSB__G1554	RnfABCDGE type electron transport complex subunit D	1	RnfABCDGE type electron transport complex subunit D	Na+-translocating ferredoxin:NAD+ oxidoreductase RNF, RnfD subunit (RnfD)	H+/Na+-translocating ferredoxin:NAD+ oxidoreductase subunit D [EC:7.1.1.11 7.2.1.2]
PSB__G3160	electron transport complex subunit RsxA	1	electron transport complex subunit RsxA (rsxA)	Na+-translocating ferredoxin:NAD+ oxidoreductase RNF, RnfA subunit (RnfA)	H+/Na+-translocating ferredoxin:NAD+ oxidoreductase subunit A
PSB__G7108	IS66 family transposase	1	DUF4338 domain-containing protein	NA	NA
PSB__G5046	hypothetical protein	1	NA	NA	NA

PSB__G443	hypothetical protein	1	NA	Alpha-tubulin suppressor ATS1 and related RCC1 domain-containing proteins (ATS1) (PDB:1A12)	regulator of chromosome condensation
PSB__G6971	hypothetical protein	1	NA	NA	NA
PSB__G1037	FprA family A-type flavoprotein	1	FprA family A-type flavoprotein	Flavorubredoxin (NorV) (PDB:4JHC)	NADH oxidase (H2O-forming) [EC:1.6.3.4]
PSB__G726	nitrogenase cofactor biosynthesis protein NifB	1	nitrogenase cofactor biosynthesis protein NifB (nifB)	Radical SAM superfamily maturase, SkfB/NifB/PqqE family (SkfB) (PDB:6Q2P) (PUBMED:23282011)!!!Predicted Fe-Mo cluster-binding protein, NifX family (NifX) (PDB:1EO1)	nitrogen fixation protein NifB
PSB__G492	SurA N-terminal domain-containing protein	1	peptidylprolyl isomerase	Peptidyl-prolyl isomerase, parvulin family (SurA) (PDB:2ZR5)	peptidyl-prolyl cis-trans isomerase D [EC:5.2.1.8]
PSB__G6599	hypothetical protein	1	NA	dTDP-4-amino-4,6-dideoxygalactose transaminase (WecE) (PDB:4PIW) (PUBMED:15271350)	NA
PSB__G3957	Co2+/Mg2+ efflux protein ApaG	1	Co2+/Mg2+ efflux protein ApaG (apaG)	Uncharacterized conserved protein ApaG affecting Mg2+/Co2+ transport (ApaG) (PDB:1TZA)	ApaG protein
PSB__G5177	LysR family transcriptional regulator	1	LysR family transcriptional regulator	DNA-binding transcriptional regulator, LysR family (LysR) (PDB:2FYI)	NA
PSB__G1908	IS630 family transposase	1	IS630 family transposase	Transposase!!!CRISPR-associated protein Csa3, CARF domain (Csa3) (PDB:2WTE) (PUBMED:21093452)	putative transposase
PSB__G4203	DUF6883 domain-containing protein	1	DUF6883 domain-containing protein	NA	NA
PSB__G2903	response regulator transcription factor	1	response regulator transcription factor	DNA-binding response regulator, NarL/FixJ family, contains REC and HTH domains (CitB) (PDB:1A04)	NA
PSB__G4445	hypothetical protein	1	NA	NA	NA
PSB__G3137	NUDIX hydrolase	1	NUDIX hydrolase	8-oxo-dGTP pyrophosphatase MutT and related house-cleaning NTP pyrophosphohydrolases, NUDIX family (MutT) (PDB:4V14)	ADP-ribose diphosphatase [EC:3.6.1.13 3.6.1.-]
PSB__G86	transglutaminase domain-containing protein	1	transglutaminase domain-containing protein	NA	NA
PSB__G2229	transposase	1	transposase	Transposase	NA
PSB__G3496	chromate resistance protein ChrB domain-containing protein	1	chromate resistance protein	Chromate resistance protein ChrB1 (ChrB1)	NA
PSB__G948	nitrogenase iron-molybdenum cofactor biosynthesis protein NifE	1	nitrogenase iron-molybdenum cofactor biosynthesis protein NifE (nifE)	Nitrogenase Mo-Fe protein NifD/coenzyme F430 biosynthesis subunit CfbD (NifD/CfbD) (PDB:1FP4) (PUBMED:28225763)	nitrogenase molybdenum-cofactor synthesis protein NifE
PSB__G3520	SoxR reducing system RseC family protein	1	SoxR reducing system RseC family protein	RseC, positive regulator of sigma E activity (RseC)	sigma-E factor negative regulatory protein RseC
PSB__G3567	NifX-associated nitrogen fixation protein	1	NifX-associated nitrogen fixation protein	NA	NA
PSB__G4319	ferredoxin III, nif-specific	1	ferredoxin III, nif-specific (fdxB)	Formate hydrogenlyase subunit 6/NADH:ubiquinone oxidoreductase 23 kD subunit (chain I) (NuoI) (PDB:1B0P)	NA
PSB__G6966	transposase	1	transposase	dTDP-4-amino-4,6-dideoxygalactose transaminase (WecE) (PDB:4PIW) (PUBMED:15271350)	NA
PSB__G3080	50S ribosomal protein L4	1	50S ribosomal protein L4 (rplD)	Ribosomal protein L4 (RplD) (PDB:1VS6)	large subunit ribosomal protein L4
PSB__G2885	50S ribosomal protein L3	1	50S ribosomal protein L3 (rplC)	Ribosomal protein L3 (RplC) (PDB:1VS6)	large subunit ribosomal protein L3
PSB__G389	elongation factor G	1	elongation factor G (fusA)	Translation elongation factor EF-G, a GTPase (FusA) (PDB:1DAR)	elongation factor G
PSB__G4020	30S ribosomal protein S12	1	30S ribosomal protein S12 (rpsL)	Ribosomal protein S12 (RpsL) (PDB:4V4H)	small subunit ribosomal protein S12
PSB__G1414	IS91 family transposase	1	IS91 family transposase	AraC-type DNA-binding domain and AraC-containing proteins (AraC) (PDB:1BL0)	NA
PSB__G3834	sulfur globule family protein	1	sulfur globule protein CV1	NA	NA
PSB__G6405	AAA family ATPase	1	MoxR family ATPase	MoxR-like ATPase (MoxR) (PDB:2R44)	MoxR-like ATPase [EC:3.6.3.-]
PSB__G2261	arsenite methyltransferase	1	arsenite methyltransferase	Ubiquinone/menaquinone biosynthesis C-methylase UbiE/MenG (UbiE) (PDB:4OBW)	arsenite methyltransferase [EC:2.1.1.137]
PSB__G6018	ISL3 family transposase	1	ISL3 family transposase	Transposase	transposase
PSB__G2281	phosphate ABC transporter ATP-binding protein PstB	1	phosphate ABC transporter ATP-binding protein	ABC-type phosphate transport system, ATPase component (PstB)	phosphate transport system ATP-binding protein [EC:7.3.2.1]
PSB__G1646	phosphate ABC transporter substrate-binding protein PstS	1	phosphate ABC transporter substrate-binding protein PstS (pstS)	ABC-type phosphate transport system, periplasmic component (PstS) (PDB:2Q9T)	phosphate transport system substrate-binding protein
PSB__G3835	hypothetical protein	1	NA	Transposase	NA
PSB__G6684	DUF6920 family protein	1	DUF6920 family protein	NA	NA
PSB__G2275	type II CAAX prenyl endopeptidase Rce1 family protein	1	CPBP family intramembrane metalloprotease	Membrane protease YdiL, CAAX protease family (YdiL)	CAAX protease family protein
PSB__G6668	ATP synthase subunit I	1	F0F1 ATP synthase assembly protein I	FoF1-type ATP synthase accessory protein AtpI (AtpI) (PUBMED:18083842;23123906)	ATP synthase protein I
PSB__G4705	hypothetical protein	1	NA	NA	NA
PSB__G143	tubulin-like doman-containing protein	1	tubulin-like doman-containing protein	NA	NA
PSB__G1709	hypothetical protein	1	NA	Nucleoside phosphorylase/nucleosidase, includes 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase MtnN and futalosine hydrolase MqnB (MtnN) (PDB:1T8R) (PUBMED:18801996)	adenosylhomocysteine nucleosidase [EC:3.2.2.9]
PSB__G4293	hypothetical protein	1	NA	NA	NA
PSB__G6532	arylesterase	1	arylesterase	Lysophospholipase L1 or related esterase. Includes spore coat protein LipC/YcsK (TesA) (PDB:1BWP)	acyl-CoA thioesterase I [EC:3.1.2.- 3.1.2.2 3.1.1.2 3.1.1.5]
PSB__G3941	sulfurtransferase complex subunit TusD	1	sulfurtransferase complex subunit TusD (tusD)	Sulfur relay (sulfurtransferase) complex TusBCD TusD component, DsrE family (DsrE) (PDB:1JX7)	tRNA 2-thiouridine synthesizing protein D [EC:2.8.1.-]
PSB__G1595	phospho-N-acetylmuramoyl-pentapeptide- transferase	1	phospho-N-acetylmuramoyl-pentapeptide-transferase	UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase (Rfe) (PDB:4J72)	phospho-N-acetylmuramoyl-pentapeptide-transferase [EC:2.7.8.13]
PSB__G3660	hypothetical protein	1	NA	NA	NA

PSB__G210	XrtA/PEP-CTERM system TPR-repeat protein PrsT	1	PEP-CTERM system TPR-repeat protein PrsT (prsT)	Type IV pilus assembly protein PilF/PilW (PilF) (PDB:2VQ2) (PUBMED:16403447)!!!Flp pilus assembly protein TadD, contains TPR repeats (TadD) (PDB:3WXX)	NA
PSB__G4689	hypothetical protein	1	NA	NA	NA
PSB__G2641	hypothetical protein	1	NA	NA	NA
PSB__G6816	pilus assembly protein PilP	1	pilus assembly protein PilP	Type IV pilus assembly protein PilP (PilP) (PDB:2IVW)	type IV pilus assembly protein PilP
PSB__G6956	hypothetical protein	1	asparagine synthase	Asparagine synthetase B (glutamine-hydrolyzing) (AsnB) (PDB:2ETD) (PUBMED:25951518)	asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4]
PSB__G4060	hypothetical protein	1	NA	NA	NA
PSB__G1902	transaldolase	1	transaldolase (tal)	Transaldolase/fructose-6-phosphate aldolase (TalA) (PDB:1F05)	transaldolase [EC:2.2.1.2]
PSB__G4383	hypothetical protein	1	NA	NA	NA
PSB__G6547	hypothetical protein	1	four helix bundle protein	NA	NA
PSB__G2983	TSUP family transporter	1	sulfite exporter TauE/SafE family protein	NA	NA
PSB__G1915	RICIN domain-containing protein	1	RICIN domain-containing protein	NA	NA
PSB__G84	hypothetical protein	1	NA	NA	NA
PSB__G2332	hypothetical protein	1	NA	NA	NA
PSB__G4825	hypothetical protein	1	NA	NA	NA
PSB__G4821	hypothetical protein	2	NA	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	NA
PSB__G2170	hypothetical protein	2	NA	NA	NA
PSB__G5365	hypothetical protein	2	NA	NA	NA
PSB__G3560	transposase	2	NA	NA	NA
PSB__G5409	transposase	2	NA	NA	NA
PSB__G5388	hypothetical protein	2	NA	NA	NA
PSB__G6800	GNAT family N-acetyltransferase	2	GNAT family N-acetyltransferase	Ribosomal protein S18 acetylase RimI and related acetyltransferases (RimI) (PDB:1GHE)	NA
PSB__G3627	hypothetical protein	2	NA	NA	NA
PSB__G5346	transposase	2	transposase	NA	NA
PSB__G6225	hypothetical protein	2	NA	dTDP-4-amino-4,6-dideoxygalactose transaminase (WecE) (PDB:4PIW) (PUBMED:15271350)	NA
PSB__G3425	DUF2452 domain-containing protein	2	DUF2452 domain-containing protein	NA	NA
PSB__G759	mechanosensitive ion channel domain-containing protein	2	mechanosensitive ion channel	Small-conductance mechanosensitive channel MscK (MscK)	NA
PSB__G3873	hypothetical protein	2	NA	NA	NA
PSB__G4340	hypothetical protein	2	NA	NA	NA
PSB__G4897	hypothetical protein	2	NA	NA	NA
PSB__G3268	CIA30 family protein	2	NADH:ubiquinone oxidoreductase	Nucleoside-diphosphate-sugar epimerase (WcaG) (PDB:2X4G)!!!Uncharacterized conserved protein YbjT, contains NAD(P)-binding and DUF2867 domains (YbjT) (PDB:5L3Z)	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex assembly factor 1
PSB__G1245	Fic family protein	2	Fic family protein	Fic family protein (PDB:3CUC)	cell filamentation protein, protein adenylyltransferase [EC:2.7.7.108]
PSB__G2898	NUDIX hydrolase	2	CoA pyrophosphatase	ADP-ribose pyrophosphatase YjhB, NUDIX family (YjhB) (PDB:3GZ5)	NA
PSB__G5104	hypothetical protein	2	NA	NA	NA
PSB__G2926	hypothetical protein	2	NA	NA	NA
PSB__G4648	hypothetical protein	2	NA	NA	NA
PSB__G960	hypothetical protein	2	NA	Lysophospholipase, alpha-beta hydrolase superfamily (PldB) (PDB:5O7G)	NA
PSB__G5333	light-harvesting antenna LH1, beta subunit	2	light-harvesting protein	NA	light-harvesting complex 1 beta chain
PSB__G4475	hypothetical protein	2	NA	NA	NA
PSB__G3902	hypothetical protein	2	NA	NA	NA
PSB__G1994	EDSAP-1 family PEP-CTERM protein	2	PEP-CTERM sorting domain-containing protein	NA	NA
PSB__G4535	RnfH family protein	2	RnfH family protein	Putative antitoxin component PasI (RatB) of the RatAB toxin-antitoxin module, ubiquitin-RnfH superfamily (PasI) (PDB:2HJ1) (PUBMED:23055930)	uncharacterized protein
PSB__G5192	hypothetical protein	2	NA	NA	NA
PSB__G1605	SGNH/GDSL hydrolase family protein	2	SGNH/GDSL hydrolase family protein	NA	NA
PSB__G6430	tyrosine-type recombinase/integrase	2	tyrosine-type recombinase/integrase	Site-specific recombinase XerD (XerD) (PDB:1A0P)	NA
PSB__G5423	efflux RND transporter permease subunit	2	efflux RND transporter permease subunit	Multidrug efflux pump subunit AcrB (AcrB) (PDB:1IWG)	NA
PSB__G2378	hypothetical protein	2	NA	NA	NA
PSB__G3086	sulfotransferase	2	sulfotransferase family protein	Predicted phosphohydrolase or phosphomutase, AlkP superfamily	NA
PSB__G4886	hypothetical protein	2	NA	NA	NA
PSB__G4744	hypothetical protein	2	NA	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB__G3903	hypothetical protein	2	NA	NA	NA
PSB__G1890	nitrogen fixation protein NifM	2	nitrogen fixation protein NifM (nifM)	Peptidyl-prolyl isomerase, parvulin family (SurA) (PDB:2ZR5)	NA
PSB__G4567	hypothetical protein	2	NA	NA	NA
PSB__G6219	alpha/beta fold hydrolase	2	alpha/beta hydrolase	Dipeptidyl aminopeptidase/acylaminoacyl peptidase (DAP2) (PDB:2JBW)	NA
PSB__G4766	hypothetical protein	2	NA	NA	NA
PSB__G4359	hypothetical protein	2	NA	NA	NA
PSB__G6514	ATP-binding cassette domain-containing protein	2	ABC transporter ATP-binding protein	ATPase components of ABC transporters with duplicated ATPase domains (Uup) (PDB:5ZXD)	NA
PSB__G17	thrombospondin type 3 repeat-containing protein	2	tandem-95 repeat protein	Periplasmic component TolB of the Tol biopolymer transport system (TolB) (PDB:1C5K)!!!Leucine-rich repeat (LRR) protein (LRR) (PDB:4OW2)	thrombospondin 2/3/4/5
PSB__G4682	hypothetical protein	2	NA	NA	NA

PSB__G6835	hypothetical protein	2	NA	NA	NA
PSB__G5113	hypothetical protein	2	NA	NA	NA
PSB__G6866	alpha/beta hydrolase	2	alpha/beta fold hydrolase	Lysophospholipase, alpha-beta hydrolase superfamily (PldB) (PDB:5O7G)	NA
PSB__G7248	hypothetical protein	2	NA	dTDP-4-amino-4,6-dideoxygalactose transaminase (WecE) (PDB:4PIW) (PUBMED:15271350)	NA
PSB__G6779	hypothetical protein	2	NA	dTDP-4-amino-4,6-dideoxygalactose transaminase (WecE) (PDB:4PIW) (PUBMED:15271350)	NA
PSB__G5071	DUF4277 domain-containing protein	2	DUF4277 domain-containing protein	NA	NA
PSB__G3578	30S ribosomal protein S7	2	30S ribosomal protein S7 (rpsG)	Ribosomal protein S7 (RpsG) (PDB:4V4H)	small subunit ribosomal protein S7
PSB__G4007	preprotein translocase subunit SecE	2	preprotein translocase subunit SecE (secE)	Preprotein translocase subunit SecE (SecE) (PDB:5MG3) (PUBMED:29684280)	preprotein translocase subunit SecE
PSB__G5093	hypothetical protein	2	NA	NA	NA
PSB__G6645	hypothetical protein	2	NA	NA	NA
PSB__G4091	hypothetical protein	2	NA	NA	NA
PSB__G3797	hypothetical protein	2	NA	NA	NA
PSB__G4791	hypothetical protein	2	NA	NA	NA
PSB__G4462	CRISPR-associated endonuclease Cas2	2	CRISPR-associated endonuclease Cas2	CRISPR/Cas system-associated endoribonuclease Cas2 (Cas2) (PDB:2I0X)	CRISPR-associated protein Cas2
PSB__G2762	hypothetical protein	2	NA	NA	NA
PSB__G954	peptidoglycan-binding protein	2	peptidoglycan-binding protein	Serine/threonine protein kinase (SPS1) (PDB:6G4J)!!!Formylglycine-generating enzyme, required for sulfatase activity, contains SUMF1/FGE domain (YfmG) (PDB:2AFT)!!!Peptidoglycan-binding (PGRP) domain of peptidoglycan hydrolases (PGRP) (PDB:4FET)	NA
PSB__G6413	hypothetical protein	2	NA	NA	NA
PSB__G3216	hypothetical protein	2	NA	NA	NA
PSB__G1487	SUMF1/EgtB/PvdO family nonheme iron enzyme	2	SUMF1/EgtB/PvdO family nonheme iron enzyme	Formylglycine-generating enzyme, required for sulfatase activity, contains SUMF1/FGE domain (YfmG) (PDB:2AFT)!!!Predicted NTPase, NACHT family domain (NACHT)	NA
PSB__G2563	hypothetical protein	2	NA	NA	NA
PSB__G275	hypothetical protein	2	NA	NA	NA
PSB__G3581	PTS IIA-like nitrogen regulatory protein PtsN	2	PTS IIA-like nitrogen regulatory protein PtsN (ptsN)	Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type) (PtsN) (PDB:2OQ3)	nitrogen PTS system EIIA component [EC:2.7.1.-]
PSB__G5319	light-harvesting antenna LH1, beta subunit	2	light-harvesting protein	NA	light-harvesting complex 1 beta chain
PSB__G4662	light-harvesting antenna LH1, alpha subunit	2	light-harvesting protein	NA	light-harvesting protein B-800-850 alpha chain
PSB__G6359	hypothetical protein	2	four helix bundle protein	NA	NA
PSB__G5178	hypothetical protein	2	NA	NA	NA
PSB__G3857	sulfurtransferase complex subunit TusC	2	sulfurtransferase complex subunit TusC (tusC)	Sulfur transfer complex TusBCD TusC component, DsrF family (DsrF) (PDB:2D1P)	tRNA 2-thiouridine synthesizing protein C
PSB__G3417	DciA family protein	2	DUF721 domain-containing protein	NA	NA
PSB__G6505	Nudix family hydrolase	2	Nudix family hydrolase	Thiamine monophosphate synthase (ThiE) (PDB:1G4T) (PUBMED:19060138)!!!8-oxo-dGTP pyrophosphatase MutT and related house-cleaning NTP pyrophosphohydrolases, NUDIX family (MutT) (PDB:4V14)	8-oxo-dGTP diphosphatase [EC:3.6.1.55]
PSB__G6850	transposase	2	transposase	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB__G4000	hypothetical protein	2	NA	NA	NA
PSB__G4665	IS4 family transposase	2	IS4 family transposase	NA	NA
PSB__G4610	hypothetical protein	2	NA	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	NA
PSB__G4106	hypothetical protein	2	NA	NA	NA
PSB__G2573	PEP-CTERM/exosortase system-associated acyltransferase	2	PEP-CTERM/exosortase system-associated acyltransferase	N-acyl-homoserine lactone synthase LasI (autoinducer biosynthesis) (LasI) (PDB:1K4J)	NA
PSB__G4156	HPF/RaiA family ribosome-associated protein	2	HPF/RaiA family ribosome-associated protein	NA	NA
PSB__G1064	IPTL-CTERM sorting domain-containing protein	2	IPTL-CTERM sorting domain-containing protein	Leucine-rich repeat (LRR) protein (LRR) (PDB:4OW2)	NA
PSB__G4519	hypothetical protein	2	NA	NA	NA
PSB__G4138	hypothetical protein	2	NA	NA	NA
PSB__G1277	glycosyltransferase family 4 protein	2	glycosyltransferase	Glycosyltransferase involved in cell wall bisynthesis (RfaB) (PDB:2IV7)	polysaccharide biosynthesis protein PslH
PSB__G869	oligosaccharide flippase family protein	2	oligosaccharide flippase family protein	Membrane protein involved in the export of O-antigen and teichoic acid (RfbX)	NA
PSB__G3584	hypothetical protein	2	NA	NA	NA
PSB__G7350	PIG-L deacetylase family protein	2	PIG-L family deacetylase	Glycosyltransferase involved in cell wall bisynthesis (WcaA) (PDB:5MLZ)!!!N-acetylglucosaminyl deacetylase, LmbE family (LmbE) (PDB:1Q74)	N-acetylglucosamine malate deacetylase 1 [EC:3.5.1.-]
PSB__G2909	hypothetical protein	2	NA	NA	NA
PSB__G2479	hypothetical protein	2	NA	NA	NA
PSB__G2030	transposase	2	transposase	Transposase, IS1182 family	transposase
PSB__G1208	NADP-dependent isocitrate dehydrogenase	3	NADP-dependent isocitrate dehydrogenase (icd)	Isocitrate dehydrogenase (Icd) (PDB:1IDD)	isocitrate dehydrogenase [EC:1.1.1.42]
PSB__G4088	NADH-quinone oxidoreductase subunit A	3	NADH-quinone oxidoreductase subunit A	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (NuoA) (PDB:3RKO)	NADH-quinone oxidoreductase subunit A [EC:7.1.1.2]
PSB__G3553	NuoB/complex I 20 kDa subunit family protein	3	NADH-quinone oxidoreductase subunit B	NADH:ubiquinone oxidoreductase 20 kD subunit (chain B) or related Fe-S oxidoreductase (NuoB) (PDB:2FUG) (PUBMED:25941396)	NADH-quinone oxidoreductase subunit B [EC:7.1.1.2]
PSB__G2638	NADH-quinone oxidoreductase subunit C	3	NADH-quinone oxidoreductase subunit C	NADH:ubiquinone oxidoreductase 27 kD subunit (chain C) (NuoC)	NADH-quinone oxidoreductase subunit C [EC:7.1.1.2]

PSB__G181	ribonucleoside-diphosphate reductase subunit alpha	3	ribonucleoside-diphosphate reductase subunit alpha	Ribonucleotide reductase alpha subunit (NrdA) (PDB:1R1R)!!!Anaerobic ribonucleoside-triphosphate reductase (NrdD) (PDB:4COI)	ribonucleoside-diphosphate reductase alpha chain [EC:1.17.4.1]
PSB__G559	carbamoyltransferase	3	carbamoyltransferase	Predicted carbamoyl transferase, NodU family (PDB:3VEN)	carbamoyltransferase [EC:2.1.3.-]
PSB__G937	bifunctional enoyl-CoA hydratase/phosphate acetyltransferase	3	bifunctional enoyl-CoA hydratase/phosphate acetyltransferase	Phosphotransacetylase (includes Pta, EutD and phosphobutyryltransferase) (Pta) (PDB:1QZT) (PUBMED:28754323)!!!Acyl-CoA dehydratase PaaZ (MaoC) (PDB:5ZY8)	phosphate butyryltransferase [EC:2.3.1.19]
PSB__G979	aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme	3	aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme	7-keto-8-aminopelargonate synthetase or related enzyme (BioF) (PDB:1BS0)	8-amino-7-oxononanoate synthase [EC:2.3.1.47]
PSB__G6951	carboxylating nicotinate-nucleotide diphosphorylase	3	carboxylating nicotinate-nucleotide diphosphorylase	Nicotinate-nucleotide pyrophosphorylase (NadC) (PDB:1O4U)	nicotinate-nucleotide pyrophosphorylase (carboxylating) [EC:2.4.2.19]
PSB__G2118	acetylglutamate kinase	3	acetylglutamate kinase (argB)	N-acetylglutamate kinase (ArgB) (PDB:3S6G)	acetylglutamate kinase [EC:2.7.2.8]
PSB__G3130	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase	3	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase (folK)	7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase (folate biosynthesis) (FolK) (PDB:1CBK)	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase [EC:2.7.6.3]
PSB__G1082	group II intron reverse transcriptase/maturase	3	group II intron reverse transcriptase/maturase (ltrA)	Retron-type reverse transcriptase (YkfC)	RNA-directed DNA polymerase [EC:2.7.7.49]
PSB__G3932	hypothetical protein	3	NA	NA	secretory phospholipase A2 [EC:3.1.1.4]
PSB__G1058	guanine deaminase	3	guanine deaminase (guaD)	Cytosine/adenosine deaminase or related metal-dependent hydrolase (SsnA) (PDB:3O7U)	guanine deaminase [EC:3.5.4.3]
PSB__G2148	archaetidylserine decarboxylase	3	phosphatidylserine decarboxylase (psd)	Phosphatidylserine decarboxylase (Psd) (PDB:6L06)	phosphatidylserine decarboxylase [EC:4.1.1.65]
PSB__G877	bifunctional 2-methylcitrate dehydratase/aconitate hydratase	3	bifunctional 2-methylcitrate dehydratase/aconitate hydratase	2-methylcitrate dehydratase PrpD (PrpD) (PDB:1SZQ)	2-methylcitrate dehydratase [EC:4.2.1.79]
PSB__G701	glucose-6-phosphate isomerase	3	glucose-6-phosphate isomerase (pgi)	Glucose-6-phosphate isomerase (Pgi) (PDB:1B0Z)	glucose-6-phosphate isomerase [EC:5.3.1.9]
PSB__G984	UDP-N-acetylmuramoyl-L-alanine--D-glutamate ligase	3	UDP-N-acetylmuramoyl-L-alanine--D-glutamate ligase	UDP-N-acetylmuramoylalanine-D-glutamate ligase (MurD) (PDB:1E0D)	UDP-N-acetylmuramoylalanine--D-glutamate ligase [EC:6.3.2.9]
PSB__G1159	phosphoribosylamine--glycine ligase	3	phosphoribosylamine--glycine ligase (purD)	Phosphoribosylamine-glycine ligase (PurD) (PDB:1GSO)	phosphoribosylamine---glycine ligase [EC:6.3.4.13]
PSB__G6451	asparagine synthase-related protein	3	asparagine synthase-related protein	Asparagine synthetase B (glutamine-hydrolyzing) (AsnB) (PDB:2ETD) (PUBMED:25951518)	asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4]
PSB__G1339	ABC transporter permease	3	ABC transporter permease	ABC-type antimicrobial peptide transport system, permease component (SalY)	putative ABC transport system permease protein
PSB__G2135	amino acid ABC transporter substrate-binding protein	3	amino acid ABC transporter substrate-binding protein	ABC-type amino acid transport/signal transduction system, periplasmic component/domain (HisJ) (PDB:1GGG)	polar amino acid transport system substrate-binding protein
PSB__G2545	pimeloyl-ACP methyl ester esterase BioH	3	pimeloyl-ACP methyl ester esterase BioH (bioH)	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase MenH and related esterases, alpha/beta hydrolase fold (MenH) (PDB:4MYS)	pimeloyl-[acyl-carrier protein] methyl ester esterase [EC:3.1.1.85]
PSB__G1910	adenosylcobinamide-phosphate synthase CbiB	3	cobalamin biosynthesis protein CobD (cobD)	Cobalamin biosynthesis protein CobD/CbiB (CbiB)	adenosylcobinamide-phosphate synthase [EC:6.3.1.10]
PSB__G392	elongation factor G	3	GTP-binding protein	Translation elongation factor EF-G, a GTPase (FusA) (PDB:1DAR)	elongation factor G
PSB__G6678	prepilin-type N-terminal cleavage/methylation domain-containing protein	3	type II secretion system protein	Type II secretory pathway, pseudopilin PulG (PulG) (PDB:1AY2)	general secretion pathway protein I
PSB__G4899	translation initiation factor IF-1	3	translation initiation factor IF-1 (infA)	Translation initiation factor IF-1 (InfA) (PDB:5FIM)	translation initiation factor IF-1
PSB__G2055	quinol dehydrogenase ferredoxin subunit NapH	3	quinol dehydrogenase ferredoxin subunit NapH (napH)	Polyferredoxin NapH (NapH) (PUBMED:27322068)	ferredoxin-type protein NapH
PSB__G731	nif-specific transcriptional activator NifA	3	nif-specific transcriptional activator NifA (nifA)	FhlA-type transcriptional regulator, contains GAF, AAA-type ATPase, and DNA-binding Fis domains (FhlA)	Nif-specific regulatory protein
PSB__G3433	ribosome maturation factor RimM	3	ribosome maturation factor RimM (rimM)	Ribosomal 30S subunit maturation factor RimM, required for 16S rRNA processing (RimM) (PDB:2QGG)	16S rRNA processing protein RimM
PSB__G3792	50S ribosomal protein L13	3	50S ribosomal protein L13 (rplM)	Ribosomal protein L13 (RplM) (PDB:1VS6)	large subunit ribosomal protein L13
PSB__G2210	signal peptidase I	3	signal peptidase I (lepB)	Signal peptidase I (LepB) (PDB:1B12) (PUBMED:1925027)	signal peptidase I [EC:3.4.21.89]
PSB__G1322	ammonium transporter	3	ammonium transporter	Ammonia channel protein AmtB (AmtB) (PDB:1U77)	ammonium transporter, Amt family
PSB__G720	type I restriction-modification system subunit M	3	type I restriction-modification system subunit M	Type I restriction-modification system, DNA methylase subunit (HsdM) (PDB:2AR0) (PUBMED:26872910)	type I restriction enzyme M protein [EC:2.1.1.72]
PSB__G4537	HU family DNA-binding protein	3	HU family DNA-binding protein	Bacterial nucleoid DNA-binding protein IHF-alpha (HimA) (PDB:1B8Z)	DNA-binding protein HU-beta
PSB__G1266	putative lipid II flippase FtsW	3	putative lipid II flippase FtsW (ftsW)	Peptodoglycan polymerase FtsW/RodA/SpoVE (FtsW) (PDB:6BAR) (PUBMED:30692671)	cell division protein FtsW
PSB__G583	translation elongation factor 4	3	elongation factor 4 (lepA)	Translation elongation factor EF-4, membrane-bound GTPase (LepA) (PDB:3DEG)	GTP-binding protein LepA
PSB__G3298	sigma-E factor negative regulatory protein	3	sigma-E factor negative regulatory protein	RseA, negative regulator of sigma E activity (RseA) (PDB:1OR7)	sigma-E factor negative regulatory protein RseA
PSB__G6698	FAD:protein FMN transferase	3	FAD:protein FMN transferase ApbE	FAD:protein FMN transferase ApbE (ApbE) (PDB:1VRM) (PUBMED:23558683;30154099)	FAD:protein FMN transferase [EC:2.7.1.180]
PSB__G3749	arsenate reductase ArsC	3	arsenate reductase ArsC	Protein-tyrosine-phosphatase (Wzb) (PDB:4D74) (PUBMED:26641313)	arsenate reductase (thioredoxin) [EC:1.20.4.4]
PSB__G3505	SoxR reducing system RseC family protein	3	SoxR reducing system RseC family protein	RseC, positive regulator of sigma E activity (RseC)	sigma-E factor negative regulatory protein RseC
PSB__G4228	Rieske (2Fe-2S) protein	3	non-heme iron oxygenase ferredoxin subunit	Ferredoxin subunit of nitrite reductase or a ring-hydroxylating dioxygenase (NirD) (PDB:1FQT) (PUBMED:23844911)	3-phenylpropionate/trans-cinnamate dioxygenase ferredoxin component

PSB__G6954	acyl-[ACP]--phospholipid O-acyltransferase	3	acyl-[ACP]--phospholipid O-acyltransferase	1-acyl-sn-glycerol-3-phosphate acyltransferase (PlsC) (PDB:1ILJQ)!!!O-succinylbenzoic acid-CoA ligase MenE or related acyl-CoA synthetase (AMP-forming) (MenE/FadK) (PDB:5EY9) (PUBMED:25151136;27455459)	1-acyl-sn-glycerol-3-phosphate acyltransferase [EC:2.3.1.51]
PSB__G2372	undecaprenyl-diphosphate phosphatase	3	undecaprenyl-diphosphate phosphatase	Undecaprenyl pyrophosphate phosphatase (UppP) (PDB:5OON)	undecaprenyl-diphosphatase [EC:3.6.1.27]
PSB__G7228	zinc metallopeptidase	3	zinc metallopeptidase	Zn-dependent membrane protease YugP (YugP)	uncharacterized protein
PSB__G4474	cupin domain-containing protein	3	cupin domain-containing protein	Predicted enzyme of the cupin superfamily (PDB:1LKN)	uncharacterized protein
PSB__G3082	YqiA/YcfP family alpha/beta fold hydrolase	3	alpha/beta fold hydrolase	Predicted esterase YcpF, UPF0227 family (YcfP) (PDB:4FLE)	uncharacterized protein
PSB__G3192	acetate uptake transporter	3	acetate uptake transporter	Succinate-acetate transporter SatP (SatP) (PDB:5YS3) (PUBMED:23844911)	uncharacterized protein
PSB__G2622	TSUP family transporter	3	sulfite exporter TauE/SafE family protein	Sulfite exporter TauE/SafE/YfcA and related permeases, UPF0721 family (TauE) (PUBMED:21183667;22797525)	uncharacterized protein
PSB__G3062	metallophosphoesterase family protein	3	metallophosphoesterase family protein	Predicted phosphodiesterase, calcineurin family (YfcE) (PDB:1SU1) (PUBMED:17586769)	uncharacterized protein
PSB__G6231	aminoglycoside phosphotransferase family protein	3	phosphotransferase	Predicted phosphotransferase, aminoglycoside/choline kinase (APH/ChoK) family	N-acetylmuramate 1-kinase [EC:2.7.1.221]
PSB__G7064	type II toxin-antitoxin system PemK/MazF family toxin	3	type II toxin-antitoxin system PemK/MazF family toxin	mRNA-degrading endonuclease MazF, toxin component of the MazEF toxin-antitoxin module (MazF) (PDB:2MF2)	mRNA interferase MazF [EC:3.1.-.-]
PSB__G4337	sulfurtransferase complex subunit TusB	3	sulfurtransferase complex subunit TusB (dsrH)	Sulfur transfer complex TusBCD TusB component, DsrH family (TusB) (PDB:1RHX)	tRNA 2-thiouridine synthesizing protein B
PSB__G358	molybdopterin cofactor-binding domain-containingprotein	3	xanthine dehydrogenase family protein molybdopterin-binding subunit	Aldehyde, CO or xanthine dehydrogenase, Mo-binding subunit (CoxL) (PDB:1DGJ) (PUBMED:10095793)	isoquinoline 1-oxidoreductase subunit beta [EC:1.3.99.16]
PSB__G1702	XdhC family protein	3	XdhC family protein	Molybdoenzyme maturation factor PaoD (Mo cofactor insertion), XdhC/CoxF family (XdhC) (PDB:2WE7) (PUBMED:24498065)	xanthine dehydrogenase accessory factor
PSB__G3370	transposase	3	transposase	Transposase and inactivated derivatives, IS5 family (IS5)	transposase, IS5 family
PSB__G1188	transposase	3	transposase	Transposase	transposase
PSB__G5427	IS110 family transposase	3	IS110 family transposase	Transposase	transposase
PSB__G4517	hypothetical protein	3	NA	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB__G1279	IS4 family transposase	3	IS4 family transposase	IS4 transposase InsG (InsG)	putative transposase
PSB__G6252	methylmalonyl Co-A mutase-associated GTPase MeaB	3	methylmalonyl Co-A mutase-associated GTPase MeaB (meaB)	GTPase of the G3E family (not a periplasmic protein kinase) (ArgK) (PDB:2QM7) (PUBMED:18950999)	GTPase [EC:3.6.5.-]
PSB__G905	PucC family protein	3	PucC family protein	Na+/melibiose symporter or related transporter (MelB) (PDB:4M64)	MFS transporter, BCD family, chlorophyll transporter
PSB__G961	BCD family MFS transporter	3	BCD family MFS transporter	Na+/melibiose symporter or related transporter (MelB) (PDB:4M64)	MFS transporter, BCD family, chlorophyll transporter
PSB__G2968	TIGR04282 family arsenosugar biosynthesis glycosyltransferase	3	glycosyltransferase	Uncharacterized conserved protein, glycosyltransferase A (GT-A) superfamily, DUF2064 family (PDB:3CGX)	uncharacterized protein
PSB__G4222	TusE/DsrC/DsvC family sulfur relay protein	3	TusE/DsrC/DsvC family sulfur relay protein	Sulfur transfer complex TusBCD TusE component, DsrC family (tRNA 2-thiouridine synthesizing protein C) (TusE) (PDB:1JI8)	dissimilatory sulfite reductase related protein
PSB__G1959	bifunctional riboflavin kinase/FAD synthetase	3	bifunctional riboflavin kinase/FAD synthetase (ribF)	FAD synthase (RibF) (PDB:1MRZ)	riboflavin kinase / FMN adenylyltransferase [EC:2.7.1.26 2.7.7.2]
PSB__G1618	SGNH/GDSL hydrolase family protein	3	SGNH/GDSL hydrolase family protein	Phospholipase/lecithinase/hemolysin (PDB:6JL1)	outer membrane lipase/esterase
PSB__G807	glycosyltransferase family 39 protein	3	glycosyltransferase family 39 protein	Uncharacterized membrane protein PF0508, contains N-terminal glycosyltransferase domain of PMT family	mannosyltransferase [EC:2.4.1.-]
PSB__G2780	tRNA (adenosine(37)-N6)-threonylcarbamoyltransferase complex dimerization subunit type 1 TsaB	3	tRNA (adenosine(37)-N6)-threonylcarbamoyltransferase complex dimerization subunit type 1 TsaB (tsaB)	tRNA A37 threonylcarbamoyladenosine modification protein TsaB (TsaB) (PDB:1OKJ)	tRNA threonylcarbamoyladenosine biosynthesis protein TsaB
PSB__G6914	AbrB/MazE/SpoVT family DNA-binding domain-containing protein	3	AbrB/MazE/SpoVT family DNA-binding domain-containing protein	Bifunctional DNA-binding transcriptional regulator of stationary/sporulation/toxin gene expression and antitoxin component of the YhaV-PrlF toxin-antitoxin module (AbrB) (PDB:2RO5)	antitoxin PrlF
PSB__G2313	hypothetical protein	3	NA	NA	NA
PSB__G5398	HGGxSTG domain-containing protein	3	HGGxSTG domain-containing protein	NA	NA
PSB__G3372	neutral zinc metallopeptidase	3	NA	NA	NA
PSB__G5149	ISNCY family transposase	3	NA	NA	NA
PSB__G1991	hypothetical protein	3	NA	NA	NA
PSB__G5000	hypothetical protein	3	NA	NA	NA
PSB__G4896	hypothetical protein	3	NA	NA	NA
PSB__G3692	type I restriction endonuclease	3	type I restriction endonuclease	Type I site-specific restriction-modification system, R (restriction) subunit and related helicases ... (PDB:2W00)	NA
PSB__G6266	REP-associated tyrosine transposase	3	REP-associated tyrosine transposase	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB__G6283	four helix bundle protein	3	four helix bundle protein	NA	NA
PSB__G6292	hypothetical protein	3	NA	NA	NA
PSB__G6591	four helix bundle protein	3	four helix bundle protein	NA	NA
PSB__G6725	four helix bundle protein	3	four helix bundle protein	NA	NA
PSB__G3559	hypothetical protein	3	NA	NA	NA
PSB__G4062	hypothetical protein	3	NA	NA	NA
PSB__G5260	transposase	3	transposase	Transposase	NA
PSB__G4559	hypothetical protein	3	NA	NA	NA
PSB__G7343	efflux RND transporter permease subunit	3	efflux RND transporter permease subunit	Multidrug efflux pump subunit AcrB (AcrB) (PDB:1IWG)	NA
PSB__G5285	RHS repeat protein	3	NA	NA	NA

PSB_G7300	hypothetical protein	3	NA	NA	NA
PSB_G5402	glycogen/starch/alpha-glucan phosphorylase	3	NA	NA	NA
PSB_G6536	hypothetical protein	3	NA	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	NA
PSB_G7097	transposase	3	transposase	Transposase	putative transposase
PSB_G6945	DUF2283 domain-containing protein	3	NA	NA	NA
PSB_G6384	molybdenum cofactor guanylyltransferase MobA	3	molybdenum cofactor guanylyltransferase MobA (mobA)	Molybdopterin-guanine dinucleotide biosynthesis protein A (MobA) (PDB:1E5K) (PUBMED:32239579)	molybdenum cofactor guanylyltransferase [EC:2.7.7.77]
PSB_G6587	aldehyde dehydrogenase family protein	3	NA	NA	NA
PSB_G6683	transposase	3	transposase	NA	NA
PSB_G6952	hypothetical protein	3	NA	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB_G818	1-deoxy-D-xylulose-5-phosphate synthase	3	1-deoxy-D-xylulose-5-phosphate synthase (dxs)	Deoxyxylulose-5-phosphate synthase (Dxs) (PDB:2O1S)	1-deoxy-D-xylulose-5-phosphate synthase [EC:2.2.1.7]
PSB_G5163	geranylgeranyl diphosphate reductase	3	NA	NA	NA
PSB_G5224	IS4 family transposase	3	NA	NA	NA
PSB_G7053	IS200/IS605 family transposase	3	IS200/IS605 family transposase (tnpA)	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB_G5518	GMP/IMP nucleotidase	3	GMP/IMP nucleotidase (yrfG)	FMN and 5-amino-6-(5-phospho-D-ribitylamino)uracil phosphatase YigB, HAD superfamily (riboflavin biosynthesis) (YigB) (PDB:4JB3) (PUBMED:24123841)	GMP/IMP 5'-nucleotidase [EC:3.1.3.-]
PSB_G5482	RNA recognition motif domain-containing protein	3	RNA recognition motif domain-containing protein	NA	NA
PSB_G5476	hypothetical protein	3	NA	NA	NA
PSB_G5297	hypothetical protein	3	NA	NA	NA
PSB_G5035	hypothetical protein	3	NA	NA	NA
PSB_G5543	ATP-binding protein	3	ATP-binding protein	Predicted ATPase, AAA+ superfamily	uncharacterized protein
PSB_G3651	hypothetical protein	3	NA	NA	NA
PSB_G5451	trypsin-like peptidase domain-containing protein	3	trypsin-like peptidase domain-containing protein	Periplasmic serine protease, S1-C subfamily, contain C-terminal PDZ domain (DegQ) (PDB:2HGA) (PUBMED:29686141)	NA
PSB_G4548	DUF4258 domain-containing protein	3	DUF4258 domain-containing protein	NA	NA
PSB_G5510	hypothetical protein	3	NA	NA	NA
PSB_G5598	hypothetical protein	3	NA	NA	NA
PSB_G1468	IS630 family transposase	3	IS630 family transposase	CRISPR-associated protein Csa3, CARF domain (Csa3) (PDB:2WTE) (PUBMED:21093452)	putative transposase
PSB_G5597	YgiT-type zinc finger protein	3	YgiT-type zinc finger protein	NA	NA
PSB_G3945	hypothetical protein	3	NA	NA	NA
PSB_G4736	hypothetical protein	3	NA	NA	NA
PSB_G28	DUF4011 domain-containing anti-phage protein Hhe	3	DUF4011 domain-containing protein	Superfamily I DNA and/or RNA helicase (DNA2) (PDB:2GJK)!!!Very-short-patch-repair endonuclease (YcjD) (PDB:3HRL)	NA
PSB_G4037	transposase	3	transposase	Transposase, IS1182 family	NA
PSB_G4113	hypothetical protein	3	NA	Transposase, IS1182 family	NA
PSB_G2779	HAD family hydrolase	3	HAD family hydrolase	Phosphoglycolate phosphatase, HAD superfamily (Gph) (PDB:2NYV)	NA
PSB_G5687	PD-(D/E)XK nuclease family protein	3	DUF3782 domain-containing protein	Uncharacterized protein, contains PD-(D/E)XK nuclease domain	NA
PSB_G6315	tetratricopeptide repeat protein	3	tetratricopeptide repeat protein	Flp pilus assembly protein TadD, contains TPR repeats (TadD) (PDB:3WXX)	NA
PSB_G2977	putative 2OG-Fe(II) oxygenase	3	putative 2OG-Fe(II) oxygenase	NA	NA
PSB_G2395	polysaccharide deacetylase family protein	3	polysaccharide deacetylase family protein	Peptidoglycan/xylan/chitin deacetylase, PgdA/NodB/CDA1 family (CDA1) (PDB:1NY1)	NA
PSB_G1583	glycosyltransferase	3	glycosyltransferase	Glycosyltransferase involved in cell wall bisynthesis (RfaB) (PDB:2IV7)	NA
PSB_G1589	glycosyltransferase	3	glycosyltransferase	Glycosyltransferase involved in cell wall bisynthesis (RfaB) (PDB:2IV7)	NA
PSB_G3038	hypothetical protein	3	NA	NA	NA
PSB_G1355	formylglycine-generating enzyme family protein	3	formylglycine-generating enzyme family protein	Formylglycine-generating enzyme, required for sulfatase activity, contains SUMF1/FGE domain (YfmG) (PDB:2AFT)	NA
PSB_G4983	hypothetical protein	3	NA	NA	NA
PSB_G3068	transglutaminase domain-containing protein	3	transglutaminase domain-containing protein	Uncharacterized conserved protein RhaS, contains 28 RHS repeats (RhsA) (PDB:4IGL)	NA
PSB_G5238	hypothetical protein	3	NA	NA	NA
PSB_G2301	hypothetical protein	3	NA	NA	NA
PSB_G4473	hypothetical protein	3	NA	NA	NA
PSB_G1093	phytoene desaturase family protein	3	NAD(P)-binding protein	Phytoene dehydrogenase-related protein (PDB:4REP)	NA
PSB_G2724	CHAP domain-containing protein	3	CHAP domain-containing protein	Glutathionylspermidine synthase, CHAP domain (Gsp) (PDB:2IO7)	NA
PSB_G1953	alpha-E domain-containing protein	3	alpha-E domain-containing protein	Uncharacterized conserved protein, Alpha-E superfamily (Alpha-E) (PUBMED:20023723)	NA
PSB_G3169	DUF4390 domain-containing protein	3	DUF4390 domain-containing protein	NA	NA
PSB_G6470	four helix bundle protein	3	four helix bundle protein	NA	NA
PSB_G4646	hypothetical protein	3	NA	NA	NA
PSB_G7051	class I SAM-dependent methyltransferase	3	class I SAM-dependent methyltransferase	SAM-dependent methyltransferase (SmtA) (PDB:5DNK)	NA
PSB_G5301	hypothetical protein	3	NA	NA	NA
PSB_G3061	hypothetical protein	3	NA	NA	NA
PSB_G4270	hypothetical protein	3	NA	NA	NA
PSB_G4528	DUF4258 domain-containing protein	3	DUF4258 domain-containing protein	NA	NA
PSB_G4833	hypothetical protein	3	NA	NA	NA
PSB_G6442	hypothetical protein	3	NA	NA	NA
PSB_G5229	hypothetical protein	3	NA	NA	NA
PSB_G2172	hypothetical protein	3	NA	NA	NA
PSB_G6912	sodium:proton antiporter NhaD	3	sodium:proton antiporter NhaD (nhaD)	Na+/H+ antiporter NhaD or related arsenite permease (ArsB)	NA

PSB__G3642	CBS domain-containing protein	3	CBS domain-containing protein	Signal-transduction protein containing cAMP-binding, CBS, and nucleotidyltransferase domains	NA
PSB__G4376	hypothetical protein	3	NA	Recombination-promoting DNA endonuclease RpnC/YadD (RpnC) (PUBMED:28096446)	NA
PSB__G3269	TrbM/KikA/MpfK family conjugal transfer protein	3	conjugal transfer protein TrbM	NA	NA
PSB__G4426	hypothetical protein	3	NA	NA	NA
PSB__G1906	MBL fold metallo-hydrolase	3	MBL fold metallo-hydrolase	Flavorubredoxin (NorV) (PDB:4JHC)	NA
PSB__G4490	hypothetical protein	3	NA	NA	NA
PSB__G1310	aminomethyltransferase family protein	3	aminomethyl transferase family protein	Glycine cleavage system protein T (aminomethyltransferase) (GcvT) (PDB:2GAH) (PUBMED:25433025;25505271)	glycine cleavage system T protein (aminomethyltransferase) [EC:2.1.2.10]
PSB__G5252	hypothetical protein	3	NA	NA	NA
PSB__G5214	hypothetical protein	3	NA	NA	NA
PSB__G4805	hypothetical protein	3	NA	NA	NA
PSB__G2610	hypothetical protein	3	NA	NA	NA
PSB__G4304	hypothetical protein	3	NA	NA	NA
PSB__G6869	hypothetical protein	3	NA	NA	NA
PSB__G4257	hypothetical protein	3	NA	NA	NA
PSB__G7221	hypothetical protein	3	NA	Transcriptional accessory protein Tex/SPT6 (Tex) (PDB:2OCE)	NA
PSB__G2653	SDR family NAD(P)-dependent oxidoreductase	3	SDR family NAD(P)-dependent oxidoreductase	Short-chain dehydrogenase (YqjQ)	NA
PSB__G4836	hypothetical protein	3	NA	Predicted ATPase, AAA+ superfamily	NA
PSB__G2484	NERD domain-containing protein	3	nuclease	DNA topoisomerase I, ssDNA-binding Zn-finger and Zn-ribbon domains (YrdD) (PDB:1YUA)	NA
PSB__G3701	DUF805 domain-containing protein	3	DUF805 domain-containing protein	Uncharacterized membrane protein YhaH, DUF805 family (yhaH)	NA
PSB__G1521	putative zinc-binding metallopeptidase	3	putative zinc-binding peptidase	Uncharacterized conserved protein, DUF2248 family	NA
PSB__G6316	hypothetical protein	3	NA	NA	NA
PSB__G4391	hypothetical protein	3	NA	NA	NA
PSB__G5344	hypothetical protein	3	NA	NA	NA
PSB__G4441	type II toxin-antitoxin system RelE/ParE family toxin	3	type II toxin-antitoxin system RelE/ParE family toxin	Plasmid stabilization system protein ParE (ParE) (PDB:3KXE)	toxin ParE1/3/4
PSB__G4799	addiction module protein	3	addiction module protein	NA	NA
PSB__G6369	hypothetical protein	3	four helix bundle protein	NA	NA
PSB__G3232	DUF427 domain-containing protein	3	DUF427 domain-containing protein	Uncharacterized conserved protein, DUF427 family (PDB:3DJM)	NA
PSB__G16	hypothetical protein	3	DUF7507 domain-containing protein	Uncharacterized conserved protein, PKD repeat domain (PDB:2KZW)!!!Uncharacterized conserved protein, UCP014979 family	NA
PSB__G5231	hypothetical protein	3	NA	NA	NA
PSB__G6269	DUF2201 family putative metallopeptidase	3	DUF2201 family putative metallopeptidase	Predicted metal-dependent peptidase	NA
PSB__G3517	type II toxin-antitoxin system VapC family toxin	3	type II toxin-antitoxin system VapC family toxin	VapC family ribonuclease, toxin component of the VapBC toxin-antitoxin module, contains PIN domain (VapC) (PUBMED:30169502)	NA
PSB__G5067	hypothetical protein	3	NA	NA	NA
PSB__G3807	hypothetical protein	3	NA	NA	NA
PSB__G3923	hypothetical protein	3	NA	NA	NA
PSB__G4763	transposase	3	transposase	Transposase	NA
PSB__G4943	hypothetical protein	3	NA	NA	NA
PSB__G7326	hypothetical protein	3	NA	NA	NA
PSB__G5191	hypothetical protein	3	NA	NA	NA
PSB__G3778	PIN domain nuclease	3	PIN domain nuclease	Ribonuclease/mRNA interferase VapC, contains PIN domain (VapC) (PDB:3TND)	NA
PSB__G5020	type II toxin-antitoxin system VapB family antitoxin	3	type II toxin-antitoxin system VapB family antitoxin	Transcription regulator of the Arc/MetJ class (VapB6) (PDB:6A7V)	NA
PSB__G5924	hypothetical protein	3	NA	NA	NA
PSB__G4967	hypothetical protein	3	NA	NA	NA
PSB__G5176	hypothetical protein	3	NA	NA	NA
PSB__G4132	NifB/NifX family molybdenum-iron cluster-bindingprotein	3	dinitrogenase iron-molybdenum cofactor biosynthesis protein	Predicted Fe-Mo cluster-binding protein, NifX family (NifX) (PDB:1EO1)	nitrogen fixation protein NifX
PSB__G4600	hypothetical protein	3	NA	NA	NA
PSB__G4747	hypothetical protein	3	NA	NA	NA
PSB__G6752	hypothetical protein	3	four helix bundle protein	NA	NA
PSB__G569	hypothetical protein	3	sensor domain-containing diguanylate cyclase	GGDEF domain, diguanylate cyclase (c-di-GMP synthetase) or its enzymatically inactive variants (GGDEF) (PDB:6D9M)	NA
PSB__G2029	hypothetical protein	3	NA	NA	NA
PSB__G3636	ArsC/Spx/MgsR family protein	3	nitrogenase-associated protein	Arsenate reductase or related protein, glutaredoxin family (ArsC) (PDB:1I9D)	NA
PSB__G5090	hypothetical protein	3	NA	NA	NA
PSB__G4496	4Fe-4S binding protein	3	4Fe-4S binding protein	Fe-S-cluster-containing hydrogenase component 2 (HycB)	NA
PSB__G2012	NF038130 family PEP-CTERM protein	3	PEP-CTERM sorting domain-containing protein	NA	NA
PSB__G1360	nucleotidyltransferase domain-containing protein	3	nucleotidyltransferase domain-containing protein	NA	NA
PSB__G4601	hypothetical protein	3	NA	CRISPR-associated protein Csa3, CARF domain (Csa3) (PDB:2WTE) (PUBMED:21093452)	NA
PSB__G2431	DUF2189 domain-containing protein	3	DUF2189 domain-containing protein	Uncharacterized membrane protein	NA
PSB__G1933	hypothetical protein	3	NA	NA	NA
PSB__G2228	transglutaminase family protein	3	transglutaminase family protein	Transglutaminase-like enzyme, putative cysteine protease (YebA) (PDB:3ISR)	NA
PSB__G7193	hypothetical protein	3	NA	NA	NA
PSB__G6387	DUF2189 domain-containing protein	3	DUF2189 domain-containing protein	Uncharacterized membrane protein	NA

PSB_G5195	hypothetical protein	3	NA	NA	NA
PSB_G667	DUF6399 domain-containing protein	3	DUF6399 domain-containing protein	NA	NA
PSB_G6679	hypothetical protein	3	NA	NA	NA
PSB_G3044	2OG-Fe(II) oxygenase	3	2OG-Fe(II) oxygenase	Predicted 2-oxoglutarate- and Fe(II)-dependent dioxygenase YbiX (PliuC) (PDB:3DKQ)	NA
PSB_G1152	hypothetical protein	3	NA	NA	NA
PSB_G5255	hypothetical protein	3	NA	NA	NA
PSB_G6321	IS4 family transposase	3	IS4 family transposase	NA	NA
PSB_G2254	esterase/lipase family protein	3	esterase/lipase family protein	Serine protease, subtilisin family (AprE) (PDB:1SCJ)	NA
PSB_G169	transglutaminase domain-containing protein	3	transglutaminase domain-containing protein	NA	NA
PSB_G4657	hypothetical protein	3	NA	NA	NA
PSB_G3387	PH domain-containing protein	3	PH domain-containing protein	Uncharacterized membrane protein YdbT, contains bPH2 (bacterial pleckstrin homology) domain (YdbT)	NA
PSB_G6749	glycosyltransferase	3	glycosyltransferase	Glycosyltransferase involved in cell wall bisynthesis (RfaB) (PDB:2IV7)	teichuronic acid biosynthesis glycosyltransferase TuaH [EC:2.4.-.-]
PSB_G272	hypothetical protein	3	NA	NA	NA
PSB_G5091	hypothetical protein	3	NA	NA	NA
PSB_G3906	hypothetical protein	3	NA	NA	NA
PSB_G7104	DUF6311 domain-containing protein	3	DUF6311 domain-containing protein	NA	NA
PSB_G1216	hypothetical protein	3	NA	NA	NA
PSB_G3451	DUF1269 domain-containing protein	3	DUF1269 domain-containing protein	Uncharacterized membrane protein	NA
PSB_G4291	hypothetical protein	3	NA	NA	NA
PSB_G4221	hypothetical protein	3	NA	NA	NA
PSB_G6999	ATP-binding cassette domain-containing protein	3	ABC-F family ATP-binding cassette domain-containing protein	ATPase components of ABC transporters with duplicated ATPase domains (Uup) (PDB:5ZXD)	NA
PSB_G3866	response regulator	3	response regulator	DNA-binding transcriptional response regulator, NtrC family, contains REC, AAA-type ATPase, and a Fis-type DNA-binding domains (AtoC) (PDB:1ZY2)	NA
PSB_G6529	hypothetical protein	3	four helix bundle protein	NA	NA
PSB_G4336	hypothetical protein	3	NA	NA	NA
PSB_G4847	hypothetical protein	3	NA	NA	NA
PSB_G3533	hypothetical protein	3	NA	NA	NA
PSB_G4279	hypothetical protein	3	NA	NA	NA
PSB_G3785	hypothetical protein	3	DNA-binding protein	NA	uncharacterized protein
PSB_G6566	UPF0158 family protein	3	UPF0158 family protein	NA	NA
PSB_G4323	hypothetical protein	3	NA	NA	NA
PSB_G5320	hypothetical protein	3	NA	NA	NA
PSB_G2036	class I SAM-dependent methyltransferase	3	class I SAM-dependent methyltransferase	O-Methyltransferase involved in polyketide biosynthesis (YktD) (PDB:1RJD)	NA
PSB_G5233	hypothetical protein	3	NA	NA	NA
PSB_G4663	light-harvesting antenna LH1, alpha subunit	3	light-harvesting protein	NA	light-harvesting protein B-800-850 alpha chain
PSB_G1449	transposase	3	IS91 family transposase	dTDP-4-amino-4,6-dideoxygalactose transaminase (WecE) (PDB:4PIW) (PUBMED:15271350)	NA
PSB_G3265	transporter substrate-binding domain-containing protein	3	transporter substrate-binding domain-containing protein	Signal transduction histidine kinase (BaeS) (PDB:1JOY)!!!Chemotaxis protein histidine kinase CheA (CheA) (PDB:1B3Q)!!!CheY-like REC (receiver) domain, includes chemotaxis protein CheY and sporulation regulator Spo0F (CheY) (PDB:6QRJ)!!!ABC-type amino acid transport/signal transduction system, periplasmic component/domain (HisJ) (PDB:1GGG)	NA
PSB_G4729	hypothetical protein	3	NA	NA	NA
PSB_G5305	hypothetical protein	3	NA	NA	NA
PSB_G6508	three-Cys-motif partner protein TcmP	3	three-Cys-motif partner protein TcmP (tcmP)	Formylglycine-generating enzyme, required for sulfatase activity, contains SUMF1/FGE domain (YfmG) (PDB:2AFT)!!!Predicted NTPase, NACHT family domain (NACHT)	NA
PSB_G3115	Uma2 family endonuclease	3	Uma2 family endonuclease	Endonuclease, Uma2 family (restriction endonuclease fold) (Uma2) (PDB:1WDJ) (PUBMED:17873039)	NA
PSB_G2531	hypothetical protein	3	NA	Transposase, IS1182 family	NA
PSB_G4171	hypothetical protein	3	NA	NA	NA
PSB_G688	cyclic peptide export ABC transporter	3	cyclic peptide export ABC transporter	ABC-type siderophore export system, fused ATPase and permease components (PvdE)	putative pyoverdin transport system ATP-binding/permease protein
PSB_G6571	NAD(P)-binding protein	3	NAD(P)-binding protein	NA	NA
PSB_G1033	alpha/beta hydrolase	3	alpha/beta hydrolase	Esterase/lipase superfamily enzyme	NA
PSB_G5270	hypothetical protein	3	NA	NA	NA
PSB_G6179	TIGR02678 family protein	3	TIGR02678 family protein	NA	NA
PSB_G3848	hypothetical protein	3	NA	NA	NA
PSB_G5078	hypothetical protein	3	NA	NA	NA
PSB_G4482	hypothetical protein	3	NA	Cation-transporting P-type ATPase (ZntA) (PDB:4BBJ)	NA
PSB_G5245	hypothetical protein	3	NA	NA	NA
PSB_G2865	protein-L-isoaspartate O-methyltransferase family protein	4	protein-L-isoaspartate O-methyltransferase	Protein-L-isoaspartate O-methyltransferase (Pcm) (PDB:3LBF) (PUBMED:24909784)	protein-L-isoaspartate(D-aspartate) O-methyltransferase [EC:2.1.1.77]
PSB_G1653	nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase	4	nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase (cobT)	NaMN:DMB phosphoribosyltransferase (CobT) (PDB:1D0S)	nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase [EC:2.4.2.21]
PSB_G1543	tRNA guanosine(34) transglycosylase Tgt	4	tRNA guanosine(34) transglycosylase Tgt (tgt)	Queuine/archaeosine tRNA-ribosyltransferase (Tgt) (PDB:1XW8) (PUBMED:15805528)	queuine tRNA-ribosyltransferase [EC:2.4.2.29]
PSB_G1061	adenosylmethionine--8-amino-7-oxononanoate transaminase	4	adenosylmethionine--8-amino-7-oxononanoate transaminase (bioA)	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (BioA) (PDB:1MLY)	adenosylmethionine---8-amino-7-oxononanoate aminotransferase [EC:2.6.1.62]

PSB__G964	argininosuccinate lyase	4	argininosuccinate lyase (argH)	Argininosuccinate lyase (ArgH) (PDB:1AOS)	argininosuccinate lyase [EC:4.3.2.1]
PSB__G2198	acetyl-CoA carboxylase, carboxyltransferase subunit beta	4	acetyl-CoA carboxylase carboxyltransferase subunit beta	Acetyl-CoA carboxylase beta subunit (AccD) (PDB:2F9I)	acetyl-CoA carboxylase carboxyl transferase subunit beta [EC:6.4.1.2 2.1.3.15]
PSB__G498	dipeptide/oligopeptide/nickel ABC transporter permease/ATP-binding protein	4	dipeptide/oligopeptide/nickel ABC transporter permease/ATP-binding protein	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component (DppD) (PDB:4FWI)!!!ABC-type dipeptide/oligopeptide/nickel transport system, permease component (DppC)	peptide/nickel transport system ATP-binding protein
PSB__G952	chromosomal replication initiator protein DnaA	4	chromosomal replication initiator protein DnaA (dnaA)	Chromosomal replication initiation ATPase DnaA (DnaA) (PDB:1L8Q)	chromosomal replication initiator protein
PSB__G1527	DNA polymerase III subunit beta	4	DNA polymerase III subunit beta	DNA polymerase III sliding clamp (beta) subunit, PCNA homolog (DnaN) (PDB:1JQJ)	DNA polymerase III subunit beta [EC:2.7.7.7]
PSB__G1832	GDP-L-fucose synthase family protein	4	GDP-L-fucose synthase	Nucleoside-diphosphate-sugar epimerase (WcaG) (PDB:2X4G)	GDP-L-fucose synthase [EC:1.1.1.271]
PSB__G912	coproporphyrinogen-III oxidase family protein	4	coproporphyrinogen III oxidase family protein	Coproporphyrinogen-III oxidase HemN (oxygen-independent) or related Fe-S oxidoreductase (HemN) (PDB:1OLT)	oxygen-independent coproporphyrinogen III oxidase [EC:1.3.98.3]
PSB__G5092	50S ribosomal protein L30	4	50S ribosomal protein L30 (rpmD)	Ribosomal protein L30/L7E (RpmD) (PDB:1VS6)	large subunit ribosomal protein L30
PSB__G1904	protein translocase subunit SecF	4	protein translocase subunit SecF (secF)	Preprotein translocase subunit SecF (SecF) (PDB:2FYH) (PUBMED:28830929)	preprotein translocase subunit SecF
PSB__G3485	signal peptidase II	4	signal peptidase II (lspA)	Lipoprotein signal peptidase (LspA) (PDB:5DIR)	signal peptidase II [EC:3.4.23.36]
PSB__G1552	glycine oxidase ThiO	4	glycine oxidase ThiO (thiO)	Glycine/D-amino acid oxidase (deaminating) (DadA) (PDB:3AWI)	glycine oxidase [EC:1.4.3.19]
PSB__G2064	bile acid:sodium symporter family protein	4	bile acid:sodium symporter family protein	Predicted Na+-dependent transporter YfeH (YfeH) (PDB:4N7X) (PUBMED:21175741)	bile acid:Na+ symporter, BASS family
PSB__G6313	LexA family protein	4	translesion error-prone DNA polymerase V autoproteolytic subunit (umuD)	SOS-response transcriptional repressor LexA (RecA-mediated autopeptidase) (LexA) (PDB:1AY9)	DNA polymerase V [EC:3.4.21.-]
PSB__G3058	Holliday junction branch migration protein RuvA	4	Holliday junction branch migration protein RuvA (ruvA)	Holliday junction resolvasome RuvABC DNA-binding subunit (RuvA) (PDB:1BDX)	holliday junction DNA helicase RuvA [EC:5.6.2.4]
PSB__G3658	ExbD/TolR family protein	4	biopolymer transporter ExbD	Biopolymer transport protein ExbD (ExbD) (PDB:2JWK)	biopolymer transport protein ExbD
PSB__G829	electron transport complex subunit RsxC	4	electron transport complex subunit RsxC (rsxC)	Na+-translocating ferredoxin:NAD+ oxidoreductase RNF, RnfC subunit (RnfC)	H+/Na+-translocating ferredoxin:NAD+ oxidoreductase subunit C [EC:7.1.1.11 7.2.1.2]
PSB__G1722	GTP 3',8-cyclase MoaA	4	GTP 3',8-cyclase MoaA (moaA)	GTP 3',8-cyclase (molybdenum cofactor biosynthesis protein MoaA) (MoaA) (PDB:1TV7)	GTP 3',8-cyclase [EC:4.1.99.22]
PSB__G2736	META domain-containing protein	4	META domain-containing protein	Uncharacterized lipoprotein YbaY (YbaY) (PDB:5WEC)!!!Heat shock protein HslJ (HslJ) (PDB:2KTS)!!!Uncharacterized membrane protein!!!Membrane-bound inhibitor of C-type lysozyme (MliC) (PDB:2F09)	heat shock protein HslJ
PSB__G2183	ferredoxin:protochlorophyllide reductase (ATP-dependent) iron-sulfur ATP-binding protein	4	ferredoxin:protochlorophyllide reductase (ATP-dependent) iron-sulfur ATP-binding protein	Nitrogenase ATPase subunit NifH/coenzyme F430 biosynthesis subunit CfbC (NifH/CfbC) (PDB:1CP2) (PUBMED:28225763)	light-independent protochlorophyllide reductase subunit L [EC:1.3.7.7]
PSB__G1305	M20 metallopeptidase family protein	4	amidohydrolase	Metal-dependent amidase/aminoacylase/carboxypeptidase (AbgB) (PDB:3RAM) (PUBMED:17307853)	NA
PSB__G428	PHB depolymerase family esterase	4	PHB depolymerase family esterase	Acetyl xylan esterase AxeA and related esterases, LpqC family (LpqC) (PDB:2D80) (PUBMED:9291122)	acetylxylan esterase [EC:3.1.1.72]
PSB__G4023	fluoride efflux transporter CrcB	4	fluoride efflux transporter CrcB (crcB)	Fluoride ion exporter CrcB/FEX, affects chromosome condensation (CrcB) (PDB:5A40)	fluoride exporter
PSB__G7025	ankyrin repeat domain-containing protein	4	ankyrin repeat domain-containing protein	Ankyrin repeat (ANKYR) (PDB:6AUN)	NA
PSB__G1569	redox-regulated ATPase YchF	4	redox-regulated ATPase YchF (ychF)	Ribosome-binding ATPase YchF, GTP1/OBG family (GTP1) (PDB:2DWQ) (PUBMED:21527254)	ribosome-binding ATPase
PSB__G72	VWA domain-containing protein	4	VWA domain-containing protein	Secreted protein containing bacterial Ig-like domain and vWFA domain (YfbK)	NA
PSB__G5126	type II toxin-antitoxin system HicA family toxin	4	type II toxin-antitoxin system HicA family toxin	Predicted RNA binding protein YcfA, dsRBD-like fold, HicA-like mRNA interferase family (YcfA) (PDB:1WHZ) (PUBMED:16895922)	mRNA interferase HicA [EC:3.1.-.-]
PSB__G6634	restriction endonuclease	4	restriction endonuclease	Predicted restriction endonuclease, Mrr-cat superfamily	restriction system protein
PSB__G2733	transposase	4	transposase	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB__G3011	transposase	4	transposase	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB__G7120	transposase	4	transposase	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB__G4816	hypothetical protein	4	NA	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB__G2685	transposase	4	transposase	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB__G7029	transposase	4	transposase	Transposase	putative transposase
PSB__G5973	transposase	4	transposase	IS4 transposase InsG (InsG)	putative transposase
PSB__G4622	light-harvesting antenna LH1, alpha subunit	4	light-harvesting protein	NA	light-harvesting protein B-800-850 alpha chain
PSB__G235	GNAT family N-acetyltransferase	4	bifunctional acetate--CoA ligase family protein/GNAT family N-acetyltransferase	Acyl-CoA synthetase (NDP forming) (PatZN) (PDB:4YAK) (PUBMED:26251518)!!!L-amino acid N-acyltransferase MnaT (MnaT) (PDB:3DR6) (PUBMED:27941785)	acetyltransferase
PSB__G6748	lipopolysaccharide transport periplasmic proteinLptA	4	lipopolysaccharide transport periplasmic protein LptA (lptA)	Lipopolysaccharide export system protein LptA (LptA) (PDB:2R19)	lipopolysaccharide export system protein LptA
PSB__G6657	YdcH family protein	4	YdcH family protein	Uncharacterized conserved protein YdcH, DUF465 family (YdcH) (PDB:1ZHC)	uncharacterized protein
PSB__G1545	glycolate oxidase subunit GlcE	4	glycolate oxidase subunit GlcE (glcE)	FAD/FMN-containing lactate dehydrogenase/glycolate oxidase (GlcD) (PDB:1AHU)	glycolate dehydrogenase FAD-binding subunit [EC:1.1.99.14]

PSB__G1437	cation diffusion facilitator family transporter	4	cation transporter	Divalent metal cation (Fe/Co/Zn/Cd) efflux pump (FieF) (PDB:2QFI)	ferrous-iron efflux pump FieF
PSB__G6629	nitrogen fixation protein NifQ	4	nitrogen fixation protein NifQ	Fe-S cluster assembly scaffold protein IscU, NifU family (IscU) (PDB:1Q48)	nitrogen fixation protein NifQ
PSB__G2364	CDP-diacylglycerol--serine O-phosphatidyltransferase	4	CDP-diacylglycerol--serine O-phosphatidyltransferase (pssA)	Phosphatidylserine synthase (PssA)	CDP-diacylglycerol---serine O-phosphatidyltransferase [EC:2.7.8.8]
PSB__G4358	type II toxin-antitoxin system RelE/ParE family toxin	4	type II toxin-antitoxin system RelE/ParE family toxin	Plasmid stabilization system protein ParE (ParE) (PDB:3KXE)	toxin ParE1/3/4
PSB__G1703	tRNA (adenosine(37)-N6)-threonylcarbamoyltransferase complex transferase subunit TsaD	4	tRNA (adenosine(37)-N6)-threonylcarbamoyltransferase complex transferase subunit TsaD (tsaD)	tRNA A37 threonylcarbamoyltransferase TsaD (TsaD) (PDB:3ENO) (PUBMED:23072323;21285948;32385138)	tRNA N6-adenosine threonylcarbamoyltransferase [EC:2.3.1.234]
PSB__G5354	hypothetical protein	4	NA	NA	NA
PSB__G3454	hypothetical protein	4	NA	NA	NA
PSB__G2169	IS66 family transposase	4	IS66 family transposase	NA	NA
PSB__G5366	IS1/IS1595 family N-terminal zinc-binding domain-containing protein	4	NA	NA	NA
PSB__G4451	hypothetical protein	4	NA	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
PSB__G4158	hypothetical protein	4	NA	NA	NA
PSB__G6524	hypothetical protein	4	NA	NA	NA
PSB__G4935	type II toxin-antitoxin system RelE/ParE family toxin	4	type II toxin-antitoxin system RelE/ParE family toxin	Plasmid maintenance system killer protein (HigB)	toxin HigB-1
PSB__G5407	carboxylesterase	4	NA	NA	NA
PSB__G5359	IS4 family transposase	4	NA	NA	NA
PSB__G6701	4-alpha-glucanotransferase	4	NA	NA	NA
PSB__G496	P-loop ATPase, Sll1717 family	4	P-loop ATPase, Sll1717 family	NA	NA
PSB__G4855	DUF2283 domain-containing protein	4	DUF2283 domain-containing protein	Predicted antitoxin component YuzE of a toxin-antitoxin system, DUF2283 family (YuzE)	NA
PSB__G4693	hypothetical protein	4	NA	NA	NA
PSB__G4013	hypothetical protein	4	NA	NA	NA
PSB__G3653	hypothetical protein	4	NA	NA	NA
PSB__G1859	helix-turn-helix domain-containing protein	4	helix-turn-helix domain-containing protein	AraC-type DNA-binding domain and AraC-containing proteins (AraC) (PDB:1BL0)	AraC family transcriptional regulator, ethanolamine operon transcriptional activator
PSB__G1679	dTDP-glucose 4,6-dehydratase	4	dTDP-glucose 4,6-dehydratase (rfbB)	dTDP-D-glucose 4,6-dehydratase (RfbB) (PDB:1G1A)	dTDP-glucose 4,6-dehydratase [EC:4.2.1.46]
PSB__G5556	hypothetical protein	4	NA	NA	NA
PSB__G1514	HNH endonuclease signature motif containing protein	4	HNH endonuclease signature motif containing protein	NA	5-methylcytosine-specific restriction enzyme A [EC:3.1.21.-]
PSB__G883	WSC domain-containing protein	4	WSC domain-containing protein	Type IV pilus assembly protein, tip-associated adhesin PilY1 (PilY1) (PDB:3HX6)	NA
PSB__G490	ATP-dependent helicase	4	ATP-dependent helicase	Superfamily I DNA or RNA helicase (UvrD) (PDB:6PPR)	ATP-dependent DNA helicase UvrD/PcrA [EC:5.6.2.4]
PSB__G4643	phage integrase N-terminal SAM-like domain-containing protein	4	phage integrase N-terminal SAM-like domain-containing protein	NA	NA
PSB__G3090	hypothetical protein	4	NA	NA	NA
PSB__G4521	hypothetical protein	4	NA	NA	NA
PSB__G2110	glycosyltransferase	4	glycosyltransferase	Glycosyltransferase, GT2 family (WcaE) (PDB:2Z86)	N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase [EC:2.4.1.289]
PSB__G5553	hypothetical protein	4	NA	NA	NA
PSB__G3334	hypothetical protein	4	NA	NA	NA
PSB__G3196	ClpX C4-type zinc finger protein	4	ClpX C4-type zinc finger protein	NA	NA
PSB__G3110	hypothetical protein	4	NA	NA	NA
PSB__G2910	hypothetical protein	4	NA	NA	NA
PSB__G2691	hypothetical protein	4	NA	NA	NA
PSB__G2585	WSC domain-containing protein	4	WSC domain-containing protein	NA	NA
PSB__G1986	hypothetical protein	4	NA	NA	NA
PSB__G1960	hypothetical protein	4	NA	NA	NA
PSB__G5554	hypothetical protein	4	NA	NA	NA
PSB__G1393	glycosyltransferase family 4 protein	4	glycosyltransferase family 4 protein	Glycosyltransferase involved in cell wall bisynthesis (RfaB) (PDB:2IV7)	NA
PSB__G3410	hypothetical protein	4	NA	NA	NA
PSB__G5521	DNA/RNA nuclease SfsA	4	DNA/RNA nuclease SfsA (sfsA)	DNA-binding protein, stimulates sugar fermentation (SfsA) (PDB:4DA2)	sugar fermentation stimulation protein A
PSB__G645	hypothetical protein	4	NA	DNA repair exonuclease SbcCD ATPase subunit (SbcC) (PDB:3QKR)	NA
PSB__G37	hypothetical protein	4	NA	NA	NA
PSB__G5454	hypothetical protein	4	NA	NA	NA
PSB__G3912	hypothetical protein	4	NA	NA	NA
PSB__G3279	hypothetical protein	4	NA	NA	NA
PSB__G22	neuraminidase-like domain-containing protein	4	neuraminidase-like domain-containing protein	Chromosome segregation ATPase Smc (Smc) (PDB:5XG3)	NA
PSB__G5182	hypothetical protein	4	NA	NA	NA
PSB__G617	thiamine pyrophosphate-binding protein	4	thiamine pyrophosphate-binding protein	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase (MenD) (PDB:5Z2P)	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase [EC:2.2.1.9]
PSB__G4339	hypothetical protein	4	NA	RNA interferase (RNase) VapC, contains PIN domain (VapC) (PDB:1V8O) (PUBMED:22539524)	NA
PSB__G3319	hypothetical protein	4	NA	NA	NA
PSB__G3255	hypothetical protein	4	NA	NA	NA
PSB__G1456	glycosyltransferase	4	glycosyltransferase	Glycosyltransferase involved in cell wall bisynthesis (RfaB) (PDB:2IV7)	NA
PSB__G1700	acyltransferase family protein	4	acyltransferase family protein	Peptidoglycan/LPS O-acetylase OafA/YrhL, contains acyltransferase and SGNH-hydrolase domains (OafA)	exopolysaccharide production protein ExoZ
PSB__G3117	hypothetical protein	4	NA	NA	NA
PSB__G2608	hypothetical protein	4	NA	Uncharacterized protein RhuM, Salmonella virulence factor (RhuM)	NA
PSB__G3306	hypothetical protein	4	NA	NA	NA
PSB__G2549	hypothetical protein	4	NA	NA	NA
PSB__G6138	IS701 family transposase	4	IS701 family transposase	SRSO17 transposase (Srso17)	NA
PSB__G5608	hypothetical protein	4	NA	NA	NA

PSB_G3382	hypothetical protein	4	NA	NA	NA
PSB_G2190	ferritin-like domain-containing protein	4	ferritin-like domain-containing protein	Uncharacterized protein, contains ferritin-like DUF455 domain	NA
PSB_G3898	diacylglycerol kinase	4	diacylglycerol kinase	Diacylglycerol kinase (DgkA) (PDB:2KDC)	NA
PSB_G1761	AAA family ATPase	4	AAA domain-containing protein	MoxR-like ATPase (MoxR) (PDB:2R44)	NA
PSB_G5683	hypothetical protein	4	NA	NA	NA
PSB_G6255	IS1 family transposase	4	IS1 family transposase	NA	NA
PSB_G5380	IS1-like element transposase	4	IS1-like element transposase	Transposase InsA (InsA)	NA
PSB_G2976	hypothetical protein	4	NA	NA	NA
PSB_G2168	hypothetical protein	4	NA	NA	NA
PSB_G2516	NRDE family protein	4	NRDE family protein	Uncharacterized stress-responsive protein, TANGO2 (Transport and Golgi organisation 2) family, contains NRDE motif (Tango2) (PUBMED:26631869)	NA
PSB_G2845	YcjF family protein	4	DUF697 domain-containing protein	Uncharacterized conserved protein, DUF697 family	NA
PSB_G1089	efflux RND transporter periplasmic adaptor subunit	4	efflux RND transporter periplasmic adaptor subunit	Multidrug resistance efflux pump EmrA (EmrA) (PDB:4TKO)	NA
PSB_G2316	hypothetical protein	4	NA	NA	NA
PSB_G5039	hypothetical protein	4	NA	NA	NA
PSB_G4874	hypothetical protein	4	NA	NA	NA
PSB_G4876	hypothetical protein	4	NA	NA	NA
PSB_G6523	hypothetical protein	4	two pore domain potassium channel family protein	NA	NA
PSB_G3221	Uma2 family endonuclease	4	Uma2 family endonuclease	Endonuclease, Uma2 family (restriction endonuclease fold) (Uma2) (PDB:1WDJ) (PUBMED:17873039)	NA
PSB_G2382	hypothetical protein	4	NA	NA	NA
PSB_G5833	type II toxin-antitoxin system VapC family toxin	4	type II toxin-antitoxin system VapC family toxin	NA	NA
PSB_G6881	MFS transporter	4	MFS transporter	Na+/melibiose symporter or related transporter (MelB) (PDB:4M64)	glycoside/pentoside/hexuronide:cation symporter, GPH family
PSB_G3875	DUF4277 domain-containing protein	4	DUF4277 domain-containing protein	Transposase	NA
PSB_G4423	hypothetical protein	4	NA	NA	NA
PSB_G3817	OsmC family protein	4	OsmC family protein	Uncharacterized OsmC-related protein (YhfA) (PDB:1LQL)	putative redox protein
PSB_G3842	UvrD-helicase domain-containing protein	4	UvrD-helicase domain-containing protein	Superfamily I DNA or RNA helicase (UvrD) (PDB:6PPR)	NA
PSB_G280	Hsp70 family protein	4	Hsp70 family protein	Molecular chaperone DnaK (HSP70) (DnaK) (PDB:3I33)!!!Tetratricopeptide (TPR) repeat (TPR) (PDB:3AS4)	NA
PSB_G2813	hypothetical protein	4	NA	NA	NA
PSB_G287	PAS domain S-box protein	4	PAS domain S-box protein	PAS domain (PAS) (PDB:2MWG)!!!GAF domain (GAF)!!!Signal transduction histidine kinase regulating C4-dicarboxylate transport system (PDB:4GCZ)	NA
PSB_G6265	transposase	4	NA	Transposase	NA
PSB_G388	hypothetical protein	4	NA	Exopolysaccharide export protein/domain GumC/Wzc1 (GumC) (PUBMED:15003255;23089617)	NA
PSB_G1638	alpha/beta fold hydrolase	4	alpha/beta hydrolase	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase MenH and related esterases, alpha/beta hydrolase fold (MenH) (PDB:4MYS)	NA
PSB_G4004	hypothetical protein	4	NA	NA	NA
PSB_G4427	DUF4405 domain-containing protein	4	DUF4405 domain-containing protein	NA	NA
PSB_G6597	phage integrase N-terminal SAM-like domain-containing protein	4	phage integrase N-terminal SAM-like domain-containing protein	Site-specific recombinase XerD (XerD) (PDB:1A0P)	NA
PSB_G4165	hypothetical protein	4	NA	NA	NA
PSB_G2474	toll/interleukin-1 receptor domain-containing protein	4	toll/interleukin-1 receptor domain-containing protein	NA	NA
PSB_G1284	hypothetical protein	4	NA	NA	NA
PSB_G710	ATP-binding protein	4	HAMP domain-containing protein	Sensor histidine kinase WalK (WalK) (PDB:4I5S)	NA
PSB_G3492	hypothetical protein	4	NA	NA	NA
PSB_G2755	YqiJ family protein	4	DUF1449 family protein	NA	NA
PSB_G5108	hypothetical protein	4	NA	NA	NA
PSB_G2196	polyprenyl synthetase family protein	4	geranylgeranyl pyrophosphate synthase	Geranylgeranyl pyrophosphate synthase (IspA) (PDB:2AZJ)	NA
PSB_G3934	DUF302 domain-containing protein	4	DUF302 domain-containing protein	Uncharacterized conserved protein, DUF302 family (PDB:1J3M)	NA
PSB_G1788	hypothetical protein	4	NA	NA	NA
PSB_G6526	ArsC family reductase	4	ArsC family reductase	Arsenate reductase or related protein, glutaredoxin family (ArsC) (PDB:1I9D)	arsenate reductase (glutaredoxin) [EC:1.20.4.1]
PSB_G4317	hypothetical protein	4	NA	NA	NA
PSB_G4493	helix-turn-helix domain-containing protein	4	helix-turn-helix domain-containing protein	NA	NA
PSB_G1200	ATP-binding protein	4	HAMP domain-containing protein	Signal transduction histidine kinase (BaeS) (PDB:1JOY)	NA
PSB_G3746	rhodanese-like domain-containing protein	4	rhodanese-like domain-containing protein	Rhodanese-related sulfurtransferase (PspE) (PDB:1TQ1)	phage shock protein E
PSB_G1889	hypothetical protein	4	NA	NA	NA
PSB_G7299	ABC transporter ATP-binding protein	4	ATP-binding cassette domain-containing protein	ABC-type Fe3+/spermidine/putrescine transport systems, ATPase component (PotA) (PDB:1Z47)	NA
PSB_G3245	chromate resistance protein ChrB domain-containing protein	4	chromate resistance protein ChrB domain-containing protein	Chromate resistance protein ChrB1 (ChrB1)	NA
PSB_G981	AAA family ATPase	4	ATP-binding protein	Predicted ATPase	putative ATP-dependent endonuclease of the OLD family
PSB_G7066	IS66 family transposase	4	NA	NA	NA
PSB_G2900	DUF4433 domain-containing protein	4	DUF4433 domain-containing protein	NA	NA
PSB_G5944	hypothetical protein	4	NA	NA	NA
PSB_G4556	hypothetical protein	4	NA	NA	NA

PSB__G6786	SPOR domain-containing protein	4	SPOR domain-containing protein	Cell division protein DamX, binds to the septal ring, contains C-terminal SPOR domain (DamX)	NA
PSB__G5194	hypothetical protein	4	NA	Phosphotransacetylase (includes Pta, EutD and phosphobutyryltransferase) (Pta) (PDB:1QZT) (PUBMED:28754323)!!!BioD-like N-terminal domain of phosphotransacetylase (PtaN) (PDB:2IOJ)	NA
PSB__G4615	addiction module protein	4	addiction module protein	NA	NA
PSB__G3628	Uma2 family endonuclease	4	Uma2 family endonuclease	Endonuclease, Uma2 family (restriction endonuclease fold) (Uma2) (PDB:1WDJ) (PUBMED:17873039)	NA
PSB__G4655	hypothetical protein	4	NA	NA	NA
PSB__G4617	hypothetical protein	4	NA	NA	NA
PSB__G5314	hypothetical protein	4	NA	NA	NA
PSB__G3043	thioredoxin family protein	4	thioredoxin family protein	Peroxiredoxin (Bcp) (PDB:5IPH)	NA
PSB__G4946	type II toxin-antitoxin system HicB family antitoxin	4	type II toxin-antitoxin system HicB family antitoxin	Antitoxin component HicB of the HicAB toxin-antitoxin system (HicB) (PDB:4P78) (PUBMED:19060138)	NA
PSB__G751	ATP-binding protein	4	response regulator	Signal transduction histidine kinase (BaeS) (PDB:1JOY)!!!DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain (OmpR) (PDB:1XHF)	NA
PSB__G3332	Rab family GTPase	4	GTP-binding protein	GTPase SAR1 family domain (Gem1) (PDB:4C0J)	NA
PSB__G4041	DUF5615 family PIN-like protein	4	DUF5615 family PIN-like protein	Predicted nuclease, contains PIN domain, potential toxin-antitoxin system component (PUBMED:19493340)	NA
PSB__G3519	hypothetical protein	4	NA	NA	NA
PSB__G3959	transposase	4	transposase	Transposase, IS1182 family	NA
PSB__G2154	ATPase	4	ATPase	Adenylate kinase or related kinase (Adk) (PDB:1AK2)	NA
PSB__G3550	Uma2 family endonuclease	4	Uma2 family endonuclease	Endonuclease, Uma2 family (restriction endonuclease fold) (Uma2) (PDB:1WDJ) (PUBMED:17873039)	NA
PSB__G6280	NAD(P)-dependent oxidoreductase	4	SDR family oxidoreductase	Putative NADH-flavin reductase (YwnB) (PDB:3EW7)	NA
PSB__G6445	hypothetical protein	4	NA	NA	NA
PSB__G3987	hypothetical protein	4	NA	NA	NA
PSB__G7236	DUF2283 domain-containing protein	4	DUF2283 domain-containing protein	Predicted antitoxin component YuzE of a toxin-antitoxin system, DUF2283 family (YuzE)	NA
PSB__G1883	lysylphosphatidylglycerol synthase transmembranedomain-containing protein	4	flippase-like domain-containing protein	NA	NA
PSB__G6836	RibD family protein	4	RibD family protein	Pyrimidine reductase, riboflavin biosynthesis (RibD) (PDB:2O7P)	diaminohydroxyphosphoribosylaminopyrimidine deaminase / 5-amino-6-(5-phosphoribosylamino)uracil reductase [EC:3.5.4.26 1.1.1.193]
PSB__G6281	hypothetical protein	4	NA	Glycosyltransferase, GT2 family (WcaE) (PDB:2Z86)	NA
PSB__G6706	glycosyltransferase	4	glycosyltransferase	Glycosyltransferase involved in cell wall bisynthesis (RfaB) (PDB:2IV7)	NA
PSB__G3072	DUF6817 domain-containing protein	4	DUF6817 domain-containing protein	NA	NA
PSB__G2772	hypothetical protein	4	NA	Glycosyltransferase involved in cell wall bisynthesis (RfaB) (PDB:2IV7)	NA
PSB__G3938	hypothetical protein	4	NA	NA	NA
PSB__G3704	hypothetical protein	4	NA	NA	NA
PSB__G4033	hypothetical protein	4	NA	NA	NA
PSB__G4048	NifB/NifX family molybdenum-iron cluster-bindingprotein	4	nitrogen fixation protein	Predicted Fe-Mo cluster-binding protein, NifX family (NifX) (PDB:1EO1)	NA
PSB__G3718	DUF6858 family protein	4	DUF6858 family protein	NA	NA
PSB__G6927	phage integrase N-terminal SAM-like domain-containing protein	4	phage integrase N-terminal SAM-like domain-containing protein	Site-specific recombinase XerD (XerD) (PDB:1A0P)	NA
PSB__G5049	hypothetical protein	4	NA	Site-specific recombinase XerD (XerD) (PDB:1A0P)	NA
PSB__G2499	sulfotransferase domain-containing protein	4	sulfotransferase domain-containing protein	NA	NA
PSB__G401	heavy metal translocating P-type ATPase	4	heavy metal translocating P-type ATPase	Cation-transporting P-type ATPase (ZntA) (PDB:4BBJ)	NA
PSB__G2405	thioesterase domain-containing protein	4	alpha/beta fold hydrolase	Acyl carrier protein (AcpP) (PDB:1ACP)!!!EntF, seryl-AMP synthase component of non-ribosomal peptide synthetase (EntF) (PDB:5ES5)!!!Thioesterase domain of type I polyketide synthase or non-ribosomal peptide synthetase (EntF2) (PDB:5JA1)	NA
PSB__G2432	AMP-binding protein	4	AMP-binding protein	EntF, seryl-AMP synthase component of non-ribosomal peptide synthetase (EntF) (PDB:5ES5)	NA
PSB__G195	CHAT domain-containing protein	4	CHAT domain-containing protein	Uncharacterized conserved protein, contains CHAT domain (PUBMED:11835511)	NA
PSB__G6450	hypothetical protein	4	NA	NA	NA
PSB__G5197	hypothetical protein	4	NA	NA	NA
PSB__G6302	hypothetical protein	4	NA	Ca2+-binding protein, RTX toxin-related (PDB:1AF0) (PUBMED:15135544)	NA
PSB__G3812	hypothetical protein	4	NA	NA	NA
PSB__G7056	ISAs1 family transposase	4	ISAs1 family transposase	Predicted transposase YbfD/YdcC associated with H repeats (YhhI)	NA
PSB__G6595	four helix bundle protein	4	four helix bundle protein	NA	NA
PSB__G4010	Rieske (2Fe-2S) protein	4	Rieske 2Fe-2S domain-containing protein	Ferredoxin subunit of nitrite reductase or a ring-hydroxylating dioxygenase (NirD) (PDB:1FQT) (PUBMED:23844911)	NA

PSB__G986	integron integrase	4	integron integrase	Site-specific recombinase XerD (XerD) (PDB:1A0P)	NA
PSB__G6421	reverse transcriptase domain-containing protein	4	reverse transcriptase domain-containing protein	NA	NA
PSB__G341	DUF2793 domain-containing protein	4	DUF2793 domain-containing protein	NA	NA
PSB__G1790	helix-turn-helix domain-containing protein	4	helix-turn-helix domain-containing protein	AraC-type DNA-binding domain and AraC-containing proteins (AraC) (PDB:1BL0)	NA
PSB__G4094	hypothetical protein	4	NA	Uncharacterized conserved protein (PDB:2HWJ)	NA
PSB__G4449	hypothetical protein	4	NA	NA	NA
PSB__G1927	acetyltransferase	4	acetyltransferase	ssDNA-specific exonuclease RecJ, DHH superfamily, may be involved in archaeal DNA replication intiation (RecJ) (PDB:1IR6)	NA
PSB__G1114	ATP-dependent 6-phosphofructokinase	4	ATP-dependent 6-phosphofructokinase	6-phosphofructokinase (PfkA) (PDB:1KZH)	6-phosphofructokinase 1 [EC:2.7.1.11]
PSB__G3672	BLUF domain-containing protein	4	BLUF domain-containing protein	Uncharacterized membrane protein, DUF373 family	NA
PSB__G1334	BPSS1780 family membrane protein	4	BPSS1780 family membrane protein	Uncharacterized membrane protein	NA
PSB__G3557	Uma2 family endonuclease	4	Uma2 family endonuclease	Endonuclease, Uma2 family (restriction endonuclease fold) (Uma2) (PDB:1WDJ) (PUBMED:17873039)	NA
PSB__G1566	NACHT domain-containing protein	4	pentapeptide repeat-containing protein	Serine/threonine protein kinase (SPS1) (PDB:6G4J)!!!Predicted NTPase, NACHT family domain (NACHT)	NA
PSB__G6953	DUF3301 domain-containing protein	4	DUF3301 domain-containing protein	NA	NA
PSB__G3849	hypothetical protein	4	NA	NA	NA
PSB__G2557	UTRA domain-containing protein	4	GntR family transcriptional regulator	NA	NA
PSB__G6833	UPF0236 family transposase-like protein	4	UPF0236 family transposase-like protein	NA	NA
PSB__G3195	hypothetical protein	4	NA	NA	NA
PSB__G1741	methyltransferase domain-containing protein	4	methyltransferase domain-containing protein	2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase (UbiG) (PDB:1BHJ)	NA
PSB__G1985	hypothetical protein	4	NA	NA	NA
PSB__G2664	potassium channel family protein	4	two pore domain potassium channel family protein	NA	NA
PSB__G356	FUSC family protein	4	FUSC family protein	Uncharacterized membrane protein YccC (YccC)	NA

Supplemental Table 3. PB-PSB1 clusters 1-4 genes, and their associated PGAP, COG20 and KOfam annotations

Gene	PSB_reference_genome	sample_024	sample_002	sample_001	Gene	PSB_reference_genome	sample_024	sample_002	sample_001	Gene	PSB_reference_genome	sample_024	sample_002	sample_001
PSB__G2865	N838_12230	ACPL57_07000	ACLI5Z_27770	ACLH7Q_18165	PSB__G6211	N838_11870	ACPL57_07375	ACLI5Z_28140	ACLH7Q_18545	PSB__G3778	N838_12780	ACPL57_06455	ACLI5Z_27210	ACLH7Q_17585
PSB__G1208	N838_07965	ACPL57_02570	ACLI5Z_32165	ACLH7Q_12995	PSB__G5451	NA	ACPL57_30935	NA	NA	PSB__G4004	N838_08585	ACPL57_01925	ACLI5Z_31515	ACLH7Q_12345
PSB__G4821	0_refound_122	ACPL57_16390	ACLI5Z_08910	ACLH7Q_20315	PSB__G496	NA	ACPL57_00005	ACLI5Z_29590	ACLH7Q_08215	PSB__G4427	N838_09080	ACPL57_01410	ACLI5Z_31005	ACLH7Q_08600
PSB__G3066	0_refound_123	ACPL57_16395	ACLI5Z_08905	ACLH7Q_20310	PSB__G4855	NA	ACPL57_29130	NA	NA	PSB__G4091	N838_23750	ACPL57_13030	ACLI5Z_04220	ACLH7Q_05735
PSB__G1653	N838_09870	ACPL57_00615	ACLI5Z_30215	ACLH7Q_07600	PSB__G4548	NA	NA	NA	ACLH7Q_32665	PSB__G3797	N838_23825	ACPL57_13105	ACLI5Z_04295	ACLH7Q_05660
PSB__G1543	N838_01285	ACPL57_15810	ACLI5Z_15420	ACLH7Q_33460	PSB__G960	N838_11980	ACPL57_07265	ACLI5Z_28020	ACLH7Q_18425	PSB__G5020	N838_12785	ACPL57_06450	ACLI5Z_27205	ACLH7Q_17580
PSB__G4756	0_refound_149	ACPL57_20700	2_refound_600	3_refound_748	PSB__G4693	NA	ACPL57_29155	NA	NA	PSB__G2261	N838_23890	ACPL57_13165	ACLI5Z_04355	ACLH7Q_05595
PSB__G2170	0_refound_156	ACPL57_22650	2_refound_605	3_refound_753	PSB__G5510	NA	NA	NA	ACLH7Q_00140	PSB__G6597	N838_09700	ACPL57_00790	ACLI5Z_30385	ACLH7Q_07425
PSB__G5365	0_refound_159	ACPL57_23475	ACLI5Z_23730	ACLH7Q_26675	PSB__G5598	NA	NA	NA	ACLH7Q_32670	PSB__G6018	N838_23940	ACPL57_13215	ACLI5Z_04405	ACLH7Q_05545
PSB__G4088	N838_20960	ACPL57_25250	ACLI5Z_01370	ACLH7Q_22955	PSB__G4013	NA	ACPL57_21230	NA	NA	PSB__G4165	N838_09735	ACPL57_00755	ACLI5Z_30350	ACLH7Q_07460
PSB__G3560	0_refound_172	ACPL57_24875	2_refound_608	3_refound_756	PSB__G1468	NA	NA	NA	ACLH7Q_32465	PSB__G2474	N838_10390	ACPL57_00090	ACLI5Z_29675	ACLH7Q_08130
PSB__G495	0_refound_173	ACPL57_24920	ACLI5Z_01705	ACLH7Q_23285	PSB__G3653	NA	ACPL57_26925	NA	ACLH7Q_00785	PSB__G2281	N838_24115	ACPL57_13400	ACLI5Z_04595	ACLH7Q_05360
PSB__G3553	N838_20955	ACPL57_25255	ACLI5Z_01365	ACLH7Q_22950	PSB__G5597	NA	NA	NA	ACLH7Q_32660	PSB__G1646	N838_24130	ACPL57_13415	ACLI5Z_04610	ACLH7Q_05345
PSB__G2638	N838_20950	ACPL57_25260	ACLI5Z_01360	ACLH7Q_22945	PSB__G3945	NA	ACPL57_06930	NA	NA	PSB__G5924	N838_13190	ACPL57_06035	ACLI5Z_26795	ACLH7Q_12095
PSB__G1061	N838_20285	ACPL57_24305	ACLI5Z_00665	ACLH7Q_22245	PSB__G4736	N838_00095	ACPL57_29280	ACLI5Z_14220	NA	PSB__G1284	N838_11470	ACPL57_07785	ACLI5Z_28550	ACLH7Q_18955
PSB__G964	N838_19210	ACPL57_21745	ACLI5Z_20745	ACLH7Q_21150	PSB__G1859	NA	NA	ACLI5Z_16105	ACLH7Q_15085	PSB__G4967	N838_13915	ACPL57_05305	ACLI5Z_26050	NA
PSB__G181	N838_07360	ACPL57_03180	ACLI5Z_32770	ACLH7Q_13560	PSB__G878	N838_12405	ACPL57_06835	ACLI5Z_27580	ACLH7Q_17970	PSB__G3835	N838_24240	ACPL57_13525	ACLI5Z_04720	ACLH7Q_05235
PSB__G4735	0_refound_206	ACPL57_30640	ACLI5Z_06540	ACLH7Q_11195	PSB__G5312	N838_12415	ACPL57_06825	ACLI5Z_27570	ACLH7Q_17960	PSB__G4791	N838_24300	1_refound_428	ACLI5Z_04785	ACLH7Q_05170
PSB__G3851	0_refound_217	NA	ACLI5Z_05620	3_refound_766	PSB__G5313	N838_12440	ACPL57_06800	ACLI5Z_27545	ACLH7Q_17935	PSB__G5176	N838_13965	ACPL57_05255	2_refound_538	NA
PSB__G559	N838_19910	ACPL57_22440	ACLI5Z_00285	ACLH7Q_21865	PSB__G5333	N838_12450	ACPL57_06790	ACLI5Z_27535	ACLH7Q_17925	PSB__G4462	N838_24575	ACPL57_13860	ACLI5Z_05060	NA
PSB__G5409	0_refound_247	NA	ACLI5Z_19020	ACLH7Q_03485	PSB__G1679	NA	NA	NA	ACLH7Q_32480	PSB__G4132	N838_13985	ACPL57_05235	ACLI5Z_25980	NA
PSB__G937	N838_24685	ACPL57_13965	ACLI5Z_05170	ACLH7Q_04885	PSB__G5556	NA	NA	NA	ACLH7Q_09065	PSB__G4600	N838_14875	1_refound_405	ACLI5Z_25160	ACLH7Q_25255
PSB__G979	N838_03395	ACPL57_09040	ACLI5Z_18605	ACLH7Q_03090	PSB__G28	N838_00115	NA	ACLI5Z_14240	NA	PSB__G710	N838_11595	ACPL57_07655	ACLI5Z_28420	ACLH7Q_18825
PSB__G2198	N838_17615	ACPL57_19960	ACLI5Z_22380	ACLH7Q_28035	PSB__G4037	N838_00395	ACPL57_28905	ACLI5Z_14525	ACLH7Q_16740	PSB__G3492	N838_11915	ACPL57_07330	ACLI5Z_28090	ACLH7Q_18495
PSB__G5331	0_refound_273	NA	2_refound_627	ACLH7Q_11950	PSB__G4113	N838_00400	ACPL57_28900	ACLI5Z_14530	ACLH7Q_16735	PSB__G2755	N838_11990	ACPL57_07255	ACLI5Z_28010	ACLH7Q_18415
PSB__G5388	0_refound_280	1_refound_483	2_refound_634	ACLH7Q_20360	PSB__G2779	N838_00405	ACPL57_28895	ACLI5Z_14535	ACLH7Q_16730	PSB__G6684	N838_24960	ACPL57_31515	ACLI5Z_05440	ACLH7Q_04615
PSB__G4249	0_refound_31	ACPL57_05240	ACLI5Z_25985	NA	PSB__G4475	N838_12935	ACPL57_06290	ACLI5Z_27045	ACLH7Q_17420	PSB__G2275	N838_24965	ACPL57_31520	ACLI5Z_05445	ACLH7Q_04610
PSB__G6951	N838_22615	ACPL57_11850	ACLI5Z_03060	ACLH7Q_24655	PSB__G1514	NA	ACPL57_21085	NA	NA	PSB__G6668	N838_25060	ACPL57_31610	ACLI5Z_05540	ACLH7Q_04520
PSB__G2118	N838_22230	ACPL57_25715	ACLI5Z_02670	ACLH7Q_24265	PSB__G883	NA	ACPL57_21220	NA	NA	PSB__G2762	N838_25145	NA	ACLI5Z_05630	ACLH7Q_12135
PSB__G3130	N838_17220	ACPL57_20350	ACLI5Z_22775	ACLH7Q_27635	PSB__G5687	N838_00625	ACPL57_27975	ACLI5Z_14760	ACLH7Q_16485	PSB__G954	N838_25160	NA	ACLI5Z_05645	ACLH7Q_12120
PSB__G6360	0_refound_326	ACPL57_29780	ACLI5Z_21610	ACLH7Q_28780	PSB__G521	N838_13305	ACPL57_05915	ACLI5Z_26675	ACLH7Q_17050	PSB__G6413	N838_25165	NA	ACLI5Z_05650	ACLH7Q_12115
PSB__G498	N838_28485	ACPL57_16300	ACLI5Z_09000	ACLH7Q_20410	PSB__G490	NA	ACPL57_29250	NA	NA	PSB__G3216	N838_25170	NA	ACLI5Z_05655	ACLH7Q_12110
PSB__G1082	N838_29400	ACPL57_19665	ACLI5Z_09930	ACLH7Q_29510	PSB__G6315	N838_01185	ACPL57_15910	ACLI5Z_15320	ACLH7Q_33360	PSB__G4747	N838_15310	ACPL57_22510	ACLI5Z_24705	ACLH7Q_25710
PSB__G3932	N838_09615	ACPL57_00875	ACLI5Z_30470	ACLH7Q_07340	PSB__G3902	N838_13590	NA	ACLI5Z_26385	3_refound_684	PSB__G1487	N838_25190	NA	ACLI5Z_05675	ACLH7Q_12090
PSB__G6800	0_refound_347	ACPL57_04795	ACLI5Z_25625	ACLH7Q_24805	PSB__G1994	N838_13805	ACPL57_05415	ACLI5Z_26160	NA	PSB__G4705	N838_25245	NA	ACLI5Z_05730	ACLH7Q_12035
PSB__G1058	N838_31635	ACPL57_27235	ACLI5Z_12225	ACLH7Q_31830	PSB__G4535	N838_13810	ACPL57_05410	ACLI5Z_26155	NA	PSB__G6752	N838_15455	ACPL57_22655	ACLI5Z_24560	ACLH7Q_25855
PSB__G2148	N838_04805	ACPL57_10455	ACLI5Z_17185	ACLH7Q_16190	PSB__G2655	N838_13815	ACPL57_05405	ACLI5Z_26150	NA	PSB__G5108	N838_12065	ACPL57_07180	ACLI5Z_27940	ACLH7Q_18340
PSB__G3627	0_refound_5	ACPL57_01575	ACLI5Z_31170	ACLH7Q_08435	PSB__G2742	N838_13820	ACPL57_05400	ACLI5Z_26145	NA	PSB__G143	N838_25280	NA	ACLI5Z_05765	ACLH7Q_12000
PSB__G877	N838_10580	NA	ACLI5Z_29440	ACLH7Q_19850	PSB__G1554	N838_13825	ACPL57_05395	ACLI5Z_26140	NA	PSB__G2563	N838_25310	NA	ACLI5Z_05795	ACLH7Q_11970
PSB__G701	N838_06155	ACPL57_04465	ACLI5Z_33990	ACLH7Q_14790	PSB__G4643	NA	ACPL57_29120	NA	NA	PSB__G1709	N838_25315	NA	ACLI5Z_05800	ACLH7Q_11965
PSB__G952	N838_00005	ACPL57_29375	ACLI5Z_14130	ACLH7Q_01365	PSB__G3160	N838_13840	ACPL57_05380	ACLI5Z_26125	NA	PSB__G275	N838_25330	NA	ACLI5Z_05815	ACLH7Q_11945
PSB__G5346	0_refound_8	ACPL57_01705	ACLI5Z_31300	ACLH7Q_08305	PSB__G2977	N838_01255	ACPL57_15840	ACLI5Z_15390	ACLH7Q_33430	PSB__G569	N838_15480	ACPL57_22680	ACLI5Z_24535	ACLH7Q_25880
PSB__G1527	N838_00010	ACPL57_29370	ACLI5Z_14135	ACLH7Q_01360	PSB__G5192	N838_13880	ACPL57_05335	ACLI5Z_26080	NA	PSB__G2196	N838_12380	ACPL57_06860	ACLI5Z_27605	ACLH7Q_17995
PSB__G1832	N838_08995	ACPL57_01490	ACLI5Z_31085	ACLH7Q_08520	PSB__G2395	N838_01815	ACPL57_15275	ACLI5Z_20225	NA	PSB__G2029	N838_15535	ACPL57_22735	ACLI5Z_24480	ACLH7Q_25935
PSB__G984	N838_27345	ACPL57_17475	ACLI5Z_07825	ACLH7Q_09915	PSB__G3090	NA	ACPL57_21060	NA	NA	PSB__G3636	N838_15570	ACPL57_22770	ACLI5Z_24445	ACLH7Q_25970

PSB__G583	N838_18490	ACPL57_29915	ACLI5Z_21475	ACLH7Q_28915	PSB__G5454	NA	ACPL57_31495	NA	NA	PSB__G3906	N838_19995	ACPL57_24605	ACLI5Z_00370	ACLH7Q_21950
PSB__G3298	N838_18510	ACPL57_29935	ACLI5Z_21455	ACLH7Q_28935	PSB__G6470	N838_04480	ACPL57_10130	ACLI5Z_17510	ACLH7Q_01990	PSB__G7104	N838_21185	ACPL57_25010	ACLI5Z_01610	ACLH7Q_23190
PSB__G6730	N838_02765	ACPL57_14320	ACLI5Z_19245	ACLH7Q_03710	PSB__G4646	N838_04710	ACPL57_10360	2_refound_516	3_refound_662	PSB__G3660	N838_28440	ACPL57_16345	ACLI5Z_08955	ACLH7Q_20365
PSB__G1424	N838_02845	ACPL57_14240	ACLI5Z_19165	ACLH7Q_03630	PSB__G6599	N838_15950	ACPL57_23160	ACLI5Z_24050	ACLH7Q_26365	PSB__G210	N838_28445	ACPL57_16340	ACLI5Z_08960	ACLH7Q_20370
PSB__G759	N838_02855	NA	ACLI5Z_19155	ACLH7Q_03620	PSB__G4567	N838_16065	ACPL57_23275	ACLI5Z_23930	ACLH7Q_26480	PSB__G3245	N838_14885	ACPL57_30460	ACLI5Z_25150	ACLH7Q_25265
PSB__G2025	N838_02910	NA	ACLI5Z_19100	ACLH7Q_03565	PSB__G7051	N838_04715	ACPL57_10365	ACLI5Z_17275	ACLH7Q_01755	PSB__G4689	N838_28500	ACPL57_16285	ACLI5Z_09015	ACLH7Q_20425
PSB__G7025	N838_29900	ACPL57_19170	ACLI5Z_10400	ACLH7Q_30005	PSB__G3912	NA	ACPL57_29160	NA	NA	PSB__G1216	N838_22550	ACPL57_11785	ACLI5Z_02995	ACLH7Q_24590
PSB__G6698	N838_09285	ACPL57_01200	ACLI5Z_30795	ACLH7Q_08820	PSB__G5301	N838_04720	ACPL57_10370	ACLI5Z_17270	ACLH7Q_01750	PSB__G4000	N838_28725	NA	ACLI5Z_09245	ACLH7Q_20655
PSB__G3873	N838_03010	NA	2_refound_512	3_refound_657	PSB__G3279	NA	ACPL57_21200	NA	NA	PSB__G3451	N838_22670	ACPL57_11905	ACLI5Z_03115	ACLH7Q_06925
PSB__G4958	N838_03015	NA	ACLI5Z_18995	ACLH7Q_03465	PSB__G3957	N838_16735	ACPL57_20835	ACLI5Z_23250	ACLH7Q_27155	PSB__G4291	N838_23025	1_refound_421	2_refound_552	3_refound_698
PSB__G4340	N838_03020	NA	ACLI5Z_18990	ACLH7Q_03460	PSB__G5177	N838_16820	ACPL57_20750	ACLI5Z_23165	ACLH7Q_27240	PSB__G4221	N838_23075	ACPL57_12340	ACLI5Z_03540	ACLH7Q_06505
PSB__G1569	N838_07450	ACPL57_03090	ACLI5Z_32680	ACLH7Q_13470	PSB__G3061	N838_04800	ACPL57_10450	ACLI5Z_17190	ACLH7Q_16195	PSB__G6999	N838_23080	ACPL57_12345	ACLI5Z_03545	ACLH7Q_06500
PSB__G4897	N838_03150	NA	ACLI5Z_18855	ACLH7Q_03325	PSB__G4270	N838_05235	ACPL57_10900	ACLI5Z_16750	ACLH7Q_15750	PSB__G4665	N838_29000	NA	ACLI5Z_09540	ACLH7Q_29120
PSB__G3749	N838_17315	ACPL57_20255	ACLI5Z_22680	ACLH7Q_27730	PSB__G1908	N838_16920	ACPL57_20645	ACLI5Z_23065	ACLH7Q_27340	PSB__G981	N838_15215	ACPL57_30125	ACLI5Z_24815	ACLH7Q_25595
PSB__G72	N838_25270	NA	ACLI5Z_05755	ACLH7Q_12010	PSB__G22	NA	ACPL57_19865	NA	ACLH7Q_28135	PSB__G3866	N838_23530	ACPL57_12790	ACLI5Z_03985	ACLH7Q_05970
PSB__G5126	N838_17675	ACPL57_19895	ACLI5Z_22320	ACLH7Q_28095	PSB__G5182	NA	ACPL57_26940	NA	ACLH7Q_00770	PSB__G7066	N838_15230	ACPL57_30090	ACLI5Z_24795	ACLH7Q_25615
PSB__G6634	N838_20360	ACPL57_24215	ACLI5Z_00750	ACLH7Q_22330	PSB__G617	NA	NA	NA	ACLH7Q_32555	PSB__G6529	N838_23875	ACPL57_13150	ACLI5Z_04340	ACLH7Q_05610
PSB__G3505	N838_18500	ACPL57_29925	ACLI5Z_21465	ACLH7Q_28925	PSB__G6219	N838_17055	ACPL57_20510	ACLI5Z_22930	ACLH7Q_27475	PSB__G2641	N838_29420	ACPL57_19645	ACLI5Z_09950	ACLH7Q_29530
PSB__G4228	N838_03490	ACPL57_09135	ACLI5Z_18510	ACLH7Q_02990	PSB__G4766	N838_17105	1_refound_410	2_refound_543	3_refound_690	PSB__G4336	N838_24215	ACPL57_13500	ACLI5Z_04695	ACLH7Q_05260
PSB__G6954	N838_13575	ACPL57_05640	ACLI5Z_26400	ACLH7Q_16795	PSB__G4528	N838_05410	ACPL57_11070	ACLI5Z_16575	ACLH7Q_15570	PSB__G4847	N838_24680	ACPL57_13960	ACLI5Z_05165	ACLH7Q_04890
PSB__G2372	N838_32665	ACPL57_27020	ACLI5Z_13275	ACLH7Q_00690	PSB__G4833	N838_05425	ACPL57_11085	ACLI5Z_16560	ACLH7Q_15555	PSB__G6816	N838_29650	ACPL57_19415	ACLI5Z_10180	ACLH7Q_29760
PSB__G7228	N838_16870	ACPL57_20695	ACLI5Z_23115	ACLH7Q_27290	PSB__G6442	N838_05735	ACPL57_11410	ACLI5Z_16230	ACLH7Q_15215	PSB__G3533	N838_25180	NA	ACLI5Z_05665	ACLH7Q_12100
PSB__G4455	N838_03610	ACPL57_09250	ACLI5Z_18390	NA	PSB__G5229	N838_06045	ACPL57_11705	ACLI5Z_15930	ACLH7Q_14905	PSB__G2900	N838_15265	ACPL57_30050	ACLI5Z_24755	ACLH7Q_25655
PSB__G4474	N838_07920	ACPL57_02610	ACLI5Z_32205	ACLH7Q_13035	PSB__G4359	N838_17490	ACPL57_20085	ACLI5Z_22515	ACLH7Q_27900	PSB__G5944	N838_15300	ACPL57_22490	ACLI5Z_24715	ACLH7Q_25700
PSB__G4760	N838_04055	ACPL57_09700	ACLI5Z_17940	ACLH7Q_02410	PSB__G4339	NA	NA	NA	ACLH7Q_32680	PSB__G6956	N838_29930	ACPL57_19135	ACLI5Z_10430	ACLH7Q_30035
PSB__G3082	N838_30620	ACPL57_18405	ACLI5Z_11165	ACLH7Q_30760	PSB__G2172	N838_06200	ACPL57_04415	ACLI5Z_33940	ACLH7Q_14740	PSB__G4556	N838_15410	ACPL57_22610	ACLI5Z_24605	ACLH7Q_25810
PSB__G2733	N838_12530	ACPL57_06705	ACLI5Z_27450	ACLH7Q_17825	PSB__G3319	NA	ACPL57_21145	NA	NA	PSB__G4610	N838_29990	ACPL57_19070	ACLI5Z_10500	ACLH7Q_30105
PSB__G3192	N838_12005	ACPL57_07240	ACLI5Z_27995	ACLH7Q_18400	PSB__G3255	NA	ACPL57_26935	NA	ACLH7Q_00775	PSB__G4106	N838_30055	ACPL57_19005	ACLI5Z_10565	ACLH7Q_30170
PSB__G3899	N838_04560	ACPL57_10210	ACLI5Z_17430	ACLH7Q_01910	PSB__G1456	NA	NA	NA	ACLH7Q_32510	PSB__G6786	N838_15665	ACPL57_22860	ACLI5Z_24350	ACLH7Q_26065
PSB__G3011	N838_14420	ACPL57_04800	ACLI5Z_25630	ACLH7Q_24800	PSB__G1700	NA	ACPL57_26850	NA	ACLH7Q_00860	PSB__G2573	N838_30295	ACPL57_18755	ACLI5Z_10810	ACLH7Q_30405
PSB__G2622	N838_23575	ACPL57_12835	ACLI5Z_04030	ACLH7Q_05925	PSB__G6912	N838_06595	ACPL57_03990	ACLI5Z_33535	ACLH7Q_14345	PSB__G4156	N838_30535	ACPL57_18490	ACLI5Z_11080	ACLH7Q_30675
PSB__G3062	N838_25630	ACPL57_28645	ACLI5Z_06115	ACLH7Q_11635	PSB__G3642	N838_06600	ACPL57_03985	ACLI5Z_33530	ACLH7Q_14340	PSB__G4060	N838_30545	ACPL57_18480	ACLI5Z_11090	ACLH7Q_30685
PSB__G6231	N838_07410	ACPL57_03130	ACLI5Z_32720	ACLH7Q_13510	PSB__G4203	N838_17715	NA	ACLI5Z_22280	NA	PSB__G5194	N838_15700	ACPL57_22890	ACLI5Z_24315	ACLH7Q_26100
PSB__G7064	N838_02670	ACPL57_14420	ACLI5Z_19355	ACLH7Q_03815	PSB__G6514	N838_17745	NA	ACLI5Z_22250	NA	PSB__G1064	N838_30575	ACPL57_18450	ACLI5Z_11120	ACLH7Q_30715
PSB__G4337	N838_27020	ACPL57_17800	ACLI5Z_07500	ACLH7Q_10240	PSB__G17	N838_17880	1_refound_412	ACLI5Z_22115	ACLH7Q_28275	PSB__G4279	N838_25250	NA	ACLI5Z_05735	ACLH7Q_12030
PSB__G3530	N838_04895	ACPL57_10545	ACLI5Z_17095	ACLH7Q_16100	PSB__G4682	N838_17915	NA	ACLI5Z_22080	ACLH7Q_28310	PSB__G3785	N838_25360	NA	ACLI5Z_05845	ACLH7Q_11910
PSB__G3268	N838_05045	ACPL57_10705	ACLI5Z_16935	ACLH7Q_15940	PSB__G2903	N838_18005	NA	ACLI5Z_21985	ACLH7Q_28405	PSB__G4615	N838_15865	ACPL57_23075	ACLI5Z_24135	ACLH7Q_26280
PSB__G1453	N838_05075	ACPL57_10735	ACLI5Z_16905	ACLH7Q_15910	PSB__G4376	N838_06680	ACPL57_03895	ACLI5Z_33440	ACLH7Q_14250	PSB__G1902	N838_30800	ACPL57_18225	ACLI5Z_11345	ACLH7Q_30940
PSB__G358	N838_01225	ACPL57_15870	ACLI5Z_15360	ACLH7Q_33400	PSB__G3269	N838_06685	ACPL57_03890	ACLI5Z_33435	ACLH7Q_14245	PSB__G4383	N838_30830	ACPL57_18195	ACLI5Z_11375	ACLH7Q_30970
PSB__G1702	N838_01220	ACPL57_15875	ACLI5Z_15355	ACLH7Q_33395	PSB__G4426	N838_06690	ACPL57_03885	ACLI5Z_33430	ACLH7Q_14240	PSB__G6566	N838_25770	ACPL57_28505	ACLI5Z_06255	ACLH7Q_11495
PSB__G3370	N838_28170	ACPL57_16630	ACLI5Z_08670	ACLH7Q_20075	PSB__G1906	N838_06835	ACPL57_03730	ACLI5Z_33280	ACLH7Q_14090	PSB__G4519	N838_31070	NA	ACLI5Z_11655	ACLH7Q_31255
PSB__G1188	N838_28425	ACPL57_16360	ACLI5Z_08940	ACLH7Q_20345	PSB__G4490	N838_07035	ACPL57_03535	ACLI5Z_33090	ACLH7Q_13890	PSB__G3628	N838_16630	ACPL57_20940	ACLI5Z_23355	ACLH7Q_27050
PSB__G7120	N838_17670	ACPL57_19900	ACLI5Z_22325	ACLH7Q_28090	PSB__G4445	N838_18280	ACPL57_29695	ACLI5Z_21695	ACLH7Q_28695	PSB__G4323	N838_25780	ACPL57_28495	ACLI5Z_06265	ACLH7Q_11485
PSB__G4939	N838_05470	ACPL57_11130	ACLI5Z_16515	ACLH7Q_15510	PSB__G3137	N838_18320	ACPL57_29740	ACLI5Z_21650	ACLH7Q_28740	PSB__G5320	N838_26775	ACPL57_31210	ACLI5Z_07255	ACLH7Q_10490
PSB__G1245	N838_05490	ACPL57_11150	ACLI5Z_16495	ACLH7Q_15490	PSB__G3117	NA	ACPL57_21105	NA	NA	PSB__G4655	N838_17020	ACPL57_20545	ACLI5Z_22965	ACLH7Q_27440
PSB__G4816	N838_25915	NA	ACLI5Z_06395	ACLH7Q_11355	PSB__G2608	NA	ACPL57_21190	NA	NA	PSB__G2036	N838_26795	ACPL57_31230	ACLI5Z_07275	ACLH7Q_10470

PSB__G1159	N838_07440	ACPL57_03100	ACLI5Z_32690	ACLH7Q_13480	PSB__G1583	N838_01820	ACPL57_15270	ACLI5Z_20220	NA	PSB__G4293	N838_25585	ACPL57_28695	ACLI5Z_06065	ACLH7Q_11685
PSB__G912	N838_16310	ACPL57_23520	ACLI5Z_23685	ACLH7Q_26720	PSB__G1605	N838_13970	ACPL57_05250	ACLI5Z_25995	NA	PSB__G5090	N838_15575	1_refound_406	ACLI5Z_24440	ACLH7Q_25975
PSB__G6225	N838_00265	ACPL57_29035	ACLI5Z_14395	NA	PSB__G1589	N838_01825	ACPL57_15265	ACLI5Z_20215	NA	PSB__G4496	N838_15580	ACPL57_22775	ACLI5Z_24435	ACLH7Q_25980
PSB__G5092	N838_21785	ACPL57_26155	ACLI5Z_02220	ACLH7Q_23815	PSB__G7108	N838_14025	ACPL57_05195	ACLI5Z_25940	NA	PSB__G2012	N838_15765	ACPL57_22965	ACLI5Z_24240	ACLH7Q_26175
PSB__G1904	N838_01270	ACPL57_15825	ACLI5Z_15405	ACLH7Q_33445	PSB__G5046	N838_14030	ACPL57_05190	ACLI5Z_25935	NA	PSB__G3934	N838_12990	ACPL57_06235	ACLI5Z_26990	ACLH7Q_17365
PSB__G6451	N838_18150	ACPL57_29555	ACLI5Z_21840	ACLH7Q_28550	PSB__G6430	N838_14060	ACPL57_05160	NA	NA	PSB__G1788	N838_13145	ACPL57_06085	ACLI5Z_26840	ACLH7Q_17215
PSB__G1339	N838_14385	ACPL57_04835	ACLI5Z_25665	ACLH7Q_24765	PSB__G5423	N838_14070	1_refound_403	NA	NA	PSB__G6526	N838_13935	ACPL57_05285	ACLI5Z_26030	NA
PSB__G2135	N838_19575	ACPL57_22110	ACLI5Z_20380	ACLH7Q_21525	PSB__G4521	NA	ACPL57_21185	NA	NA	PSB__G3581	N838_26175	ACPL57_30755	ACLI5Z_06655	ACLH7Q_11080
PSB__G3485	N838_22780	ACPL57_12020	ACLI5Z_03235	ACLH7Q_06805	PSB__G2110	NA	NA	NA	ACLH7Q_32520	PSB__G4317	N838_14115	ACPL57_05105	NA	NA
PSB__G1552	N838_11520	ACPL57_07730	ACLI5Z_28495	ACLH7Q_18900	PSB__G5553	NA	NA	NA	ACLH7Q_09050	PSB__G6532	N838_26405	ACPL57_31735	ACLI5Z_06885	ACLH7Q_10850
PSB__G2545	N838_27620	ACPL57_17205	ACLI5Z_08095	ACLH7Q_09635	PSB__G2378	N838_14280	ACPL57_04940	ACLI5Z_25770	NA	PSB__G1360	N838_15910	ACPL57_23120	ACLI5Z_24090	ACLH7Q_26325
PSB__G5728	N838_00755	ACPL57_28100	ACLI5Z_14895	ACLH7Q_16360	PSB__G3086	N838_14285	ACPL57_04935	ACLI5Z_25765	NA	PSB__G5319	N838_26765	ACPL57_31200	ACLI5Z_07245	ACLH7Q_10500
PSB__G2064	N838_21670	ACPL57_26270	ACLI5Z_02105	ACLH7Q_23700	PSB__G4886	N838_14290	ACPL57_04930	ACLI5Z_25760	NA	PSB__G4662	N838_26770	ACPL57_31205	ACLI5Z_07250	ACLH7Q_10495
PSB__G6313	N838_24695	ACPL57_13975	ACLI5Z_05180	ACLH7Q_04875	PSB__G3334	NA	ACPL57_21055	NA	NA	PSB__G4601	N838_15920	ACPL57_23130	ACLI5Z_24080	ACLH7Q_26335
PSB__G4588	N838_00880	ACPL57_28235	ACLI5Z_15020	ACLH7Q_33065	PSB__G3038	N838_01935	ACPL57_15155	ACLI5Z_20105	NA	PSB__G2431	N838_16080	ACPL57_23290	ACLI5Z_23915	ACLH7Q_26495
PSB__G3058	N838_24010	ACPL57_13295	ACLI5Z_04480	ACLH7Q_05470	PSB__G3196	NA	ACPL57_21130	NA	NA	PSB__G1933	N838_16365	ACPL57_23575	ACLI5Z_23630	ACLH7Q_26775
PSB__G3658	N838_20150	ACPL57_24440	ACLI5Z_00530	ACLH7Q_22110	PSB__G4744	N838_14555	1_refound_404	2_refound_539	3_refound_685	PSB__G2228	N838_17225	ACPL57_20345	ACLI5Z_22770	ACLH7Q_27640
PSB__G5171	N838_01080	ACPL57_16010	ACLI5Z_15215	ACLH7Q_33260	PSB__G3903	N838_14800	NA	ACLI5Z_25235	ACLH7Q_25190	PSB__G7193	N838_17440	ACPL57_20150	ACLI5Z_22575	ACLH7Q_27840
PSB__G153	N838_01130	ACPL57_15965	ACLI5Z_15265	ACLH7Q_33305	PSB__G443	N838_14850	NA	ACLI5Z_25185	NA	PSB__G4493	N838_14130	ACPL57_05090	NA	NA
PSB__G1910	N838_20200	ACPL57_24390	ACLI5Z_00580	ACLH7Q_22160	PSB__G3110	NA	ACPL57_21210	NA	NA	PSB__G6387	N838_17530	ACPL57_20045	ACLI5Z_22465	ACLH7Q_27950
PSB__G392	N838_29500	ACPL57_19565	ACLI5Z_10030	ACLH7Q_29610	PSB__G1355	N838_02035	ACPL57_15060	ACLI5Z_20005	NA	PSB__G5195	N838_17700	NA	ACLI5Z_22295	NA
PSB__G6678	N838_32820	ACPL57_26810	ACLI5Z_13430	ACLH7Q_00900	PSB__G2910	NA	ACPL57_22495	NA	NA	PSB__G667	N838_17705	NA	ACLI5Z_22290	NA
PSB__G4899	N838_07985	ACPL57_02550	ACLI5Z_32145	ACLH7Q_12975	PSB__G2691	NA	ACPL57_21065	NA	NA	PSB__G6679	N838_18035	NA	ACLI5Z_21955	ACLH7Q_28435
PSB__G829	N838_13830	ACPL57_05390	ACLI5Z_26135	NA	PSB__G2585	NA	ACPL57_21225	NA	NA	PSB__G3044	N838_18040	NA	ACLI5Z_21950	ACLH7Q_28440
PSB__G1722	N838_07685	ACPL57_02845	ACLI5Z_32435	ACLH7Q_13225	PSB__G1986	NA	ACPL57_29255	NA	NA	PSB__G6359	N838_26955	ACPL57_17860	ACLI5Z_07440	ACLH7Q_10305
PSB__G2736	N838_29950	ACPL57_19115	ACLI5Z_10455	ACLH7Q_30060	PSB__G1960	NA	ACPL57_21070	NA	NA	PSB__G5178	N838_26960	1_refound_434	2_refound_563	ACLH7Q_10300
PSB__G4024	N838_01310	ACPL57_15785	ACLI5Z_15445	ACLH7Q_33485	PSB__G5554	NA	NA	NA	ACLH7Q_09055	PSB__G1152	N838_18110	ACPL57_29510	ACLI5Z_21880	ACLH7Q_28510
PSB__G2183	N838_12655	ACPL57_06585	ACLI5Z_27330	ACLH7Q_17705	PSB__G4983	N838_02415	ACPL57_14685	ACLI5Z_19620	ACLH7Q_04085	PSB__G5255	N838_18145	ACPL57_29550	ACLI5Z_21845	ACLH7Q_28545
PSB__G3425	N838_01585	ACPL57_15510	ACLI5Z_15720	ACLH7Q_33765	PSB__G1393	NA	NA	NA	ACLH7Q_32515	PSB__G3857	N838_27025	ACPL57_17795	ACLI5Z_07505	ACLH7Q_10235
PSB__G1305	N838_23970	ACPL57_13250	ACLI5Z_04435	ACLH7Q_05515	PSB__G6971	N838_15450	1_refound_500	ACLI5Z_24565	ACLH7Q_25850	PSB__G3941	N838_27030	ACPL57_17790	ACLI5Z_07510	ACLH7Q_10230
PSB__G2768	N838_01695	ACPL57_15400	ACLI5Z_15830	ACLH7Q_33875	PSB__G3068	N838_02660	ACPL57_14430	ACLI5Z_19365	ACLH7Q_03825	PSB__G6321	N838_18670	NA	ACLI5Z_21295	NA
PSB__G3472	N838_01790	ACPL57_15300	ACLI5Z_20250	ACLH7Q_33975	PSB__G5238	N838_02700	ACPL57_14390	ACLI5Z_19315	ACLH7Q_03775	PSB__G2254	N838_18705	NA	ACLI5Z_21260	NA
PSB__G2055	N838_26780	ACPL57_31215	ACLI5Z_07260	ACLH7Q_10485	PSB__G2301	N838_02945	NA	ACLI5Z_19070	ACLH7Q_03535	PSB__G1595	N838_27340	ACPL57_17480	ACLI5Z_07820	ACLH7Q_09920
PSB__G731	N838_13850	ACPL57_05370	ACLI5Z_26115	NA	PSB__G3410	NA	ACPL57_21150	NA	NA	PSB__G169	N838_18865	ACPL57_21385	ACLI5Z_21100	ACLH7Q_20790
PSB__G3433	N838_23430	ACPL57_12690	ACLI5Z_03885	ACLH7Q_06070	PSB__G1037	N838_15565	ACPL57_22765	ACLI5Z_24450	ACLH7Q_25965	PSB__G4657	N838_19505	ACPL57_22040	ACLI5Z_20450	ACLH7Q_21455
PSB__G428	N838_21550	ACPL57_26395	ACLI5Z_34205	ACLH7Q_23575	PSB__G4473	N838_03170	NA	ACLI5Z_18835	ACLH7Q_03305	PSB__G3417	N838_27400	ACPL57_17420	ACLI5Z_07880	ACLH7Q_09860
PSB__G3792	N838_30720	ACPL57_18305	ACLI5Z_11265	ACLH7Q_30860	PSB__G1093	N838_03410	ACPL57_09055	ACLI5Z_18590	ACLH7Q_03075	PSB__G6505	N838_27420	ACPL57_17400	ACLI5Z_07900	ACLH7Q_09840
PSB__G2210	N838_18485	ACPL57_29910	ACLI5Z_21480	ACLH7Q_28910	PSB__G2724	N838_03435	ACPL57_09080	ACLI5Z_18565	ACLH7Q_03050	PSB__G3387	N838_19580	ACPL57_22115	ACLI5Z_20375	ACLH7Q_21530
PSB__G1322	N838_22985	ACPL57_12245	ACLI5Z_03450	ACLH7Q_06590	PSB__G726	N838_15585	ACPL57_22780	ACLI5Z_24430	ACLH7Q_25985	PSB__G6749	N838_19600	ACPL57_22135	ACLI5Z_20355	ACLH7Q_21550
PSB__G6182	N838_02395	ACPL57_14705	ACLI5Z_19640	ACLH7Q_04105	PSB__G1953	N838_03715	ACPL57_09355	ACLI5Z_18285	ACLH7Q_02755	PSB__G1200	N838_14270	ACPL57_04950	ACLI5Z_25780	NA
PSB__G720	N838_06675	ACPL57_03900	ACLI5Z_33445	ACLH7Q_14255	PSB__G492	N838_15630	ACPL57_22825	ACLI5Z_24385	ACLH7Q_26030	PSB__G3746	N838_14370	ACPL57_04850	ACLI5Z_25680	ACLH7Q_24750
PSB__G4371	N838_02465	ACPL57_14635	ACLI5Z_19570	ACLH7Q_04035	PSB__G5521	NA	NA	NA	ACLH7Q_00195	PSB__G1889	N838_14865	ACPL57_30475	ACLI5Z_25170	ACLH7Q_25245
PSB__G4023	N838_32710	ACPL57_26975	ACLI5Z_13320	ACLH7Q_00735	PSB__G645	NA	ACPL57_21075	NA	NA	PSB__G7299	N838_14880	ACPL57_30465	ACLI5Z_25155	ACLH7Q_25260
PSB__G2058	N838_02500	ACPL57_14600	ACLI5Z_19535	ACLH7Q_04000	PSB__G3169	N838_04275	ACPL57_09930	ACLI5Z_17710	ACLH7Q_02190	PSB__G272	N838_19675	ACPL57_22200	ACLI5Z_00050	NA
PSB__G4537	N838_15615	ACPL57_22810	ACLI5Z_24400	ACLH7Q_26015	PSB__G37	NA	ACPL57_29140	NA	NA	PSB__G6850	N838_28010	ACPL57_16795	ACLI5Z_08505	ACLH7Q_09225
PSB__G1266	N838_27350	ACPL57_17470	ACLI5Z_07830	ACLH7Q_09910	PSB__G1890	N838_15785	ACPL57_22990	ACLI5Z_24220	ACLH7Q_26195	PSB__G5091	N838_19905	ACPL57_22435	ACLI5Z_00280	ACLH7Q_21860

PSB_G5427	N838_33215	NA	NA	NA	PSB_G1310	N838_07050	ACPL57_03520	ACLI5Z_33075	ACLH7Q_13875	PSB_G4617	N838_17025	ACPL57_20540	ACLI5Z_22960	ACLH7Q_27445
PSB_G2685	N838_31725	ACPL57_31930	ACLI5Z_12315	ACLH7Q_31920	PSB_G5252	N838_07270	ACPL57_03275	ACLI5Z_32855	ACLH7Q_13655	PSB_G4138	N838_31810	1_refound_447	ACLI5Z_12400	ACLH7Q_32005
PSB_G7029	N838_33510	NA	ACLI5Z_13965	ACLH7Q_01525	PSB_G5214	N838_07380	ACPL57_03160	ACLI5Z_32750	ACLH7Q_13540	PSB_G5314	N838_17030	ACPL57_20535	ACLI5Z_22955	ACLH7Q_27450
PSB_G4517	N838_25550	NA	NA	NA	PSB_G4805	N838_07680	ACPL57_02850	ACLI5Z_32440	ACLH7Q_13230	PSB_G6547	N838_31920	NA	ACLI5Z_12515	ACLH7Q_32120
PSB_G1279	N838_12275	NA	ACLI5Z_27725	ACLH7Q_18120	PSB_G86	N838_18630	NA	ACLI5Z_21335	NA	PSB_G3043	N838_17520	ACPL57_20055	ACLI5Z_22475	ACLH7Q_27940
PSB_G6252	N838_05255	ACPL57_10920	ACLI5Z_16730	ACLH7Q_15730	PSB_G2610	N838_07820	ACPL57_02705	ACLI5Z_32300	ACLH7Q_13090	PSB_G4946	N838_17680	ACPL57_19890	ACLI5Z_22315	ACLH7Q_28100
PSB_G2180	N838_06185	ACPL57_04430	ACLI5Z_33955	ACLH7Q_14755	PSB_G4304	N838_08120	ACPL57_02415	ACLI5Z_32000	ACLH7Q_12840	PSB_G751	N838_18335	ACPL57_29755	ACLI5Z_21635	ACLH7Q_28755
PSB_G905	N838_12665	ACPL57_06575	ACLI5Z_27320	ACLH7Q_17695	PSB_G3306	NA	ACPL57_21170	NA	NA	PSB_G3332	N838_18340	ACPL57_29760	ACLI5Z_21630	ACLH7Q_28760
PSB_G2898	N838_06420	ACPL57_04180	ACLI5Z_33720	ACLH7Q_14525	PSB_G6869	N838_08125	ACPL57_02410	ACLI5Z_31995	ACLH7Q_12835	PSB_G5233	N838_26805	ACPL57_31240	ACLI5Z_07285	ACLH7Q_10460
PSB_G961	N838_26840	ACPL57_31275	ACLI5Z_07320	ACLH7Q_10425	PSB_G6835	N838_19000	ACPL57_21535	ACLI5Z_20960	ACLH7Q_20930	PSB_G4663	N838_26815	ACPL57_31250	ACLI5Z_07295	ACLH7Q_10450
PSB_G2968	N838_12085	ACPL57_07155	ACLI5Z_27915	ACLH7Q_18315	PSB_G2229	N838_19015	ACPL57_21555	ACLI5Z_20940	ACLH7Q_20950	PSB_G4041	N838_18805	ACPL57_21320	ACLI5Z_21160	ACLH7Q_20730
PSB_G5973	N838_23695	ACPL57_12960	ACLI5Z_04150	ACLH7Q_05805	PSB_G4257	N838_08235	ACPL57_02295	ACLI5Z_31880	ACLH7Q_12720	PSB_G3519	N838_19275	ACPL57_21810	ACLI5Z_20680	ACLH7Q_21215
PSB_G4222	N838_27015	ACPL57_17805	ACLI5Z_07495	ACLH7Q_10245	PSB_G2549	NA	ACPL57_21205	NA	NA	PSB_G1277	N838_32265	NA	ACLI5Z_12860	ACLH7Q_00275
PSB_G1959	N838_24355	ACPL57_13650	ACLI5Z_04840	ACLH7Q_05115	PSB_G6138	NA	NA	NA	ACLH7Q_06305	PSB_G869	N838_32275	NA	ACLI5Z_12870	ACLH7Q_00285
PSB_G1618	N838_28290	ACPL57_16505	ACLI5Z_08795	ACLH7Q_20200	PSB_G5608	NA	NA	NA	ACLH7Q_32870	PSB_G3584	N838_32365	NA	ACLI5Z_12960	ACLH7Q_00375
PSB_G807	N838_29755	ACPL57_19310	ACLI5Z_10285	ACLH7Q_29865	PSB_G7221	N838_08395	ACPL57_02120	ACLI5Z_31710	ACLH7Q_12540	PSB_G2983	N838_32530	NA	ACLI5Z_13135	ACLH7Q_00550
PSB_G2780	N838_11200	ACPL57_08070	ACLI5Z_28825	ACLH7Q_19235	PSB_G3382	N838_00125	NA	ACLI5Z_14250	NA	PSB_G1449	N838_26850	ACPL57_31285	ACLI5Z_07330	ACLH7Q_10415
PSB_G6914	N838_25545	ACPL57_28725	ACLI5Z_06030	ACLH7Q_11725	PSB_G2190	N838_00280	ACPL57_29020	ACLI5Z_14410	NA	PSB_G3959	N838_19385	ACPL57_21925	ACLI5Z_20570	ACLH7Q_21325
PSB_G2313	0_refound_111	ACPL57_14425	ACLI5Z_19360	ACLH7Q_03820	PSB_G2653	N838_09025	ACPL57_01460	ACLI5Z_31055	ACLH7Q_08550	PSB_G7350	N838_32750	ACPL57_26885	ACLI5Z_13360	ACLH7Q_00825
PSB_G5398	0_refound_17	ACPL57_03040	ACLI5Z_32630	ACLH7Q_13420	PSB_G4836	N838_09030	ACPL57_01455	ACLI5Z_31050	ACLH7Q_08555	PSB_G3265	N838_26860	ACPL57_31295	ACLI5Z_07340	ACLH7Q_10405
PSB_G3372	0_refound_182	ACPL57_26770	2_refound_611	3_refound_759	PSB_G2484	N838_09040	ACPL57_01445	ACLI5Z_31040	ACLH7Q_08565	PSB_G2154	N838_19515	ACPL57_22050	ACLI5Z_20440	ACLH7Q_21465
PSB_G5149	0_refound_184	ACPL57_27595	ACLI5Z_11870	ACLH7Q_31475	PSB_G3701	N838_09095	ACPL57_01395	ACLI5Z_30990	ACLH7Q_08615	PSB_G3550	N838_20115	ACPL57_24475	ACLI5Z_00495	ACLH7Q_22075
PSB_G1991	0_refound_194	ACPL57_29670	ACLI5Z_21720	ACLH7Q_28670	PSB_G3496	N838_19845	ACPL57_22375	ACLI5Z_00220	ACLH7Q_21800	PSB_G4729	N838_27175	ACPL57_17650	ACLI5Z_07655	ACLH7Q_10090
PSB_G5000	0_refound_238	NA	ACLI5Z_13995	ACLH7Q_01500	PSB_G948	N838_19870	ACPL57_22400	ACLI5Z_00245	ACLH7Q_21825	PSB_G6280	N838_20270	ACPL57_24320	ACLI5Z_00650	ACLH7Q_22230
PSB_G1376	N838_07415	ACPL57_03125	ACLI5Z_32715	ACLH7Q_13505	PSB_G3520	N838_19885	ACPL57_22415	ACLI5Z_00260	ACLH7Q_21840	PSB_G6445	N838_20290	ACPL57_24300	ACLI5Z_00670	ACLH7Q_22250
PSB_G4896	0_refound_256	1_refound_466	ACLI5Z_26000	NA	PSB_G3567	N838_19890	ACPL57_22420	ACLI5Z_00265	ACLH7Q_21845	PSB_G3987	N838_21025	ACPL57_25190	ACLI5Z_01440	ACLH7Q_23025
PSB_G4622	N838_26835	ACPL57_31270	ACLI5Z_07315	ACLH7Q_10430	PSB_G4319	N838_19900	ACPL57_22430	ACLI5Z_00275	ACLH7Q_21855	PSB_G5305	N838_27245	ACPL57_17580	ACLI5Z_07725	ACLH7Q_10015
PSB_G235	N838_15780	ACPL57_22985	ACLI5Z_24225	ACLH7Q_26190	PSB_G1521	N838_09630	ACPL57_00860	ACLI5Z_30455	ACLH7Q_07355	PSB_G7236	N838_21525	ACPL57_24655	ACLI5Z_34180	ACLH7Q_23550
PSB_G6748	N838_26155	ACPL57_30735	ACLI5Z_06635	ACLH7Q_11100	PSB_G6316	N838_10135	ACPL57_00345	ACLI5Z_29925	ACLH7Q_07870	PSB_G6508	N838_27550	ACPL57_17280	ACLI5Z_08020	ACLH7Q_09710
PSB_G6657	N838_19570	ACPL57_22105	ACLI5Z_20385	ACLH7Q_21520	PSB_G4391	N838_10140	ACPL57_00340	ACLI5Z_29920	ACLH7Q_07875	PSB_G3115	N838_27785	ACPL57_17035	ACLI5Z_08260	ACLH7Q_09470
PSB_G3692	0_refound_258	NA	ACLI5Z_27690	ACLH7Q_18085	PSB_G3898	N838_00290	ACPL57_29010	ACLI5Z_14420	NA	PSB_G1883	N838_21605	ACPL57_26330	ACLI5Z_02045	ACLH7Q_23640
PSB_G1545	N838_11535	ACPL57_07715	ACLI5Z_28480	ACLH7Q_18885	PSB_G1761	N838_00515	ACPL57_27860	ACLI5Z_14645	ACLH7Q_16610	PSB_G6836	N838_22055	ACPL57_25885	ACLI5Z_02490	ACLH7Q_24085
PSB_G1437	N838_09565	ACPL57_00925	ACLI5Z_30520	ACLH7Q_07295	PSB_G5344	N838_10320	ACPL57_00165	ACLI5Z_29740	ACLH7Q_08060	PSB_G2531	N838_28560	ACPL57_16215	ACLI5Z_09075	ACLH7Q_20485
PSB_G6266	0_refound_319	ACPL57_01215	ACLI5Z_30810	ACLH7Q_08805	PSB_G5683	N838_00580	ACPL57_27930	ACLI5Z_14715	ACLH7Q_16540	PSB_G4171	N838_28735	NA	ACLI5Z_09255	ACLH7Q_20665
PSB_G6283	0_refound_320	ACPL57_10915	ACLI5Z_16735	ACLH7Q_15735	PSB_G6255	N838_00810	ACPL57_28160	ACLI5Z_14950	ACLH7Q_32995	PSB_G6281	N838_22105	ACPL57_25845	ACLI5Z_02540	ACLH7Q_24135
PSB_G6292	0_refound_321	ACPL57_09860	ACLI5Z_17780	3_refound_772	PSB_G5380	N838_00815	ACPL57_28165	ACLI5Z_14955	ACLH7Q_33000	PSB_G6706	N838_22390	ACPL57_31465	ACLI5Z_02835	ACLH7Q_24430
PSB_G6591	0_refound_338	ACPL57_20705	ACLI5Z_23120	ACLH7Q_27285	PSB_G2976	N838_01020	ACPL57_28375	ACLI5Z_15155	ACLH7Q_33200	PSB_G3072	N838_22405	ACPL57_31450	ACLI5Z_02850	ACLH7Q_24445
PSB_G5104	N838_08040	ACPL57_02495	ACLI5Z_32090	ACLH7Q_12920	PSB_G5113	N838_20390	ACPL57_24185	ACLI5Z_00780	3_refound_692	PSB_G1915	NA	NA	NA	ACLH7Q_16345
PSB_G6629	N838_15550	ACPL57_22750	ACLI5Z_24465	ACLH7Q_25950	PSB_G4441	N838_10325	ACPL57_00160	ACLI5Z_29735	ACLH7Q_08065	PSB_G2772	N838_22475	ACPL57_31380	ACLI5Z_02920	ACLH7Q_24515
PSB_G2364	N838_11920	ACPL57_07325	ACLI5Z_28085	ACLH7Q_18490	PSB_G6866	N838_20645	ACPL57_23910	ACLI5Z_01050	ACLH7Q_22630	PSB_G3938	N838_22575	ACPL57_11810	ACLI5Z_03020	ACLH7Q_24615
PSB_G6725	0_refound_344	ACPL57_15835	ACLI5Z_15395	ACLH7Q_33435	PSB_G4799	N838_10330	ACPL57_00155	ACLI5Z_29730	ACLH7Q_08070	PSB_G3704	N838_22860	ACPL57_12110	ACLI5Z_03320	ACLH7Q_06720
PSB_G3559	0_refound_4	ACPL57_01570	ACLI5Z_31165	ACLH7Q_08440	PSB_G6369	N838_10510	NA	ACLI5Z_29510	ACLH7Q_19920	PSB_G688	N838_28740	NA	ACLI5Z_09260	NA
PSB_G535	N838_08140	ACPL57_02395	ACLI5Z_31980	ACLH7Q_12820	PSB_G3232	N838_10630	ACPL57_08630	ACLI5Z_29390	ACLH7Q_19800	PSB_G4033	N838_23005	ACPL57_12265	ACLI5Z_03470	ACLH7Q_06570
PSB_G3516	N838_08145	ACPL57_02390	ACLI5Z_31975	ACLH7Q_12815	PSB_G7248	N838_21015	NA	ACLI5Z_01430	ACLH7Q_23015	PSB_G4048	N838_23010	ACPL57_12270	ACLI5Z_03475	ACLH7Q_06565
PSB_G6891	N838_08150	ACPL57_02385	ACLI5Z_31970	ACLH7Q_12810	PSB_G6966	N838_21020	ACPL57_25195	ACLI5Z_01435	ACLH7Q_23020	PSB_G84	NA	ACPL57_29145	NA	NA
PSB_G2926	N838_08220	ACPL57_02310	ACLI5Z_31895	ACLH7Q_12735	PSB_G2168	N838_01070	ACPL57_16020	ACLI5Z_15205	ACLH7Q_33250	PSB_G3718	N838_23015	ACPL57_12275	ACLI5Z_03480	ACLH7Q_06560

PSB__G4062	0_refound_48	ACPL57_06740	ACLI5Z_27485	ACLH7Q_17865	PSB__G6779	N838_21180	ACPL57_25015	ACLI5Z_01605	ACLH7Q_23185	PSB__G6571	N838_28750	NA	ACLI5Z_09270	NA
PSB__G4358	N838_13405	ACPL57_05815	ACLI5Z_26575	ACLH7Q_16950	PSB__G16	N838_10935	ACPL57_08330	ACLI5Z_29075	ACLH7Q_19500	PSB__G1033	N838_28830	NA	ACLI5Z_09355	NA
PSB__G5260	0_refound_60	ACPL57_07460	ACLI5Z_28225	ACLH7Q_18630	PSB__G2516	N838_01265	ACPL57_15830	ACLI5Z_15400	ACLH7Q_33440	PSB__G6927	N838_23020	ACPL57_12285	ACLI5Z_03485	ACLH7Q_06555
PSB__G4648	N838_08410	ACPL57_02105	ACLI5Z_31695	ACLH7Q_12525	PSB__G2845	N838_01400	ACPL57_15695	ACLI5Z_15535	ACLH7Q_33575	PSB__G5270	N838_29180	NA	ACLI5Z_09715	ACLH7Q_29295
PSB__G1703	N838_01610	ACPL57_15485	ACLI5Z_15745	ACLH7Q_33790	PSB__G5071	N838_21595	ACPL57_26340	ACLI5Z_02035	ACLH7Q_23630	PSB__G5049	N838_23125	ACPL57_12390	ACLI5Z_03590	ACLH7Q_06455
PSB__G5354	0_refound_100	ACPL57_13280	ACLI5Z_04465	ACLH7Q_05485	PSB__G1089	N838_01910	ACPL57_15180	ACLI5Z_20130	NA	PSB__G2499	N838_23195	ACPL57_12460	ACLI5Z_03660	ACLH7Q_06385
PSB__G4559	0_refound_68	ACPL57_08310	ACLI5Z_29055	ACLH7Q_19480	PSB__G2316	N838_02495	ACPL57_14605	ACLI5Z_19540	ACLH7Q_04005	PSB__G401	N838_23200	ACPL57_12465	ACLI5Z_03665	ACLH7Q_06380
PSB__G7343	N838_01970	ACPL57_15125	ACLI5Z_20070	NA	PSB__G5039	N838_03140	NA	ACLI5Z_18865	ACLH7Q_03335	PSB__G2405	N838_23205	ACPL57_12470	ACLI5Z_03670	ACLH7Q_06375
PSB__G5285	N838_02665	1_refound_372	2_refound_510	3_refound_655	PSB__G3080	N838_21870	ACPL57_26070	ACLI5Z_02305	ACLH7Q_23900	PSB__G2432	N838_23210	1_refound_422	2_refound_553	3_refound_699
PSB__G3454	0_refound_127	ACPL57_16895	ACLI5Z_08400	ACLH7Q_09330	PSB__G2885	N838_21875	ACPL57_26065	ACLI5Z_02310	ACLH7Q_23905	PSB__G195	N838_23745	ACPL57_13025	ACLI5Z_04215	ACLH7Q_05740
PSB__G7300	N838_03290	ACPL57_08940	ACLI5Z_18705	ACLH7Q_03180	PSB__G389	N838_21890	ACPL57_26050	ACLI5Z_02325	ACLH7Q_23920	PSB__G6179	N838_29435	ACPL57_19630	ACLI5Z_09965	ACLH7Q_29545
PSB__G5402	N838_06060	ACPL57_11715	ACLI5Z_15920	ACLH7Q_14895	PSB__G3578	N838_21895	ACPL57_26045	ACLI5Z_02330	ACLH7Q_23925	PSB__G6450	N838_23915	ACPL57_13190	ACLI5Z_04380	ACLH7Q_05570
PSB__G4531	N838_09290	ACPL57_01195	ACLI5Z_30790	ACLH7Q_08825	PSB__G4020	N838_21900	ACPL57_26040	ACLI5Z_02335	ACLH7Q_23930	PSB__G2909	NA	ACPL57_21120	NA	NA
PSB__G2709	N838_09295	ACPL57_01190	ACLI5Z_30785	ACLH7Q_08830	PSB__G4007	N838_21940	ACPL57_26000	ACLI5Z_02375	ACLH7Q_23970	PSB__G5197	N838_24225	ACPL57_13510	ACLI5Z_04705	ACLH7Q_05250
PSB__G838	N838_09315	ACPL57_01170	ACLI5Z_30765	ACLH7Q_08850	PSB__G4874	N838_03190	NA	ACLI5Z_18810	ACLH7Q_03280	PSB__G6302	N838_24790	ACPL57_14070	ACLI5Z_05275	ACLH7Q_04780
PSB__G2169	0_refound_128	ACPL57_16900	ACLI5Z_08395	ACLH7Q_09335	PSB__G4876	N838_03230	ACPL57_08875	ACLI5Z_18770	ACLH7Q_03240	PSB__G2479	NA	ACPL57_21125	NA	NA
PSB__G6536	N838_07130	ACPL57_03435	ACLI5Z_32995	ACLH7Q_13795	PSB__G5231	N838_10960	ACPL57_08305	ACLI5Z_29050	ACLH7Q_19475	PSB__G2332	NA	ACPL57_21115	NA	NA
PSB__G7097	N838_11885	ACPL57_07360	ACLI5Z_28125	ACLH7Q_18530	PSB__G1414	N838_22275	ACPL57_25670	ACLI5Z_02715	ACLH7Q_24310	PSB__G3812	N838_25460	ACPL57_28810	ACLI5Z_05945	ACLH7Q_11810
PSB__G5366	0_refound_187	ACPL57_28170	2_refound_612	3_refound_760	PSB__G6523	N838_03240	ACPL57_08890	ACLI5Z_18755	ACLH7Q_03225	PSB__G7056	N838_25895	NA	ACLI5Z_06380	ACLH7Q_11375
PSB__G4451	0_refound_190	ACPL57_29095	ACLI5Z_14335	NA	PSB__G3221	N838_04435	ACPL57_10090	ACLI5Z_17550	ACLH7Q_02030	PSB__G6595	N838_26335	ACPL57_30915	ACLI5Z_06815	ACLH7Q_10920
PSB__G4158	0_refound_266	NA	2_refound_620	ACLH7Q_01475	PSB__G2382	N838_04650	ACPL57_10300	ACLI5Z_17340	ACLH7Q_01820	PSB__G4010	N838_27635	ACPL57_17190	ACLI5Z_08110	ACLH7Q_09620
PSB__G6945	N838_12775	ACPL57_06460	ACLI5Z_27215	ACLH7Q_17590	PSB__G6269	N838_11580	ACPL57_07670	ACLI5Z_28435	ACLH7Q_18840	PSB__G986	N838_27655	ACPL57_17170	ACLI5Z_08130	ACLH7Q_09600
PSB__G6384	N838_16875	ACPL57_20690	ACLI5Z_23110	ACLH7Q_27295	PSB__G5833	N838_05445	ACPL57_11105	ACLI5Z_16540	ACLH7Q_15535	PSB__G6421	N838_27660	ACPL57_17165	ACLI5Z_08135	ACLH7Q_09595
PSB__G4633	N838_10265	NA	2_refound_523	ACLH7Q_08005	PSB__G3834	N838_22610	ACPL57_11845	ACLI5Z_03055	ACLH7Q_24650	PSB__G341	N838_27705	ACPL57_17115	ACLI5Z_08180	ACLH7Q_09550
PSB__G6587	N838_19170	ACPL57_21705	ACLI5Z_20785	ACLH7Q_21110	PSB__G3517	N838_11665	ACPL57_07585	ACLI5Z_28350	ACLH7Q_18755	PSB__G3848	N838_30995	ACPL57_18000	ACLI5Z_11570	ACLH7Q_31165
PSB__G6683	N838_20455	ACPL57_24115	ACLI5Z_00845	ACLH7Q_22420	PSB__G5093	N838_22640	ACPL57_11875	ACLI5Z_03085	ACLH7Q_06955	PSB__G1790	N838_29015	NA	ACLI5Z_09555	ACLH7Q_29135
PSB__G6952	N838_24230	ACPL57_13515	ACLI5Z_04710	ACLH7Q_05245	PSB__G5067	N838_11670	ACPL57_07580	ACLI5Z_28345	ACLH7Q_18750	PSB__G4825	NA	NA	ACLI5Z_25905	NA
PSB__G6524	0_refound_336	ACPL57_08880	ACLI5Z_18765	ACLH7Q_03235	PSB__G6881	N838_05730	ACPL57_11405	ACLI5Z_16235	ACLH7Q_15220	PSB__G4094	N838_29195	NA	2_refound_566	3_refound_711
PSB__G818	N838_26725	ACPL57_31160	ACLI5Z_07205	ACLH7Q_10540	PSB__G3875	N838_05745	ACPL57_11420	ACLI5Z_16220	ACLH7Q_15205	PSB__G4449	N838_29895	ACPL57_19175	ACLI5Z_10395	ACLH7Q_30000
PSB__G5163	N838_26845	ACPL57_31280	ACLI5Z_07325	ACLH7Q_10420	PSB__G3807	N838_11760	ACPL57_07485	ACLI5Z_28250	ACLH7Q_18655	PSB__G5078	N838_31355	ACPL57_27515	ACLI5Z_11950	ACLH7Q_31550
PSB__G5224	N838_26905	NA	ACLI5Z_07385	ACLH7Q_10360	PSB__G6405	N838_22995	ACPL57_12255	ACLI5Z_03460	ACLH7Q_06580	PSB__G1927	N838_30100	ACPL57_18955	ACLI5Z_10610	ACLH7Q_30215
PSB__G5107	N838_10845	ACPL57_08420	ACLI5Z_29180	ACLH7Q_19590	PSB__G4423	N838_06010	ACPL57_11670	ACLI5Z_15965	ACLH7Q_14945	PSB__G4482	N838_31365	ACPL57_27505	ACLI5Z_11960	ACLH7Q_31560
PSB__G3055	N838_10895	ACPL57_08370	ACLI5Z_29130	ACLH7Q_19540	PSB__G3817	N838_06605	ACPL57_03980	ACLI5Z_33525	ACLH7Q_14335	PSB__G1114	N838_30555	ACPL57_18470	ACLI5Z_11100	ACLH7Q_30695
PSB__G6810	N838_10900	ACPL57_08365	ACLI5Z_29125	ACLH7Q_19535	PSB__G3842	N838_07460	ACPL57_03080	ACLI5Z_32670	ACLH7Q_13460	PSB__G3672	N838_30740	ACPL57_18285	ACLI5Z_11285	ACLH7Q_30880
PSB__G7053	N838_31990	ACPL57_08680	ACLI5Z_12585	ACLH7Q_32190	PSB__G280	N838_07465	ACPL57_03075	ACLI5Z_32665	ACLH7Q_13455	PSB__G1334	N838_31195	ACPL57_27685	ACLI5Z_11780	ACLH7Q_31380
PSB__G5518	NA	NA	NA	ACLH7Q_00180	PSB__G3923	N838_11765	ACPL57_07480	ACLI5Z_28245	ACLH7Q_18650	PSB__G3557	N838_31440	ACPL57_27435	ACLI5Z_12030	ACLH7Q_31630
PSB__G5482	NA	NA	ACLI5Z_34080	NA	PSB__G4763	N838_11970	ACPL57_07275	ACLI5Z_28030	ACLH7Q_18435	PSB__G5245	N838_31985	ACPL57_08685	ACLI5Z_12580	ACLH7Q_32185
PSB__G4935	0_refound_69	ACPL57_08770	ACLI5Z_12475	ACLH7Q_32080	PSB__G4943	N838_12010	ACPL57_07235	2_refound_527	ACLH7Q_18395	PSB__G1566	N838_31915	NA	ACLI5Z_12510	ACLH7Q_32115
PSB__G5407	N838_02915	NA	2_refound_511	3_refound_656	PSB__G2813	N838_07470	ACPL57_03070	ACLI5Z_32660	ACLH7Q_13450	PSB__G6953	N838_31945	ACPL57_08725	ACLI5Z_12540	ACLH7Q_32145
PSB__G5359	N838_24760	ACPL57_14040	ACLI5Z_05245	ACLH7Q_04805	PSB__G287	N838_07690	ACPL57_02840	ACLI5Z_32430	ACLH7Q_13220	PSB__G3849	N838_31950	ACPL57_08720	ACLI5Z_12545	ACLH7Q_32150
PSB__G5476	NA	NA	ACLI5Z_01935	NA	PSB__G6265	N838_08090	ACPL57_02445	ACLI5Z_32030	ACLH7Q_12870	PSB__G2557	N838_31980	ACPL57_08690	ACLI5Z_12575	ACLH7Q_32180
PSB__G6701	N838_31940	ACPL57_08730	ACLI5Z_12535	ACLH7Q_32140	PSB__G388	N838_08095	ACPL57_02440	ACLI5Z_32025	ACLH7Q_12865	PSB__G6833	N838_32250	NA	ACLI5Z_12845	ACLH7Q_00260
PSB__G5297	NA	ACPL57_07305	ACLI5Z_28065	NA	PSB__G1638	N838_08335	ACPL57_02190	ACLI5Z_31780	ACLH7Q_12615	PSB__G3195	N838_32255	NA	ACLI5Z_12850	ACLH7Q_00265
PSB__G5035	NA	ACPL57_02125	ACLI5Z_31715	NA	PSB__G6645	N838_23290	ACPL57_12550	ACLI5Z_03750	NA	PSB__G1741	N838_33145	ACPL57_26485	ACLI5Z_13755	ACLH7Q_01220
PSB__G5543	NA	NA	NA	ACLH7Q_07090	PSB__G7326	N838_12300	ACPL57_06935	ACLI5Z_27695	ACLH7Q_18090	PSB__G2030	NA	NA	ACLI5Z_30035	NA
PSB__G3651	NA	ACPL57_18050	ACLI5Z_11520	ACLH7Q_31115	PSB__G5191	N838_12330	ACPL57_06915	ACLI5Z_27660	ACLH7Q_18055	PSB__G1985	N838_33210	NA	NA	NA

					PSB__G356	N838_33585	NA	ACLI5Z_14050	ACLH7Q_01445	PSB__G2664	N838_33330	NA	NA	ACLH7Q_01705

Supplemental Table 5. PB-PSB1 gene ID's related to individual composite PB-PSB1 bins from metagenome database as well as our 2011 reference genome: 2011 PB-PSB1 reference genome (CP050890.1), sample001 bin, sample002 bin, sample024 bin

Gene	product	cluster	NCBI_PGAP	COG20_FUNCTION	KOfam
SRB__G4384	hypothetical protein	1	NA	NA	NA
SRB__G4353	CBS domain-containing protein	1	CBS domain-containing protein	CBS domain (CBS) (PDB:1B3O)	NA
SRB__G4340	sulfatase-like hydrolase/transferase	1	sulfatase-like hydrolase/transferase	Arylsulfatase A or related enzyme, AlkP superfamily (AslA) (PDB:1E1Z)	NA
SRB__G4333	cytochrome d ubiquinol oxidase subunit II	1	cytochrome d ubiquinol oxidase subunit II (cydB)	Cytochrome bd-type quinol oxidase, subunit 2 (AppB) (PDB:6RKO)	cytochrome bd ubiquinol oxidase subunit II [EC:7.1.1.7]
SRB__G4294	IS1634 family transposase	1	IS1634 family transposase	Transposase	NA
SRB__G4257	transposase zinc-binding domain-containing protein	1	transposase zinc-binding domain-containing protein	NA	NA
SRB__G4215	hypothetical protein	1	NA	NA	NA
SRB__G4210	protein-L-isoaspartate(D-aspartate) O-methyltransferase	1	protein-L-isoaspartate(D-aspartate) O-methyltransferase	Protein-L-isoaspartate O-methyltransferase (Pcm) (PDB:3LBF) (PUBMED:24909784)	protein-L-isoaspartate(D-aspartate) O-methyltransferase [EC:2.1.1.77]
SRB__G4184	VacJ family lipoprotein	1	VacJ family lipoprotein	Lipoprotein subunit MlaA of the ABC-type intermembrane phospholipid transporter Mla (MlaA) (PDB:5NUQ)	phospholipid-binding lipoprotein MlaA
SRB__G4172	PAS domain-containing protein	1	PAS domain-containing protein	NA	NA
SRB__G4147	sulfite exporter TauE/SafE family protein	1	sulfite exporter TauE/SafE family protein	Sulfite exporter TauE/SafE/YfcA and related permeases, UPF0721 family (TauE) (PUBMED:21183667;22797525)	uncharacterized protein
SRB__G4115	MFS transporter	1	MFS transporter	Predicted arabinose efflux permease AraJ, MFS family (AraJ) (PDB:4LDS)	NA
SRB__G4101	cardiolipin synthase	1	cardiolipin synthase (cls)	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase (Cls) (PDB:1BYR)	cardiolipin synthase A/B [EC:2.7.8.-]
SRB__G4090	homocysteine biosynthesis protein	1	homocysteine biosynthesis protein	Sulfur incorporation enzyme MA1821 (anaerobic homocysteine biosynthesis), HcyBio/DUF39 family (HcyBio) (PUBMED:25315403)	L-aspartate semialdehyde sulfurtransferase [EC:2.8.1.16]
SRB__G4085	sigma-54-dependent transcriptional regulator	1	sigma-54-dependent Fis family transcriptional regulator	DNA-binding transcriptional response regulator, NtrC family, contains REC, AAA-type ATPase, and a Fis-type DNA-binding domains (AtoC) (PDB:1ZY2)	NA
SRB__G4082	YqaA family protein	1	DedA family protein	Uncharacterized membrane protein YqaA, VTT domain (YgaA)	NA
SRB__G4054	AI-2E family transporter	1	AI-2E family transporter	Predicted PurR-regulated permease PerM (PerM)	NA
SRB__G4052	glycosyltransferase family 4 protein	1	glycosyltransferase family 4 protein	Glycosyltransferase involved in cell wall bisynthesis (RfaB) (PDB:2IV7)	colanic acid/amylovoran/stewartan biosynthesis glycosyltransferase WcaL/AmsK/CpsK [EC:2.4.-.-]
SRB__G4047	hypothetical protein	1	NA	NA	NA
SRB__G4041	AmpG family muropeptide MFS transporter	1	MFS transporter	ABC-type Fe3+-siderophore transport system, permease component (FepD) (PDB:2NQ2)	MFS transporter, PAT family, beta-lactamase induction signal transducer AmpG
SRB__G4015	tripartite tricarboxylate transporter TctB family protein	1	tripartite tricarboxylate transporter TctB family protein	NA	putative tricarboxylic transport membrane protein
SRB__G4000	RelA/SpoT family protein	1	bifunctional (p)ppGpp synthetase/guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase	(p)ppGpp synthase/hydrolase, HD superfamily (SpoT) (PDB:1TDJ)	GTP diphosphokinase / guanosine-3',5'-bis(diphosphate) 3'-diphosphatase [EC:2.7.6.5 3.1.7.2]
SRB__G3976	AI-2E family transporter	1	AI-2E family transporter	Predicted PurR-regulated permease PerM (PerM)	NA
SRB__G3958	site-specific tyrosine recombinase XerD	1	site-specific tyrosine recombinase XerD (xerD)	Site-specific recombinase XerD (XerD) (PDB:1A0P)	integrase/recombinase XerD
SRB__G3948	hypothetical protein	1	NA	NA	NA
SRB__G3943	U32 family peptidase C-terminal domain-containing protein	1	U32 family peptidase	23S rRNA C2501 and tRNA U34 5'-hydroxylation protein RlhA/YrrN/YrrO, U32 peptidase family (RlhA) (PDB:5D88) (PUBMED:31253794;31358606)	U32 family peptidase [EC:3.4.-.-]
SRB__G3812	IS1380 family transposase	1	IS1380 family transposase	NA	NA
SRB__G3645	hypothetical protein	1	NA	NA	NA
SRB__G3643	hypothetical protein	1	NA	NA	NA
SRB__G3628	hypothetical protein	1	NA	NA	NA
SRB__G3620	hypothetical protein	1	NA	Transposase InsE and inactivated derivatives (InsE) (PDB:2JN6) (PUBMED:23818957)	NA
SRB__G3543	hypothetical protein	1	NA	NA	NA
SRB__G3527	hypothetical protein	1	NA	NA	NA
SRB__G3461	cold-shock protein	1	cold-shock protein	Cold shock protein, CspA family (CspC) (PDB:1C9O)	cold shock protein
SRB__G3437	AbrB/MazE/SpoVT family DNA-binding domain-containing protein	1	AbrB/MazE/SpoVT family DNA-binding domain-containing protein	Bifunctional DNA-binding transcriptional regulator of stationary/sporulation/toxin gene expression and antitoxin component of the YhaV-PrlF toxin-antitoxin module (AbrB) (PDB:2RO5)	antitoxin PrlF
SRB__G3419	hypothetical protein	1	NA	NA	NA
SRB__G3407	hypothetical protein	1	NA	NA	NA
SRB__G3400	type II toxin-antitoxin system Phd/YefM family antitoxin	1	type II toxin-antitoxin system Phd/YefM family antitoxin	NA	NA
SRB__G3379	RNA-binding S4 domain-containing protein	1	RNA-binding S4 domain-containing protein	Ribosome-associated protein YbcJ, S4-like RNA-binding protein (YbcJ) (PDB:1P9K)	ribosome-associated protein
SRB__G3350	hypothetical protein	1	NA	NA	NA
SRB__G3338	acyl carrier protein	1	acyl carrier protein (acpP)	Acyl carrier protein (AcpP) (PDB:1ACP)	acyl carrier protein

SRB__G3326	dissimilatory sulfite reductase D family protein	1	dissimilatory sulfite reductase D family protein	NA	NA
SRB__G3281	type II toxin-antitoxin system Phd/YefM family antitoxin	1	type II toxin-antitoxin system Phd/YefM family antitoxin	Antitoxin component YafN of the YafNO toxin-antitoxin module, PHD/YefM family (StbD) (PDB:2A6Q)	antitoxin StbD
SRB__G3261	hypothetical protein	1	NA	NA	NA
SRB__G3210	carboxymuconolactone decarboxylase family protein	1	carboxymuconolactone decarboxylase family protein	NA	NA
SRB__G3160	type II toxin-antitoxin system RelE family toxin	1	type II toxin-antitoxin system RelE/ParE family toxin	mRNA-degrading endonuclease RelE, toxin component of the RelBE toxin-antitoxin system (RelE) (PDB:1WMI)	mRNA interferase RelE/StbE
SRB__G3106	hypothetical protein	1	NA	NA	NA
SRB__G3070	hypothetical protein	1	NA	NA	NA
SRB__G3068	hypothetical protein	1	NA	NA	NA
SRB__G3053	hypothetical protein	1	NA	Putative sterol carrier protein, contains SCP2 domain (SCP2)	NA
SRB__G3044	chaperone modulator CbpM	1	chaperone modulator CbpM	DNA-binding transcriptional regulator, MerR family (SoxR) (PDB:2VZ4)	chaperone modulatory protein CbpM
SRB__G3023	FtsB family cell division protein	1	septum formation initiator family protein	Cell division protein FtsB (FtsB) (PDB:4IFF)	NA
SRB__G3012	thioredoxin	1	thioredoxin (trxA)	Chaperedoxin CnoX, contains thioredoxin-like and TPR-like domains, YbbN/TrxSC family (CnoX) (PDB:3QDN) (PUBMED:21498507;29754824)	thioredoxin 1
SRB__G3004	hypothetical protein	1	NA	NA	NA
SRB__G2978	tRNA-binding protein	1	tRNA-binding protein	tRNA-binding EMAP/Myf domain (EMAP) (PDB:1MKH)	tRNA-binding protein
SRB__G2976	putative signal transducing protein	1	DUF2007 domain-containing protein	NA	NA
SRB__G2922	hypothetical protein	1	NA	NA	NA
SRB__G2820	cupin domain-containing protein	1	cupin domain-containing protein	Mannose-6-phosphate isomerase, cupin superfamily (ManC) (PDB:1IPJ)	NA
SRB__G2819	cupin domain-containing protein	1	cupin domain-containing protein	Cupin domain protein related to quercetin dioxygenase (QdoI) (PDB:2H0V)	NA
SRB__G2793	zf-TFIIB domain-containing protein	1	zf-TFIIB domain-containing protein	Predicted nucleic acid-binding protein, contains Zn-finger domain (PUBMED:22120742)	uncharacterized protein
SRB__G2786	PqqD family protein	1	PqqD family protein	NA	pyrroloquinoline quinone biosynthesis protein D
SRB__G2783	secondary thiamine-phosphate synthase enzyme YjbQ	1	YjbQ family protein	Thiamin phosphate synthase YjbQ, UPF0047 family (YjbQ) (PDB:1VE0)	NA
SRB__G2782	hypothetical protein	1	NA	Predicted regulator of Ras-like GTPase activity, Roadblock/LC7/MglB family (PDB:1J3W)	NA
SRB__G2777	type III-B CRISPR module-associated protein Cmr5	1	type III-B CRISPR module-associated protein Cmr5 (cmr5)	CRISPR-Cas system type III CMR-effector complex small subunit Cmr5 (Cmr5) (PDB:2OEB)	CRISPR-associated protein Cmr5
SRB__G2768	hypothetical protein	1	NA	NA	NA
SRB__G2765	VOC family protein	1	VOC family protein	Catechol 2,3-dioxygenase or related enzyme, vicinal oxygen chelate (VOC) family (GloA) (PDB:3R4Q) (PUBMED:11830595)	NA
SRB__G2745	methylmalonyl-CoA epimerase	1	methylmalonyl-CoA epimerase (mce)	Catechol 2,3-dioxygenase or related enzyme, vicinal oxygen chelate (VOC) family (GloA) (PDB:3R4Q) (PUBMED:11830595)	methylmalonyl-CoA/ethylmalonyl-CoA epimerase [EC:5.1.99.1]
SRB__G2741	hypothetical protein	1	NA	NA	NA
SRB__G2733	hypothetical protein	1	NA	NA	NA
SRB__G2732	hypothetical protein	1	NA	NA	NA
SRB__G2677	flavodoxin domain-containing protein	1	FprA family A-type flavoprotein	Flavorubredoxin (NorV) (PDB:4JHC)	NAD(P)H dehydrogenase (quinone) [EC:1.6.5.2]
SRB__G2654	pyridoxamine 5'-phosphate oxidase family protein	1	pyridoxamine 5'-phosphate oxidase family protein	General stress protein 26 (function unknown) (YzzA) (PDB:2I02)	NA
SRB__G2636	Smr/MutS family protein	1	Smr/MutS family protein	NA	NA
SRB__G2626	hypothetical protein	1	NA	Uncharacterized conserved protein	NA
SRB__G2619	NUDIX domain-containing protein	1	NUDIX domain-containing protein	ADP-ribose pyrophosphatase YjhB, NUDIX family (YjhB) (PDB:3GZ5)	8-oxo-dGTP diphosphatase [EC:3.6.1.55]
SRB__G2551	NUDIX hydrolase	1	8-oxo-dGTP diphosphatase	ADP-ribose pyrophosphatase YjhB, NUDIX family (YjhB) (PDB:3GZ5)	NA
SRB__G2548	flavodoxin family protein	1	flavodoxin family protein	Flavodoxin (FldA) (PDB:1AG9)	NA
SRB__G2533	hypothetical protein	1	NA	NA	NA
SRB__G2527	Rap1a/Tai family immunity protein	1	Rap1a/Tai family immunity protein	NA	NA
SRB__G2518	isoprenylcysteine carboxyl methyltransferase family protein	1	isoprenylcysteine carboxyl methyltransferase family protein	Uncharacterized conserved protein YpbQ, isoprenylcysteine carboxyl methyltransferase (ICMT) family (YpbQ)	methyltransferase
SRB__G2509	tripartite tricarboxylate transporter TctB family protein	1	tripartite tricarboxylate transporter TctB family protein	NA	putative tricarboxylic transport membrane protein
SRB__G2472	response regulator	1	response regulator	DNA-binding transcriptional response regulator, NtrC family, contains REC, AAA-type ATPase, and a Fis-type DNA-binding domains (AtoC) (PDB:1ZY2)	NA
SRB__G2447	rhodanese-like domain-containing protein	1	rhodanese-like domain-containing protein	Rhodanese-related sulfurtransferase (PspE) (PDB:1TQ1)	NA
SRB__G2445	cold-shock protein	1	HPF/RaiA family ribosome-associated protein	Cold shock protein, CspA family (CspC) (PDB:1C9O)	NA
SRB__G2435	rhodanese-like domain-containing protein	1	rhodanese-like domain-containing protein	Rhodanese-related sulfurtransferase (PspE) (PDB:1TQ1)	NA

SRB_G2430	OmpH family outer membrane protein	1	OmpH family outer membrane protein	Periplasmic chaperone for outer membrane proteins, Skp family (HlpA) (PDB:1SG2)	outer membrane protein
SRB_G2418	metal-dependent transcriptional regulator	1	metal-dependent transcriptional regulator	Mn-dependent transcriptional regulator MntR, DtxR family (MntR) (PDB:1B1B)	DtxR family transcriptional regulator, manganese transport regulator
SRB_G2411	dTDP-4-dehydrorhamnose 3,5-epimerase	1	dTDP-4-dehydrorhamnose 3,5-epimerase (rfbC)	dTDP-4-dehydrorhamnose 3,5-epimerase or related enzyme (RfbC) (PDB:1DZR)	dTDP-4-dehydrorhamnose 3,5-epimerase [EC:5.1.3.13]
SRB_G2409	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	1	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (pgsA)	Phosphatidylglycerophosphate synthase (PgsA) (PDB:4MND)	CDP-diacylglycerol---glycerol-3-phosphate 3-phosphatidyltransferase [EC:2.7.8.5]
SRB_G2406	hypoxanthine phosphoribosyltransferase	1	hypoxanthine phosphoribosyltransferase (hpt)	Hypoxanthine-guanine phosphoribosyltransferase (HptA) (PDB:1G9S)	hypoxanthine phosphoribosyltransferase [EC:2.4.2.8]
SRB_G2401	hypothetical protein	1	NA	NA	NA
SRB_G2388	PepSY-associated TM helix domain-containing protein	1	PepSY-associated TM helix domain-containing protein	Uncharacterized conserved protein	uncharacterized protein
SRB_G2374	metal-dependent hydrolase	1	metal-dependent hydrolase	Membrane-bound metal-dependent hydrolase YbcI, DUF457 family (YbcI)	inner membrane protein
SRB_G2367	hypothetical protein	1	C-GCAxxG-C-C family protein	NA	NA
SRB_G2345	electron transport complex protein RnfA	1	RnfABCDGE type electron transport complex subunit A	Na+-translocating ferredoxin:NAD+ oxidoreductase RNF, RnfA subunit (RnfA)	H+/Na+-translocating ferredoxin:NAD+ oxidoreductase subunit A
SRB_G2335	rubrerythrin	1	rubrerythrin family protein	Rubrerythrin (YotD) (PDB:1B71)	NA
SRB_G2315	YeeE/YedE thiosulfate transporter family protein	1	YeeE/YedE family protein	Uncharacterized membrane protein YedE/YeeE, contains two sulfur transport domains (YedE)	uncharacterized protein
SRB_G2303	class I SAM-dependent methyltransferase	1	class I SAM-dependent methyltransferase	Cyclopropane fatty-acyl-phospholipid synthase and related methyltransferases (Cfa) (PDB:5Z9O)	tellurite methyltransferase [EC:2.1.1.265]
SRB_G2293	peroxiredoxin	1	peroxiredoxin	Alkyl hydroperoxide reductase subunit AhpC (peroxiredoxin) (AhpC) (PDB:2RII)	peroxiredoxin 2/4 [EC:1.11.1.24]
SRB_G2287	membrane integrity-associated transporter subunit PqiC	1	membrane integrity-associated transporter subunit PqiC	Intermembrane transporter PqiABC lipoprotein subunit PqiC (PqiC) (PDB:6OSX) (PUBMED:27795327)	uncharacterized protein
SRB_G2281	cytochrome c oxidase subunit 3	1	cytochrome c oxidase subunit 3	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (CyoC) (PDB:1QLE)	cytochrome c oxidase subunit III [EC:7.1.1.9]
SRB_G2221	electron transport complex subunit RsxE	1	electron transport complex subunit E	Na+-translocating ferredoxin:NAD+ oxidoreductase RNF, RnfE subunit (RnfE)	H+/Na+-translocating ferredoxin:NAD+ oxidoreductase subunit E
SRB_G2211	LysE family translocator	1	LysE family translocator	Threonine/homoserine/homoserine lactone efflux protein (RhtB)	homoserine/homoserine lactone efflux protein
SRB_G2206	REP-associated tyrosine transposase	1	transposase	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
SRB_G2201	DsbA family protein	1	DsbA family oxidoreductase	Predicted dithiol-disulfide isomerase, DsbA/YjbH family (virulence, stress resistance) (FrnE) (PDB:6GHB)	putative protein-disulfide isomerase
SRB_G2193	2-dehydro-3-deoxy-6-phosphogalactonate aldolase	1	2-dehydro-3-deoxy-6-phosphogalactonate aldolase	2-keto-3-deoxy-6-phosphogluconate aldolase (Eda) (PDB:2V81)	2-dehydro-3-deoxyphosphogalactonate aldolase [EC:4.1.2.21]
SRB_G2189	site-2 protease family protein	1	site-2 protease family protein	Zn-dependent protease (includes sporulation protein SpoIVFB) (SpoIVFB) (PDB:3B4R)	NA
SRB_G2162	LutC/YkgG family protein	1	lactate utilization protein	L-lactate utilization protein LutC, contains LUD domain (LutC) (PDB:2G40)	L-lactate dehydrogenase complex protein LldG
SRB_G2156	DUF3786 domain-containing protein	1	DUF3786 domain-containing protein	NA	NA
SRB_G2153	class I SAM-dependent methyltransferase	1	class I SAM-dependent methyltransferase	2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase (UbiG) (PDB:1BHJ)	NA
SRB_G2151	TVP38/TMEM64 family protein	1	VTT domain-containing protein	Uncharacterized membrane protein YdjX, related to fungal oxalate transporter, TVP38/TMEM64 family (TVP38) (PUBMED:20889782)	membrane-associated protein
SRB_G2150	tetratricopeptide repeat protein	1	tetratricopeptide repeat protein	Cytochrome c-type biogenesis protein CcmH/NrfG (NrfG)	NA
SRB_G2127	GntR family transcriptional regulator	1	GntR family transcriptional regulator	DNA-binding transcriptional regulator, GntR family (GntR) (PDB:3FMS)	NA
SRB_G2120	alpha/beta hydrolase	1	carboxylesterase	Predicted esterase (YpfH)	phospholipase/carboxylesterase
SRB_G2082	hypothetical protein	1	NA	NA	NA
SRB_G2081	ubiquinone/menaquinone biosynthesis methyltransferase	1	ubiquinone/menaquinone biosynthesis methyltransferase	Ubiquinone/menaquinone biosynthesis C-methylase UbiE/MenG (UbiE) (PDB:4OBW)	demethylmenaquinone methyltransferase / 2-methoxy-6-polyprenyl-1,4-benzoquinol methylase [EC:2.1.1.163 2.1.1.201]
SRB_G2080	HAD family hydrolase	1	HAD hydrolase-like protein	FMN and 5-amino-6-(5-phospho-D-ribitylamino)uracil phosphatase YigB, HAD superfamily (riboflavin biosynthesis) (YigB) (PDB:4JB3) (PUBMED:24123841)	GMP/IMP 5'-nucleotidase [EC:3.1.3.-]
SRB_G2067	transposase	1	transposase	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
SRB_G2060	tetratricopeptide repeat protein	1	tetratricopeptide repeat protein	NA	NA
SRB_G2032	fumarylacetoacetate hydrolase family protein	1	fumarylacetoacetate hydrolase family protein	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase (catechol pathway) (YcgM) (PDB:1GTT)	NA
SRB_G2030	cytidylyltransferase domain-containing protein	1	acylneuraminate cytidylyltransferase family protein	CMP-N-acetylneuraminic acid synthetase, NeuA/PseF family (NeuA) (PDB:1EYR)	CMP-N,N'-diacetyllegionaminic acid synthase [EC:2.7.7.82]

SRB__G2005	YkgJ family cysteine cluster protein	1	YkgJ family cysteine cluster protein	Uncharacterized protein YkgJ, contains CxxCxxCC motif (YkgJ)	uncharacterized protein
SRB__G2001	succinate dehydrogenase iron-sulfur subunit	1	succinate dehydrogenase iron-sulfur subunit	Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit (SdhB/FrdB) (PDB:6LUM)	succinate dehydrogenase iron-sulfur subunit [EC:1.3.5.1]
SRB__G1988	NAD(P)-dependent oxidoreductase	1	SDR family oxidoreductase	Putative NADH-flavin reductase (YwnB) (PDB:3EW7)	NA
SRB__G1974	amino acid ABC transporter ATP-binding protein	1	amino acid ABC transporter ATP-binding protein	ABC-type polar amino acid transport system, ATPase component (GlnQ) (PDB:4YMS)	glutamate/aspartate transport system ATP-binding protein [EC:7.4.2.1]
SRB__G1966	SDR family NAD(P)-dependent oxidoreductase	1	SDR family NAD(P)-dependent oxidoreductase	Short-chain dehydrogenase involved in D-alanine esterification of teichoic acids (DltE)	uncharacterized oxidoreductase [EC:1.-.-.-]
SRB__G1954	hypothetical protein	1	NA	Multimeric flavodoxin WrbA, includes NAD(P)H:quinone oxidoreductase (WrbA) (PDB:4DY4)	NA
SRB__G1950	3-oxoacyl-ACP reductase FabG	1	3-oxoacyl-ACP reductase FabG (fabG)	NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family (FabG) (PDB:6L1H)	3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100]
SRB__G1947	RluA family pseudouridine synthase	1	RluA family pseudouridine synthase	Pseudouridine synthase RluA, 23S rRNA- or tRNA-specific (RluA) (PDB:2I82)	NA
SRB__G1938	histone deacetylase family protein	1	histone deacetylase family protein	Acetoin utilization deacetylase AcuC or a related deacetylase (AcuC) (PDB:1C3P)	NA
SRB__G1934	energy-coupling factor ABC transporter ATP-binding protein	1	ABC transporter ATP-binding protein	Energy-coupling factor transporter ATP-binding protein EcfA2 (EcfA2) (PDB:4HUQ)	cobalt/nickel transport system ATP-binding protein
SRB__G1898	ArsR/SmtB family transcription factor	1	SRPBCC domain-containing protein	DNA-binding transcriptional regulator, ArsR family (ArsR) (PDB:3F6V)!!!Chalcone/flavanone-binding protein YndB, AHSA1/START/SRPBCC domain (YndB) (PDB:2KTE) (PUBMED:20818668)	NA
SRB__G1856	class I SAM-dependent methyltransferase	1	class I SAM-dependent methyltransferase	tRNA 5-hydroxyU34 O-methylase TrmR/YrrM (TrmR) (PDB:1H1D) (PUBMED:29982645)	NA
SRB__G1803	DUF1269 domain-containing protein	1	DUF1269 domain-containing protein	Uncharacterized membrane protein	NA
SRB__G1778	phosphatidate cytidylyltransferase	1	phosphatidate cytidylyltransferase	CDP-diglyceride synthetase (CdsA) (PDB:4Q2E)	phosphatidate cytidylyltransferase [EC:2.7.7.41]
SRB__G1755	MotA/TolQ/ExbB proton channel family protein	1	MotA/TolQ/ExbB proton channel family protein	Biopolymer transport protein ExbB/TolQ (TolQ) (PDB:5SV0)	NA
SRB__G1750	ABC transporter permease	1	ABC transporter permease	ABC-type polysaccharide/teichoic acid/polyol phosphate export permease (TagG) (PDB:6JBH)	lipopolysaccharide transport system permease protein
SRB__G1749	bifunctional hydroxymethylpyrimidine kinase/phosphomethylpyrimidine kinase	1	bifunctional hydroxymethylpyrimidine kinase/phosphomethylpyrimidine kinase (thiD)	Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase (ThiD) (PDB:1Y9X) (PUBMED:17351295)	hydroxymethylpyrimidine/phosphomethylpyrimidine kinase [EC:2.7.1.49 2.7.4.7]
SRB__G1730	shikimate dehydrogenase	1	shikimate dehydrogenase	Shikimate 5-dehydrogenase (AroE) (PDB:1NPD)	shikimate dehydrogenase [EC:1.1.1.25]
SRB__G1724	ribonuclease H-like domain-containing protein	1	ribonuclease H-like domain-containing protein	Uncharacterized conserved protein YprB, contains RNaseH-like and TPR domains (YprB)	uncharacterized protein
SRB__G1721	hypothetical protein	1	NA	NA	NA
SRB__G1720	LuxR C-terminal-related transcriptional regulator	1	LuxR C-terminal-related transcriptional regulator	ATP-, maltotriose- and DNA-dependent transcriptional regulator MalT (MalT)	LuxR family transcriptional regulator, maltose regulon positive regulatory protein
SRB__G1699	transporter substrate-binding domain-containing protein	1	transporter substrate-binding domain-containing protein	ABC-type amino acid transport/signal transduction system, periplasmic component/domain (HisJ) (PDB:1GGG)	polar amino acid transport system substrate-binding protein
SRB__G1692	ABC transporter permease	1	proline/glycine betaine ABC transporter permease	ABC-type proline/glycine betaine transport system, permease component (ProW)	glycine betaine/proline transport system permease protein
SRB__G1660	aminotransferase class IV	1	aminotransferase class IV	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase (IlvE) (PDB:1A3G)	branched-chain amino acid aminotransferase [EC:2.6.1.42]
SRB__G1620	1,4-dihydroxy-2-naphthoate polyprenyltransferase	1	1,4-dihydroxy-2-naphthoate polyprenyltransferase	1,4-dihydroxy-2-naphthoate polyprenyltransferase (MenA)	1,4-dihydroxy-2-naphthoate polyprenyltransferase [EC:2.5.1.74]
SRB__G1584	REP-associated tyrosine transposase	1	transposase	REP element-mobilizing transposase RayT (RAYT) (PDB:4ER8)	REP-associated tyrosine transposase
SRB__G1579	phytoene/squalene synthase family protein	1	phytoene/squalene synthase family protein	Phytoene/squalene synthetase (ERG9) (PDB:3WEK)	15-cis-phytoene synthase [EC:2.5.1.32]
SRB__G1577	NAD(P)-dependent oxidoreductase	1	NAD(P)-dependent oxidoreductase	3-hydroxyisobutyrate dehydrogenase or related beta-hydroxyacid dehydrogenase (MmsB) (PDB:1I36)	3-hydroxyisobutyrate dehydrogenase [EC:1.1.1.31]
SRB__G1565	DMT family transporter	1	DMT family transporter	Permease of the drug/metabolite transporter (DMT) superfamily (RhaT) (PDB:5I20)	drug/metabolite transporter, DME family
SRB__G1562	homocysteine S-methyltransferase family protein	1	homocysteine S-methyltransferase family protein	Homocysteine/selenocysteine methylase (S-methylmethionine-dependent) (MHT1) (PDB:5DML)	NA
SRB__G1549	DMT family transporter	1	DMT family transporter	Permease of the drug/metabolite transporter (DMT) superfamily (RhaT) (PDB:5I20)	NA
SRB__G1512	DMT family transporter	1	EamA family transporter	Permease of the drug/metabolite transporter (DMT) superfamily (RhaT) (PDB:5I20)	NA
SRB__G1501	restriction endonuclease	1	restriction endonuclease	NA	NA
SRB__G1490	2-hydroxyacid dehydrogenase	1	glyoxylate/hydroxypyruvate reductase A	Phosphoglycerate dehydrogenase or related dehydrogenase (SerA) (PDB:3OET)	glyoxylate/hydroxypyruvate reductase [EC:1.1.1.79 1.1.1.81]

SRB_G1449	electron transport complex protein RnfA	1	NADH-quinone reductase	Na+-translocating ferredoxin:NAD+ oxidoreductase RNF, RnfA subunit (RnfA)	H+/Na+-translocating ferredoxin:NAD+ oxidoreductase subunit A
SRB_G1446	hypothetical protein	1	NA	NA	NA
SRB_G1411	aldo/keto reductase	1	aldo/keto reductase	Pyridoxal reductase PdxI or related oxidoreductase, aldo/keto reductase family (PdxI) (PDB:1LQA) (PUBMED:32253339)	1-deoxyxylulose-5-phosphate synthase [EC:1.1.-.-]
SRB_G1398	DUF2232 domain-containing protein	1	DUF2232 domain-containing protein	NA	NA
SRB_G1354	ABC transporter permease	1	ABC transporter permease	ABC-type dipeptide/oligopeptide/nickel transport system, permease component (DppB)	peptide/nickel transport system permease protein
SRB_G1333	TRAP transporter substrate-binding protein	1	TRAP transporter substrate-binding protein DctP (dctP)	TRAP-type C4-dicarboxylate transport system, periplasmic component (DctP) (PDB:2CEX)	TRAP-type transport system periplasmic protein
SRB_G1298	NAD(P)-dependent oxidoreductase	1	hydroxyacid dehydrogenase	Phosphoglycerate dehydrogenase or related dehydrogenase (SerA) (PDB:3OET)	D-3-phosphoglycerate dehydrogenase / 2-oxoglutarate reductase [EC:1.1.1.95 1.1.1.399]
SRB_G1284	complex I subunit 1/NuoH family protein	1	NADH-quinone oxidoreductase subunit H	NADH:ubiquinone oxidoreductase subunit 1 (chain H) (NuoH) (PDB:6HUM)	NADH-quinone oxidoreductase subunit H [EC:7.1.1.2]
SRB_G1275	peptidylprolyl isomerase	1	peptidylprolyl isomerase	FKBP-type peptidyl-prolyl cis-trans isomerase (FkpA) (PDB:1A7X)!!!Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family (PpiB) (PDB:4DGD)	peptidylprolyl isomerase [EC:5.2.1.8]
SRB_G1245	energy-coupling factor ABC transporter ATP-binding protein	1	ABC transporter ATP-binding protein	ABC-type sugar transport system, ATPase component MalK (MalK) (PDB:2ONK)	NA
SRB_G1242	branched-chain amino acid ABC transporter permease	1	branched-chain amino acid ABC transporter permease	ABC-type branched-chain amino acid transport system, permease component (LivM)	branched-chain amino acid transport system permease protein
SRB_G1237	N-acetyl-gamma-glutamyl-phosphate reductase	1	N-acetyl-gamma-glutamyl-phosphate reductase	N-acetyl-gamma-glutamylphosphate reductase (ArgC) (PDB:3DR3)	N-acetyl-gamma-glutamyl-phosphate reductase [EC:1.2.1.38]
SRB_G1227	recombinase RecA	1	recombinase RecA (recA)	RecA/RadA recombinase (RecA) (PDB:1B22)	recombination protein RecA
SRB_G1206	N-acetylneuraminate synthase family protein	1	N-acetylneuraminate synthase family protein	Sialic acid synthase SpsE, contains C-terminal SAF domain (SpsE) (PDB:1VLI)	N,N'-diacetyllegionaminate synthase [EC:2.5.1.101]
SRB_G1201	permease	1	permease	Uncharacterized membrane protein YraQ, UPF0718 family (YraQ)	uncharacterized protein
SRB_G1187	radical SAM protein	1	radical SAM protein	Radical SAM superfamily maturase, SkfB/NifB/PqqE family (SkfB) (PDB:6Q2P) (PUBMED:23282011)	NA
SRB_G1184	histone deacetylase	1	histone deacetylase	Acetoin utilization deacetylase AcuC or a related deacetylase (AcuC) (PDB:1C3P)	NA
SRB_G1182	DUF933 domain-containing protein	1	redox-regulated ATPase YchF (ychF)	Ribosome-binding ATPase YchF, GTP1/OBG family (GTP1) (PDB:2DWQ) (PUBMED:21527254)	NA
SRB_G1159	type IV pilus twitching motility protein PilT	1	type IV pilus twitching motility protein PilT	Type IV pilus assembly protein PilT, pilus retraction ATPase (PilT) (PDB:5FL3)	twitching motility protein PilT
SRB_G1158	LptF/LptG family permease	1	LptF/LptG family permease	Lipopolysaccharide export LptBFGC system, permease protein LptF (LptF) (PDB:6S8H)	lipopolysaccharide export system permease protein
SRB_G1115	tRNA preQ1(34) S-adenosylmethionine ribosyltransferase-isomerase QueA	1	tRNA preQ1(34) S-adenosylmethionine ribosyltransferase-isomerase QueA (queA)	S-adenosylmethionine:tRNA-ribosyltransferase-isomerase (queuine synthetase) (QueA) (PDB:1VKY)	S-adenosylmethionine:tRNA ribosyltransferase-isomerase [EC:2.4.99.17]
SRB_G1095	NADH-quinone oxidoreductase subunit D	1	NADH-quinone oxidoreductase subunit D	NADH:ubiquinone oxidoreductase 49 kD subunit (chain D) (NuoD) (PDB:2FUG)	NADH-quinone oxidoreductase subunit D [EC:7.1.1.2]
SRB_G1085	alginate O-acetyltransferase AlgX-related protein	1	alginate O-acetyltransferase AlgX-related protein	NA	alginate O-acetyltransferase complex protein AlgJ
SRB_G1045	ATP-binding protein	1	ATP-binding protein	Predicted ATPase, AAA+ superfamily	uncharacterized protein
SRB_G1041	glycosyltransferase	1	glycosyltransferase	Glycosyltransferase involved in cell wall bisynthesis (RfaB) (PDB:2IV7)	NA
SRB_G1040	2-oxoacid:acceptor oxidoreductase subunit alpha	1	2-oxoacid:acceptor oxidoreductase subunit alpha	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, alpha subunit (PorA) (PDB:5C4I)	2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunit alpha [EC:1.2.7.3 1.2.7.11]
SRB_G1029	acyl-CoA dehydrogenase family protein	1	acyl-CoA dehydrogenase family protein	Acyl-CoA dehydrogenase related to the alkylation response protein AidB (CaiA) (PDB:1BUC)	acyl-CoA dehydrogenase [EC:1.3.8.7]
SRB_G1028	UDP-N-acetylglucosamine 2-epimerase	1	UDP-N-acetylglucosamine 2-epimerase (hydrolyzing) (neuC)	UDP-N-acetylglucosamine 2-epimerase (WecB) (PDB:1F6D) (PUBMED:26641313;30389844)	GDP/UDP-N,N'-diacetylbacillosamine 2-epimerase (hydrolysing) [EC:3.2.1.184]
SRB_G1021	radical SAM protein	1	radical SAM protein	Radical SAM superfamily enzyme YgiQ, UPF0313 family (YgiQ) (PDB:6FD2)	NA
SRB_G1020	MBL fold metallo-hydrolase	1	MBL fold metallo-hydrolase	Metal-dependent hydrolase, beta-lactamase superfamily II (PDB:2P4Z)	7,8-dihydropterin-6-yl-methyl-4-(beta-D-ribofuranosyl)aminobenzene 5'-phosphate synthase [EC:2.5.1.105]
SRB_G1019	MFS transporter	1	MFS transporter	Predicted arabinose efflux permease AraJ, MFS family (AraJ) (PDB:4LDS)	MFS transporter, FSR family, fosmidomycin resistance protein
SRB_G1015	trans-sulfuration enzyme family protein	1	PLP-dependent transferase	Cystathionine beta-lyase/cystathionine gamma-synthase (MetC) (PDB:1CL1)	NA

SRB_G1010	sodium ion-translocating decarboxylase subunit beta	1	sodium ion-translocating decarboxylase subunit beta	Na+-transporting oxaloacetate/methylmalonyl-CoA decarboxylase, beta subunit (OadB)	carboxybiotin decarboxylase [EC:7.2.4.1]
SRB_G1009	glycosyltransferase	1	glycosyltransferase	Glycosyltransferase involved in cell wall bisynthesis (WcaA) (PDB:5MLZ)!!!PMT family glycosyltransferase ArnT/Agl22, involved in glycosylation of proteins and lipid IVA (ArnT) (PDB:5EZM) (PUBMED:16226890;31233657;32706435)!!!Putative flippase GtrA (transmembrane translocase of bactoprenol-linked glucose) (GtrA) (PDB:5MLZ) (PUBMED:24710389)	dolichol-phosphate mannosyltransferase [EC:2.4.1.83]
SRB_G989	sulfate reduction electron transfer complex DsrMKJOP subunit DsrP	1	polysulfide reductase NrfD (nrfD)	Ni/Fe-hydrogenase 2 integral membrane subunit HybB (HybB)	NA
SRB_G981	NrfD/PsrC family molybdoenzyme membrane anchor subunit	1	polysulfide reductase NrfD (nrfD)	Ni/Fe-hydrogenase 2 integral membrane subunit HybB (HybB)	NA
SRB_G980	restriction endonuclease subunit S	1	restriction endonuclease subunit S	Restriction endonuclease S subunit (HsdS) (PDB:1YDX)	type I restriction enzyme, S subunit [EC:3.1.21.3]
SRB_G948	CaiB/BaiF CoA transferase family protein	1	CoA transferase	Crotonobetainyl-CoA:carnitine CoA-transferase CaiB and related acyl-CoA transferases (CaiB) (PDB:1P5H)	NA
SRB_G927	NADH:ubiquinone reductase (Na(+)-transporting) subunit B	1	NADH:ubiquinone reductase (Na(+)-transporting) subunit B	Na+-transporting NADH:ubiquinone oxidoreductase, subunit NqrB (NqrB) (PDB:4P6V)	Na+-transporting NADH:ubiquinone oxidoreductase subunit B [EC:7.2.1.1]
SRB_G893	serine hydrolase	1	serine hydrolase	CubicO group peptidase, beta-lactamase class C family (AmpC) (PDB:1BLS)	NA
SRB_G885	efflux RND transporter periplasmic adaptor subunit	1	efflux RND transporter periplasmic adaptor subunit	Multidrug efflux pump subunit AcrA (membrane-fusion protein) (AcrA) (PDB:1T5E)	membrane fusion protein, multidrug efflux system
SRB_G883	OmpP1/FadL family transporter	1	outer membrane protein transport protein	Long-chain fatty acid transport protein (FadL) (PDB:1T16)	long-chain fatty acid transport protein
SRB_G882	response regulator	1	response regulator	DNA-binding transcriptional response regulator, NtrC family, contains REC, AAA-type ATPase, and a Fis-type DNA-binding domains (AtoC) (PDB:1ZY2)	two-component system, response regulator, stage 0 sporulation protein F
SRB_G866	M48 family metallopeptidase	1	M48 family metallopeptidase	Zn-dependent protease with chaperone function (HtpX) (PDB:2YPT)	STE24 endopeptidase [EC:3.4.24.84]
SRB_G864	tagaturonate epimerase family protein	1	tagaturonate epimerase family protein	NA	tagaturonate epimerase [EC:5.1.2.7]
SRB_G843	histidine--tRNA ligase	1	histidine--tRNA ligase	Histidyl-tRNA synthetase (HisS) (PDB:1KMM)	histidyl-tRNA synthetase [EC:6.1.1.21]
SRB_G838	methyltransferase domain-containing protein	1	methyltransferase domain-containing protein	Peptidyl-prolyl cis-trans isomerase, FKBP type (SlpA) (PDB:2M2A)!!!Ubiquinone/menaquinone biosynthesis C-methylase UbiE/MenG (UbiE) (PDB:4OBW)	NA
SRB_G835	HD-GYP domain-containing protein	1	HD domain-containing protein	HD-GYP domain, c-di-GMP phosphodiesterase class II (or its inactivated variant) (HDGYP) (PDB:5Z7C) (PUBMED:23475961)	NA
SRB_G831	nickel-dependent lactate racemase	1	nickel-dependent lactate racemase (larA)	Nickel-dependent lactate racemase (LarA) (PDB:2YJG) (PUBMED:24710389)	lactate racemase [EC:5.1.2.1]
SRB_G784	NCS2 family permease	1	NCS2 family permease	Xanthine/guanine/uracil/vitamin C permease GhxP/GhxQ, nucleobase:cation symporter 2 (NCS2) family (NCS2) (PUBMED:24169576;24214977)	adenine/guanine/hypoxanthine permease
SRB_G746	magnesium transporter	1	magnesium transporter (mgtE)	Mg/Co/Ni transporter MgtE (contains CBS domain) (MgtE) (PDB:2YVX)	magnesium transporter
SRB_G736	TRAP transporter large permease	1	TRAP transporter large permease subunit	TRAP-type mannitol/chloroaromatic compound transport system, large permease component (FcbT3)	C4-dicarboxylate transporter, DctM subunit
SRB_G729	uracil-xanthine permease family protein	1	uracil-xanthine permease	Xanthine/uracil permease (UraA) (PDB:3QE7)	uracil permease
SRB_G713	amidohydrolase	1	amidohydrolase	Cytosine/adenosine deaminase or related metal-dependent hydrolase (SsnA) (PDB:3O7U)	5-methylthioadenosine/S-adenosylhomocysteine deaminase [EC:3.5.4.31 3.5.4.28]
SRB_G704	16S rRNA (cytosine(967)-C(5))-methyltransferase RsmB	1	16S rRNA (cytosine(967)-C(5))-methyltransferase RsmB (rsmB)	16S rRNA C967 or C1407 C5-methylase, RsmB/RsmF family (RsmB) (PDB:2FRX)!!!Transcription antitermination protein NusB (NusB) (PDB:1EY1) (PUBMED:21652641)	16S rRNA (cytosine967-C5)-methyltransferase [EC:2.1.1.176]
SRB_G678	nucleotide 5'-monophosphate nucleosidase PpnN	1	LOG family protein	Nucleotide monophosphate nucleosidase PpnN/YdgH, Lonely Guy (LOG) family (PpnN) (PDB:2PMB) (PUBMED:25728768;27941785;31023582;31420341)	pyrimidine/purine-5'-nucleotide nucleosidase [EC:3.2.2.10 3.2.2.-]
SRB_G677	cobyrinate a,c-diamide synthase	1	cobyrinate a,c-diamide synthase	Cobyrinic acid a,c-diamide synthase (CobB)	cobyrinic acid a,c-diamide synthase [EC:6.3.5.9 6.3.5.11]
SRB_G662	hypothetical protein	1	NA	NA	NA
SRB_G649	aspartate kinase	1	aspartate kinase	Aspartate kinase (MetL1) (PDB:2CDQ)	NA
SRB_G635	23S rRNA (uracil(1939)-C(5))-methyltransferase RlmD	1	23S rRNA (uracil(1939)-C(5))-methyltransferase RlmD (rlmD)	tRNA/tmRNA/rRNA uracil-C5-methylase, TrmA/RlmC/RlmD family (TrmA) (PDB:1UWV)	23S rRNA (uracil1939-C5)-methyltransferase [EC:2.1.1.190]
SRB_G627	biotin carboxylase N-terminal domain-containing protein	1	acetyl-CoA carboxylase biotin carboxylase subunit	Acetyl/propionyl-CoA carboxylase, alpha subunit (PccA) (PDB:3U9S)	NA

SRB__G614	heavy metal sensor histidine kinase	1	heavy metal sensor histidine kinase	Signal transduction histidine kinase (BaeS) (PDB:1JOY)	two-component system, OmpR family, sensor kinase [EC:2.7.13.3]
SRB__G609	sigma-54 interaction domain-containing protein	1	sigma 54-interacting transcriptional regulator	RocR-type transcriptional regulator, contains PAS, AAA-type ATPase, and DNA-binding Fis domains (RocR)	NA
SRB__G598	two-component system sensor histidine kinase CreC	1	two-component system sensor histidine kinase CreC (creC)	Signal transduction histidine kinase (BaeS) (PDB:1JOY)	two-component system, OmpR family, sensor histidine kinase CreC [EC:2.7.13.3]
SRB__G545	sigma-54 interaction domain-containing protein	1	sigma 54-interacting transcriptional regulator	PAS domain (PAS) (PDB:2MWG)!!!FhlA-type transcriptional regulator, contains GAF, AAA-type ATPase, and DNA-binding Fis domains (FhlA)	formate hydrogenlyase transcriptional activator
SRB__G536	DUF4129 domain-containing protein	1	DUF4129 domain-containing protein	NA	NA
SRB__G520	tripartite tricarboxylate transporter permease	1	tripartite tricarboxylate transporter permease	TctA family transporter (PUBMED:14499931)	putative tricarboxylic transport membrane protein
SRB__G505	fumarate hydratase	1	fumarate hydratase	Tartrate dehydratase beta subunit/Fumarate hydratase class I, C-terminal domain (FumA) (PDB:5L2R)!!!Tartrate dehydratase alpha subunit/Fumarate hydratase class I, N-terminal domain (TtdA) (PDB:5L2R) (PUBMED:15613473)	fumarate hydratase, class I [EC:4.2.1.2]
SRB__G497	F0F1 ATP synthase subunit alpha	1	F0F1 ATP synthase subunit alpha	FoF1-type ATP synthase, alpha subunit (AtpA) (PDB:3OAA)	F-type H+/Na+-transporting ATPase subunit alpha [EC:7.1.2.2 7.2.2.1]
SRB__G478	acyl-CoA carboxylase subunit beta	1	methylmalonyl-CoA carboxyltransferase	Acetyl-CoA carboxylase, carboxyltransferase component (MmdA) (PDB:1OD2)	methylmalonyl-CoA decarboxylase subunit alpha [EC:7.2.4.3]
SRB__G463	sigma-54 interaction domain-containing protein	1	sigma 54-interacting transcriptional regulator	FhlA-type transcriptional regulator, contains GAF, AAA-type ATPase, and DNA-binding Fis domains (FhlA)	Nif-specific regulatory protein
SRB__G462	thio(seleno)oxazole modification radical SAM maturase SbtM	1	selenobiotic family peptide radical SAM maturase (sbtM)	Radical SAM superfamily maturase, SkfB/NifB/PqqE family (SkfB) (PDB:6Q2P) (PUBMED:23282011)	NA
SRB__G417	putative ABC transporter permease subunit	1	putative ABC transporter permease subunit	NA	NA
SRB__G405	cbb3-type cytochrome c oxidase subunit I	1	cbb3-type cytochrome c oxidase subunit I	Heme/copper-type cytochrome/quinol oxidase, subunit 1 (CyoB) (PDB:1AR1)	cytochrome c oxidase subunit I [EC:7.1.1.9]
SRB__G396	ArnT family glycosyltransferase	1	glycosyltransferase family 39 protein	NA	NA
SRB__G391	AarF/UbiB family protein	1	AarF/UbiB family protein	Predicted protein kinase regulating ubiquinone biosynthesis, AarF/ABC1/UbiB family (AarF) (PDB:4PED)	NA
SRB__G389	membrane protein insertase YidC	1	membrane protein insertase YidC (yidC)	Membrane protein insertase Oxa1/YidC/SpoIIIJ (YidC) (PDB:3WO6)	YidC/Oxa1 family membrane protein insertase
SRB__G369	biosynthetic-type acetolactate synthase large subunit	1	biosynthetic-type acetolactate synthase large subunit (ilvB)	Acetolactate synthase large subunit or other thiamine pyrophosphate-requiring enzyme (IlvB) (PDB:1BFD)	acetolactate synthase I/II/III large subunit [EC:2.2.1.6]
SRB__G344	succinate dehydrogenase flavoprotein subunit	1	succinate dehydrogenase flavoprotein subunit	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (SdhA) (PDB:1KF6)	succinate dehydrogenase flavoprotein subunit [EC:1.3.5.1]
SRB__G325	sulfoacetaldehyde acetyltransferase	1	sulfoacetaldehyde acetyltransferase (xsc)	Acetolactate synthase large subunit or other thiamine pyrophosphate-requiring enzyme (IlvB) (PDB:1BFD)	sulfoacetaldehyde acetyltransferase [EC:2.3.3.15]
SRB__G320	DnaB-like helicase C-terminal domain-containing protein	1	AAA family ATPase	Replicative DNA helicase (DnaB) (PDB:1B79)	replicative DNA helicase [EC:5.6.2.3]
SRB__G311	phospholipase D family protein	1	phospholipase D family protein	NA	NA
SRB__G304	ATP-dependent zinc metalloprotease FtsH	1	ATP-dependent zinc metalloprotease FtsH (ftsH)	ATP-dependent Zn proteases (HflB) (PDB:1LV7)	NA
SRB__G299	glycerophosphodiester phosphodiesterase family protein	1	glycerophosphoryl diester phosphodiesterase membrane domain-containing protein	Glycerophosphoryl diester phosphodiesterase (UgpQ) (PDB:1O1Z)	glycerophosphoryl diester phosphodiesterase [EC:3.1.4.46]
SRB__G296	dihydroxy-acid dehydratase	1	dihydroxy-acid dehydratase (ilvD)	Dihydroxyacid dehydratase/phosphogluconate dehydratase (IlvD) (PDB:2GP4)	dihydroxy-acid dehydratase [EC:4.2.1.9]
SRB__G287	molecular chaperone HtpG	1	molecular chaperone HtpG (htpG)	Molecular chaperone, HSP90 family (HtpG) (PDB:1Y4S) (PUBMED:30150682)	TNF receptor-associated protein 1
SRB__G285	fructose-bisphosphatase class III	1	fructose-bisphosphatase class III	Fructose-1,6-bisphosphatase (Fbp2)	fructose-1,6-bisphosphatase III [EC:3.1.3.11]
SRB__G276	aminodeoxychorismate synthase component I	1	aminodeoxychorismate synthase component I (pabB)	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase (IlvE) (PDB:1A3G)!!!Anthranilate/para-aminobenzoate synthases component I (TrpE) (PDB:1I1Q)	para-aminobenzoate synthetase / 4-amino-4-deoxychorismate lyase [EC:2.6.1.85 4.1.3.38]
SRB__G273	sigma-54-dependent Fis family transcriptional regulator	1	sigma 54-interacting transcriptional regulator	RocR-type transcriptional regulator, contains PAS, AAA-type ATPase, and DNA-binding Fis domains (RocR)	sigma-54 dependent transcriptional regulator, acetoin dehydrogenase operon transcriptional activator AcoR
SRB__G268	aspartate-alanine antiporter	1	aspartate-alanine antiporter (aspT)	Uncharacterized membrane protein YbjL, putative transporter (YbjL)	putative transport protein
SRB__G261	fumarate reductase/succinate dehydrogenase flavoprotein subunit	1	fumarate reductase/succinate dehydrogenase flavoprotein subunit	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (SdhA) (PDB:1KF6)	succinate dehydrogenase flavoprotein subunit [EC:1.3.5.1]

SRB_G243	pyruvate carboxylase subunit B	1	pyruvate carboxylase subunit B	Acetyl/propionyl-CoA carboxylase, alpha subunit (PccA) (PDB:3U9S)!!!Pyruvate/oxaloacetate carboxyltransferase (OadA1) (PDB:2NX9) (PUBMED:18681940)	NA
SRB_G211	elongation factor G	1	elongation factor G (fusA)	Translation elongation factor EF-G, a GTPase (FusA) (PDB:1DAR)	elongation factor G
SRB_G183	DUF3141 domain-containing protein	1	DUF3141 domain-containing protein	Poly-beta-hydroxybutyrate synthase (PhaC) (PDB:5XAV)	NA
SRB_G165	penicillin acylase family protein	1	penicillin acylase family protein	Acyl-homoserine lactone (AHL) acylase PvdQ (PvdQ) (PDB:4HST)	penicillin G amidase [EC:3.5.1.11]
SRB_G147	gluconeogenesis factor YvcK family protein	1	YvcK family protein	Archaeal 2-phospho-L-lactate transferase/Bacterial gluconeogenesis factor, CofD/UPF0052 family (CofD) (PDB:1ZX8) (PUBMED:11888293;18252724)	NA
SRB_G140	carbamoyltransferase HypF	1	carbamoyltransferase HypF (hypF)	Hydrogenase maturation factor HypF (carbamoyltransferase) (HypF) (PDB:3TSP)	hydrogenase maturation protein HypF
SRB_G124	DNA topoisomerase (ATP-hydrolyzing) subunit B	1	DNA topoisomerase (ATP-hydrolyzing) subunit B (gyrB)	DNA gyrase/topoisomerase IV, subunit B (GyrB) (PDB:1EI1)	DNA gyrase subunit B [EC:5.6.2.2]
SRB_G119	PEP/pyruvate-binding domain-containing protein	1	PEP/pyruvate-binding domain-containing protein	Phosphoenolpyruvate synthase/pyruvate phosphate dikinase (PpsA) (PDB:5HV6)	pyruvate, water dikinase [EC:2.7.9.2]
SRB_G115	DNA gyrase subunit A	1	DNA gyrase subunit A (gyrA)	DNA gyrase/topoisomerase IV, subunit A (GyrA) (PDB:1SUU)	DNA gyrase subunit A [EC:5.6.2.2]
SRB_G103	preprotein translocase subunit SecA	1	preprotein translocase subunit SecA (secA)	Preprotein translocase subunit SecA (ATPase, RNA helicase) (SecA) (PDB:2FSF)	preprotein translocase subunit SecA [EC:7.4.2.8]
SRB_G97	protein translocase subunit SecD	1	protein translocase subunit SecD (secD)	Preprotein translocase subunit SecF (SecF) (PDB:2FYH) (PUBMED:28830929)!!!Preprotein translocase subunit SecD (SecD) (PDB:3AQP) (PUBMED:26100040)	SecD/SecF fusion protein
SRB_G89	site-specific DNA-methyltransferase	1	site-specific DNA-methyltransferase	Adenine specific DNA methylase Mod (Mod) (PDB:4ZCF)	adenine-specific DNA-methyltransferase [EC:2.1.1.72]
SRB_G74	DNA methyltransferase	1	DNA methylase	DNA modification methylase (YhdJ) (PDB:1BOO)!!!Adenine-specific DNA methylase, contains a Zn-ribbon domain	NA
SRB_G46	efflux RND transporter permease subunit	1	multidrug efflux RND transporter permease subunit	Multidrug efflux pump subunit AcrB (AcrB) (PDB:1IWG)	multidrug efflux pump
SRB_G39	efflux RND transporter permease subunit	1	efflux RND transporter permease subunit	Multidrug efflux pump subunit AcrB (AcrB) (PDB:1IWG)	NA
SRB_G28	pyruvate:ferredoxin (flavodoxin) oxidoreductase	1	pyruvate:ferredoxin (flavodoxin) oxidoreductase (nifJ)	Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, alpha subunit (PorA) (PDB:5C4I)!!!Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, beta subunit (PorB) (PDB:1B0P)!!!Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, gamma subunit (PorG) (PDB:1B0P)!!!Pyruvate:ferredoxin oxidoreductase or related 2-oxoacid:ferredoxin oxidoreductase, delta subunit (PorD) (PDB:5C4I)	pyruvate-ferredoxin/flavodoxin oxidoreductase [EC:1.2.7.1 1.2.7.-]
SRB_G4426	hypothetical protein	1	NA	NA	NA
SRB_G4407	class I SAM-dependent methyltransferase	1	RsmD family RNA methyltransferase	tRNA G37 N-methylase Trm5 (Trm5) (PDB:2YX1)	tRNA (guanine37-N1)-methyltransferase [EC:2.1.1.228]
SRB_G4390	DUF3299 domain-containing protein	1	DUF3299 domain-containing protein	Uncharacterized conserved protein, DUF3299 family (PDB:4HSP) (PUBMED:21378183)	nickel/cobalt transporter (NicO) family protein
SRB_G4385	glutamine-hydrolyzing carbamoyl-phosphate synthase small subunit	1	glutamine-hydrolyzing carbamoyl-phosphate synthase small subunit (carA)	Carbamoylphosphate synthase small subunit (CarA) (PDB:1C30)	carbamoyl-phosphate synthase small subunit [EC:6.3.5.5]
SRB_G4351	pyridoxine 5'-phosphate synthase	1	pyridoxine 5'-phosphate synthase	ssRNA-specific RNase YbeY, 16S rRNA maturation enzyme (YbeY) (PDB:1XM5) (PUBMED:17407324;23543739)!!!Pyridoxine 5'-phosphate synthase PdxJ (PdxJ) (PDB:1HO1)	pyridoxine 5-phosphate synthase [EC:2.6.99.2]
SRB_G4279	RHS repeat-associated core domain-containing protein	1	RHS repeat-associated core domain-containing protein	Uncharacterized conserved protein RhaS, contains 28 RHS repeats (RhsA) (PDB:4IGL)	NA
SRB_G3996	DeoR/GlpR family DNA-binding transcription regulator	1	DeoR/GlpR transcriptional regulator	DNA-binding transcriptional regulator of sugar metabolism, DeoR/GlpR family (GlpR)	DeoR family transcriptional regulator, glycerol-3-phosphate regulon repressor
SRB_G3962	DNA methyltransferase	1	site-specific DNA-methyltransferase	Adenine specific DNA methylase Mod (Mod) (PDB:4ZCF)	adenine-specific DNA-methyltransferase [EC:2.1.1.72]
SRB_G3515	hypothetical protein	1	NA	NA	NA
SRB_G3189	acylphosphatase	1	acylphosphatase	Acylphosphatase (AcyP) (PDB:1APS)	acylphosphatase [EC:3.6.1.7]
SRB_G2875	PIN domain-containing protein	1	PIN domain-containing protein	Ribonuclease/mRNA interferase VapC, contains PIN domain (VapC) (PDB:3TND)	NA
SRB_G2439	arsenic resistance N-acetyltransferase ArsN2	1	arsenic resistance N-acetyltransferase ArsN2 (arsN2)	N-acetylglutamate synthase or related acetyltransferase, GNAT family (ArgA) (PDB:2R98)	NA
SRB_G2407	hypothetical protein	1	NA	NA	NA
SRB_G2302	cysteine hydrolase family protein	1	cysteine hydrolase	Nicotinamidase-related amidase (PncA) (PDB:1ILW)	NA

SRB__G2266	uracil-DNA glycosylase family protein	1	uracil-DNA glycosylase family protein	Uracil-DNA glycosylase (Udg4) (PDB:3IKB)	uracil-DNA glycosylase [EC:3.2.2.27]
SRB__G1961	oxygen-insensitive NADPH nitroreductase	1	oxygen-insensitive NADPH nitroreductase (nfsA)	Nitroreductase (NfnB) (PDB:1F5V)	FMN reductase (NADPH) [EC:1.5.1.38]
SRB__G1854	DUF166 family protein	1	DUF166 family protein	Uncharacterized conserved protein, DUF166 domain	thymidylate synthase [EC:2.1.1.-]
SRB__G1662	carbohydrate ABC transporter permease	1	sugar ABC transporter permease	ABC-type sugar transport system, permease component (UgpA) (PDB:3FH6)	glycerol transport system permease protein
SRB__G1641	cation diffusion facilitator family transporter	1	cation transporter	Divalent metal cation (Fe/Co/Zn/Cd) efflux pump (FieF) (PDB:2QFI)	ferrous-iron efflux pump FieF
SRB__G1539	TIGR01777 family oxidoreductase	1	TIGR01777 family oxidoreductase	NAD dependent epimerase/dehydratase family enzyme (YfcH) (PDB:4B4O)	uncharacterized protein
SRB__G1292	elongator complex protein 3	1	radical SAM protein	tRNA U34 5-carboxymethylaminomethylation enzyme Elp3 (RNA elongator complex protein 3), contains radical SAM and acetyltransferase domains (ELP3) (PDB:6IA8) (PUBMED:25151136;27455459)	NA
SRB__G1166	ABC transporter ATP-binding protein	1	ABC transporter ATP-binding protein	ABC-type sugar transport system, ATPase component MalK (MalK) (PDB:2ONK)	glycerol transport system ATP-binding protein
SRB__G1075	NifB/NifX family molybdenum-iron cluster-bindingprotein	1	radical SAM protein	GTP 3',8-cyclase (molybdenum cofactor biosynthesis protein MoaA) (MoaA) (PDB:1TV7)	nitrogen fixation protein NifB
SRB__G673	nitrogenase molybdenum-iron protein subunit beta	1	nitrogenase molybdenum-iron protein subunit beta (nifK)	Nitrogenase Mo-Fe protein NifD/coenzyme F430 biosynthesis subunit CfbD (NifD/CfbD) (PDB:1FP4) (PUBMED:28225763)	nitrogenase molybdenum-iron protein beta chain [EC:1.18.6.1]
SRB__G610	hypothetical protein	1	NA	NA	NA
SRB__G204	elongation factor G	1	elongation factor G	Translation elongation factor EF-G, a GTPase (FusA) (PDB:1DAR)	elongation factor G
SRB__G145	CHC2 zinc finger domain-containing protein	1	toprim domain-containing protein	DNA primase (bacterial type) (DnaG) (PDB:3B39) (PUBMED:28128549)	DNA primase [EC:2.7.7.101]
SRB__G113	S8 family peptidase	1	S8 family peptidase	Serine protease, subtilisin family (AprE) (PDB:1SCJ)	NA
SRB__G4736	tyrosine-type recombinase/integrase	1	tyrosine-type recombinase/integrase	Site-specific recombinase XerD (XerD) (PDB:1A0P)	NA
SRB__G4602	hypothetical protein	1	NA	NA	NA
SRB__G4594	type II toxin-antitoxin system VapC family toxin	1	type II toxin-antitoxin system VapC family toxin	RNA interferase (RNase) VapC, contains PIN domain (VapC) (PDB:1V8O) (PUBMED:22539524)	uncharacterized protein
SRB__G4556	von Willebrand factor type A domain-containing protein	1	von Willebrand factor type A domain-containing protein	Secreted protein containing bacterial Ig-like domain and vWFA domain (YfbK)	Ca-activated chloride channel homolog
SRB__G4454	IS3 family transposase	1	NA	NA	NA
SRB__G4447	transposase	1	transposase	NA	NA
SRB__G3754	hypothetical protein	1	NA	NA	NA
SRB__G3739	hypothetical protein	1	NA	Retron-type reverse transcriptase (YkfC)	RNA-directed DNA polymerase [EC:2.7.7.49]
SRB__G3734	PKD domain-containing protein	1	PKD domain-containing protein	Ca2+-binding protein, RTX toxin-related (PDB:1AF0) (PUBMED:15135544)	NA
SRB__G3668	IS1380 family transposase	1	IS1380 family transposase	NA	NA
SRB__G3332	hypothetical protein	1	NA	NA	NA
SRB__G3150	IS66 family insertion sequence element accessoryprotein TnpA	1	IS66 family insertion sequence element accessory protein TnpA (tnpA)	NA	NA
SRB__G2880	hypothetical protein	1	NA	NA	NA
SRB__G2823	putative toxin-antitoxin system toxin component,PIN family	1	putative toxin-antitoxin system toxin component, PIN family	Predicted nucleic acid-binding protein, contains PIN domain	uncharacterized protein
SRB__G2464	AAA family ATPase	1	AAA family ATPase	ATP-, maltotriose- and DNA-dependent transcriptional regulator MalT (MalT)	LuxR family transcriptional regulator, maltose regulon positive regulatory protein
SRB__G2322	RHS repeat-associated core domain-containing protein	1	RHS repeat-associated core domain-containing protein	Uncharacterized conserved protein RhaS, contains 28 RHS repeats (RhsA) (PDB:4IGL)	NA
SRB__G873	hypothetical protein	1	NA	NA	NA
SRB__G152	ThiF family adenylyltransferase	1	ThiF family adenylyltransferase	tRNA A37 threonylcarbamoyladenosine dehydratase (TcdA) (PDB:4D7A) (PUBMED:27913733)	NA
SRB__G96	sigma 54-interacting transcriptional regulator	1	sigma 54-interacting transcriptional regulator	PAS domain (PAS) (PDB:2MWG)!!!FhlA-type transcriptional regulator, contains GAF, AAA-type ATPase, and DNA-binding Fis domains (FhlA)	formate hydrogenlyase transcriptional activator
SRB__G4393	type II toxin-antitoxin system Phd/YefM family antitoxin	2	type II toxin-antitoxin system Phd/YefM family antitoxin	Antitoxin component of toxin-antitoxin stability system, DNA-binding transcriptional repressor (Phd) (PDB:3DBO)	NA
SRB__G4386	cereblon family protein	2	cereblon family protein	NA	NA
SRB__G4296	hypothetical protein	2	NA	NA	NA
SRB__G4259	transposase	2	transposase	NA	transposase
SRB__G4242	type II toxin-antitoxin system VapC family toxin	2	type II toxin-antitoxin system VapC family toxin	PIN domain nuclease, a component of toxin-antitoxin system (PIN domain)	NA
SRB__G4238	HigA family addiction module antitoxin	2	HigA family addiction module antidote protein	Plasmid maintenance system antidote protein VapI, contains XRE-type HTH domain (VapI) (PDB:4MCT)	antitoxin HigA-1
SRB__G4208	translocation/assembly module TamB domain-containing protein	2	translocation/assembly module TamB domain-containing protein	Autotransporter translocation and assembly protein TamB (TamB) (PDB:5VTG)	translocation and assembly module TamB

SRB__G4204	GTP 3',8-cyclase MoaA	2	GTP 3',8-cyclase MoaA (moaA)	GTP 3',8-cyclase (molybdenum cofactor biosynthesis protein MoaA) (MoaA) (PDB:1TV7)	GTP 3',8-cyclase [EC:4.1.99.22]
SRB__G4196	histone deacetylase family protein	2	NA	NA	NA
SRB__G4170	VOC family protein	2	VOC family protein	Catechol 2,3-dioxygenase or related enzyme, vicinal oxygen chelate (VOC) family (GloA) (PDB:3R4Q) (PUBMED:11830595)	lactoylglutathione lyase [EC:4.4.1.5]
SRB__G4153	5' nucleotidase, NT5C type	2	5' nucleotidase, NT5C type	Uncharacterized conserved protein YqfW, HAD superfamily (YqfW)	uncharacterized protein
SRB__G4123	GTP-binding protein	2	elongation factor 4 (lepA)	Translation elongation factor EF-4, membrane-bound GTPase (LepA) (PDB:3DEG)	GTP-binding protein LepA
SRB__G4119	type II toxin-antitoxin system RelE/ParE family toxin	2	type II toxin-antitoxin system RelE/ParE family toxin	Uncharacterized conserved protein	NA
SRB__G4065	hypothetical protein	2	NA	NA	NA
SRB__G4049	FAD-dependent oxidoreductase	2	hydrogenase iron-sulfur subunit	Heterodisulfide reductase, subunit A (polyferredoxin) (HdrA) (PDB:5ODC)!!!Coenzyme F420-reducing hydrogenase, delta subunit (FrhD) (PDB:5ODC)	quinone-modifying oxidoreductase, subunit QmoB
SRB__G4028	ElyC/SanA/YdcF family protein	2	ElyC/SanA/YdcF family protein	NA	NA
SRB__G4023	DUF4350 domain-containing protein	2	DUF4350 domain-containing protein	NA	NA
SRB__G3869	transposase	2	transposase	NA	NA
SRB__G3591	hypothetical protein	2	NA	NA	NA
SRB__G3590	hypothetical protein	2	NA	NA	NA
SRB__G3584	hypothetical protein	2	NA	NA	NA
SRB__G3561	hypothetical protein	2	NA	NA	NA
SRB__G3546	hypothetical protein	2	NA	NA	NA
SRB__G3529	hypothetical protein	2	NA	NA	NA
SRB__G3519	hypothetical protein	2	NA	NA	NA
SRB__G3509	hypothetical protein	2	NA	ADP-ribose pyrophosphatase YjhB, NUDIX family (YjhB) (PDB:3GZ5)	NA
SRB__G3480	hypothetical protein	2	NA	NA	NA
SRB__G3452	hypothetical protein	2	NA	NA	NA
SRB__G3398	hypothetical protein	2	NA	NA	NA
SRB__G3359	type II toxin-antitoxin system antitoxin VapB	2	antitoxin	Virulence-associated protein VagC (function unknown) (VagC)	antitoxin VapB
SRB__G3351	hypothetical protein	2	NA	NA	NA
SRB__G3328	type II toxin-antitoxin system BrnA family antitoxin	2	CopG family transcriptional regulator	NA	NA
SRB__G3239	DUF2442 domain-containing protein	2	DUF2442 domain-containing protein	NA	NA
SRB__G3196	hypothetical protein	2	NA	NA	NA
SRB__G3107	type II toxin-antitoxin system RelE/ParE family toxin	2	type II toxin-antitoxin system RelE/ParE family toxin	Plasmid maintenance system killer protein (HigB)	toxin HigB-1
SRB__G3091	hypothetical protein	2	NA	NA	NA
SRB__G3051	hypothetical protein	2	NA	NA	NA
SRB__G3049	hypothetical protein	2	NA	Methionine synthase II (cobalamin-independent) (MetE) (PDB:1T7L)	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase [EC:2.1.1.14]
SRB__G2973	hypothetical protein	2	NA	NA	NA
SRB__G2960	hydrogenase maturation nickel metallochaperone HypA	2	hydrogenase maturation nickel metallochaperone HypA	Hydrogenase maturation factor HypA/HybF, metallochaperone involved in Ni insertion (HybF) (PDB:3A43) (PUBMED:16817932)	hydrogenase nickel incorporation protein HypA/HybF
SRB__G2955	ribonuclease P protein component	2	ribonuclease P protein component (rnpA)	RNase P protein component (RnpA) (PDB:1A6F)	ribonuclease P protein component [EC:3.1.26.5]
SRB__G2952	hypothetical protein	2	NA	NA	NA
SRB__G2899	hypothetical protein	2	NA	NA	NA
SRB__G2863	PH domain-containing protein	2	PH domain-containing protein	Uncharacterized membrane protein YdbT, contains bPH2 (bacterial pleckstrin homology) domain (YdbT)	NA
SRB__G2855	response regulator	2	response regulator	DNA-binding transcriptional response regulator, NtrC family, contains REC, AAA-type ATPase, and a Fis-type DNA-binding domains (AtoC) (PDB:1ZY2)	NA
SRB__G2839	redoxin domain-containing protein	2	redoxin domain-containing protein	Peroxiredoxin (Bcp) (PDB:5IPH)	NA
SRB__G2821	PilZ domain-containing protein	2	PilZ domain-containing protein	NA	NA
SRB__G2813	OsmC family protein	2	OsmC family protein	Uncharacterized OsmC-related protein (YhfA) (PDB:1LQL)	putative redox protein
SRB__G2766	acyl-CoA thioesterase	2	thioesterase family protein	Acyl-CoA thioesterase FadM (FadM) (PDB:2FUJ)	acyl-CoA thioester hydrolase [EC:3.1.2.]
SRB__G2757	PaaI family thioesterase	2	PaaI family thioesterase	Acyl-CoA thioesterase PaaI, contains HGG motif (PaaI) (PDB:1PSU)	NA
SRB__G2743	MaoC family dehydratase	2	MaoC family dehydratase	Acyl-CoA dehydratase PaaZ (MaoC) (PDB:5ZY8)	3-hydroxybutyryl-CoA dehydratase [EC:4.2.1.55]
SRB__G2694	MOSC domain-containing protein	2	MOSC domain-containing protein	N-hydroxylaminopurine reductase YiiM, contains MOSC domain (YiiM) (PDB:1O65) (PUBMED:18312271;29459651)	NA
SRB__G2648	YtoQ family protein	2	YtoQ family protein	Nucleoside 2-deoxyribosyltransferase (RCL) (PDB:1F8X)	NA
SRB__G2646	porin	2	porin	NA	NA

SRB__G2616	MarR family winged helix-turn-helix transcriptional regulator	2	MarR family transcriptional regulator	DNA-binding transcriptional regulator, MarR family (MarR) (PDB:1JGS)	MarR family transcriptional regulator, 2-MHQ and catechol-resistance regulon repressor
SRB__G2559	rhodanese-like domain-containing protein	2	rhodanese-like domain-containing protein	Molybdopterin or thiamine biosynthesis adenylyltransferase (ThiF) (PDB:1ZUD) (PUBMED:32239579)!!!Rhodanese-related sulfurtransferase (PspE) (PDB:1TQ1)	phage shock protein E
SRB__G2557	transcription elongation factor GreA	2	transcription elongation factor GreA (greA)	Transcription elongation factor, GreA/GreB family (GreA) (PDB:1GRJ)	transcription elongation factor GreA
SRB__G2529	thiol peroxidase	2	thiol peroxidase (tpx)	Peroxiredoxin (Tpx) (PDB:4AF2)	thioredoxin-dependent peroxiredoxin [EC:1.11.1.24]
SRB__G2522	dihydrofolate reductase	2	dihydrofolate reductase	Dihydrofolate reductase (FolA) (PDB:1DDR)	dihydrofolate reductase [EC:1.5.1.3]
SRB__G2440	TRAP transporter small permease	2	TRAP transporter small permease subunit	TRAP-type C4-dicarboxylate transport system, small permease component YiaM (DctM)	TRAP-type transport system small permease protein
SRB__G2372	flavodoxin family protein	2	NAD(P)H-dependent oxidoreductase	Multimeric flavodoxin WrbA, includes NAD(P)H:quinone oxidoreductase (WrbA) (PDB:4DY4)	NAD(P)H dehydrogenase (quinone) [EC:1.6.5.2]
SRB__G2360	HDIG domain-containing metalloprotein	2	HDIG domain-containing protein	Predicted hydrolase, HD superfamily	uncharacterized protein
SRB__G2317	LPS export ABC transporter periplasmic protein LptC	2	LPS export ABC transporter periplasmic protein LptC (lptC)	NA	lipopolysaccharide export system protein LptC
SRB__G2239	plasmid pRiA4b ORF-3 family protein	2	plasmid pRiA4b ORF-3 family protein	NA	NA
SRB__G2175	flavodoxin family protein	2	flavodoxin family protein	Multimeric flavodoxin WrbA, includes NAD(P)H:quinone oxidoreductase (WrbA) (PDB:4DY4)	NA
SRB__G2130	HAD family hydrolase	2	HAD family hydrolase	Phosphoglycolate phosphatase, HAD superfamily (Gph) (PDB:2NYV)	phosphoglycolate phosphatase [EC:3.1.3.18]
SRB__G2084	endonuclease III domain-containing protein	2	endonuclease III	Endonuclease III (Nth) (PDB:4UOB)!!!Exonuclease III (XthA) (PDB:1AKO)	endonuclease III [EC:3.2.2.- 4.2.99.18]
SRB__G2071	phosphonate C-P lyase system protein PhnL	2	phosphonate C-P lyase system protein PhnL (phnL)	Alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnL (PhnL)	alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnL [EC:2.7.8.37]
SRB__G2069	(d)CMP kinase	2	(d)CMP kinase	Cytidylate kinase (Cmk) (PDB:3R20) (PUBMED:19111750;25691594;28928205)	CMP/dCMP kinase [EC:2.7.4.25]
SRB__G2055	alpha/beta hydrolase	2	alpha/beta hydrolase	Dienelactone hydrolase (DLH) (PDB:4ZV9)	NA
SRB__G2009	AzlC family ABC transporter permease	2	AzlC family ABC transporter permease	Predicted branched-chain amino acid permease (azaleucine resistance) (AzlC)	branched chain amino acid efflux pump
SRB__G1964	PmeII family type II restriction endonuclease	2	PmeII family type II restriction endonuclease	DNA modification methylase (YhdJ) (PDB:1BOO)	NA
SRB__G1907	cytochrome c biogenesis CcdA family protein	2	cytochrome c biogenesis CcdA family protein	Cytochrome c biogenesis protein CcdA (CcdA) (PDB:5VKV)	cytochrome c-type biogenesis protein
SRB__G1882	DUF1883 domain-containing protein	2	DUF1883 domain-containing protein	Chromosomal replication initiation ATPase DnaA (DnaA) (PDB:1L8Q)	NA
SRB__G1836	lysophospholipid acyltransferase family protein	2	1-acyl-sn-glycerol-3-phosphate acyltransferase	1-acyl-sn-glycerol-3-phosphate acyltransferase (PlsC) (PDB:1IUQ)	1-acyl-sn-glycerol-3-phosphate acyltransferase [EC:2.3.1.51]
SRB__G1711	tetratricopeptide repeat protein	2	tetratricopeptide repeat protein	NA	NA
SRB__G1685	isochorismatase family protein	2	isochorismatase family protein	Nicotinamidase-related amidase (PncA) (PDB:1ILW)	NA
SRB__G1672	hypothetical protein	2	NA	NA	NA
SRB__G1472	GDP-L-fucose synthase family protein	2	GDP-L-fucose synthase	Nucleoside-diphosphate-sugar epimerase (WcaG) (PDB:2X4G)	GDP-L-fucose synthase [EC:1.1.1.271]
SRB__G1439	DMT family transporter	2	DMT family transporter	Permease of the drug/metabolite transporter (DMT) superfamily (RhaT) (PDB:5I20)	NA
SRB__G1432	transposase	2	transposase	NA	NA
SRB__G1421	tryptophan--tRNA ligase	2	tryptophan--tRNA ligase (trpS)	Tryptophanyl-tRNA synthetase (TrpS) (PDB:3N9I)	tryptophanyl-tRNA synthetase [EC:6.1.1.2]
SRB__G1410	tripartite tricarboxylate transporter substrate binding protein	2	tripartite tricarboxylate transporter substrate binding protein	Tripartite-type tricarboxylate transporter, extracytoplasmic receptor component TctC (TctC) (PDB:2DVZ)	putative tricarboxylic transport membrane protein
SRB__G1378	ATP-grasp domain-containing protein	2	ATP-grasp domain-containing protein	D-alanine-D-alanine ligase or related ATP-grasp enzyme (DdlA) (PDB:1E4E)	D-alanine-D-alanine ligase [EC:6.3.2.4]
SRB__G1356	MDR family oxidoreductase	2	oxidoreductase	NADPH:quinone reductase or related Zn-dependent oxidoreductase (Qor) (PDB:2VCY)	acrylyl-CoA reductase (NADPH) [EC:1.3.1.-]
SRB__G1279	TAXI family TRAP transporter solute-binding subunit	2	TAXI family TRAP transporter solute-binding subunit	TRAP-type uncharacterized transport system, periplasmic component (Imp) (PDB:1US4)	uncharacterized protein
SRB__G1169	class III poly(R)-hydroxyalkanoic acid synthase subunit PhaC	2	class III poly(R)-hydroxyalkanoic acid synthase subunit PhaC (phaC)	Poly-beta-hydroxybutyrate synthase (PhaC) (PDB:5XAV)	poly[(R)-3-hydroxyalkanoate] polymerase subunit PhaC [EC:2.3.1.304]
SRB__G1152	6-phosphofructokinase	2	6-phosphofructokinase	6-phosphofructokinase (PfkA) (PDB:1KZH)	ATP-dependent phosphofructokinase / diphosphate-dependent phosphofructokinase [EC:2.7.1.11 2.7.1.90]
SRB__G1107	histidinol-phosphate transaminase	2	histidinol-phosphate transaminase	Histidinol-phosphate/aromatic aminotransferase or cobyric acid decarboxylase (HisC) (PDB:4R2N)	histidinol-phosphate aminotransferase [EC:2.6.1.9]
SRB__G1070	dissimilatory-type sulfite reductase subunit beta	2	dissimilatory-type sulfite reductase subunit beta (dsrB)	Dissimilatory sulfite reductase (desulfoviridin), alpha and beta subunits (DsrA)	dissimilatory sulfite reductase beta subunit [EC:1.8.99.5]
SRB__G1022	lipid-A-disaccharide synthase	2	lipid-A-disaccharide synthase (lpxB)	Lipid A disaccharide synthetase (LpxB) (PDB:5W8N)	lipid-A-disaccharide synthase [EC:2.4.1.182]

SRB__G940	30S ribosomal protein S1	2	30S ribosomal protein S1 (rpsA)	Ribosomal protein S1 (RpsA) (PDB:4NNH)	small subunit ribosomal protein S1
SRB__G919	chemotaxis protein CheB	2	response regulator	Chemotaxis response regulator CheB, contains REC and protein-glutamate methylesterase domains (CheB) (PDB:1A2O)	two-component system, chemotaxis family, protein-glutamate methylesterase/glutaminase [EC:3.1.1.61 3.5.1.44]
SRB__G913	MFS transporter	2	YbfB/YjiJ family MFS transporter	Sugar phosphate permease (UhpC) (PDB:1PW4)	MFS transporter, CP family, cyanate transporter
SRB__G905	LL-diaminopimelate aminotransferase	2	LL-diaminopimelate aminotransferase	Aspartate/methionine/tyrosine aminotransferase (AspB) (PDB:2O0R)	LL-diaminopimelate aminotransferase [EC:2.6.1.83]
SRB__G899	tyrosine--tRNA ligase	2	tyrosine--tRNA ligase	Tyrosyl-tRNA synthetase (TyrS) (PDB:4OUD)	tyrosyl-tRNA synthetase [EC:6.1.1.1]
SRB__G803	phosphopyruvate hydratase	2	phosphopyruvate hydratase (eno)	Enolase (Eno) (PDB:1E9I)	enolase [EC:4.2.1.11]
SRB__G744	TRAP transporter large permease	2	TRAP transporter large permease	TRAP-type C4-dicarboxylate transport system, large permease component (DctQ)	TRAP-type transport system large permease protein
SRB__G590	GTPase/DUF3482 domain-containing protein	2	GTPase/DUF3482 domain-containing protein	GTPase Era, involved in 16S rRNA processing (Era) (PDB:1EGA)	NA
SRB__G524	tripartite tricarboxylate transporter permease	2	tripartite tricarboxylate transporter permease	TctA family transporter (PUBMED:14499931)	putative tricarboxylic transport membrane protein
SRB__G488	leucyl aminopeptidase	2	leucyl aminopeptidase	Leucyl aminopeptidase (PepB) (PDB:3H8E) (PUBMED:19553197;23825549)	leucyl aminopeptidase [EC:3.4.11.1]
SRB__G458	uroporphyrinogen-III C-methyltransferase	2	uroporphyrinogen-III C-methyltransferase (cobA)	Uroporphyrinogen-III methylase (siroheme synthase) (CysG) (PDB:1PJQ)!!!Uroporphyrinogen-III synthase (HemD) (PDB:1JR2)	uroporphyrinogen-III synthase [EC:4.2.1.75]
SRB__G446	acyl CoA:acetate/3-ketoacid CoA transferase	2	acyl CoA:acetate/3-ketoacid CoA transferase	Acyl CoA:acetate/3-ketoacid CoA transferase (YdiF) (PDB:5VIT)	propionate CoA-transferase [EC:2.8.3.1]
SRB__G428	IS66 family transposase	2	IS66 family transposase	Transposase	NA
SRB__G350	lipid A export permease/ATP-binding protein MsbA	2	lipid A export permease/ATP-binding protein MsbA (msbA)	ABC-type multidrug transport system, ATPase and permease component (MdlB) (PDB:2HYD)	ATP-binding cassette, subfamily B, bacterial
SRB__G326	phosphoribosylamine--glycine ligase	2	phosphoribosylamine--glycine ligase (purD)	Phosphoribosylcarboxyaminoimidazole (NCAIR) mutase (PurE) (PDB:1D7A)!!!Phosphoribosylamine-glycine ligase (PurD) (PDB:1GSO)	5-(carboxyamino)imidazole ribonucleotide mutase [EC:5.4.99.18]
SRB__G310	ABC transporter ATP-binding protein	2	ABC transporter ATP-binding protein	ABC-type multidrug transport system, ATPase and permease component (MdlB) (PDB:2HYD)	ATP-binding cassette, subfamily B, multidrug efflux pump
SRB__G279	transketolase C-terminal domain-containing protein	2	transketolase	Transketolase, C-terminal subunit (TktA2)!!!Transketolase, N-terminal subunit (TktA1)	NA
SRB__G205	glycogen debranching protein	2	glycogen-debranching protein	Pullulanase/glycogen debranching enzyme (PulA) (PDB:2WSK)	isoamylase [EC:3.2.1.68]
SRB__G164	mechanosensitive ion channel family protein	2	mechanosensitive ion channel family protein	Small-conductance mechanosensitive channel (MscS) (PDB:2OAU)	moderate conductance mechanosensitive channel
SRB__G163	type IV-A pilus assembly ATPase PilB	2	type IV-A pilus assembly ATPase PilB (pilB)	cAMP-binding domain of CRP or a regulatory subunit of cAMP-dependent protein kinases (Crp) (PDB:1FT9)!!!Type II secretory pathway ATPase GspE/PulE or T4P pilus assembly pathway ATPase PilB (PulE) (PDB:4KSR)	type IV pilus assembly protein PilB
SRB__G143	cysteine--tRNA ligase	2	cysteine--tRNA ligase	Cysteine synthase (CysK) (PDB:1FCJ)!!!Cysteinyl-tRNA synthetase (CysS) (PDB:1LI5)	cysteinyl-tRNA synthetase [EC:6.1.1.16]
SRB__G60	DUF294 nucleotidyltransferase-like domain-containing protein	2	cache domain-containing protein	Signal-transduction protein containing cAMP-binding, CBS, and nucleotidyltransferase domains	CBS domain-containing protein
SRB__G59	L-glutamate gamma-semialdehyde dehydrogenase	2	L-glutamate gamma-semialdehyde dehydrogenase (pruA)	Proline dehydrogenase (PutA) (PDB:5KF6)!!!Acyl-CoA reductase or other NAD-dependent aldehyde dehydrogenase (AdhE) (PDB:1A4S)	RHH-type transcriptional regulator, proline utilization regulon repressor / proline dehydrogenase / delta 1-pyrroline-5-carboxylate dehydrogenase [EC:1.5.5.2 1.2.1.88]
SRB__G4361	glycosyltransferase family 4 protein	2	glycosyltransferase family 1 protein	Glycosyltransferase involved in cell wall bisynthesis (RfaB) (PDB:2IV7)	phosphatidylinositol alpha 1,6-mannosyltransferase [EC:2.4.1.-]
SRB__G4355	sucrose synthase	2	sucrose synthase	Glycosyltransferase involved in cell wall bisynthesis (RfaB) (PDB:2IV7)	sucrose synthase [EC:2.4.1.13]
SRB__G4298	DNA integrity scanning protein DisA nucleotide-binding domain protein	2	DNA integrity scanning protein DisA nucleotide-binding domain protein	NA	NA
SRB__G3650	hypothetical protein	2	NA	NA	NA
SRB__G3642	hypothetical protein	2	NA	NA	NA
SRB__G3604	hypothetical protein	2	NA	NA	NA
SRB__G3578	hypothetical protein	2	NA	NA	NA
SRB__G3314	hypothetical protein	2	NA	NA	NA
SRB__G2650	hypothetical protein	2	NA	NA	NA
SRB__G2592	hypothetical protein	2	NA	NA	NA
SRB__G2531	nucleoside deaminase	2	nucleoside deaminase	tRNA(Arg) A34 adenosine deaminase TadA (TadA) (PDB:2A8N)	NA
SRB__G2336	CRISPR-associated protein Cas4	2	CRISPR-associated protein Cas4 (cas4)	CRISPR/Cas system-associated exonuclease Cas4, RecB family (Cas4) (PDB:4R5Q)	CRISPR-associated exonuclease Cas4 [EC:3.1.12.1]
SRB__G2007	FadR/GntR family transcriptional regulator	2	FadR/GntR family transcriptional regulator	DNA-binding transcriptional regulator, FadR family (FadR) (PDB:1E2X)	GntR family transcriptional regulator, transcriptional repressor for pyruvate dehydrogenase complex
SRB__G1794	indole-3-glycerol phosphate synthase TrpC	2	indole-3-glycerol phosphate synthase TrpC (trpC)	Indole-3-glycerol phosphate synthase (TrpC) (PDB:6BMA)	indole-3-glycerol phosphate synthase [EC:4.1.1.48]

SRB__G1767	zinc-dependent peptidase	2	zinc-dependent peptidase	Mlc titration factor MtfA, regulates ptsG expression (MtfA) (PDB:3DL1) (PUBMED:16855233)	MtfA peptidase
SRB__G1757	UDP-2,3-diacylglucosamine diphosphatase	2	UDP-2,3-diacylglucosamine diphosphatase	UDP-2,3-diacylglucosamine pyrophosphatase LpxH (LpxH) (PDB:5WLY)	UDP-2,3-diacylglucosamine hydrolase [EC:3.6.1.54]
SRB__G1605	phosphoribosylaminoimidazolesuccinocarboxamide synthase	2	phosphoribosylaminoimidazolesuccinocarboxamide synthase	Phosphoribosylaminoimidazole-succinocarboxamide synthase (PurC) (PDB:1A48)	phosphoribosylaminoimidazole-succinocarboxamide synthase [EC:6.3.2.6]
SRB__G1514	GTPase Era	2	GTPase Era (era)	GTPase Era, involved in 16S rRNA processing (Era) (PDB:1EGA)	GTPase
SRB__G1269	radical SAM protein	2	radical SAM protein	Radical SAM superfamily maturase, SkfB/NifB/PqqE family (SkfB) (PDB:6Q2P) (PUBMED:23282011)	NA
SRB__G923	hypothetical protein	2	NA	NA	NA
SRB__G886	hypothetical protein	2	NA	DNA-directed RNA polymerase, subunit RPC12/RpoP, contains C4-type Zn-finger (RPC10) (PDB:1I3Q)	NA
SRB__G208	glycine--tRNA ligase subunit beta	2	glycine--tRNA ligase subunit beta	Glycyl-tRNA synthetase, beta subunit (GlyS)	glycyl-tRNA synthetase beta chain [EC:6.1.1.14]
SRB__G4641	hypothetical protein	2	NA	NA	NA
SRB__G4595	hypothetical protein	2	NA	NA	NA
SRB__G4542	RHS repeat-associated core domain-containing protein	2	RHS repeat-associated core domain-containing protein	Uncharacterized conserved protein RhaS, contains 28 RHS repeats (RhsA) (PDB:4IGL)	NA
SRB__G3732	type II toxin-antitoxin system HicB family antitoxin	2	type II toxin-antitoxin system HicB family antitoxin	Antitoxin component HicB of the HicAB toxin-antitoxin system (HicB) (PDB:4P78) (PUBMED:19060138)	antitoxin HicB
SRB__G3703	hypothetical protein	2	NA	NA	NA
SRB__G3702	DUF488 family protein	2	DUF488 family protein	Uncharacterized conserved protein, DUF488 family	NA
SRB__G3701	hypothetical protein	2	NA	NA	NA
SRB__G3696	hypothetical protein	2	NA	NA	NA
SRB__G3626	transposase zinc-binding domain-containing protein	2	NA	NA	NA
SRB__G3430	hypothetical protein	2	NA	Transposase InsE and inactivated derivatives (InsE) (PDB:2JN6) (PUBMED:23818957)	NA
SRB__G3007	transposase	2	transposase	SRSO17 transposase (Srso17)	NA
SRB__G2991	hypothetical protein	2	NA	NA	NA
SRB__G2990	hypothetical protein	2	NA	NA	NA
SRB__G2988	DUF86 domain-containing protein	2	DUF86 domain-containing protein	HEPN domain protein, predicted toxin of MNT-HEPN system (HEPN) (PUBMED:26112399)	uncharacterized protein
SRB__G2324	Maf family nucleotide pyrophosphatase	2	Maf family nucleotide pyrophosphatase	7-methyl-GTP pyrophosphatase and related NTP pyrophosphatases, Maf/HAM1 superfamily (Maf) (PDB:2P5X) (PUBMED:24210219)	nucleoside triphosphate pyrophosphatase [EC:3.6.1.-]
SRB__G2223	hypothetical protein	2	NA	Retron-type reverse transcriptase (YkfC)	NA
SRB__G1941	RHS repeat-associated core domain-containing protein	2	RHS repeat-associated core domain-containing protein	Uncharacterized conserved protein RhaS, contains 28 RHS repeats (RhsA) (PDB:4IGL)	NA
SRB__G1418	glycosyltransferase	2	glycosyltransferase	Glycosyltransferase involved in cell wall bisynthesis (WcaA) (PDB:5MLZ)	NA
SRB__G218	toprim domain-containing protein	2	toprim domain-containing protein	DNA primase (bacterial type) (DnaG) (PDB:3B39) (PUBMED:28128549)	DNA primase [EC:2.7.7.101]
SRB__G33	RHS repeat-associated core domain-containing protein	2	RHS repeat-associated core domain-containing protein	Uncharacterized conserved protein RhaS, contains 28 RHS repeats (RhsA) (PDB:4IGL)	NA

Supplemental Table 6. PB-SRB1 clusters 1-2 genes, and their associated PGAP, COG20 and KOfam annotations

Gene	SRB_reference_genome	sample002	sample024	Gene	SRB_reference_genome	sample002	sample024	Gene	SRB_reference_genome	sample002	sample024
SRB__G4384	JKP90_18675	ACLI63_16370	ACLI64_07760	SRB__G1490	JKP90_01400	ACLI63_13150	ACLI64_15670	SRB__G96	JKP90_15845	NA	NA
SRB__G4353	JKP90_07425	ACLI63_08660	ACLI64_14710	SRB__G1449	JKP90_14720	ACLI63_03815	ACLI64_09545	SRB__G4393	JKP90_17655	ACLI63_14555	ACLI64_05925
SRB__G4340	JKP90_19040	ACLI63_16010	ACLI64_07400	SRB__G1446	JKP90_13430	ACLI63_07895	ACLI64_10960	SRB__G4386	JKP90_20145	ACLI63_14905	ACLI64_06300
SRB__G4333	JKP90_21170	ACLI63_12215	ACLI64_16605	SRB__G1411	JKP90_19100	ACLI63_15950	ACLI64_07340	SRB__G4296	JKP90_12350	ACLI63_06815	ACLI64_12045
SRB__G4294	JKP90_07735	ACLI63_13810	ACLI64_15020	SRB__G1398	JKP90_22785	ACLI63_00785	ACLI64_18100	SRB__G4259	JKP90_02755	ACLI63_10425	ACLI64_03480
SRB__G4257	JKP90_12360	ACLI63_06825	ACLI64_12035	SRB__G1354	JKP90_15750	ACLI63_17190	ACLI64_08535	SRB__G4242	JKP90_17660	ACLI63_14550	ACLI64_05920
SRB__G4215	JKP90_04475	ACLI63_01155	ACLI64_01770	SRB__G1333	JKP90_11770	ACLI63_12065	ACLI64_05075	SRB__G4238	JKP90_03380	ACLI63_09800	ACLI64_02865
SRB__G4210	JKP90_17875	ACLI63_14330	ACLI64_05710	SRB__G1298	JKP90_19415	ACLI63_15635	ACLI64_07020	SRB__G4208	JKP90_05235	ACLI63_01885	ACLI64_01025
SRB__G4184	JKP90_05030	ACLI63_01685	ACLI64_01225	SRB__G1284	JKP90_05425	ACLI63_02085	ACLI64_00830	SRB__G4204	JKP90_02690	ACLI63_10505	ACLI64_03555
SRB__G4172	JKP90_01915	ACLI63_11300	ACLI64_04310	SRB__G1275	JKP90_05285	ACLI63_01935	ACLI64_00975	SRB__G4196	JKP90_19185	ACLI63_15865	ACLI64_07250
SRB__G4147	JKP90_08300	ACLI63_05640	ACLI64_13230	SRB__G1245	JKP90_14540	ACLI63_03630	ACLI64_09730	SRB__G4170	JKP90_02885	ACLI63_10295	ACLI64_03355
SRB__G4115	JKP90_01540	ACLI63_13010	ACLI64_15805	SRB__G1242	JKP90_05835	ACLI63_02520	ACLI64_00425	SRB__G4153	JKP90_02785	ACLI63_10390	ACLI64_03450
SRB__G4101	JKP90_08615	ACLI63_05260	ACLI64_13530	SRB__G1237	JKP90_05460	ACLI63_02120	ACLI64_00795	SRB__G4123	JKP90_21715	ACLI63_09130	ACLI64_17050
SRB__G4090	JKP90_21895	ACLI63_09305	ACLI64_17225	SRB__G1227	JKP90_06180	ACLI63_02850	ACLI64_00060	SRB__G4119	JKP90_07530	ACLI63_14015	ACLI64_14815
SRB__G4085	JKP90_03000	ACLI63_10180	ACLI64_03240	SRB__G1206	JKP90_19400	ACLI63_15650	ACLI64_07035	SRB__G4065	JKP90_11865	ACLI63_12155	ACLI64_05160
SRB__G4082	JKP90_17880	ACLI63_14325	ACLI64_05705	SRB__G1201	JKP90_01460	ACLI63_13090	ACLI64_15730	SRB__G4049	JKP90_09215	ACLI63_04655	ACLI64_14120
SRB__G4054	JKP90_13520	ACLI63_07985	ACLI64_10870	SRB__G1187	JKP90_09655	ACLI63_08815	ACLI64_14535	SRB__G4028	JKP90_03540	ACLI63_09635	ACLI64_02705
SRB__G4052	JKP90_03470	ACLI63_09705	ACLI64_02775	SRB__G1184	JKP90_18785	ACLI63_16265	ACLI64_07655	SRB__G4023	JKP90_14780	ACLI63_03875	ACLI64_09485
SRB__G4047	JKP90_03130	ACLI63_10050	ACLI64_03110	SRB__G1182	JKP90_02505	ACLI63_10690	ACLI64_03740	SRB__G3869	JKP90_02365	ACLI63_10825	ACLI64_03875
SRB__G4041	JKP90_01700	ACLI63_12855	ACLI64_15965	SRB__G1159	JKP90_07575	ACLI63_13970	ACLI64_14860	SRB__G3591	JKP90_20270	ACLI63_14775	ACLI64_06175
SRB__G4015	JKP90_21355	ACLI63_12390	ACLI64_16415	SRB__G1158	JKP90_09590	ACLI63_08880	ACLI64_14470	SRB__G3590	JKP90_18140	ACLI63_14045	1_refound_64
SRB__G4000	JKP90_09560	ACLI63_08910	ACLI64_14440	SRB__G1115	JKP90_01180	ACLI63_13355	ACLI64_15460	SRB__G3584	JKP90_12365	ACLI63_06830	ACLI64_12030
SRB__G3976	JKP90_15740	ACLI63_17200	ACLI64_08545	SRB__G1095	JKP90_05430	ACLI63_02090	ACLI64_00825	SRB__G3561	2_refound_135	ACLI63_01200	1_refound_58
SRB__G3958	JKP90_02825	ACLI63_10350	ACLI64_03410	SRB__G1085	JKP90_14405	ACLI63_03495	ACLI64_09860	SRB__G3546	JKP90_04630	ACLI63_01315	ACLI64_01615
SRB__G3948	JKP90_12140	ACLI63_06605	ACLI64_12255	SRB__G1045	JKP90_17215	ACLI63_06270	ACLI64_12600	SRB__G3529	JKP90_12220	ACLI63_06685	ACLI64_12175
SRB__G3943	JKP90_06200	ACLI63_02870	ACLI64_00040	SRB__G1041	JKP90_19380	ACLI63_15670	ACLI64_07055	SRB__G3519	JKP90_04725	ACLI63_01385	ACLI64_01525
SRB__G3812	JKP90_15270	ACLI63_09795	ACLI64_08995	SRB__G1040	JKP90_21970	ACLI63_09380	ACLI64_17300	SRB__G3509	JKP90_02780	ACLI63_10395	ACLI64_03455
SRB__G3645	2_refound_153	ACLI63_09410	ACLI64_02480	SRB__G1029	JKP90_05060	ACLI63_01715	ACLI64_01195	SRB__G3480	2_refound_139	ACLI63_03595	1_refound_59
SRB__G3643	2_refound_144	ACLI63_06910	ACLI64_11950	SRB__G1028	JKP90_19405	ACLI63_15645	ACLI64_07030	SRB__G3452	JKP90_03530	ACLI63_09645	ACLI64_02715
SRB__G3628	JKP90_14875	ACLI63_03970	ACLI64_09390	SRB__G1021	JKP90_10800	ACLI63_03285	ACLI64_19285	SRB__G3398	JKP90_19020	ACLI63_16030	ACLI64_07420
SRB__G3620	JKP90_17355	0_refound_37	1_refound_100	SRB__G1020	2_refound_160	ACLI63_11320	ACLI64_04330	SRB__G3359	JKP90_17415	ACLI63_06495	ACLI64_12365
SRB__G3543	JKP90_11110	ACLI63_11455	ACLI64_04465	SRB__G1019	JKP90_07335	ACLI63_08750	ACLI64_14620	SRB__G3351	JKP90_03495	ACLI63_09680	ACLI64_02750
SRB__G3527	JKP90_10740	ACLI63_03225	ACLI64_19225	SRB__G1015	JKP90_03765	ACLI63_09415	ACLI64_02485	SRB__G3328	JKP90_13495	ACLI63_07960	ACLI64_10895
SRB__G3461	JKP90_22605	ACLI63_00605	ACLI64_17925	SRB__G1010	JKP90_17765	ACLI63_14435	ACLI64_05815	SRB__G3239	JKP90_08490	ACLI63_05470	ACLI64_13410
SRB__G3437	JKP90_12440	ACLI63_06905	ACLI64_11955	SRB__G1009	JKP90_10850	ACLI63_03335	ACLI64_19335	SRB__G3196	JKP90_12210	ACLI63_06675	ACLI64_12185
SRB__G3419	JKP90_15015	ACLI63_04110	ACLI64_09250	SRB__G989	JKP90_11890	ACLI63_12175	ACLI64_05180	SRB__G3107	JKP90_17090	ACLI63_06145	ACLI64_12725
SRB__G3407	JKP90_08935	ACLI63_04935	ACLI64_13840	SRB__G981	JKP90_03615	ACLI63_09570	ACLI64_02640	SRB__G3091	JKP90_10970	ACLI63_11335	ACLI64_04345
SRB__G3400	JKP90_01350	0_refound_11	ACLI64_15620	SRB__G980	JKP90_13465	ACLI63_07930	ACLI64_10925	SRB__G3051	JKP90_14390	ACLI63_03480	ACLI64_09875
SRB__G3379	JKP90_22675	ACLI63_00675	ACLI64_17990	SRB__G948	JKP90_11715	ACLI63_12010	ACLI64_05020	SRB__G3049	JKP90_01860	0_refound_29	1_refound_87
SRB__G3350	JKP90_03760	ACLI63_09420	ACLI64_02490	SRB__G927	JKP90_12925	ACLI63_07380	ACLI64_11475	SRB__G2973	JKP90_03545	0_refound_19	1_refound_76
SRB__G3338	JKP90_05080	ACLI63_01735	ACLI64_01175	SRB__G893	JKP90_12260	ACLI63_06725	ACLI64_12135	SRB__G2960	JKP90_14940	ACLI63_04035	ACLI64_09325
SRB__G3326	JKP90_07990	ACLI63_05950	ACLI64_12920	SRB__G885	JKP90_18950	ACLI63_16100	ACLI64_07490	SRB__G2955	JKP90_02650	ACLI63_10545	ACLI64_03595
SRB__G3281	JKP90_15730	ACLI63_17210	ACLI64_08555	SRB__G883	JKP90_02410	ACLI63_10785	ACLI64_03835	SRB__G2952	JKP90_21775	ACLI63_09190	ACLI64_17110
SRB__G3261	JKP90_08155	ACLI63_05785	ACLI64_13085	SRB__G882	JKP90_03050	ACLI63_10130	ACLI64_03190	SRB__G2899	JKP90_00970	ACLI63_13570	ACLI64_15245
SRB__G3210	JKP90_06005	ACLI63_02690	ACLI64_00240	SRB__G866	JKP90_05380	ACLI63_02040	ACLI64_00875	SRB__G2863	JKP90_14385	ACLI63_03475	ACLI64_09880
SRB__G3160	JKP90_15735	ACLI63_17205	ACLI64_08550	SRB__G864	JKP90_22280	ACLI63_00280	ACLI64_17600	SRB__G2855	JKP90_03080	ACLI63_10100	ACLI64_03160
SRB__G3106	JKP90_00625	ACLI63_03065	ACLI64_19070	SRB__G843	JKP90_14810	ACLI63_03905	ACLI64_09455	SRB__G2839	JKP90_08930	ACLI63_04940	ACLI64_13835
SRB__G3070	JKP90_15010	ACLI63_04105	ACLI64_09255	SRB__G838	JKP90_03135	ACLI63_10045	ACLI64_03105	SRB__G2821	JKP90_12175	ACLI63_06640	ACLI64_12220
SRB__G3068	2_refound_134	ACLI63_01050	ACLI64_01875	SRB__G835	JKP90_02730	ACLI63_10465	ACLI64_03515	SRB__G2813	JKP90_19905	ACLI63_15140	ACLI64_06535
SRB__G3053	JKP90_02395	ACLI63_10795	ACLI64_03845	SRB__G831	JKP90_07655	ACLI63_13890	ACLI64_14940	SRB__G2766	JKP90_03245	ACLI63_09935	ACLI64_02995
SRB__G3044	JKP90_11220	ACLI63_11565	ACLI64_04575	SRB__G784	JKP90_11290	ACLI63_11635	ACLI64_04645	SRB__G2757	JKP90_03550	ACLI63_09630	ACLI64_02700
SRB__G3023	JKP90_03315	ACLI63_09865	ACLI64_02925	SRB__G746	JKP90_11365	ACLI63_11705	ACLI64_04715	SRB__G2743	JKP90_11180	ACLI63_11525	ACLI64_04535
SRB__G3012	JKP90_08060	ACLI63_05880	ACLI64_12990	SRB__G736	JKP90_21390	ACLI63_12425	ACLI64_16380	SRB__G2694	JKP90_02695	ACLI63_10500	ACLI64_03550
SRB__G3004	JKP90_02675	ACLI63_10520	ACLI64_03570	SRB__G729	JKP90_14295	ACLI63_03385	ACLI64_09970	SRB__G2648	JKP90_02080	ACLI63_11100	ACLI64_04150
SRB__G2978	JKP90_18105	ACLI63_14080	ACLI64_05490	SRB__G713	JKP90_08535	ACLI63_05425	ACLI64_13455	SRB__G2646	JKP90_13190	ACLI63_07655	ACLI64_11200
SRB__G2976	JKP90_05865	ACLI63_02550	ACLI64_00395	SRB__G704	JKP90_14635	ACLI63_03725	ACLI64_09635	SRB__G2616	JKP90_17345	ACLI63_06425	ACLI64_12435
SRB__G2922	JKP90_08265	ACLI63_05675	ACLI64_13195	SRB__G678	JKP90_00980	ACLI63_13560	ACLI64_15255	SRB__G2559	JKP90_11380	ACLI63_11715	ACLI64_04725
SRB__G2820	JKP90_17170	ACLI63_06225	ACLI64_12645	SRB__G677	JKP90_01705	ACLI63_12850	ACLI64_15970	SRB__G2557	JKP90_21735	ACLI63_09150	ACLI64_17070
SRB__G2819	JKP90_17130	ACLI63_06185	ACLI64_12685	SRB__G662	JKP90_19395	ACLI63_15655	ACLI64_07040	SRB__G2529	JKP90_17755	ACLI63_14445	ACLI64_05825
SRB__G2793	JKP90_01795	ACLI63_12760	ACLI64_16060	SRB__G649	JKP90_08225	ACLI63_05715	ACLI64_13155	SRB__G2522	JKP90_14695	ACLI63_03785	ACLI64_09575
SRB__G2786	JKP90_03405	ACLI63_09770	ACLI64_02840	SRB__G635	JKP90_11385	ACLI63_11720	ACLI64_04730	SRB__G2440	JKP90_19320	ACLI63_15730	ACLI64_07115
SRB__G2783	JKP90_08115	ACLI63_05825	ACLI64_13045	SRB__G627	JKP90_13025	ACLI63_07490	ACLI64_11365	SRB__G2372	JKP90_04485	ACLI63_01165	ACLI64_01760
SRB__G2782	JKP90_14575	ACLI63_03665	ACLI64_09695	SRB__G614	JKP90_11490	ACLI63_11825	ACLI64_04835	SRB__G2360	JKP90_05280	ACLI63_01930	ACLI64_00980
SRB__G2777	JKP90_20060	ACLI63_14980	ACLI64_06375	SRB__G609	JKP90_03120	ACLI63_10060	ACLI64_03120	SRB__G2317	JKP90_18715	ACLI63_16330	ACLI64_07720
SRB__G2768	JKP90_14520	ACLI63_03610	ACLI64_09750	SRB__G598	JKP90_02030	ACLI63_11150	ACLI64_04200	SRB__G2239	JKP90_19900	ACLI63_15145	ACLI64_06540
SRB__G2765	JKP90_09660	ACLI63_08810	ACLI64_14540	SRB__G545	JKP90_14105	ACLI63_08600	ACLI64_10280	SRB__G2175	JKP90_08385	ACLI63_05560	ACLI64_13310
SRB__G2745	JKP90_17760	ACLI63_14440	ACLI64_05820	SRB__G536	JKP90_14775	ACLI63_03870	ACLI64_09490	SRB__G2130	JKP90_00975	ACLI63_13565	ACLI64_15250
SRB__G2741	JKP90_13535	ACLI63_08000	ACLI64_10855	SRB__G520	JKP90_00865	ACLI63_13670	ACLI64_15140	SRB__G2084	JKP90_02625	ACLI63_10570	ACLI64_03620
SRB__G2733	JKP90_18570	ACLI63_16480	ACLI64_07865	SRB__G505	JKP90_06285	ACLI63_02930	ACLI64_18935	SRB__G2071	JKP90_00565	ACLI63_03130	ACLI64_19130
SRB__G2732	JKP90_19175	ACLI63_15875	ACLI64_07260	SRB__G497	JKP90_02585	ACLI63_10610	ACLI64_03660	SRB__G2069	JKP90_18765	ACLI63_16285	ACLI64_07675
SRB__G2677	JKP90_01510	ACLI63_13040	ACLI64_15775	SRB__G478	JKP90_17780	ACLI63_14420	ACLI64_05800	SRB__G2055	JKP90_02330	ACLI63_10860	ACLI64_03910

SRB__G2654	JKP90_04360	ACLI63_01040	ACLI64_01885	SRB__G463	JKP90_18850	ACLI63_16200	ACLI64_07590	SRB__G2009	JKP90_07355	ACLI63_08730	ACLI64_14640
SRB__G2636	JKP90_09565	ACLI63_08905	ACLI64_14445	SRB__G462	JKP90_08905	ACLI63_04965	ACLI64_13810	SRB__G1964	JKP90_15030	ACLI63_04125	ACLI64_09235
SRB__G2626	JKP90_03125	ACLI63_10055	ACLI64_03115	SRB__G417	JKP90_08255	ACLI63_05685	ACLI64_13185	SRB__G1907	JKP90_18135	ACLI63_14050	ACLI64_05460
SRB__G2619	JKP90_04300	ACLI63_00980	ACLI64_01950	SRB__G405	JKP90_06130	ACLI63_02795	ACLI64_00115	SRB__G1882	JKP90_01235	ACLI63_13300	ACLI64_15515
SRB__G2551	JKP90_01090	ACLI63_13445	ACLI64_15365	SRB__G396	JKP90_17930	ACLI63_14275	ACLI64_05655	SRB__G1836	JKP90_18100	ACLI63_14085	ACLI64_05495
SRB__G2548	JKP90_12870	ACLI63_07325	ACLI64_11530	SRB__G391	JKP90_02085	ACLI63_11095	ACLI64_04145	SRB__G1711	JKP90_04540	ACLI63_01225	ACLI64_01705
SRB__G2533	JKP90_17180	ACLI63_06235	ACLI64_12635	SRB__G389	JKP90_02660	ACLI63_10535	ACLI64_03585	SRB__G1685	JKP90_01615	ACLI63_12940	ACLI64_15880
SRB__G2527	JKP90_11165	ACLI63_11510	ACLI64_04520	SRB__G369	JKP90_05190	ACLI63_01845	ACLI64_01065	SRB__G1672	JKP90_02515	ACLI63_10680	ACLI64_03730
SRB__G2518	JKP90_11810	ACLI63_12105	ACLI64_05115	SRB__G344	JKP90_02090	ACLI63_11090	ACLI64_04140	SRB__G1472	JKP90_19470	ACLI63_15575	ACLI64_06960
SRB__G2509	JKP90_00870	ACLI63_13665	ACLI64_15145	SRB__G325	JKP90_13395	ACLI63_07860	ACLI64_10995	SRB__G1439	JKP90_02990	ACLI63_10190	ACLI64_03250
SRB__G2472	JKP90_03100	ACLI63_10080	ACLI64_03140	SRB__G320	JKP90_08335	ACLI63_05605	ACLI64_13265	SRB__G1432	JKP90_00885	ACLI63_13650	ACLI64_15160
SRB__G2447	JKP90_13070	ACLI63_07535	ACLI64_11320	SRB__G311	JKP90_16160	ACLI63_16830	ACLI64_08195	SRB__G1421	JKP90_05275	ACLI63_01925	ACLI64_00985
SRB__G2445	JKP90_09075	ACLI63_04795	ACLI64_13980	SRB__G304	JKP90_04340	ACLI63_01020	ACLI64_01905	SRB__G1410	JKP90_21350	ACLI63_12385	ACLI64_16420
SRB__G2435	JKP90_14830	ACLI63_03925	ACLI64_09435	SRB__G299	JKP90_14885	ACLI63_03980	ACLI64_09380	SRB__G1378	JKP90_19195	ACLI63_15855	ACLI64_07240
SRB__G2430	JKP90_01135	ACLI63_13400	ACLI64_15415	SRB__G296	JKP90_19805	ACLI63_15240	ACLI64_06635	SRB__G1356	JKP90_07995	ACLI63_05945	ACLI64_12925
SRB__G2418	JKP90_08600	ACLI63_05275	ACLI64_13515	SRB__G287	JKP90_17865	ACLI63_14340	ACLI64_05720	SRB__G1279	JKP90_17695	ACLI63_14505	ACLI64_05885
SRB__G2411	JKP90_04890	ACLI63_01545	ACLI64_01365	SRB__G285	JKP90_05885	ACLI63_02570	ACLI64_00375	SRB__G1169	JKP90_11140	ACLI63_11485	ACLI64_04495
SRB__G2409	JKP90_19565	ACLI63_15480	ACLI64_06870	SRB__G276	JKP90_01820	ACLI63_12735	ACLI64_16085	SRB__G1152	JKP90_18045	ACLI63_14145	ACLI64_05545
SRB__G2406	JKP90_19305	ACLI63_15745	ACLI64_07130	SRB__G273	JKP90_16015	ACLI63_16980	ACLI64_08345	SRB__G1107	JKP90_18770	ACLI63_16280	ACLI64_07670
SRB__G2401	JKP90_13655	ACLI63_08120	ACLI64_10735	SRB__G268	JKP90_19675	ACLI63_15370	ACLI64_06765	SRB__G1070	JKP90_07985	ACLI63_05955	ACLI64_12915
SRB__G2388	JKP90_21510	ACLI63_12540	ACLI64_16265	SRB__G261	JKP90_21950	ACLI63_09360	ACLI64_17280	SRB__G1022	JKP90_22470	ACLI63_00470	ACLI64_17790
SRB__G2374	JKP90_12835	ACLI63_07290	ACLI64_11565	SRB__G243	JKP90_14990	ACLI63_04085	ACLI64_09275	SRB__G940	JKP90_01160	ACLI63_13375	ACLI64_15440
SRB__G2367	JKP90_11615	ACLI63_11910	ACLI64_04920	SRB__G211	2_refound_132	ACLI63_00655	ACLI64_17970	SRB__G919	JKP90_22730	ACLI63_00730	ACLI64_18045
SRB__G2345	JKP90_15810	ACLI63_17130	ACLI64_08475	SRB__G183	JKP90_11215	ACLI63_11560	ACLI64_04570	SRB__G913	JKP90_13310	ACLI63_07775	ACLI64_11080
SRB__G2335	JKP90_02495	ACLI63_10700	ACLI64_03750	SRB__G165	JKP90_19965	ACLI63_15080	ACLI64_06475	SRB__G905	JKP90_21935	ACLI63_09345	ACLI64_17265
SRB__G2315	JKP90_18985	ACLI63_16065	ACLI64_07455	SRB__G147	JKP90_22830	ACLI63_00830	ACLI64_18145	SRB__G899	JKP90_01115	ACLI63_13420	ACLI64_15390
SRB__G2303	JKP90_22175	ACLI63_00180	ACLI64_17500	SRB__G140	JKP90_14915	ACLI63_04010	ACLI64_09350	SRB__G803	JKP90_12215	ACLI63_06680	ACLI64_12180
SRB__G2293	JKP90_06165	ACLI63_02835	ACLI64_00075	SRB__G124	JKP90_18825	ACLI63_16225	ACLI64_07615	SRB__G744	JKP90_08885	ACLI63_05000	ACLI64_13790
SRB__G2287	JKP90_19630	ACLI63_15420	ACLI64_06810	SRB__G119	JKP90_03010	ACLI63_10170	ACLI64_03230	SRB__G590	JKP90_14435	ACLI63_03525	ACLI64_09830
SRB__G2281	JKP90_06135	ACLI63_02800	ACLI64_00110	SRB__G115	JKP90_18830	ACLI63_16220	ACLI64_07610	SRB__G524	JKP90_13590	ACLI63_08055	ACLI64_10800
SRB__G2221	JKP90_15815	ACLI63_17125	ACLI64_08470	SRB__G103	JKP90_01745	ACLI63_12810	ACLI64_16010	SRB__G488	JKP90_22610	ACLI63_00610	ACLI64_17930
SRB__G2211	JKP90_21385	ACLI63_12420	ACLI64_16385	SRB__G97	JKP90_08030	ACLI63_05910	ACLI64_12960	SRB__G458	2_refound_167	ACLI63_14910	ACLI64_06305
SRB__G2206	JKP90_19955	ACLI63_15090	ACLI64_06485	SRB__G89	JKP90_12970	ACLI63_07435	ACLI64_11420	SRB__G446	JKP90_19010	ACLI63_16040	ACLI64_07430
SRB__G2201	JKP90_04510	ACLI63_01190	ACLI64_01735	SRB__G74	JKP90_13810	ACLI63_08300	ACLI64_10580	SRB__G428	JKP90_09680	ACLI63_08790	ACLI64_14560
SRB__G2193	JKP90_15920	ACLI63_17095	ACLI64_08440	SRB__G46	JKP90_18955	ACLI63_16095	ACLI64_07485	SRB__G350	JKP90_00990	ACLI63_13550	ACLI64_15265
SRB__G2189	JKP90_02815	ACLI63_10360	ACLI64_03420	SRB__G39	JKP90_01620	ACLI63_12935	ACLI64_15885	SRB__G326	JKP90_21745	ACLI63_09160	ACLI64_17080
SRB__G2162	JKP90_21450	ACLI63_12485	ACLI64_16320	SRB__G28	JKP90_19280	ACLI63_15770	ACLI64_07155	SRB__G310	JKP90_21380	ACLI63_12415	ACLI64_16390
SRB__G2156	JKP90_08150	ACLI63_05790	ACLI64_13080	SRB__G4426	JKP90_21590	NA	ACLI64_16930	SRB__G279	JKP90_13320	ACLI63_07785	ACLI64_11070
SRB__G2153	JKP90_02245	ACLI63_10940	ACLI64_03990	SRB__G4407	JKP90_07125	NA	ACLI64_18465	SRB__G205	JKP90_02795	ACLI63_10380	ACLI64_03440
SRB__G2151	JKP90_07405	ACLI63_08680	ACLI64_14690	SRB__G4390	JKP90_17440	NA	ACLI64_05370	SRB__G164	JKP90_13085	ACLI63_07550	ACLI64_11305
SRB__G2150	JKP90_12530	ACLI63_06990	ACLI64_11870	SRB__G4385	JKP90_18160	NA	ACLI64_05440	SRB__G163	JKP90_22920	ACLI63_00920	ACLI64_18235
SRB__G2127	JKP90_13380	ACLI63_07845	ACLI64_11010	SRB__G4351	JKP90_21615	NA	ACLI64_16955	SRB__G143	JKP90_12100	ACLI63_06565	ACLI64_12295
SRB__G2120	JKP90_02955	ACLI63_10225	ACLI64_03285	SRB__G4279	JKP90_10365	NA	ACLI64_18655	SRB__G60	JKP90_04575	ACLI63_01260	ACLI64_01670
SRB__G2082	JKP90_09115	ACLI63_04755	ACLI64_14020	SRB__G3996	JKP90_15470	NA	ACLI64_08795	SRB__G59	JKP90_17925	ACLI63_14280	ACLI64_05660
SRB__G2081	JKP90_05500	ACLI63_02160	ACLI64_00755	SRB__G3962	JKP90_10710	NA	ACLI64_16825	SRB__G4361	JKP90_14265	NA	ACLI64_10000
SRB__G2080	JKP90_17830	ACLI63_14375	ACLI64_05755	SRB__G3515	JKP90_15105	NA	1_refound_106	SRB__G4355	JKP90_21600	NA	ACLI64_16940
SRB__G2067	JKP90_04305	ACLI63_00985	ACLI64_01945	SRB__G3189	JKP90_10555	NA	ACLI64_16685	SRB__G4298	JKP90_04150	NA	ACLI64_02090
SRB__G2060	JKP90_21855	ACLI63_09265	ACLI64_17185	SRB__G2875	JKP90_15220	NA	ACLI64_09045	SRB__G3650	NA	ACLI63_14360	ACLI64_05740
SRB__G2032	JKP90_12435	ACLI63_06900	ACLI64_11960	SRB__G2439	NA	ACLI63_12630	ACLI64_16180	SRB__G3642	2_refound_182	NA	ACLI64_08895
SRB__G2030	JKP90_19410	ACLI63_15640	ACLI64_07025	SRB__G2407	JKP90_15965	NA	ACLI64_08395	SRB__G3604	NA	ACLI63_11310	ACLI64_04320
SRB__G2005	JKP90_08025	ACLI63_05915	ACLI64_12955	SRB__G2302	JKP90_09980	NA	ACLI64_16905	SRB__G3578	JKP90_14200	NA	ACLI64_10180
SRB__G2001	JKP90_02095	ACLI63_11085	ACLI64_04135	SRB__G2266	JKP90_03775	NA	ACLI64_02470	SRB__G3314	JKP90_15620	NA	ACLI64_08645
SRB__G1988	JKP90_07325	ACLI63_08760	ACLI64_14610	SRB__G1961	JKP90_07135	NA	ACLI64_18475	SRB__G2650	JKP90_20835	NA	ACLI64_18725
SRB__G1974	JKP90_02890	ACLI63_10290	ACLI64_03350	SRB__G1854	JKP90_17595	NA	ACLI64_05990	SRB__G2592	JKP90_23360	NA	ACLI64_18835
SRB__G1966	JKP90_07415	ACLI63_08670	ACLI64_14700	SRB__G1662	JKP90_15455	NA	ACLI64_08810	SRB__G2531	JKP90_04060	NA	ACLI64_02180
SRB__G1954	JKP90_06160	ACLI63_02830	ACLI64_00080	SRB__G1641	JKP90_15600	NA	ACLI64_08665	SRB__G2336	NA	ACLI63_11350	ACLI64_04360
SRB__G1950	JKP90_22545	ACLI63_00545	ACLI64_17865	SRB__G1539	JKP90_07090	NA	ACLI64_18435	SRB__G2007	NA	ACLI63_11305	ACLI64_04315
SRB__G1947	JKP90_17735	ACLI63_14465	ACLI64_05845	SRB__G1292	JKP90_04250	NA	ACLI64_02000	SRB__G1794	JKP90_10635	NA	ACLI64_16765
SRB__G1938	JKP90_12145	ACLI63_06610	ACLI64_12250	SRB__G1166	JKP90_15460	NA	ACLI64_08805	SRB__G1767	JKP90_15095	NA	ACLI64_09170
SRB__G1934	JKP90_02215	ACLI63_10970	ACLI64_04020	SRB__G1075	JKP90_07210	NA	ACLI64_18550	SRB__G1757	JKP90_14270	NA	ACLI64_09995
SRB__G1898	JKP90_13425	ACLI63_07890	ACLI64_10965	SRB__G673	2_refound_194	NA	ACLI64_18555	SRB__G1605	2_refound_193	NA	ACLI64_18440
SRB__G1856	JKP90_17725	ACLI63_14475	ACLI64_05855	SRB__G610	NA	ACLI63_11315	ACLI64_04325	SRB__G1514	JKP90_04185	NA	ACLI64_02065
SRB__G1803	JKP90_13745	ACLI63_08235	ACLI64_10645	SRB__G204	JKP90_07440	NA	ACLI64_14725	SRB__G1269	JKP90_04065	NA	ACLI64_02175
SRB__G1778	JKP90_09530	ACLI63_08940	ACLI64_14410	SRB__G145	JKP90_20890	NA	ACLI64_18780	SRB__G923	JKP90_04175	NA	ACLI64_02075
SRB__G1755	JKP90_11440	ACLI63_11775	ACLI64_04785	SRB__G113	JKP90_16250	NA	ACLI64_08110	SRB__G886	JKP90_04055	NA	ACLI64_02185
SRB__G1750	JKP90_14375	ACLI63_03465	ACLI64_09890	SRB__G4736	JKP90_06625	NA	NA	SRB__G208	JKP90_04120	NA	ACLI64_02120
SRB__G1749	JKP90_02210	ACLI63_10975	ACLI64_04025	SRB__G4602	JKP90_23205	NA	NA	SRB__G4641	JKP90_20375	NA	NA
SRB__G1730	JKP90_11730	ACLI63_12025	ACLI64_05035	SRB__G4594	JKP90_23215	NA	NA	SRB__G4595	JKP90_23255	NA	NA
SRB__G1724	JKP90_01415	ACLI63_13135	ACLI64_15685	SRB__G4556	JKP90_10210	NA	NA	SRB__G4542	JKP90_16855	NA	NA
SRB__G1721	JKP90_04670	ACLI63_01330	ACLI64_01580	SRB__G4454	JKP90_16910	NA	NA	SRB__G3732	JKP90_20975	NA	NA
SRB__G1720	JKP90_19610	ACLI63_15440	ACLI64_06830	SRB__G4447	JKP90_16820	NA	NA	SRB__G3703	JKP90_20780	NA	NA
SRB__G1699	JKP90_05875	ACLI63_02560	ACLI64_00385	SRB__G3754	JKP90_21560	NA	NA	SRB__G3702	JKP90_20775	NA	NA
SRB__G1692	JKP90_03565	ACLI63_09615	ACLI64_02685	SRB__G3739	JKP90_21090	NA	NA	SRB__G3701	JKP90_20760	NA	NA

SRB__G1660	JKP90_17720	ACLI63_14480	ACLI64_05860	SRB__G3734	JKP90_21010	NA	NA	SRB__G3696	JKP90_00470	NA	NA
SRB__G1620	JKP90_13900	ACLI63_08390	ACLI64_10485	SRB__G3668	NA	ACLI63_09850	NA	SRB__G3626	JKP90_21065	NA	NA
SRB__G1584	JKP90_08845	ACLI63_05040	ACLI64_13750	SRB__G3332	JKP90_09955	NA	NA	SRB__G3430	JKP90_12045	NA	NA
SRB__G1579	JKP90_07555	ACLI63_13990	ACLI64_14840	SRB__G3150	NA	NA	ACLI64_18915	SRB__G3007	JKP90_00905	NA	NA
SRB__G1577	JKP90_13355	ACLI63_07820	ACLI64_11035	SRB__G2880	JKP90_21100	NA	NA	SRB__G2991	JKP90_18665	NA	NA
SRB__G1565	JKP90_03595	ACLI63_09590	ACLI64_02660	SRB__G2823	NA	ACLI63_11185	NA	SRB__G2990	JKP90_20685	NA	NA
SRB__G1562	JKP90_05785	ACLI63_02470	ACLI64_00475	SRB__G2464	NA	ACLI63_08135	NA	SRB__G2988	NA	NA	ACLI64_05325
SRB__G1549	JKP90_22875	ACLI63_00875	ACLI64_18190	SRB__G2322	JKP90_20715	NA	NA	SRB__G2324	NA	ACLI63_04210	NA
SRB__G1512	JKP90_08160	ACLI63_05780	ACLI64_13090	SRB__G873	NA	ACLI63_05350	NA	SRB__G2223	JKP90_15700	NA	NA
SRB__G1501	JKP90_13460	ACLI63_07925	ACLI64_10930	SRB__G152	NA	ACLI63_16715	NA	SRB__G1941	NA	ACLI63_13715	NA
SRB__G33	NA	ACLI63_13725	NA	SRB__G218	JKP90_07280	NA	NA	SRB__G1418	NA	NA	ACLI64_18345

Supplemental Table 7. PB-SRB1 gene ID's related to individual composite PB-SRB1 bins from metagenome database as well as our 2011 reference genome: 2011 PB-SRB1 reference genome (JAEQMT010000000.1), sample002 bin, sample024 bin

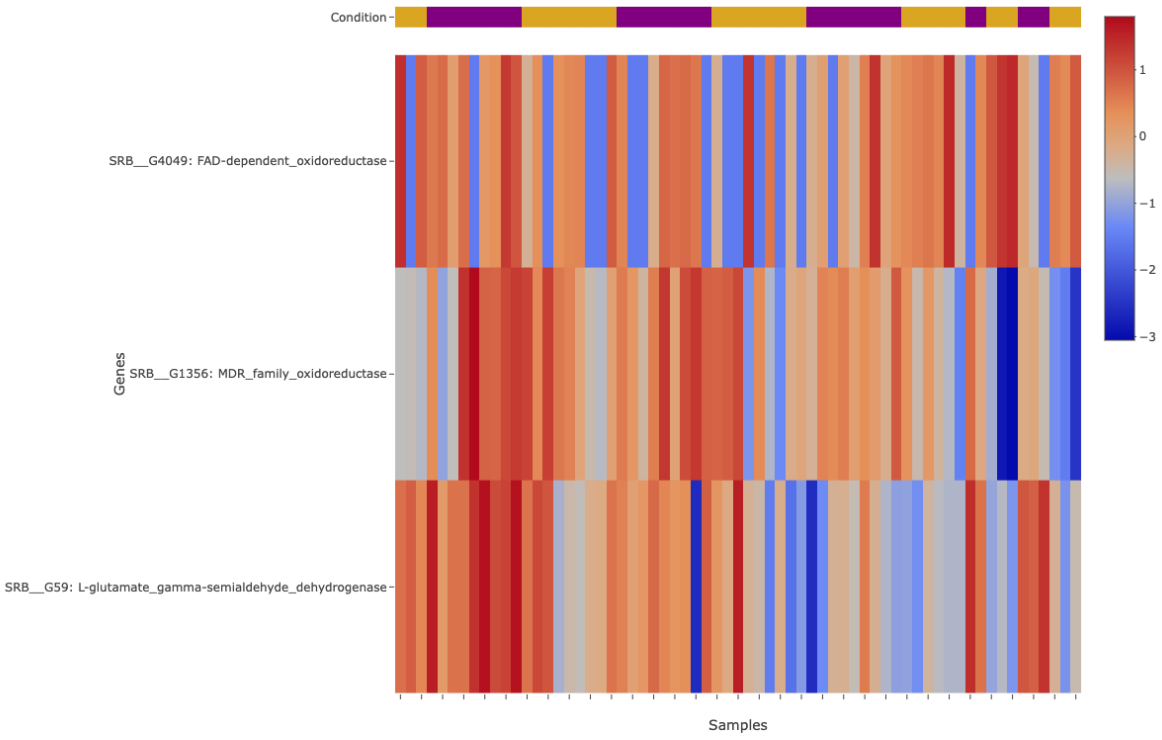


Figure 2-D. Dehydrogenase genes in PB-SRB1 that were significantly upregulated in dark conditions within some form of light-dark comparison. Shrinkage adjusted, row-scaled, and visualized with plot_ly. From left to right we see experiment progression with time, gold bars correspond to light time points and purple bars to dark time points.

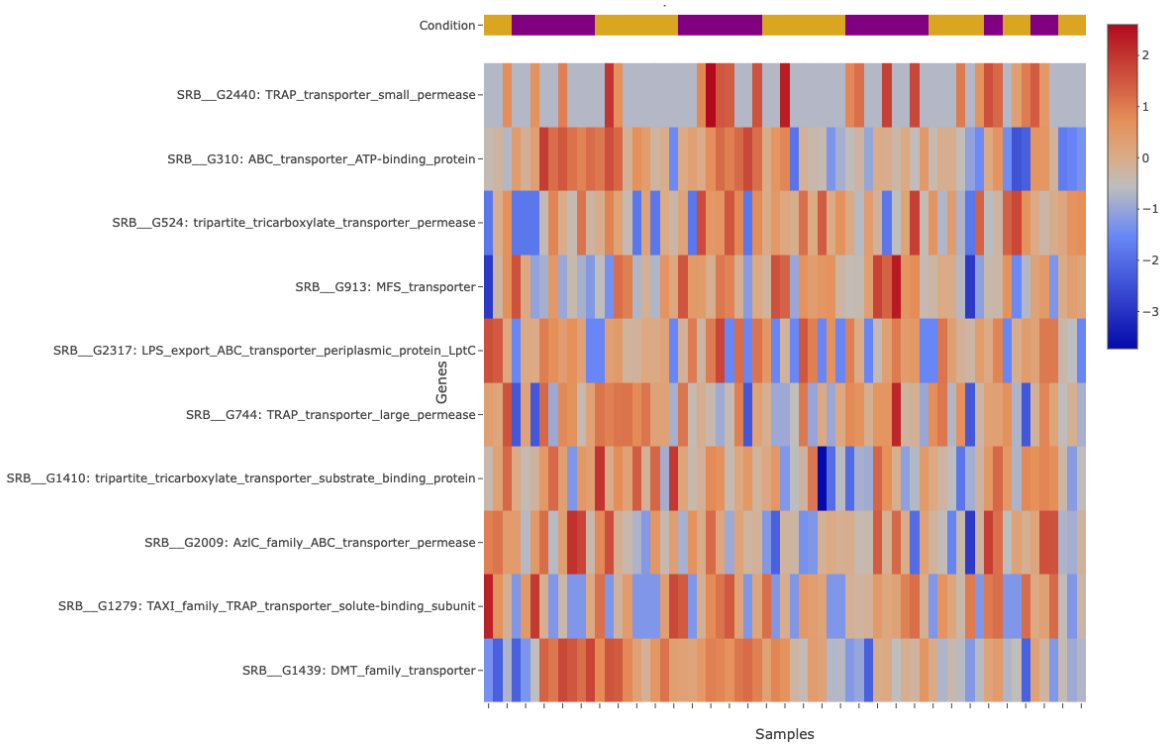


Figure 2-E. Transporter genes in PB-SRB1 that were significantly upregulated in dark conditions within some form of light-dark comparison. Shrinkage adjusted, row-scaled, and visualized with plot_ly. From left to right we see experiment progression with time, gold bars correspond to light time points and purple bars to dark time points.

Chapter III Supplemental Information

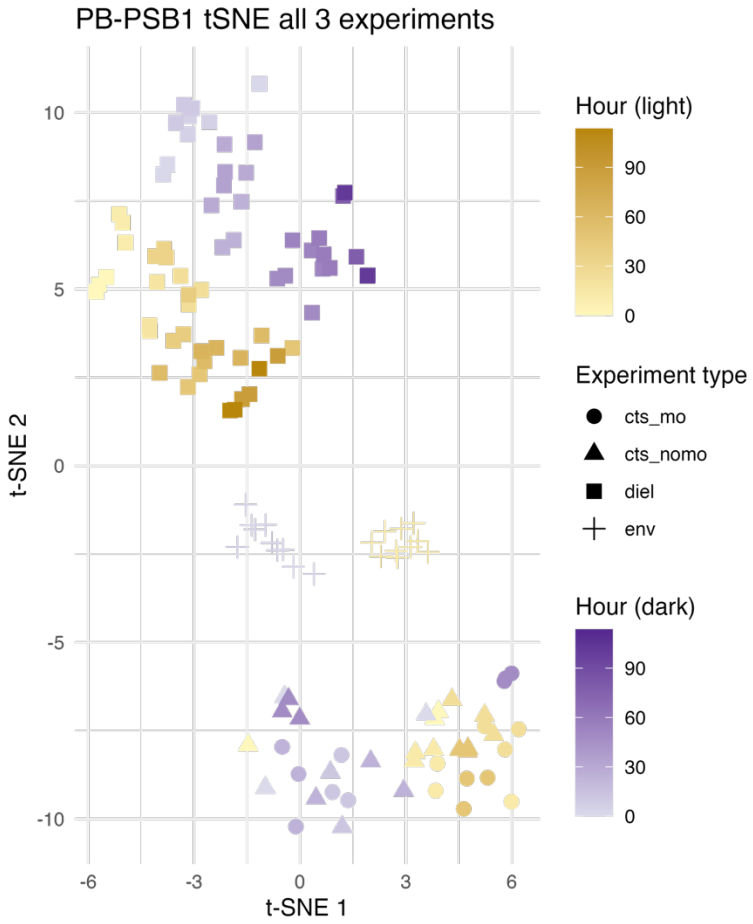


Figure 3-A. Similarities between PB-PSB1 transcripts across all our different incubations was visualized using a t-distributed stochastic neighbor embedding (t-SNE) analysis. Colors correspond to light-dark conditions (dark=purple, light=gold) and are shaded with darker tones corresponding to incubation progression. Shape corresponds to experiment type (circle=molybdate and continuous light or continuous dark conditions, triangle= continuous light or continuous dark conditions, square=lab incubated samples over a diel cycle, cross=environmental samples at peak light or peak dark conditions).

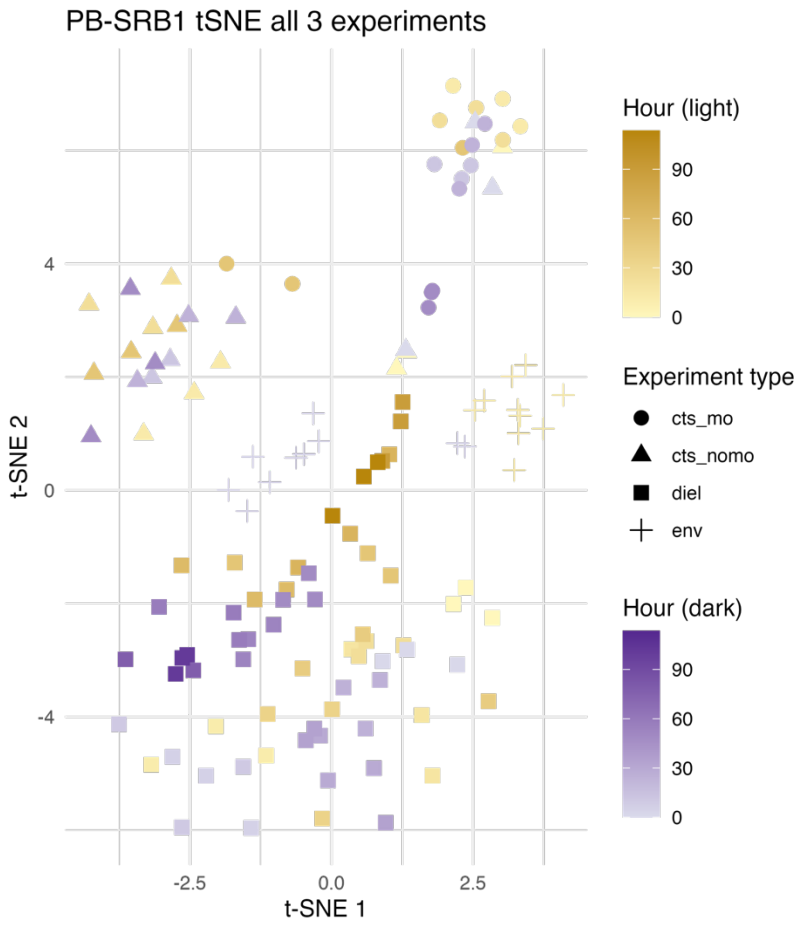


Figure 3-B. Similarities between PB-SRB1 transcripts across all our different incubations was visualized using a t-distributed stochastic neighbor embedding (t-SNE) analysis. Colors correspond to light-dark conditions (dark=purple, light=gold) and are shaded with darker tones corresponding to incubation progression. Shape corresponds to experiment type (circle=molybdate and continuous light or continuous dark conditions, triangle= continuous light or continuous dark conditions, square=lab incubated samples over a diel cycle, cross=environmental samples at peak light or peak dark conditions).



Figure 3-C. Transcriptional activity for pink berry bins across our continuous lighting, molybdate v non-molybdate incubations. Visualized is total TPM per bin summed across samples, $\log_{10}$-scaled, and ordered by distribution's median in descending order. salmon=molybdate incubated samples, teal=non-molybdate incubated samples

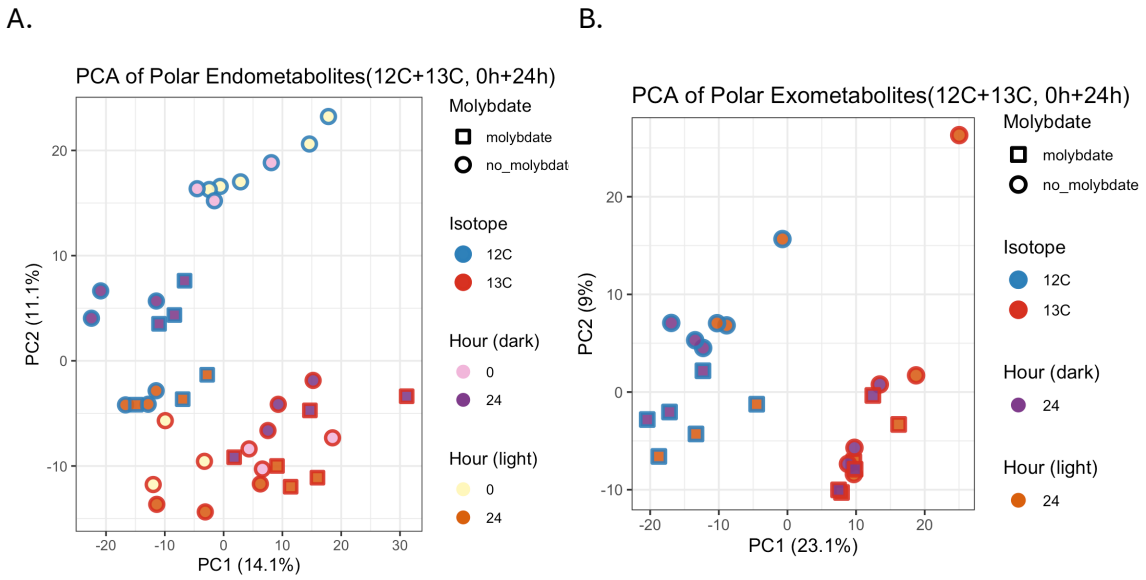


Figure 3-D. Differences in ^{12}C and ^{13}C incubated metabolite samples after exposure to the same amount of continuous light or dark. Red tones correspond to continuous light exposure and purple tones correspond to continuous dark exposure. Stronger shades indicate longer incubation time. Squares are molybdate incubated samples and circles were not incubated with molybdate. Blue outlines are ^{12}C incubations and red outlines indicate ^{13}C incubations. (A) Endometabolome (B) Exometabolome.

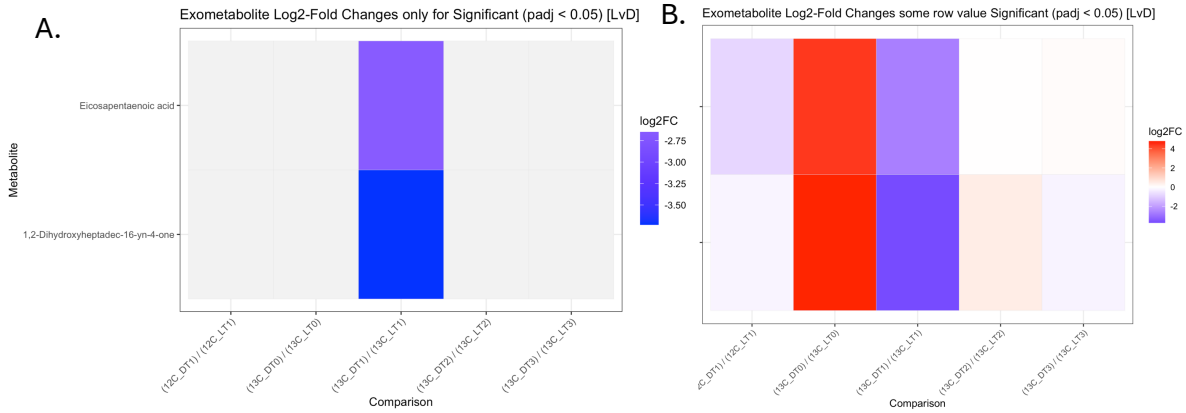


Figure 3-E. Exometabolites compared for significant $\log_2$ (fold change) between light-dark conditions within the no-molybdate exposure treatment. Shown are metabolites in which any comparison across samples was significant ($p_{adj}<0.05$), and magnitude of $\log_2$ (fold change) determines intensity of heatmap. (A) Cells are plotted specifically for the significant comparisons, and (B) all cells for a significant metabolite are plotted.

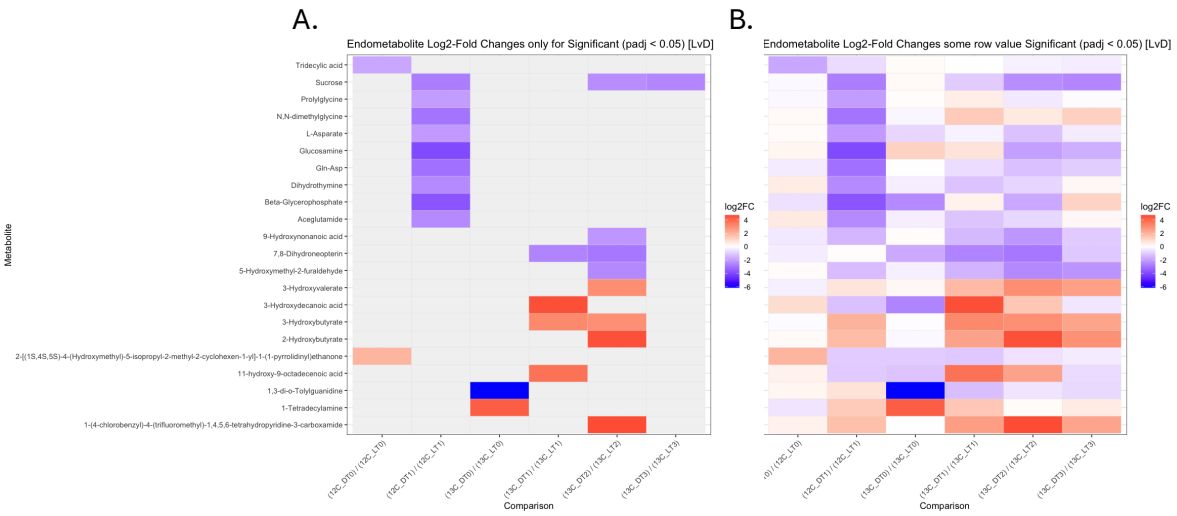


Figure 3-F. Endometabolites compared for significant $\log_2$ (fold change) between light-dark conditions within the no-molybdate exposure treatment. Shown are metabolites in which any comparison across samples was significant ($p_{adj}<0.05$), and magnitude of $\log_2$ (fold change) determines intensity of heatmap. (A) Cells are plotted specifically for the significant comparisons, and (B) all cells for a significant metabolite are plotted.

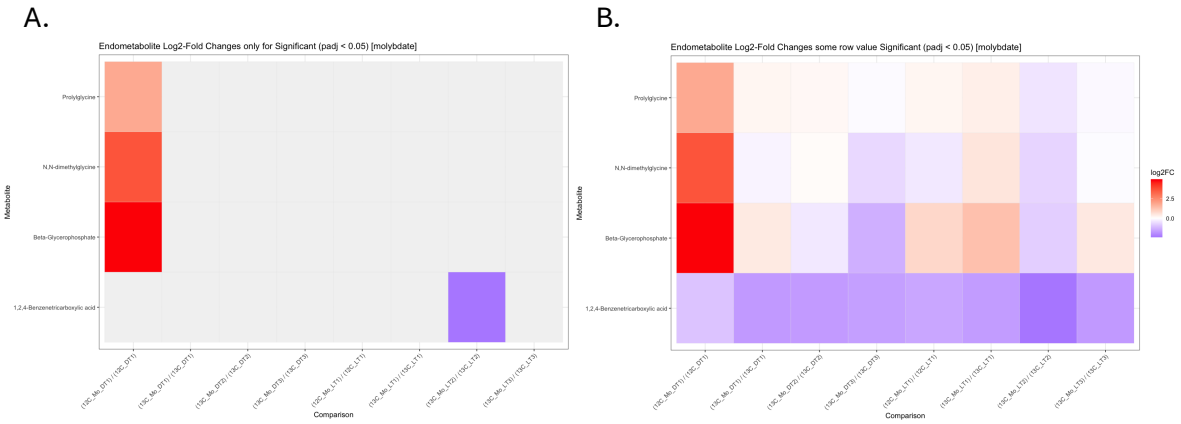


Figure 3-G. Endometabolites compared for significant $\log_2$ (fold change) between molybdate-no molybdate conditions across light/dark treatment. Shown are metabolites in which any comparison across samples was significant ($p_{adj}<0.05$), and magnitude of $\log_2$ (fold change) determines intensity of heatmap. (A) Cells are plotted specifically for the significant comparisons, and (B) all cells for a significant metabolite are plotted.

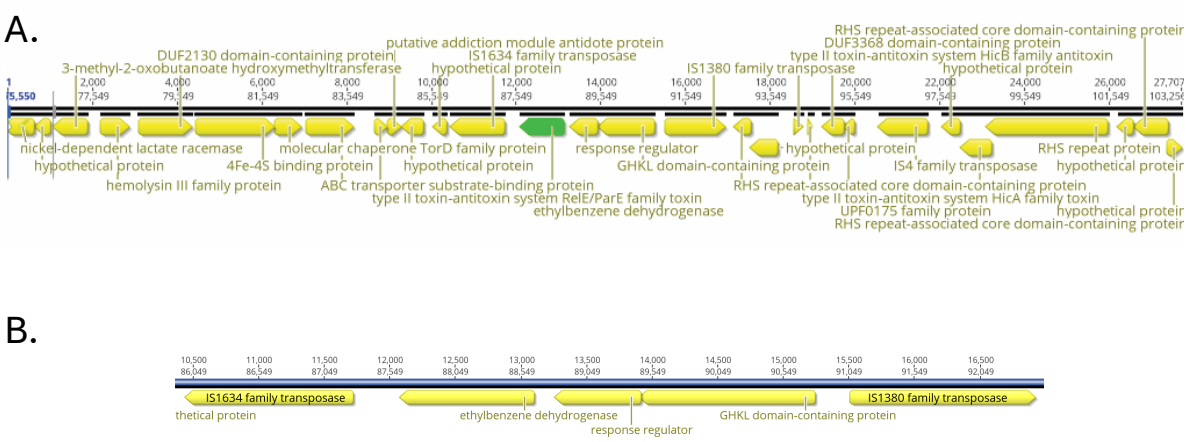


Figure 3-H. (A) PB-SRB1 neighborhood around gene coding for an ethylbenzene dehydrogenase-related protein. Gene of interest colored green, PB-SRB1 reference genome (JAEQMT000000000.1) visualized on Geneious. (B) Zoomed in view of (A)

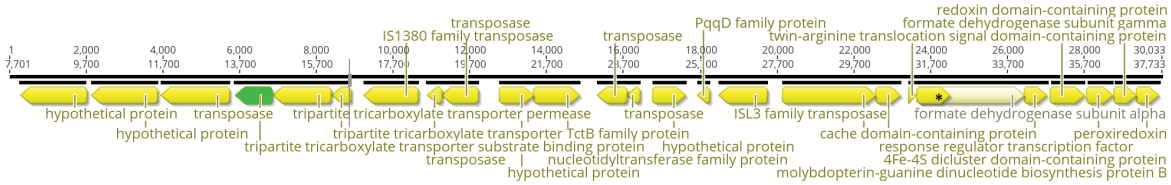


Figure 3-I. PB-SRB1 neighborhood around gene coding for Bug family tripartite tricarboxylate transporter. Gene of interest colored green, PB-SRB1 reference genome (JAEQMT000000000.1) visualized on Geneious.

Chapter IV Supplemental Information

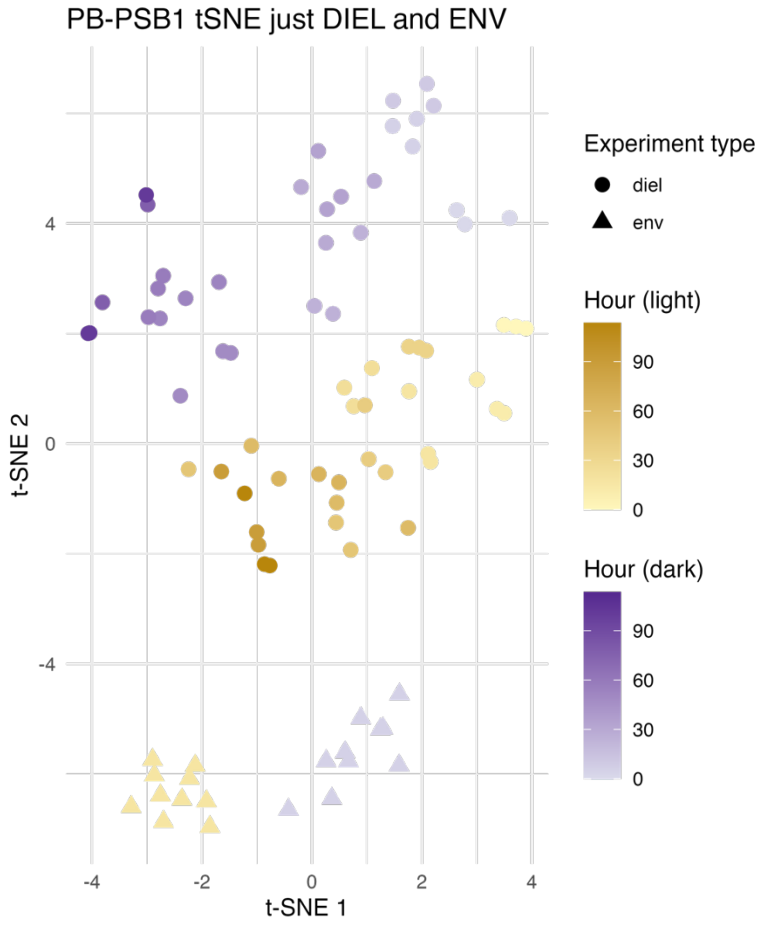


Figure 4-A. Similarities between PB-PSB1 transcripts across our different diel incubations was visualized using a t-distributed stochastic neighbor embedding (t-SNE) analysis. Colors correspond to light-dark conditions (dark=purple, light=gold) and are shaded with darker tones corresponding to incubation progression. Shape corresponds to experiment type (circle=lab incubated samples over a diel cycle, triangle=environmental samples at peak light or peak dark conditions).

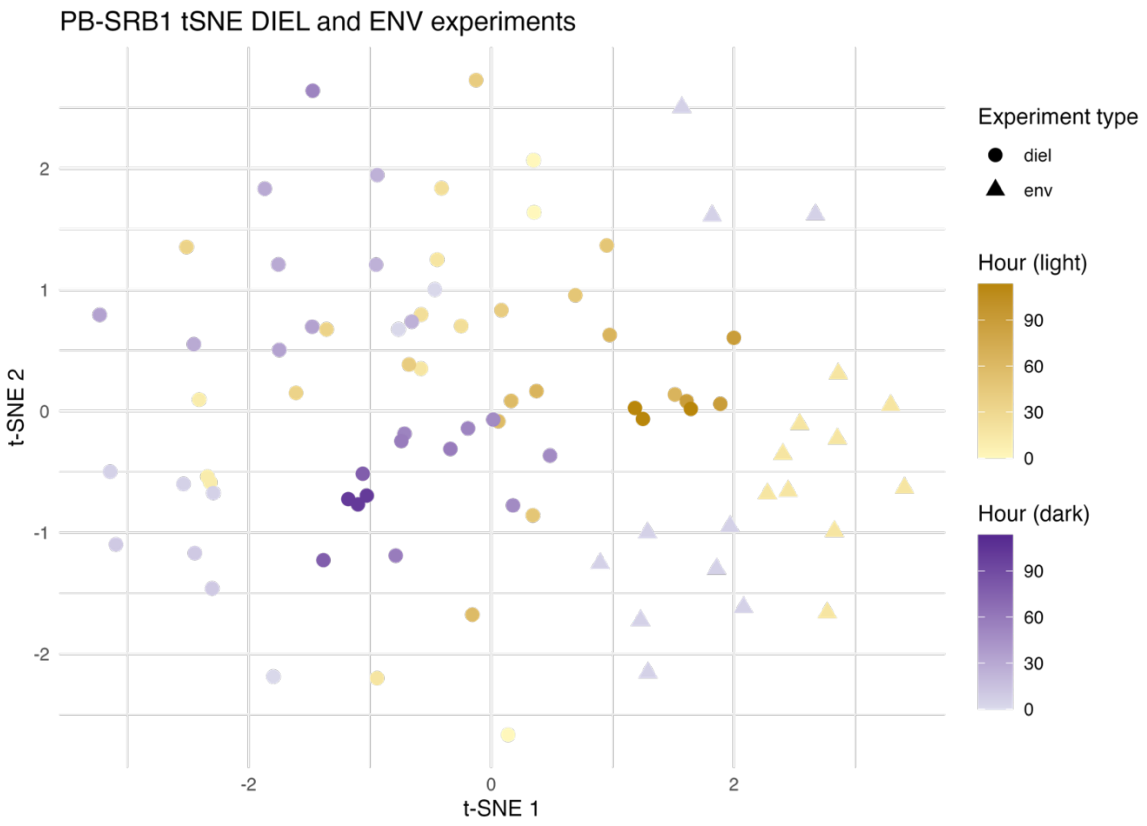


Figure 4-B. Similarities between PB-SRB1 transcripts across our different diel incubations was visualized using a t-distributed stochastic neighbor embedding (t-SNE) analysis. Colors correspond to light-dark conditions (dark=purple, light=gold) and are shaded with darker tones corresponding to incubation progression. Shape corresponds to experiment type (circle=lab incubated samples over a diel cycle, triangle=environmental samples at peak light or peak dark conditions).