TRANSCRIPTIONAL MARKERS OF ORGANIC SUBSTRATE AVAILABILITY IN A COASTAL MARINE BACTERIUM

Jessica Boulton

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Marine Sciences.

Chapel Hill
2021

Approved by:

Scott Gifford

Alecia Septer

Marc Alperin

ABSTRACT

Jessica Boulton: TRANSCRIPTIONAL MARKERS OF ORGANIC SUBSTRATE
AVAILABILITY IN A COASTAL MARINE BACTERIUM
(Under the direction of Scott Gifford)

Microbial activities are essential for the cycling of dissolved organic carbon (DOC) through the ocean. DOC is difficult to measure chemically due to its diversity, low concentrations, and spatial and temporal variability. An alternate method for investigating DOC composition and flux is to examine the transcriptional responses of the bacteria consuming it. However, there is a lack of experimentally validated data linking transcripts to specific carbon sources. In this study, we grew a model marine bacterium, *Ruegeria pomeroyi* DSS-3, in 42 carbon-limited, continuous cultures to test its transcriptomic response to 12 carbon substrates. Growth on different carbon substrates produced transcriptomic signals unique to each carbon source. Upregulated transcripts included transporters and some members of known substrate degradation pathways; however, many genes not associated with these degradation pathways were also upregulated. In this study, we identify specific transcripts that could be used as indicators of the presence of those carbon substrates.

ACKNOWLEDGEMENTS

Over the course of this project, I'm thankful to have received a great deal of support and assistance.

I would first like to thank my advisor, Scott Gifford, whose mentorship and research expertise have been invaluable. I've learned so much from being a part of this lab, and your guidance as an advisor has been instrumental in helping me grow as a researcher over the past few years. The scope, design, and implementation of this study are all substantially better due to your perceptive feedback and innovative ideas, and I deeply appreciate all of your support.

I would next like to thank Marc Alperin and Alecia Septer, my committee members, for their valuable insight and thoughtful comments- your knowledge has sharpened my thinking and brought my research to a higher level.

I would also like to thank my labmates: Garrett Sharpe, Laura Fisch, Melanie Cohn, and Acacia Zhao, for all of your support, great conversations, and much needed assistance over the years. Garrett and Laura, thank you for all of your training and assistance when I started this project, and for welcoming me into the lab. Melanie, thank you so much for all of your support while I've been running this project; the guava is now your responsibility. Acacia, thank you especially for your assistance with the bioinformatics side of this project, in particular with UNIX and Galaxy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The cycling of organic matter in the ocean is a complex and dynamic process; it represents one of the largest fluxes of reduced carbon on the planet, links primary production to all trophic levels, and affects the rate of carbon sequestration in the ocean. Organic matter enters marine ecosystems from multiple sources. These include external inputs from rivers and estuaries, which are enriched in terrestrially-derived plant organic matter, and autochthonous inputs from phytoplankton production released as photosynthate, or cell leakage and lysis (Moran 2016). The source and composition of organic matter entering the ocean vary depending on season, community composition, and plankton physiology (Morán et al. 2013). Once organic matter enters the ocean, its largest sink is bacterioplankton via the microbial loop. Bacteria typically consume a very large percentage of primary productivity in the ocean – in some environments, bacteria consume >91% of dissolved primary productivity (Morán et al. 2002). Identifying the taxa and mechanisms by which microbial communities are metabolizing specific compounds of the DOC pool is crucial for predicting carbon flux, but this identification is hampered by the fact that both the microbial community and the DOC pool are highly diverse. Chemically analyzing the DOC pool requires not only testing for myriads of different molecules with dissimilar properties, but also requires many of these compounds to be identified at very low concentrations (Moran et al., 2016). As such, it is difficult to directly measure the composition of the DOC pool chemically. This limits our understanding of the compounds and

mechanisms involved in organic matter cycling in the ocean and how those processes might change due to altered ocean conditions.

In addition, understanding the carbon pool requires knowledge not only of which compounds are present, but also which compounds are actively cycling through the environment. Identifying the former doesn't necessarily give insight into the latter; many of the carbon substrates with the highest flux are labile compounds quickly metabolized by bacteria (and are therefore present only at low concentrations at any given time), while many carbon compounds with high standing stocks are refractory compounds that build up in the water column when bacteria fail to consume them (Moran et al. 2016).

A reverse ecological approach to looking at DOC composition and flux is to examine the transcriptional responses of the bacteria consuming it. The transcript (i.e. mRNA) pool is a composite record of all the genes organisms are expressing at a given time. Bacteria are known to regulate transcription in response to environmental conditions, and the short half-lives of RNA molecules make transcription particularly well suited to examining rapid shifts in the immediate chemical environment (Moran et al., 2013). Looking at changes in gene expression could indicate which carbon compounds are actively being consumed by a bacterial community at a given time, even if those substrates are at low concentrations or the composition of the available DOC changes quickly. A further advantage of using mRNA as a molecular sensor of the chemical environment is that it provides insight into the mechanisms being used to chemically degrade those compounds.

Many studies have previously used microbial communities as biosensors of environmental conditions. Previous studies of bacterial transcriptional responses to DOC have focused mostly on environmental metatranscriptomes, which includes the expression of genes *in*

situ at the community level (Moran 2016). Poretsky et al. (2010), for example, searched libraries of expressed genes from coastal bacterial communities for carbon transporters to draw conclusions about the composition of labile coastal DOC pools.

The central assumption of microbial biosensor studies is that if bacteria alter their gene expression when different carbon substrates become available in the environment, then the upregulation of certain transcripts should act as indicators of the presence of those carbon substrates. However, there is a lack of quantitative information about which transcripts correlate to the presence of carbon substrates. Furthermore, several other factors may confound interpretation of transcriptome DOC relationships. Bacterial transcriptional responses to environmental stimuli can be complex. Genes may function in multiple pathways at once, may be constitutively expressed, or may not be regulated at a transcriptional level**.** As such, identifying which transcripts are most abundant in a certain sample can still leave many unanswered questions about the conditions of the bacterial community. To accurately predict the content of the DOC pool, it is therefore necessary to validate the relationship between transcriptomic response and carbon substrate availability. Furthermore, in addition to substrates driven transcriptional shifts, the physiological state of cells may influence mRNA composition. Growth phase has been shown to alter the type and abundance of RNA recovered in cultivation experiments (Gifford et al. 2016).

In this study, we sought to increase the utility of gene expression and metatranscriptomics data by quantitatively examining the relationship between organic substrate availability and gene expression in a model marine bacterioplankton. *Ruegeria pomeroyi* DSS-3 was grown in carbon-limited chemostats with media containing a defined, environmentally relevant substrate as the carbon source (including carboxylic acids, polyols, saccharides, aromatic compounds, and

methylated amines). Cells were collected after at least four complete turnovers of the media, and RNA-Seq analyzed to identify the transcriptomic response to each substrate. The chemostat approach enabled us to 1) keep *R. pomeroyi* at a defined growth rate, thus reducing the amount of growth-phase driven transcriptional signals, and 2) to reproduce the low concentration, high-flux DOC conditions cells may experience in the environment. We selected *R. pomeroyi* as our model organism given its known reliance on transcriptional regulation, and generalist metabolism that utilizes a wide variety of organic substrates (including both terrestrially and phytoplankton derived substrates), and given it is representative of the marine Roseobacter clade that is highly abundant in coastal environments.

We aimed to answer three questions: 1**)** Does *R. pomeroyi* exhibit unique transcriptional profiles when grown on different carbon substrates? 2) Are genes in known substrate degradation pathways upregulated when cells are grown on that substrate, or alternatively are transcripts enriched or depleted in pathways not previously identified with the substrate's metabolism. 3) Does *R. pomeroyi's* transcriptomic response change when grown on different concentrations of a particular substrate? We hypothesized that *R. pomeroyi* would show unique transcriptional signals for each substrate treatment, with increased transcription of genes in pathways known to be involved in a particular substrate's degradation. We further hypothesized that the strength of these signals would depend on the length of substrate's degradation pathway before incorporation into the TCA cycle, with substrates having to pass through metabolic modules many steps away from the TCA cycle having stronger or more unique transcriptional signals than substrates quickly incorporated into the TCA cycle.

RESULTS and DISCUSSION

   Chemostat design. *Ruegeria pomeroyi* DSS-3 was grown in carbon-limited chemostats on defined salts-Marine Basal Medium (dsMBM). In steady state, biomass in the chemostat remains constant – cells washed out of the chemostat are replaced by new cells, and the growth rate is equal to the rate at which cells are removed. As such, the dilution rate (D, units = time^{-1}) must be equal to the growth rate (μ) when the chemostat is at steady state. The dilution rate is calculated as D = F/V, where F = flow rate (the rate at which media is continuously supplied and removed, units= volume/time) and V= volume of media in the culture vessel. The chemostat turnover time (τ) represents the amount of time it takes for the media in a chemostat culture vessel to be completely replaced by new media, and is equal to V/F (or 1/D) (Kubitschek 1970). For our experiments, the dilution rate was set to D = 0.04 hour^{-1}, with V = 140 mL and F = 93.3 μL min^{-1} (5.6 mL hr^{-1}), resulting in τ = 25 hours. In our experiments, the growth rate was set to 0.96 day^{-1} , mimicking growth rates of Roseobacers *in situ* (0.70-1.64 day^{-1}; Teira et al. 2009, Chan et al 2012).

   Ensuring carbon limitation. An initial set of experiments was run to ensure that growth in the chemostats was carbon limited. *R. pomeroyi* was grown in the chemostats with a reservoir media of ds-MBM and 500 μM acetate. After several turnovers in steady state, the media concentrations in the reservoirs were stepped up with an additional 500 μM acetate, bringing the reservoir concentration up to 1 mM. Before the carbon spike-in, the chemostats had an average OD$_{600}$ of 0.010 and 0.008 (Fig. 1A). After acetate was spiked in, the OD$_{600}$ increased to 0.024 and 0.019, respectively, and the cells eventually entered a new steady state confirming carbon was the limiting growth factor. These experiments were also performed with glucose to ensure carbon limitation on a high yield carbon substrate (Fig. 1B)

Rationale for acetate background and choice of carbon substrates: To characterize *R. pomeroyi's* transcriptional profile on different organic substrates, we grew *R. pomeroyi* in continuous cultures with 500 µM of the target substrate and 500 µM acetate. For three substrates (glucose, glycerol, and benzoate) more than one target substrate concentration was tested. Acetate was used as background substrate for two reasons. The first reason is that the addition of a background substrate allowed us to maintain sufficient biomass and growth to prevent cells from washing out of the chemostat, even if the target substrate is not highly labile or easily metabolized (although such substrates still can be used by the cell). For example, glycine betaine and benzoate may not be used for synthesizing biomass, but still can support energy transduction. Secondly, acetate is a basal metabolite, with just one step (addition of CoA to acetyl-CoA) before entering the TCA cycle. As such, it is unlikely to spur changes in the transcription of metabolic modules of more complex substrates

For the target substrates, we choose twelve compounds that represent a range of chemical structures (aromatics, alkanes, polyols), have varying degrees of steps before entering the TCA cycle, and have previously been shown to be environmentally relevant (Table 1). Growth in the chemostats showed *R. pomeroyi* had distinct responses to the different carbon substrates (Fig. 2). Yields across samples were approximately similar when ODs were normalized by carbon number. Most treatments had a 0.005 to 0.007 OD_{600} carbon^{-1}. Notable exceptions were tween 40, which has more than 60 carbons, but had only small increases in cell density compared to the acetate only treatments, and the two aromatic substrates, benzoate and 4-hydroxybenzoate, which have slightly lower OD_{600} carbon^{-1} (0.002 and 0.004, respectively). The OD_{600} carbon^{-1} for glycerol was higher than the other substrates (0.009). Aromatics and complex

polyols were unsurprisingly less efficiently metabolized, while glycerol seems to be efficiently utilized.

Most chemostat runs reached steady state (which we defined as OD_{600} remaining within 10% variance for at least one turnover) within one chemostat volumetric turnover. However, several carbon substrates did not, namely the glucose and benzoate treatments (Fig. 2). The chemostat experiments that did not reach steady state were treatments 500 µM glucose & 500 µM acetate, 50 µM glucose & 500 µM acetate, 500 µM succinate & 500 µM acetate and two 500 µM benzoate & 500 µM acetate treatments. There are several potential explanations that could account for the failure of some experiments to reach steady state. For the 500 µM glucose & 500 µM acetate treatment (FN#14), the experiment appears to have been going into steady state just before it ended, suggesting that if the run was extended the culture may have entered steady state. The 50 µM glucose & 500 µM acetate treatment (FN # 30) had significant variability in OD_{600} but was otherwise trending towards steady state. For both 500 µM benzoate & 500 µM acetate cultures (FN#s 23 and 36), the OD_{600} plateaued before dramatically increasing in the last few turnovers of the experiment. The cells may have been primarily subsisting on acetate for the first few turnovers before adapting to take full advantage of benzoate availability. Chemostat runs with lower concentrations of benzoate (50 µM benzoate & 500 µM acetate) had yields and growth curves similar to the 500 µM acetate-only runs. Dramatic drops and spikes in OD in 500 µM pyruvate & 500 µM acetate, 500 µM propionate & 500 µM acetate, and 500 µM N-acetyl-D-glucosamine & 500 µM acetate treatments around the 300-hour mark are due to a stir plate malfunction, which caused cells to collect at the bottom of the vessel before becoming resuspended after the stir plate was fixed. Except for some benzoate runs, the maximum ODs reached in chemostat experiments were similar to those reached in batch culture.

Sequencing results. Samples were sequenced in two batches. The first batch consisted of 18 samples, and included glucose, glycerol, acetate, and benzoate treatments. The second batch consisted of 21 samples, and included glucose, succinate, propionate, betaine, acetate, N-acetyl-D-glucosamine, 4-hydroxybenzoate, pyruvate, glycolic acid, and Tween 40 treatments. An rRNA subtraction was performed before sequencing for the second batch of samples, but not the first. This difference in rRNA depletion likely resulted in undersequencing of transcripts for the first batch. For the first batch of samples, an average of 16 ± 3 ($\pm sd$) million reads were recovered from each sample. On average, R. pomeroyi reads consisted of 96.01% rRNA and 3.46% mRNA. Read composition varied from 93.6 to 98.8% rRNA and from 1.1 to 5.9% mRNA. In total, 9.2 million R. pomeroyi mRNA reads were recovered. Of the 4371 known genes in the *R. pomeroyi* DSS-3 genome, 4346 unique transcripts were recovered across all samples (99.4 % of the genome). Sequencing data for all samples can be found Table 2.

Sulfolobus internal standards were identified after mapping. Standard recovery was log-linear for all samples sequenced. However, standards 13 and 7 (both from the same pooled standard group) were recovered at 88% and 75% below expected values, respectively, and were not used for calculation of the conversion factor. Since the third standard in that pooled group, standard 15, was recovered at a rate similar to the rest of the standards (if slightly higher), there may have been an error in the quantification and/or pooling of these two standards before their addition to samples. Alternately, recovery of this group of standards may have been affected by sequencing bias; standard 15 is the longest sequence used for any standard and this may have influenced recovery rates of the other standards in this group. The other 7 standards (excluding 13 and 7) were used to calculate a conversion factor (standards added/standards recovered) for each sample. This conversion factor was then used to convert the recovered transcript reads to

transcripts cell^{-1}. These calculations suggest that on average 207 ± 83 total RNAs were present per cell, and 7 ± 3 mRNAs were present per cell.

For the second batch of samples, an average of average of 15 ± 3 ($\pm$ sd) million reads were recovered from each sample. On average, *R. pomeroyi* reads consisted of 74.0% rRNA and 18.8% mRNA. Read composition varied from 94.6 to 31.6% rRNA and from 42.6 to 2.25% mRNA. In total, 111 million mRNA reads were mapped to *R. pomeroyi's* genome. A total of 1.13 million internal standard reads were recovered, and internal standards accounted for 1.56% of non-rRNA reads on average. One sample, FN100 (a pyruvate treatment) was excluded from further analysis for low quality sequencing results. Sequencing data for each sample can be found in Table 2.

Transcriptome-wide fingerprints of carbon substrate availability. Expression profiles of all genes were compared across substrate treatments in a principal components analysis. Samples grouped into two main clusters based on their sequencing run. This was likely due to the differences in sample prep (an rRNA subtraction was not performed on the first group of samples sequenced, but was performed on the second group). However, within each sequencing run, there was clear grouping of samples based on their substrate treatment (Fig. 3A). The glycine betaine samples (part of the second group) were significantly different from the other samples due to the high number of differently expressed genes we observed with this treatment (see discussion below). Figure 3A shows a PCA of all samples, including both groups of samples and betaine. When the second group of samples excluded betaine the samples clustered by carbon source (Fig. 3B). demonstrating that the carbon substrate present in the media leaves a distinct, genome-wide footprint on DSS-3's transcriptome.

Transporters. We next examined whether the availability of a carbon substrate corresponded with increased transcription of transporters specific to that substrate's uptake. Uptake is an essential first step in bacterial metabolism of a compound. Given their critical role, transporter genes are frequently some of the most abundant transcripts in transcriptomes and metatranscriptomes. We hypothesized that transporter transcription levels would be higher than the average *R. pomeroyi* gene transcript abundance when the target substrate was present. In addition, many transport systems consist of multiple components encoded by different genes, with specific components for binding to the target substrate. For example, in ABC transporters the substrate binding protein binds to the target compound which is then transported through the membrane by the other subunits (permease proteins, ATP-binding proteins). We hypothesized that the substrate binding components would have the most substrate-specific transcriptional enrichment.

Of the 12 carbon substrates examined, four have transporter operons reliably annotated in *R. pomeroyi* for their uptake: glucose, betaine, succinate, and N-acetyl-D-glucosamine. The remaining eight substrates are taken up by transport systems not yet identified in *R. pomeroyi*. For each of the five substrates with annotated transporters we examined if their transcription was greater than 1) the average *R. pomeroyi* gene transcription in that treatment, 2) the average *R. pomeroyi* transporter transcript abundance, and 3) the substrate binding components of *R. pomeroyi's* transporters.

R. pomeroyi takes up glucose via an ABC transporter (SPO0861-0863) annotated as a xylose transporter. In every non-glucose amended treatment, these three transporter genes were less abundant than the upper 95% confidence interval of the average gene abundance (Fig 4). Only in the glucose amended treatments were the three genes greater than mean average. The

periplasmic binding protein was particularly enriched in the glucose treatments, having greater abundance than the mean of all gene transcripts, transporter transcripts, and all substrate binding.

Succinate is taken up via a TRAP transporter (SPO2626-2630). The non-periplasmic binding components of this transporter were all transcribed lower than the mean *R. pomeroyi* gene abundance for all treatments, including succinate (Fig 5). By contrast, the succinate transporter's DctP solute receptor protein had transcript abundances higher than the mean gene transcription in many of the treatments, but transcription was highest in the succinate treatments, with normalized abundance greater than the mean of all gene, all transporters, and all substrate binding proteins.

In *R. pomeroyi*, there are several different transporters capable of taking up glycine betaine. These include at least three ABC transporters (locus tags= SPO1131-1133; SPO2441-2443; SPOA0231-A0233). Two of these transporters had consistently low transcript abundances in all treatments, including glycine betaine (Fig 6). The SPO1131-1133 ABC transporter tended to have higher expression, particularly the substrate binding protein, but this was true across all substrate treatments. However, *R. pomeroyi* contains a single gene annotated as a glycine transporter (SPO3186) that exhibited a 100-fold increase in transcript abundance in the glycine betaine treatments compared to all other treatments.

N-acetyl-D-glucosamine is taken up by the sugar ABC transporter SPO1835-1839. In the N-acetyl-D-glucosamine treatments, transcription of the periplasmic binding protein, permease protein, and ATP-binding protein was higher than the mean of all other genes and transporters (Fig 7).

In summary, the carbon substrates elicited a substantial transcriptional response in transporter expression that was specific to the target substrate. Of the subunits making up the

transporter, the substrate binding proteins were the most reliable transcriptional indicators of substrate availability.

Transcriptional responses to target substrates We next examined how genes in metabolic pathways downstream of transport transcriptionally responded to the different carbon substrate treatments. Genome-wide comparison of transcription under the carbon substrates were compared and significantly differentially expressed genes were identified using a Likelihood Ratio Test (LRT) which analyzed multiple carbon substrates at once. Sample group 1 (which includes benzoate, glucose, and glycerol treatments) and sample group 2 were analyzed separately. Genes were considered significantly differentially expressed at p-value < 0.05. Profiles of gene expression were examined using DEGpatterns, and genes were selected for specificity to a single carbon source. After performing differential expression analysis, we grouped genes into two categories for discussion: "expected" genes, which are annotated in KEGG as being in metabolic pathways specific to the target carbon substrate, and "emergent" genes, which are upregulated in a substrate treatment despite not being known to be involved in that substrate's metabolism.

Transcriptional response to acetate background. Acetate is a simple carboxylic acid with a short metabolic path. After the addition of a Coenzyme A (CoA) it directly enters the basal metabolism of the TCA cycle. In the marine environment, sources of acetate may include photolysis of DOC in the euphotic zone, photosynthetic leakage of phytoplankton such as cyanobacteria, or bacterial fermentation in anoxic micro-environments (such as fecal pellets or marine snow; Zhuang et al. 2019). Acetate is known to be regularly metabolized by heterotrophic bacteria, including members of the Roseobacter clade (Zhuang et al. 2019). In this study, acetate was used as a background substrate to ensure cell yields were high enough to be

detectable by optical density measurements as well as our hypothesis that its short metabolic path to the TCA cycle would result in minor effect on the transcription of higher-level carbon degradation pathway genes.

To check the effect of acetate on gene transcription, we searched the results for genes upregulated in acetate only treatments. Genes annotated to acetate degradation pathways were not significantly enriched in the transcriptome when cells were grown on acetate alone. *R. pomeroyi* has four acetyl-CoA synthetase genes, none of which are differentially expressed in acetate-only treatments. The results were also checked for the presence of any other genes associated with acetate treatments. There were three genes significantly enriched ($p < 0.05$) in acetate-only treatments, when compared to at least two treatments. These are a NnrU family protein (SPO0353), methyl-malonyl-coA mutase (SPO0368), and Isobutyryl-CoA dehydrogenase (SPO0693). However, the association between the upregulation of these genes and the presence of acetate was relatively weak.

Pyruvate. Pyruvate, a small monocarboxylic acid, is an intermediary metabolite in several biologically important pathways. Glycolysis produces two pyruvate molecules for each glucose molecule, and reversing this process for gluconeogenesis requires pyruvate as an initial substrate. Pyruvate can also be shunted into fatty acid biosynthesis. In the marine environment, pyruvate is produced in surface waters by photochemical degradation of high molecular weight dissolved organic carbon (Obernosterer et al. 1999). In addition, there is evidence that several different groups of bacteria will excrete pyruvate when growing on other substrates, particularly in carbon-rich and nitrogen-limited environments (Benzon et al. 2016, Madden et al. 1996, Ruby and Nealson 1977).

No genes were significantly ($p < 0.05$) upregulated in pyruvate treatments compared to other treatments. Pyruvate has a very short path before it enters the TCA cycle, and is an intermediary metabolite in several biological pathways. As such, genes related to its uptake and metabolism may be expressed constitutively, meaning the cell does not have to alter its gene expression to metabolize pyruvate when it becomes available in the environment.

4-hydroxybenzoate. 4- hydroxybenzoate consists of an aromatic ring with a carboxyl substituent and a deprotonated hydroxyl group. 4-hydroxybenzoate enters the marine environment largely through anthropogenic pollution from terrestrial sources: 4-hydroxybenzoate is used commercially for polymer production in the chemical and electrical industries (Peng et al. 2006). In addition, alkyl esters of 4-hydroxybenzoic acid (often referred to as parabens) are used as preservatives in cosmetics and are common pollutants; parabens are largely biodegraded to 4-hydroxybenzoic acid in wastewater treatment plants (Juliano and Magrini 2017). Parabens and 4-hydroxybenzoic acid are thought to exist in high concentrations in the marine environment and have been detected in the tissues of diverse marine vertebrates (including fish, marine mammals, and marine birds; Juliano and Magrini 2017). In addition, bacteria of the genus *Microbulbifer* have been found to excrete 4-hydroxybenzoate (Peng et al. 2006). *R. pomeroyi* contains genes to degrade 4-hydroxybenzoate via the β-ketoadipate pathway (Yan 2009).

When grown on 4-hydroxybenzoate, *R. pomeroyi* significantly upregulated 27 genes, 19 of which were located on *R. pomeroyi's* megaplasmid and 8 of which were located on its main chromosome. The 19 megaplasmid genes can be divided into three groups. The first group (SPOA0040-SPOA0043) included two genes annotated as fluoride ion transporter CrcB proteins and two genes annotated as parts of the beta-ketoadipate pathway. The second group

(SPOA0341-SPOA0346) contains a gene annotated as an AfsA domain killing gene and an immunity gene, as well as a hypothetical protein and two carbon metabolism genes. The third group (SPOA0400-SPOA0409) contains several carbon metabolism genes (including a benzoate-coenzyme A ligase) as well as a MarR family transcriptional regulator (SPOA0405). The 8 genes on *R. pomeroyi*'s main chromosome mostly encode transporters and metabolic genes and can also be largely grouped by location. These genes include SPO1771 and SPO1774, a TRAP dicarboxylate transporter DctM subunit and an oxidoreductase, two carbon monoxide dehydrogenase proteins (SPO2394-SPO2395), and SPO2681-SPO2683, all of which are annotated as twin-arginine translocation proteins. Also upregulated was a kynureninase (SPO2824).

The upregulation of the Afsa domain killing gene, which encodes for a diffusible killing mechanism, and adjacent immunity gene suggests that 4-hydroxybenzoate may be available to *R. pomeroyi* in cell-dense environments where competition is important for community structure (Sharpe et al. 2020). Degradation of 4-hydroxybenzoate is likely carried out via the beta-ketoadipate pathway, which converts protocatechuate (a 4-hydroxybenzoate derivative) to beta-ketoadipate before entering the TCA cycle (Harwood & Parales, 1996). The various upregulated transport proteins on the main chromosome are likely used to transport 4-hydroxybenzoate into the cell.

N-Acetyl-D-Glucosamine One of the largest pools of amino sugars in the ocean consists of dissolved N-acetyl-D-Glucosamine (GlcNAc), a monosaccharide derivative of glucose (Riemann & Azam 2002). GlcNAc polymerizes to form chitin, the second most abundant carbohydrate after cellulose and a structural component in fungi, phytoplankton, and the shells of marine invertebrates (Chen et al. 2010). In addition, N-acetyl-D-Glucosamine is a major

component of cell wall peptidoglycan in both gram positive and gram negative bacteria (Riemann & Azam 2002). As such, N-Acetyl-D-Glucosamine is present in the marine environment at high concentrations and is readily consumed by various heterotrophic bacteria.

When grown on N-Acetyl-D-Glucosamine, *R. pomeroyi* significantly upregulated 28 genes ($p < 0.05$). Five of these genes were transport-related: a three-gene operon (SPO1835-1837) on the main chromosome coding for sugar ATP transporter proteins and a Gfo/Idh/MocA oxidoreductase, and two genes on DSS3's megaplasmid. One of the megaplasmid genes (SPOA0052) was an ATP-binding protein (NosF) for an ABC transporter, and the other megaplasmid gene (SPOA0278) was annotated as the DctM subunit of a TRAP dicarboxylate transporter. Other upregulated genes mostly fell into three general categories: hypothetical proteins (7 genes), carbon metabolism (8 genes), and signal transduction (2 genes). Signal transduction genes included a calcium-binding domain protein (SPO0490) and a fatty acid desaturase (SPO2327) annotated as a low-temperature signal transduction gene. Other upregulated genes were diverse in function; they included genes annotated for ubiquinone biosynthesis, tyrosine biosynthesis via the Shikimate pathway, two different carbon monoxide dehydrogenase proteins, a M16 family zinc protease, and a RNA polymerase sigma-70 factor. In addition, the self-catalyzing ribozyme glmS (SPOA0140) was significantly depleted in N-acetyl-D-Glucosamine treatments, suggesting that it plays some regulatory role in N-acetyl-D-Glucosamine degradation.

The sugar ATP transporter proteins in the 3-gene operon (SPO1835-1837) are utilized for N-acetyl-D-Glucosamine uptake; the other transport proteins may be part of previously un-annotated N-acetyl-D-Glucosamine uptake systems. Carbon monoxide dehydrogenase genes are used for carbon fixation from CO. While the carbon metabolism genes are likely used for N-

acetyl-D-Glucosamine degradation, the utility of other upregulated genes for growth on N-acetyl-D-Glucosamine remains obscure.

Betaine Glycine Betaine is an amino acid derivative ubiquitous in marine environments. Both prokaryotes and eukaryotes synthesize it as an osmolyte for protection from osmotic stress, and even organisms incapable of synthesizing it will import and store it in high salt environments (Jones et al. 2019) (Oren 1990). Betaine is excreted into the environment in response to changes in salinity, or it may leak into the environment from lysed or unhealthy cells (Oren 1990). Betaine can be degraded aerobically or anaerobically, and many microbes can grow on betaine (Oren 1990). Anaerobic degradation of betaine is particularly important for methane cycling, as up to 90% of methane emissions in coastal environments can be linked to betaine degradation (Jones et al. 2019).

Betaine treatments had the most differentially expressed genes of any substrate, with 753 genes significantly upregulated in betaine compared to other treatments. Most of these were hypothetical proteins (135 genes), followed in number by genes categorized as amino acid metabolism genes (115 genes), carbon metabolism genes (67 genes), and transport related genes (66 genes). Many of these genes are involved with C1 and folate metabolism (Fig. 4), suggesting that the presence of betaine spurred a remodeling of the cell's genome for one-carbon and amino acid metabolism. The strength of this response suggests that betaine may be ecologically important for *R. pomeroyi*; its presence possibly heralding a change in regime or growth stage for the phytoplankton blooms with which *R. pomeroyi* is associated *in situ*. Betaine is an osmolyte released by dying cells, meaning that betaine might be released in large quantities at the end of a phytoplankton bloom or during an osmotic stress event (Oren 1990). *R. pomeroyi* might shift its gene expression to one carbon metabolism to focus on energy transduction instead of

synthesizing biomass when this occurs, thereby conserving energy for when other carbon substrates become less available.

Glycolic Acid Glycolic acid, a monocarboxylic acid, is produced by phytoplankton during photorespiration (Wright et al. 1975). Once produced, most glycolic acid is diffusively released into the surrounding environment. Glycolic acid is often the most common and abundant phytoplankton exudate (Wright et al. 1975). Carbon compounds exuded by phytoplankton are ecologically important for phytoplankton-bacteria interactions: up to 25% of heterotrophic bacterial production in marine environments has been linked to uptake of phytoplankton-exuded compounds (Fog et al. 1983). As the most common phytoplankton exudate, glycolic acid is no exception. Glycolic acid is rapidly taken up by heterotrophic bacteria, with turnover times on the scale of hours, and the ability to take up glycolic acid is widespread among marine bacteria (Fogg et al. 1983).

When grown with glycolic acid, *R. pomeroyi* upregulated four genes. All four genes were located on *R. pomeroyi*'s megaplasmid instead of its primary chromosome, and all were annotated as part of the β-hydroxyaspartate cycle (BHAC) used by marine Proteobacteria to assimilate glycolate (Borzyskowski et al. 2019). The upregulated genes are a beta-hydroxyaspartate dehydratase (locus tag = SPOA0144), aspartate-glyoxylate aminotransferase (SPOA0145), beta-hydroxyasparate aldolase (SPOA0146), and iminosuccinate reductase (SPOA0147). No genes outside of this pathway were significantly enriched.

Propionate. Propionate is a monocarboxylic volatile fatty acid. It is an important metabolic intermediary in marine sediments, where low concentrations of propionic acid are produced via fermentation and degradation of higher molecular weight organic matter (Glombitza 2014) (Glombitza 2015). Propionate can be produced by DMSP

(dimethylsulfoniopropionate) metabolism, an abundant algal osmolyte (Yoch 2002). In surface waters, propionic acid can be produced through microbial degradation of fatty acids and photo-oxidation of anthropogenic compounds (Gad & Gad 2005). Industrial waste from coal and shale oil fuel manufacturing, textile mill waste, wastewater, ship paints, and gasoline exhaust all can also release propionate into the environment (Gad & Gad 2005).

Only one gene, propionyl-CoA carboxylase, alpha subunit (SPO1101) was upregulated in propionate treatments, but just below the significance threshold ($p = 0.056$). Propionyl-CoA carboxylase catalyzes the carboxylation of propionyl CoA to (S)-methylmalonyl-CoA, which can then be converted to succinyl-coA to enter the TCA cycle. This gene shunts propionate directly into the TCA cycle for energy production. Other genes used to bring propionate into the TCA cycle are also used for succinate degradation.

Succinic acid Succinic acid, a dicarboxylic acid, is an essential component of the TCA cycle. In addition, succinic acid is produced in the environment through anaerobic fermentation of glucose and anaerobic respirations with fumarate, malate, or aspartate as electron acceptors (Schink and Pfennig 1982). Anaerobic succinate production is carried out by a variety of organisms: prokaryotes, protozoa, and multicellular organisms such as annelids and mollusks (Hammen 1975) (Schink and Pfennig 1982). Succinic production via fermentation is also carried out commercially- succinic acid is used as a precursor molecule for synthetics in the pharmaceutical, food, and chemical industries (Song and Lee 2006). Heterotrophic bacteria grown on succinate often anaerobically convert it to propionate (Schink and Pfennig 1982).

No genes were significantly ($p < 0.05$) upregulated in succinate treatments compared to other treatments. This may be due to succinate's central role in the TCA cycle; genes related to its uptake and metabolism may be expressed constitutively. In addition, succinate and propionate

require several of the same genes for entry into the TCA cycle– the gene expression of cells grown on these substrates may be highly similar, precluding the appearance of a significantly unique expression signal.

Tween 40 Tween 40 (also known as polysorbate 40 or polyoxyethylene (20) sorbitan monopalmitate) consists of ethoxylated sorbitan esterified with palmitic acid (a fatty acid) (Schiweck et al. 2012). It is a surfactant and is used commercially as a solubilizer and emulsifier in the chemical and pharmaceutical industries and as a food additive (Schiweck et al. 2012). Tween 40 is synthetically produced and is not known to occur in the environment in large quantities. However, its fatty acid component, palmitic acid, is abundant in the environment and an important component of bacterial cell walls (Cho et al. 1966).

No genes were significantly ($p < 0.05$) upregulated in Tween 40 treatments compared to other treatments. While *R. pomeroyi* is known to have the ability to grown on Tween 40, there are no annotated pathways for its degradation. *R. pomeroyi* showed a slight increase in OD when grown on Tween 40 with acetate when compared to growth on acetate alone, which suggests that DSS-3 was taking up Tween 40 at least in small amounts. It is likely that DSS-3 grew preferentially on the background substrate, acetate, dampening any Tween 40-specific expression response.

Glucose. Glucose is a highly labile monosaccharide produced by photosynthetic organisms. Cellulose (which occurs in the cell walls of most plants and algae) consists of linked glucose monomers, and many of the other carbohydrates secreted by phytoplankton contain glucose monomers (Mühlenbruch et al. 2018). While the standing stock of dissolved glucose in the ocean may be low at any given point, there are large fluxes of glucose through the ocean and it has been reported to support 15-47% of marine bacterial production (Rich et al. 1996).

A comparison of the two glucose treatments with different concentrations showed no significant differences in gene expression between them. The same genes were upregulated in low glucose concentration treatments (50 µM Glucose & 500 µM Acetate) as in high glucose concentration treatments (500 µM Glucose + 500 µM Acetate), and glucose concentration seemed to have no significant effect on the level of transcriptional response. While some of these upregulated genes fit into known glucose degradation pathways, many have functions seemingly unrelated to glucose.

DSS3's genome contains genes for three main glucose pathways: glycolysis, the pentose phosphate pathway, and the Entner-Doundoroff pathway. Initially, we hypothesized that genes associated with these metabolic pathways would have significantly higher transcript abundance in the glucose treatments compared to all other treatments. This was true for several glycolysis genes, most notably SPO0864, galM (SPO0857), and SPO1508. SP00864 is an unnamed putative glucokinase referred to here by its organism-specific locus tag. It facilitates the phosphorylation of glucose molecules at the start of glycolysis. galM is an aldose-1-epimerase that catalyzes the conversion of α-D-Glucose to β-D-Glucose. galM, while not generally considered an essential glycolysis gene, often acts at the start of glucose degradation. SPO1508 is a quinoprotein ethanol dehydrogenase, a core metabolic gene.

Two genes we expected to be upregulated in glucose, gap-1 (SPO0701) and zw-2 (SPO3033), were expressed at relatively lower levels in glucose treatments. Gap-1, glyceraldehyde 3-phosphate dehydrogenase, catalyzes an important step in glycolysis. However, Gap-1 has also been shown in both bacteria and eukaryotes to fulfill a diverse array of functions; if this is also the case in *R. pomeroyi*, the regulation of Gap-1 transcripts may depend on factors other than the presence of glucose (Ferreira et al. 2015). Zw-2, glucose-6-phosphate 1-

dehydrogenase, is associated with the Enter-Doundoroff pathway for glucose degradation (KEGG). Its downregulation in glucose treatments indicates that DSS3 may not be degrading glucose via this pathway in the experimental environment. Other genes associated with glucose degradation did not show significant levels of differential expression – suggesting that these genes may be expressed at relatively constitutive levels or are not transcriptionally regulated in response to the presence of glucose.

The transcriptome of the glucose treatment cells also contained several other upregulated genes, most of which had functions not specifically linked to known pathways of glucose degradation. These "emergent genes" included putative transporters, a peroxidase, a carboxylase, an oxidoreductase, and a hypothetical protein. The most compelling result was the upregulation of a xylose operon, including the genes xylFHG (SPO0861, SPO0862, SPO0863), which are components of a xylose ABC transporter system known to transport glucose (see Transporter section above). Also upregulated are two genes located next to those on the genome – SPO0864, the putative glucokinase discussed above, and SPO0865, an oxidoreductase of the Gfo/Idh/MocA family. Other emergent genes upregulated in glucose were scattered around the genome. One, a putative MauG gene (SPO0858) is annotated as a diheme cytochrome- c peroxidase (although it does not appear to be located in a cluster of Mau genes, as MauG is in other organisms). Others were SPO1094, propionyl-CoA carboxylase beta subunit, SPO2721, a hypothetical protein, and SPO3152, a MaoC domain protein. The functional relevance of these genes to glucose degradation is not immediately apparent; their upregulation in response to glucose indicates that DSS3 responds to the presence of DOC in highly complex ways. These results suggest that the presence of glucose in the environment could be identified by the

upregulation of genes other than those already assigned to glucose degradation pathways, and not all genes in known glucose pathways are reliable indicators of the presence of glucose.

Benzoate. Benzoate is a simple aromatic compound consisting of an aromatic ring and a carboxyl group. Aromatic compounds can be common in the ocean but are often resistant to microbial degradation (Valderrama et al. 2012). As benzoate occurs as an intermediate in the degradation pathways of several aromatic compounds, it has been used as a model compound for the examination of bacterial degradation of aromatics (Valderrama et al. 2012). *R. pomeroyi's* genome contains the genes necessary for aerobic benzoate oxidation (Newton et al 2010).

Benzoate was tested at two concentrations (500 µM and 50 uM), both with an acetate background. There were significant differences in gene expression between benzoate treatments of different concentrations. Only 3 genes were significantly upregulated in all benzoate treatments, and two of those have been identified as components of the aerobic benzoate oxidation (box) pathway (the third is a hypothetical protein). The box pathway genes were enoyl–coA hydratase (SPO3700) and boxA (SPO3703).   A hypothetical gene (SPO3702) was located close to the box pathway genes, inside a putative 11-gene operon spanning locus tags SPO3694 – SPO3704. The rest of the operon genes were significantly upregulated in the high concentration benzoate treatments, but not in the 50 µM benzoate treatments. Along with BoxA and enoyl-coA hydratase, the operon contains BoxB, BzdR, putative hydrolases, transporters and hypothetical proteins.

There is significant variety in the function and type of other genes upregulated in the benzoate treatment. Eleven genes classified as oxidoreductases were upregulated in benzoate treatments. The previously-discussed *boxA* is the only oxidoreductase upregulated in both 500 µM and 50 µM benzoate treatments. Six oxidoreductases were upregulated in 500 µM benzoate

+ 500 μM acetate treatments alone. One of these, is the Benzoyl-CoA Oxygenase B subunit (SPO3701; BoxA's counterpart). Two hydrolases and three transcriptional regulators were also upregulated in 500 μM Benzoate. Many other upregulated genes were hypothetical proteins. Upregulated genes from 500 μM Benzoate treatments included a putative cyclase, a putative inositol-1-monophosphatase, two synthases, two transferases, and the TRAP transporter component in the 11-gene operon. In 50 μM Benzoate, upregulated genes include a putative BCAA ABC transporter and a 4- carboxymuconolactone decarboxylase.

The majority of these genes do not fit neatly into currently known of benzoate degradation pathways, but are highly likely to have functions associated with benzoate degradation – catabolic pathways generally include several redox reactions, and multiple oxidoreductases are upregulated, for example. The upregulation of different transcripts in different concentrations of benzoate suggests that the specificity of these responses vary by concentration- if only small concentrations of a difficult-to-degrade compound are detected, the cell may have evolved to increase expression of proteins present in multiple pathways. In addition, we observed that several branched chain amino acid ABC transporter genes (SPO3291, SPO3294-SPO3296) were significantly downregulated in Benzoate treatments.

Glycerol. Glycerol forms the backbone of most cellular membrane lipids in both prokaryotic and eukaryotic cells. Eukaryotic phytoplankton (such as diatoms and dinoflagellates) primarily store energy as glycerol derivative molecules known as triacylglycerols (TAGs; Becker 2018). In a study growing Synechococcus and Roseobacters in co-culture (Christie-Oleza et. al 2017), the Roseobacters upregulated transporters for phosphorylated glycerol molecules (glycerol-3-phosphate, G3P); suggesting that DOC released by phytoplankton may include glycerol derivatives. In addition, many bacteria can utilize glycerol as a primary source of carbon

(Da Silva et al. 2007). Glycerol taken up from the environment can be shunted into either glycolysis or anaerobic fermentation pathways, and it is an important precursor for membrane lipid biosynthesis.

R. pomeroyi's transcriptomic response to glycerol was tested at 4 different concentrations: 500, 50, 5, and 0.05 μM. Two 500 μM glycerol only treatments were run in the chemostat without an acetate background; the other seven samples (two of each concentration, except for 0.05 μM) were run with 500 μM Acetate as a background. *R. pomeroyi* has a higher growth yield on 500 μM glycerol than it does on 500 μM acetate (see table). When acetate was supplemented with glycerol, steady state OD_{600} increased with glycerol concentration. (No significant increase in OD was observed in treatments with 0.05 μM glycerol.)

Although *R. pomeroyi* has known genes for glycerol degradation pathways, these genes were mostly not upregulated in glycerol treatments. One exception is the aasadh gene (Alpha-aminoadipic semialdehyde dehydrogenase, SPO0235), which is part of a pathway feeding glycerol into glycolysis. Another exception is GlpT (SPO0610), one component of DSS3's glycerol ABC transporter. GlpT was significantly upregulated across all glycerol treatments, but the effect was more significant in treatments with higher concentrations of glycerol. The glpK gene (glycerol kinase, SPO0104) is also weakly upregulated, but only in high concentration glycerol treatments. Glycerol kinase converts glycerol to glycerol-3-phosphate (G3P), which is the first step in pathways that use glycerol to synthesize fatty acids and/or membrane lipids (specifically cardiolipin).

Other genes upregulated in glycerol treatments fall into two major categories: genes associated with amino acid metabolism and HK97 phage genes. Glycerol treatments also showed

elevated transcript numbers in select genes associated with quorum sensing, flagella construction, or metal ion transport, as well as several hypothetical proteins.

When comparing the glycerol treatments to the acetate only treatments, 9 of the 31 genes with that had a log fold change > 2 were associated with various amino acid synthesis or degradation pathways. The most significantly differentially expressed of these was the autoinducer synthesis protein (SPO0372), which is also associated with quorum sensing. KEGG annotations suggest this gene also plays a role in cysteine and methionine metabolism. Alanine dehydrogenase (SPO0222) and 2-amino-3-ketobutyrate coenzyme A ligase (SPO3360) are two other amino-acid associated genes significantly upregulated in glycerol treatments. Alanine dehydrogenase shows up in taurine and hypotaurine metabolic pathways, as well as in alanine, aspartate, and glutamate metabolism pathways. 2-amino-3-ketobutyrate coenzyme A ligase plays a role in glycine, serine, and threonine metabolism.

Seven of the most upregulated genes in glycerol are annotated as HK97 phage genes, and may be part of a GTA (gene transfer agent) cluster, which are known to occur in Roseobacter species (Zhao et al. 2009). These genes include a putative terminase large subunit (SPO2266), a putative head-tail adaptor protein (SPO2259;), and a HK97 family portal protein. Other upregulated genes of interest include imelysin (SPO0086), which may be involved in iron uptake (Xu et al. 2011), and an iron-binding component of DSS3's Iron (III) ABC transporter (SPO3287).

CONCLUSIONS

The goal of this study was to determine if there is a link between carbon substrate availability and transcript abundances in the model marine bacterium *R. pomeroyi* DSS-3, in order to improve the utility of information about the DOC pool gained from environmental gene expression data. Our results show *R. pomeroyi* does exhibit unique transcriptional profiles when grown on different carbon substrates. Most substrate treatments show patterns of significant enrichment of transcripts that are unique to growth on that substrate, and the expression profiles of same-substrate replicates cluster together in principal components analysis. In some treatments, such as glycolic acid and benzoate, known degradation pathway genes were upregulated; but that was not the case in all treatments. For many carbon substrates the primary transcriptional signal comprised enrichment of genes not previously identified with the substrate's metabolism. While most substrates were run at the same concentration (500 µM substrate with a 500 µM acetate background), we found differing concentrations of glucose and glycerol altered DSS3's transcriptomic response. There were significant differences in both the genes upregulated and the strength of the response in different-concentration benzoate treatments, but this effect was not as significant in glucose treatments. Glycerol treatments showed some differences in the strength of the upregulated signal across concentrations, but the identity of the upregulated genes did not vary as much between high and low glycerol concentration treatments.

We hypothesize that some substrate treatments such as acetate, succinate, and pyruvate that lacked a unique transcriptional signal did so because they are metabolically in proximity to the TCA cycle (and the genes associated with their degradation are likely constitutively expressed), However, the substrate with the strongest and most distinct transcriptional signal was glycine betaine with 753 genes significantly upregulated, many of which are related to C1 metabolism and amino acid metabolism genes. Betaine seemed to spur a remodeling of the cell's genome for one-carbon metabolism not seen with the other tested carbon substrates. The strength of this response suggests that betaine may be ecologically important for *R. pomeroyi*, its presence possibly heralding a change in regime or growth stage for the phytoplankton blooms with which DSS3 is associated *in situ*.

For other carbon substrates, we identified transcripts that could be used as markers of that substrate's presence. Three genes (2 of which were box genes) were upregulated across all benzoate concentrations, for example. In glucose treatments, glucokinase and a xylFHG operon were both significantly upregulated. For glycerol, known glycerol degradation pathway genes were not significantly enriched with the exception of aasadh and glpT. Most of the genes significantly differentially expressed in glycerol treatments were HK97 phage genes (likely part of a GTA cluster) or amino acid degradation pathway genes. 4-hydroxybenzoate treatments showed upregulation of three apparent operons (a 4-gene cluster containing fluoride ion transporters and beta-ketoadipate genes, a 5-gene cluster containing an Afsa domain killing gene and an immunity gene, and a 6-gene cluster with several carbon metabolism genes) on the megaplasmid, as well as some transporters (TatA/E, TatB, and TatC twin-arginine translocation proteins and a TRAP transporter DctM subunit) on the main chromosome. For N-acetyl-D-Glucosamine treatments, most upregulated genes were transport related, such as NosF. Two carbon monoxide dehydrogenase genes were also enriched in N-acetyl-D-Glucosamine

treatments. 4 known β-hydroxyaspartate cycle (BHAC) genes were significantly upregulated in glycolic acid treatments, confirming the importance of the BHAC for glycolate assimilation. Only one gene, propionyl-CoA carboxylase, alpha subunit, was upregulated in propionate, and three substrates (Succinate, pyruvate, and Tween 40, had no significantly differentially expressed genes.

By identifying distinct transcriptional signals of carbon substrate availability, this work will enable us to better interpret environmental transcriptomes, providing insight into the carbon sources being used by that microbial community and the composition of the DOC pool. Understanding which communities are degrading which carbon sources is important for understanding how carbon cycles through the environment, and for predicting how the carbon cycle might change in the future. Bacteria alter their gene expression when different carbon substrates become available, and these transcriptional responses are specific to the carbon source. As such, the upregulation of certain transcripts could be used as indicators of the presence of those carbon substrates. However, identifying known degradation pathway genes in a transcriptome is not necessarily sufficient for establishing that a microbe is using a particular carbon source. Bacterial transcriptional responses to environmental stimuli are complex. *R. pomeroyi* failed to upregulate known pathway degradation genes for several carbon sources, even when several emergent genes were enriched. Using quantitative information about which transcripts correlate to the presence of defined carbon substrates to help calibrate environmental transcript data thus will increase our understanding of microbe-mediated carbon cycling in the ocean environment.

METHODS

     Chemostat design and conditions. *Ruegeria pomeroyi* DSS-3 was grown in carbon-limited chemostats on defined salts-Marine Basal Medium. These experiments used a 140 mL culture vessel upon a stir plate, which was connected to two (one-liter or two-liter) bottles by pre-autoclaved tubing. The reservoir contained fresh ds-MBM media with a carbon substrate and flowed into the chemostat vessel. Culture media flowing from the chemostat was collected in a waste bottle. Peristaltic pumps were used to deliver and remove media. The carbon substrates and concentrations tested are shown in Table 1. Substrates were prepared by adding the appropriate mass of carbon to 45 mL of MilliQ water and filter-sterilizing the resulting solution with a 0.2 µM filter. These substrate solutions were refrigerated before use. Each sample ID (FN#) represents a separate biological replicate.

     Culture and chemostat setup. All glassware was acid washed and combusted. A DSS-3 starter culture was grown from cryo-stocks on rich media (½ strength YTSS (Yeast, Tryptone Sea Salts)), incubated overnight at 29°C in the dark, and then diluted into carbon-free ds-MBM to an OD_{600} of c.a. 0.03. One hundred µL of this diluted culture was used to inoculate the chemostat vessel, which contained 140 mL of ds-MBM media and carbon substrate. After inoculation, the pump was left off for ~40 hours or until the OD_{600} measured 0.01 in order to prevent cells from washing out of the chemostat vessel before they could become established. The chemostat vessel temperature was maintained at 29°C via a water jacket and circulating water bath. Media was stirred inside the vessel using a sterilized magnetic stir rod. After 40

hours or when the OD threshold was reached, the pumps were turned on and media began flowing into and out of the vessel. The culture was determined to be in steady state when the OD_{600} remained within 10% variance after 3 turnovers. Every experiment ran for at least 4 turnovers after the pump was turned on. Cell counted by staining with SYBR green (ThermoFisher Scientific, Massachusetts, USA) and enumerated with flow cytometry (EMD Millipore Corporation, Massachusetts, USA) and fluorescence microscopy on an Olympus BX51 Fluorescence microscope with.

Associated batch culture For chemostats 20 & 22, 20 mL of carbon-containing media was reserved from each chemostat reservoir before the start of the experiment, added to batch culture tubes, and placed in a 29 degree incubator. The ODs of the batch culture were taken along with the chemostat cultures, and the batch growth curves are in supplemental.

Cell and media collection. While the reservoir pump was still running, a serological pipet was used to remove 50-60 mL of culture from the chemostat vessel. This culture was then added to a sterile, 60 mL syringe and passed through a 0.22 μm 25 mm filter. The filter was immediately placed in a cryotube and flash frozen in liquid nitrogen for at least 5 minutes before transferring to -80°C for storage. Total time from culture collection to flash freezing averaged 6 minutes. No filtrate was saved for the first 7 samples (FN#s 5,6,7,9,11,12,13); for the rest of the samples, 40-50 mL filtrate per chemostat was frozen at -20°C for downstream chemical analysis. Duplicate technical replicates were obtained by repeating the cell collection and filtrate sampling twice for each vessel.

RNA extraction. RNA was extracted using the MirVana MiRNA isolation kit (Life Technologies, California, USA) following the manufacturer's instructions with the addition of internal RNA standard spike ins. The RNA standards were constructed as described in Gifford *et*

al., 2016. The individual standards were combined at different concentrations into three groups (see table at right). Just before starting the extraction, 20 µL of each of three pooled RNA standard groups were added to the tube containing the filter and lysis buffer. Further extraction steps were performed following the kit manufacturer's protocol. RNA was eluted from the columns in 100 uL of nuclease-free water (in two elutions of 50 uL each). RNA yield was determined using an Eppendorf BioSpectrometer Basic (Eppendorf AG, Germany) as well as a fluorescent RiboGreen assay (ThermoFisher Scientific, Massachusetts, USA). RNA fragment lengths were checked with an Agilent 2100 Bioanalyzer (Agilent Technologies, California, USA) and RNA 6000 Pico chips (Agilent Technologies, California, USA).

Residual DNA was removed with Turbo DNAse (ThermoFisher Scientific, Massachusetts, USA) and remaining RNA quantities were checked on the spectrophotometer. cDNA libraries were prepared using the ScriptSeq v2 RNA-seq Library Preparation Kit (Illumina, California, USA) following the manufacturer's protocol and barcoded with ScriptSeq Index Primers. The input RNA amount was approximately 68 ng on average, with a standard deviation of 28 ng. 15 cycles of 95/55/68 PCR were performed for FN#s 19 and 21; 13 cycles of 95/55/68 PCR were performed for FN#s 9, 11, 12, 13, 23, 30, 34, and 40; 12 cycles of 95/55/68 PCR were performed for FN#s 5,7, 32, 38, 6, 14, 27, and 36. The Agencourt AMPure XP system (Beckman Coulter, Indiana, USA) was used for cDNA library purification. The resulting cDNA library fragments were analyzed using an Agilent High Sensitivity DNA Chip (Agilent Technologies, California, USA) on the 2100 Bioanalyzer. Each cDNA library concentration was quantified fluorescently using PicoGreen (Thermofisher Scientific, Massachusetts, USA) and then equimolarly pooled to ~20 nM. The pooled samples were sequenced on 1 lane of a 50 bp PE HiSeq 4000 (Illumina, California, USA) run.

Sequence processing. Raw reads were processed using a local linux server. Read quality was first checked with FastQC (v0.11.9) using default parameters. Trimmomatic (version 0.39) was run with simple mode settings for single-end reads, with a sliding window size of 5 bp and a required quality score of >=20. FastQC was used to check the trimmed reads before they were converted from fastq to fasta format (using the fastq_to_fasta command). A homology search was then performed using Megablast against a custom database containing rRNA nucleotide sequences from *R. pomeroyi* DSS3, *V. fischeri* ES114, *V. fischeri* MJ11, and *Sulfolobus solfataricus*. rRNA sequences with >=90% identity and bitscores >50 were counted and removed from further analysis. Internal standards were identified with a Megablast homology search against the genome of *Sulfolobus solfataricus*, with quality cutoffs >=95 percent identity, >=50 bitscore, and >50% alignment length. Sequences identified as internal standards were removed from further sample processing. Bowtie2 was used to map reads to the *R. pomeroyi* genome (Rivers 2014), using the end-to-end alignment mode with default parameters (--sensitive: -D 15 - R 2 -N 0 -L 22 -i S,1,1.15). HTSeq count was used to count mapped reads, using stranded settings and a minimum alignment quality score of 10. HTSeq was run in intersection-nonempty mode for overlap resolution, with nonunique- none settings (overlapping reads were marked as ambiguous and not included in counts). HTSeq was set to ignore supplementary alignments. Gene expression between treatments was determined in R using the DeSeq2 Bioconductor package (Release 3.10) with the LRT test. (Love MI, Huber W, Anders S (2014)) The program DEGreport was also used to identify patterns of gene expression (Pantano L (2021). The pcaExplorer Bioconductor package was used for PCA analysis (Marini F, Binder H (2019)).

Internal Standards. Internal standards were identified with a Megablast homology search against the genome of Sulfolobus solfataricus, with quality cutoffs >=95 percent identity,

>=50 bitscore, and >50% alignment length. The reads for nine specific internal standard sequences were recovered and counted. The reads recovered were compared to the known amounts of internal standard added for each standard sequence, and an average conversion factor (equal to the number of standard molecules added/ the number of standard reads recovered) was calculated for each sample. This conversion factor was then used to convert the recovered transcript reads to transcripts/cell.

REFERENCES

Becker, K.W., Collins, J.R., Durham, B.P. et al. Daily changes in phytoplankton lipidomes reveal mechanisms of energy storage in the open ocean. Nat Commun 9, 5179 (2018). https://doi.org/10.1038/s41467-018-07346-z

Benson PJ, Purcell-Meyerink D, Hocart CH, Truong TT, James GO, Rourke L, Djordjevic MA, Blackburn SI and Price GD (2016) Factors Altering Pyruvate Excretion in a Glycogen Storage Mutant of the Cyanobacterium, Synechococcus PCC7942. Front. Microbiol. 7:475. doi: 10.3389/fmicb.2016.00475

Chan Leong-Keat, Newton Ryan, Sharma Shalabh, Smith Christa, Rayapati Pratibha, Limardo Alexander, Meile Christof, Moran Mary Ann. 2012. Stoichiometry Shifts in a Marine Heterotrophic Bacterium. Frontiers in Microbiology. 3: 159. doi: 10.3389/fmicb.2012.00159

Chen JK, Shen CR, Liu CL. N-acetylglucosamine: production and applications. Mar Drugs. 2010;8(9):2493-2516. Published 2010 Sep 15. doi:10.3390/md8092493

Cho, K. Y., & Salton, M. R. J. (1966). Fatty acid composition of bacterial membrane and wall lipids. Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism, 116(1), 73–79. doi:10.1016/0005-2760(66)90093-2

Christie-Oleza JA, Sousoni D, Lloyd M, Armengaud J, Scanlan DJ. Nutrient recycling facilitates long-term stability of marine microbial phototroph-heterotroph interactions. Nat Microbiol. 2017 Jun 26;2:17100. doi: 10.1038/nmicrobiol.2017.100. PMID: 28650444; PMCID: PMC5495174.

Da Silva, Gervásio, Mack, Matthias, Contiero, Jonas. 2009. Glycerol: A promising and abundant carbon source for industrial microbiology. Biotechnology Advances. 27: 30-39. Doi: 10.1016/j.biotechadv.2008.07.006

Federico Marini, Harald Binder (2019). pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components URL http://bioconductor.org/packages/pcaExplorer/, https://doi.org/10.1186/s12859-019-2879-1

Fogg, G. E. (1983). The Ecological Significance of Extracellular Products of Phytoplankton Photosynthesis. Botanica Marina, 26(1). doi:10.1515/botm.1983.26.1.3

Gad, S. C., & Gad, S. E. (2005). Propionic Acid*. Encyclopedia of Toxicology, 536–538. doi:10.1016/b0-12-369400-0/00802-4

Glombitza C, Jaussi M, Røy H, Seidenkrantz M-S, Lomstein BA and Jørgensen BB (2015) Formate, acetate, and propionate as substrates for sulfate reduction in sub-arctic sediments of Southwest Greenland. Front. Microbiol. 6:846. doi: 10.3389/fmicb.2015.00846

Glombitza C., Pedersen J., Røy H., Barker Jørgensen B. 2014. Direct analysis of volatile fatty acids in marine sediment porewater by two-dimensional ion chromatography-mass spectrometry. Limnol. Oceanogr. Methods 12: 455-468. https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.4319/lom.2014.12.455

Hammen, C. S. (1975). Succinate and lactate oxidoreductases of bivalve mollusks. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry, 50(3), 407–412. doi:10.1016/0305-0491(75)90250-3

Harwood, C. S., & Parales, R. E. (1996). THE β-KETOADIPATE PATHWAY AND THE BIOLOGY OF SELF-IDENTITY. Annual Review of Microbiology, 50(1), 553–590. doi:10.1146/annurev.micro.50.1.553

Jones, H.J., Kröber, E., Stephenson, J. et al. A new family of uncultivated bacteria involved in methanogenesis from the ubiquitous osmolyte glycine betaine in coastal saltmarsh sediments. Microbiome 7, 120 (2019). https://doi.org/10.1186/s40168-019-0732-4

Juliano, C.; Magrini, G.A. Cosmetic Ingredients as Emerging Pollutants of Environmental and Health Concern. A Mini-Review. Cosmetics 2017, 4, 11. https://doi.org/10.3390/cosmetics4020011.

Kim, J., Hirasawa, T., Sato, Y. et al. Effect of odhA overexpression and odhA antisense RNA expression on Tween-40-triggered glutamate production by Corynebacterium glutamicum . Appl Microbiol Biotechnol 81, 1097–1106 (2009). https://doi.org/10.1007/s00253-008-1743-4

Kubitschek, Herbert Ernest. Introduction to Research with Continuous Cultures. 1970, Prentice-Hall.

Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014). https://doi.org/10.1186/s13059-014-0550-8

Madden T., Ward J., Ison A. 1996. Organic acid excretion by Streptomyces lividans TK24 during growth on defined carbon and nitrogen sources. Microbiology. 142,3181-3185 doi: https://doi.org/10.1099/13500872-142-11-3181

Mary Ann Moran, Elizabeth B. Kujawinski, Aron Stubbins, Rob Fatland, Lihini I. Aluwihare, Alison Buchan, Byron C. Crump, Pieter C. Dorrestein, Sonya T. Dyhrman, Nancy J. Hess, Bill Howe, Krista Longnecker, Patricia Medeiros, Jutta Niggemann, Ingrid Obernosterer, Daniel J. Repeta, Jacob R. Waldbauer. 2016. Deciphering ocean carbon in a changing world. Proceedings of the National Academy of Sciences, 113 (12) 3143-3151; DOI: 10.1073/pnas.1514645113

Monod, J. 1942. The Growth of Bacterial Cultures. Annual Review of Microbiology. 3:1, 371-394

Morán, Xosé Anxelu & Ducklow, Hugh & Erickson, Matthew. (2013). Carbon fluxes through estuarine bacteria reflect coupling with phytoplankton. Marine Ecology Progress Series. 489. 10.3354/meps10428.

Mühlenbruch, M., Grossart, H.□P., Eigemann, F. and Voss, M. (2018), Mini□review: Phytoplankton□derived polysaccharides in the marine environment and their interactions with heterotrophic bacteria. Environ Microbiol, 20: 2671-2685. doi:10.1111/1462-2920.14302

Newton R. J., Griffin L. E., Bowles K. M., Meile C., Gifford S., Givens C. E., Howard E. C., King E., Oakley C. A., Reisch C. R., Rinta-Kanto J. M., Sharma S., Sun S., Varaljay V., Vila-Costa M., Westrich J. R., Moran M. A. (2010). Genome characteristics of a generalist marine bacterial lineage. ISME J. 4, 784–79810.1038/ismej.2009.150

Obernosterer I., Kraay G., de Ranitz E., Herndl G. 1999. Concentrations of low molecular weight carboxylic acids and carbonyl compounds in the Aegean Sea (Eastern Mediterranean) and the turnover of pyruvate. Aquatic Microbial Ecology. Vol. 20: 147-156.

Oren, A. Formation and breakdown of glycine betaine and trimethylamine in hypersaline environments. Antonie van Leeuwenhoek 58, 291–298 (1990). https://doi.org/10.1007/BF00399342

Pantano L (2021). DEGreport: Report of DEG analysis. R package version 1.28.0, http://lpantano.github.io/DEGreport/)

Peng X, Adachi K, Chen C, et al. Discovery of a marine bacterium producing 4-hydroxybenzoate and its alkyl esters, parabens. Appl Environ Microbiol. 2006;72(8):5556-5561. doi:10.1128/AEM.00494-06

Rich, J.H., Ducklow, H.W., Kirchman, D.L., 1996. Concentrations and uptake of neutral monosaccharides along 1401W in the Equatorial Pacific: contribution of glucose to heterotrophic bacterial activity and the DOM flux. Limnology and Oceanography 41, 595–604

Riemann L, Azam F. Widespread N-acetyl-D-glucosamine uptake among pelagic marine bacteria and its ecological implications. Appl Environ Microbiol. 2002;68(11):5554-5562. doi:10.1128/AEM.68.11.5554-5562.2002

Rivers, A. R., Smith, C. B., & Moran, M. A. (2014). An Updated genome annotation for the model marine bacterium Ruegeria pomeroyi DSS-3. Standards in genomic sciences, 9, 11. https://doi.org/10.1186/1944-3277-9-11

Ruby EG, Nealson KH. Pyruvate production and excretion by the luminous marine bacteria. Appl Environ Microbiol. 1977;34(2):164-169. doi:10.1128/aem.34.2.164-169.1977

Schada von Borzyskowski, L., Severi, F., Krüger, K. et al. Marine Proteobacteria metabolize glycolate via the β-hydroxyaspartate cycle. Nature 575, 500–504 (2019). https://doi.org/10.1038/s41586-019-1748-4

Schada von Borzyskowski, L., Severi, F., Krüger, K. et al. Marine Proteobacteria metabolize glycolate via the β-hydroxyaspartate cycle. Nature 575, 500–504 (2019). https://doi.org/10.1038/s41586-019-1748-4

Schink B and Pfennig N. 1982. Propionigenium modestum gen. nov. sp. nov. a New Strictly Anaerobic,Nonsporing Bacterium Growing on Succinate. Archives of Microbiology. 133: 209-216. https://kops.uni-konstanz.de/bitstream/handle/123456789/8671/Propionigenium_modestum_gen._nov._sp._nov._1982.pdf?sequence=1&isAllowed=y

Schiweck, H., Bär, A., Vogel, R., Schwarz, E., Kunz, M., Dusautois, C., … Peters, S. (2012). Sugar Alcohols. Ullmann's Encyclopedia of Industrial Chemistry. doi:10.1002/14356007.a25_413.pub3

Sharpe GC, Gifford SM, Septer AN. 2020. A model Roseobacter, Rugeria pomeroyi DSS-3, employs a diffusible killing mechanism to eliminate competitors. mSystems 5:4. doi: 10.1128/mSystems.00443-20Song H and Lee SY. 2006. Production of succinic acid by bacterial fermentation. Enzyme and Microbial Technology. 39: 352-361. https://doi.org/10.1016/j.enzmictec.2005.11.043

Teira, E., Martínez-García, S., Lønborg, C. and Álvarez-Salgado, X.A. (2009), Growth rates of different phylogenetic bacterioplankton groups in a coastal upwelling system. Environmental Microbiology Reports, 1: 545-554. https://doi.org/10.1111/j.1758-2229.2009.00079.x

Valderrama JA, Durante-Rodríguez G, Blázquez B, García JL, Carmona M, Díaz E. Bacterial degradation of benzoate: cross-regulation between aerobic and anaerobic pathways. J Biol Chem. 2012 Mar 23;287(13):10494-10508. doi: 10.1074/jbc.M111.309005. Epub 2012 Feb 2. PMID: 22303008; PMCID: PMC3322966.

Wright, R.T., Shah, N.M. The trophic role of glycolic acid in coastal seawater. I. Heterotrophic metabolism in seawater and bacterial cultures. Marine Biology 33, 175–183 (1975). https://doi.org/10.1007/BF00390723

X. A. G. Morán, Estrada, M., J. M. Gasol, & Pedrós-Alió, C. (2002). Dissolved Primary Production and the Strength of Phytoplankton-Bacterioplankton Coupling in Contrasting Marine Regions. Microbial Ecology, 44(3), 217-223. http://www.jstor.org/stable/4287650

Yan, D., Kang, J. & Liu, DQ. Genomic analysis of the aromatic catabolic pathways from Silicibacter pomeroyi DSS-3. Ann. Microbiol. 59, 789–800 (2009). https://doi.org/10.1007/BF03179225

Yoch D. C. (2002). Dimethylsulfoniopropionate: its sources, role in the marine food web, and biological degradation to dimethylsulfide. Applied and environmental microbiology, 68(12), 5804–5815. https://doi.org/10.1128/AEM.68.12.5804-5815.2002

Zhao, Y., Wang, K., Budinoff, C. *et al.* Gene transfer agent (GTA) genes reveal diverse and dynamic *Roseobacter* and *Rhodobacter* populations in the Chesapeake Bay. *ISME J* **3,** 364–373 (2009). https://doi.org/10.1038/ismej.2008.115

Zhuang, G.-C., Peña-Montenegro, T. D., Montgomery, A., Montoya, J. P., & Joye, S. B. (2019). Significance of acetate as a microbial carbon and energy source in the water column of Gulf of Mexico: implications for marine carbon cycling. Global Biogeochemical Cycles. doi:10.1029/2018gb006129

Zou, H., Chen, N., Shi, M. et al. The metabolism and biotechnological application of betaine in microorganism. Appl Microbiol Biotechnol 100, 3865–3876 (2016). https://doi.org/10.1007/s00253-016-7462-3

TABLES

Table 1. Carbon substrates added to the chemostats. 'Experimental concentration' is the concentration of a compound in the reservoir media. '#Cs' is the number of carbons in the compound. 'KEGG structure' and 'KEGG pathway' are the accession number in the KEGG database for the chemical compound and its respective metabolic degradation pathway. 'Steps to TCA' is the approximate number of metabolic steps that the compound must pass through before entering the TCA cycle as acetyl-CoA. 'OD' is the OD600 of the chemostat with the compound at the time of cell collection, with 1 and 2 being the duplicate chemostat runs of that treatment. 'Cell#' is the theoretically number of cells based on the OD600 calculation for the duplicates.

substrate	experimental concentrations	#Cs	KEGG structure	compound type	KEGG pathway	Steps to TCA (acetyl-CoA)	OD 1	OD 2	Cell# yield	Cell# yield	avg cell number
acetate	500 µM	2	C00033	carboxylic acid	sil00010	0	0.006	0.009	9.0E+07	1.3E+08	1.11E+08
pyruvate	500 µM + 500 µM acetate	3	C00022	Alpha-keto acid	sil00620	2	0.019	0.018	1.0E+08	1.7E+08	1.36E+08
Succinate	500 µM + 500 µM acetate	4	C00042	dicarboxylic acid	sil00640	3	0.025	0.022	2.4E+08	6.0E+08	4.15E+08
Glycolic Acid	500 µM + 500 µM acetate	2	C00160	Alpha-hydroxyacid	sil00630	3	0.011	0.014	1.4E+08	1.5E+08	1.44E+08
Propionate	500 µM + 500 µM acetate	3	C00163	carboxylic acid	sil00640	4	0.024	0.02	1.5E+08	2.3E+08	1.91E+08
4-hydroxybenzoate (PHBA)	500 µM + 500 µM acetate	7	C00156	hydroxy aromatic carboxylic acid	sil00362	7	0.015	0.2	1.6E+08	2.1E+08	1.85E+08
glycine betaine	500 µM + 500 µM acetate	5	C00719	trimethylated amino acid	sil00260	7	ND	ND	4.6E+08	3.8E+08	4.18E+08
benzoate	500 µM + 500 µM acetate	7	C00180	aromatic carboxylic acid	sil00362	10	ND	ND	1.7E+08	8.8E+07	1.29E+08
glucose*	1000 µM + 500 µM acetate	6	C00267	monosaccharide	sil00010	11	0.066	0.003	2.8E+08	ND	2.82E+08
N-acetyl-D-Glucosamine	500 µM + 500 µM acetate	8	C00140	Amide derivative of glucose	sil00520	14	0.052	0.052	2.5E+08	6.7E+08	4.58E+08
glycerol	500 µM + 500 µM acetate	3	C00116	polyol	sil00561	14	ND	ND	2.7E+08	2.6E+08	2.62E+08
polysorbate 40	500 µM + 500 µM acetate	62	D05566	polysorbate	NA	16	0.01	0.013	1.7E+08	2.2E+08	1.96E+08

Table 2. Sequencing results. Samples were sequenced in two batches, or 'Sample groups'. An rRNA subtraction was performed before sequencing for sample group 2, but not sample group 1. This difference in rRNA depletion likely resulted in undersequencing of transcripts for the first sample group. 'Treatment' is the concentration and identity of carbon substrates in the media of a single chemostat experiment. 'FN' is the sample identification number. 'Raw reads' is the number of unprocessed RNA reads sequenced '% of reads mapped to DSS3' represents the percentage of reads mapped to *R. pomeroyi DSS3*. *Sulfolobus solfataricus* transcripts were added into samples as an internal standard. The *Sulfolobus* reads recovered were compared to the known amounts of *Sulfolobus* internal standard added for each standard sequence, and an average conversion factor (equal to the number of standard molecules added/ the number of standard reads recovered) was calculated for each sample. This conversion factor was then used to convert the recovered transcript reads to mRNA transcripts/cell. 'Cell yield' represents the calculated number of cells in the chemostat vessel when harvested.

Sample group	Treatment	Steady state	FN	Raw reads	% reads mapped to Sulfolobus	% reads mapped to DSS3	% rRNA reads	mapped mRNA reads	% of mRNA reads	Alignment Rate	Transcripts/read (conversion number)	Cell yield (cells/mL)	mRNA transcripts /cell
1	500 µM Acetate and 500 µM Glycerol	Yes	5	1.54E+07	0.14	6.23E-06	93.20	3.76E+05	2.45	76.42	5.54E+05	1.60E+10	12.99631
1	500 µM Glycerol	Yes	6	1.42E+07	0.28	6.12E-06	83.38	4.71E+05	3.32	88.65	2.28E+05	1.05E+10	10.2608
1	50 µM Glycerol and 500 µM Acetate	Yes	7	1.49E+07	0.19	6.30E-06	92.15	2.49E+05	1.67	51.61	3.06E+05	7.24E+09	10.50471
1	5 µM Glycerol and 500 µM Acetate	Yes	9	1.56E+07	0.21	6.16E-06	94.61	2.29E+05	1.47	66.03	2.58E+05	6.42E+09	9.194577
1	500 µM Glycerol	Yes	11	1.72E+07	0.23	5.39E-06	87.88	8.55E+05	4.96	87.71	2.26E+05	9.68E+09	19.91756
1	50 µM Glycerol and 500 µM Acetate	Yes	12	1.56E+07	0.17	6.13E-06	93.63	2.62E+05	1.68	65.80	3.29E+05	6.52E+09	13.18342
1	50 µM Benzoate and 500 µM Acetate	Yes	13	1.77E+07	0.38	5.38E-06	92.66	4.63E+05	2.62	78.10	1.47E+05	7.56E+09	8.989389
1	500 µM Glucose and 500 µM Acetate	No	14	1.48E+07	0.14	6.24E-06	86.19	9.38E+05	6.33	83.74	4.09E+05	2.92E+10	13.14919
1	500 µM Glycerol and 500 µM Acetate	Yes	19	1.84E+07	0.29	5.01E-06	85.40	1.25E+06	6.82	89.83	1.86E+05	1.28E+10	18.19808
1	50 µM Benzoate and 500 µM Acetate	Yes	21	2.08E+07	0.38	4.56E-06	90.98	7.51E+05	3.62	80.65	1.16E+05	6.88E+09	12.67057
1	500 µM Benzoate and 500 µM Acetate	No	23	9.31E+06	0.26	9.99E-06	87.67	4.98E+05	5.35	83.77	3.67E+05	1.01E+10	18.0157
1	500 µM Acetate	Yes	27	1.65E+07	0.31	5.72E-06	89.73	7.37E+05	4.47	84.57	1.79E+05	5.74E+09	22.94468
1	50 µM Glucose and 500 µM Acetate	No	30	1.90E+07	0.15	4.82E-06	87.00	8.32E+05	4.39	69.04	2.87E+05	7.10E+09	33.65012
1	500 µM Acetate	Yes	32	1.57E+07	0.43	6.07E-06	91.80	5.23E+05	3.33	82.58	1.35E+05	4.82E+09	14.69138
1	500 µM Acetate	Yes	34	1.35E+07	0.53	6.85E-06	86.99	7.74E+05	5.72	86.62	1.24E+05	4.34E+09	22.1832
1	500 µM Benzoate and 500 µM Acetate	No	36	1.22E+07	0.34	7.78E-06	90.76	5.25E+05	4.30	88.18	2.23E+05	5.29E+09	22.10735
1	5 µM Glycerol and 500 µM Acetate	Yes	38	1.61E+07	0.39	5.90E-06	91.79	4.98E+05	3.10	78.90	1.59E+05	3.75E+09	21.05671
1	0.05 µM Glycerol and 500 µM Acetate	Yes	40	1.74E+07	0.33	5.33E-06	88.14	7.78E+05	4.48	74.43	1.40E+05	4.36E+09	24.97295
2	50 µM Glucose	Yes	44	1.41E+07	0.40	6.58E-06	68.22	3.53E+06	24.93	99.10	121668.4795	1.88E+09	227.7263
2	500 µM Succinate + 500 µM Acetate	No	48	1.40E+07	0.07	6.62E-06	63.89	4.00E+06	28.62	98.53	900289.3996	1.41E+10	255.7641
2	500 µM Propionate + 500 µM Acetate	Yes	52	1.51E+07	0.07	5.94E-06	55.28	5.16E+06	34.25	99.34	650114.0782	1.37E+10	244.9225
2	500 µM Acetate	Yes	58	1.33E+07	0.10	6.87E-06	62.59	3.87E+06	29.02	99.42	2548198.722	7.95E+09	1240.944
2	500 µM Betaine + 500 µM Acetate	Yes	60	1.52E+07	0.03	6.15E-06	65.73	4.15E+06	27.41	99.33	1244740.752	2.73E+10	189.5063
2	500 µM Succinate + 500 µM Acetate	Yes	62	1.07E+07	0.06	8.59E-06	66.89	2.66E+06	24.87	98.95	1368828.184	3.57E+10	101.7491
2	500 µM N-acetyl-D-Glucosamine + 500 µM Acetate	Yes	64	1.09E+07	0.03	8.59E-06	67.61	2.78E+06	25.61	99.12	1141598.136	4.01E+10	79.12104
2	500 µM 4-hydroxybenzoate + 500 µM Acetate	Yes	67	1.23E+07	0.16	7.45E-06	59.93	3.85E+06	31.38	99.38	718112.7803	1.23E+10	224.4034
2	500 µM Acetate	Yes	68	1.78E+07	0.10	5.25E-06	69.51	4.32E+06	24.23	99.30	392825.3555	5.51E+09	308.2635
2	500 µM Glycolic Acid + 500 µM Acetate	Yes	70	1.71E+07	0.35	5.47E-06	73.90	3.38E+06	19.75	99.04	127829.3224	8.42E+09	51.34505
2	500 µM Tween 40 + 500 µM Acetate	Yes	72	2.07E+07	0.05	2.93E-06	37.69	4.79E+06	23.11	99.59	734199.6303	1.04E+10	337.5182
2	500 µM Pyruvate + 500 µM Acetate	Yes	74	2.02E+07	0.07	2.86E-06	31.57	5.30E+06	26.21	99.53	551813.9955	1.01E+10	289.3961
2	500 µM 4-hydroxybenzoate + 500 µM Acetate	Yes	76	1.61E+07	0.04	4.74E-06	56.35	3.26E+06	20.19	99.52	1393573.724	9.86E+09	460.5439
2	500 µM Acetate	Yes	86	2.07E+07	0.09	3.08E-06	46.56	3.60E+06	17.35	99.49	451807.4951	5.42E+09	299.9758
2	500 µM Tween 40 + 500 µM Acetate	Yes	88	1.21E+07	0.19	7.61E-06	64.48	3.35E+06	27.69	99.42	314047.7272	1.31E+10	80.41735
2	500 µM Glycolic Acid + 500 µM Acetate	Yes	90	1.38E+07	0.09	6.82E-06	71.00	3.15E+06	22.86	98.93	680810.5682	8.91E+09	240.4452
2	500 µM Betaine + 500 µM Acetate	Yes	92	1.61E+07	0.07	5.69E-06	53.50	6.18E+06	38.30	99.23	1102015.662	2.28E+10	298.3077
2	500 µM N-acetyl-D-Glucosamine + 500 µM Acetate	Yes	96	1.40E+07	0.03	6.74E-06	72.07	3.09E+06	22.11	98.88	1984255.234	1.49E+10	411.0782
2	500 µM Propionate + 500 µM Acetate	No	98	1.46E+07	0.05	5.98E-06	66.38	3.04E+06	20.81	99.19	1076166.156	9.16E+09	356.8808
2	500 µM Pyruvate + 500 µM Acetate	No	100	1.57E+07	0.00	4.45E-06	53.36	2.56E+06	16.35	99.48	1308753076	6.1E+09	549955.4
2	1 mM Glucose + 500 µM Acetate	No	102	1.46E+07	0.03	6.18E-06	65.32	3.63E+06	24.88	99.32	2487215.277	1.69E+10	534.5404

Table 3. Number of differentially expressed genes by carbon substrate treatment. 'DE genes' are significantly differentially expressed genes. "# of DE genes (p< 0.05)' is the number of genes that were significantly upregulated or downregulated for a particular carbon substrate. '# of DE genes upregulated (p< 0.05)' excludes downregulated genes to show only the genes upregulated in a carbon substrate's transcriptional signal.

Substrate	# of DE Genes (p < 0.05)	# of DE Genes upregulated (p < 0.05)
N-acetyl-D-Glucosamine	42	28
Glycolic Acid	4	4
Propionate*	1	1
4-hydroxybenzoate	27	27
Betaine	1319	753
Succinate	0	0
Pyruvate	0	0
Tween 40	0	0

*= The gene upregulated in propionate treatments was differentially expressed only at an adjusted p-value of 0.056, while significance for the other treatments was at a p-value of < 0.05

Table 4. Sample ID (FN#) and corresponding carbon treatment

Sample ID (FN)	Treatment
FN 5	500 µM Acetate and 500 µM Glycerol
FN 6	500 µM Glycerol
FN 7	50 µM Glycerol and 500 µM Acetate
FN 9	5 µM Glycerol and 500 µM Acetate
FN 11	500 µM Glycerol
FN 12	50 µM Glycerol and 500 µM Acetate
FN 13	50 µM Benzoate and 500 µM Acetate
FN 14	500 µM Glucose and 500 µM Acetate
FN 19	500 µM Glycerol and 500 µM Acetate
FN 21	50 µM Benzoate and 500 µM Acetate
FN 23	500 µM Benzoate and 500 µM Acetate
FN 27	500 µM Acetate
FN 30	50 µM Glucose and 500 µM Acetate
FN 32	500 µM Acetate
FN 34	500 µM Acetate
FN 36	500 µM Benzoate and 500 µM Acetate
FN 38	5 µM Glycerol and 500 µM Acetate
FN 40	0.05 µM Glycerol and 500 µM Acetate
FN 42	500 µM Acetate
FN 44	50 µM Glucose
FN 48	500 µM Succinate + 500 µM Acetate
FN 50	500 µM Acetate
FN 52	500 µM Propionate + 500 µM Acetate
FN58	500 µM Acetate
FN 60	500 µM Betaine + 500 µM Acetate
FN 62	500 µM Succinate + 500 µM Acetate
FN 64	500 µM N-acetyl-D-Glucosamine + 500 µM Acetate
FN 67	500 µM 4-hydroxybenzoate + 500 µM Acetate
FN 68	500 µM Acetate
FN 70	500 µM Glycolic Acid + 500 µM Acetate
FN 72	500 µM Tween 40 + 500 µM Acetate
FN 74	500 µM Pyruvate + 500 µM Acetate
FN 76	500 µM 4-hydroxybenzoate + 500 µM Acetate
FN 86	500 µM Acetate
FN 88	500 µM Tween 40 + 500 µM Acetate
FN 90	500 µM Glycolic Acid + 500 µM Acetate
FN 92	500 µM Betaine + 500 µM Acetate
FN 94	500 µM Acetate
FN 96	500 µM N-acetyl-D-Glucosamine + 500 µM Acetate
FN 98	500 µM Propionate + 500 µM Acetate
FN 100	500 µM Pyruvate + 500 µM Acetate
FN 102	1 mM Glucose + 500 µM Acetate

Table 5. Internal Standard additions.

IMG_ID	Standard	Group	Molecules
638162349	s5	1	1.97E+09
638162599	s6	1	2.35E+08
638162344	s3	1	1.90E+07
638162924	s11	2	1.10E+09
638162545	s4	2	2.31E+08
638163035	s10	2	1.33E+07
638162483	s7	3	1.72E+09
638163115	s13	3	3.07E+08
638164210	s15	3	8.92E+06

FIGURES
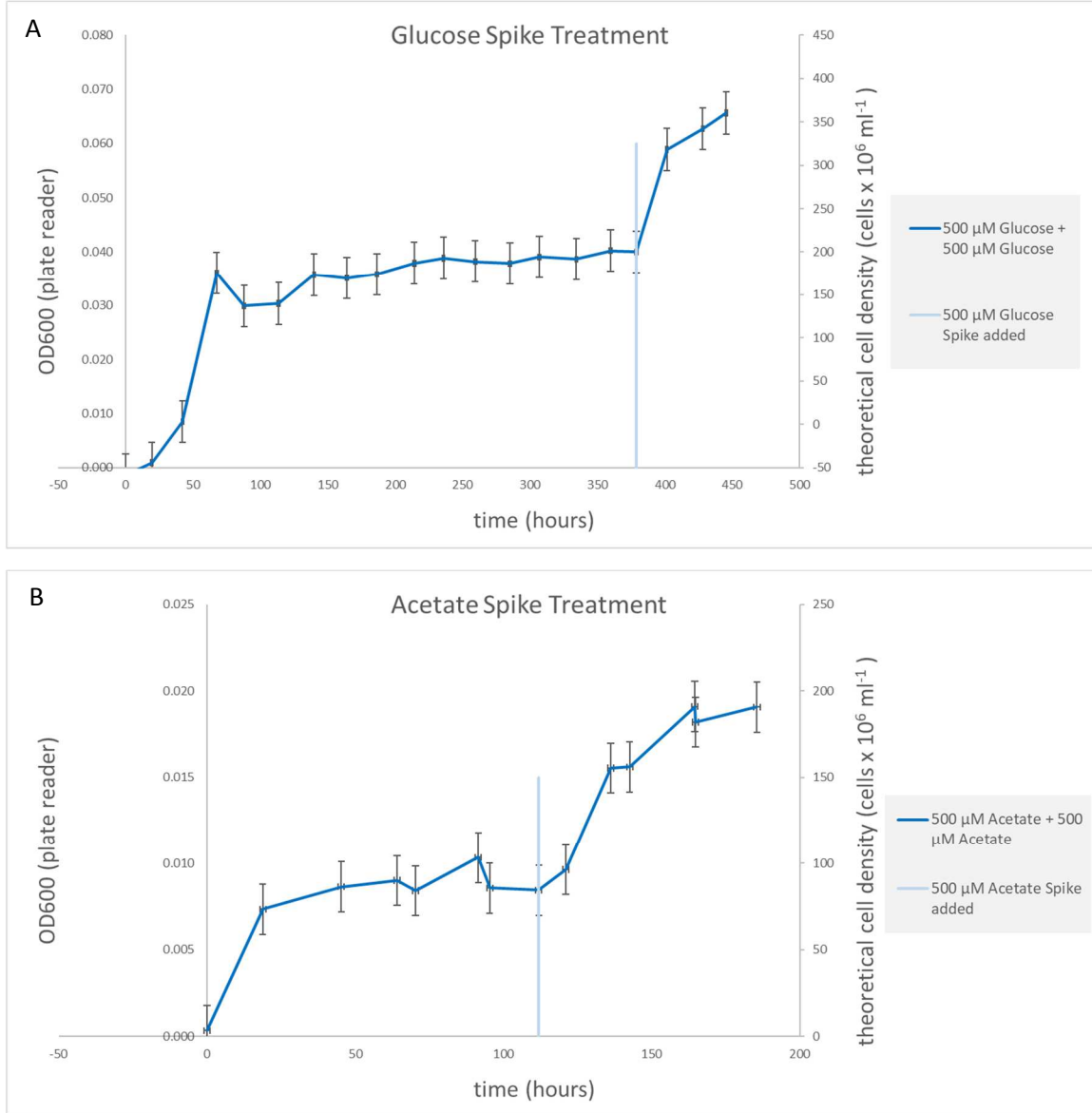
Figure 1. Carbon spike-in treatments to establish carbon limitation. The chemostat was allowed to enter steady state with 500 µM of the substrate medium, then an additional 500 µM substrate was spiked into the reservoir. A) Glucose used as the carbon limiting substrate and spike-in. B) Acetate used as the carbon limiting substrate and spike in.
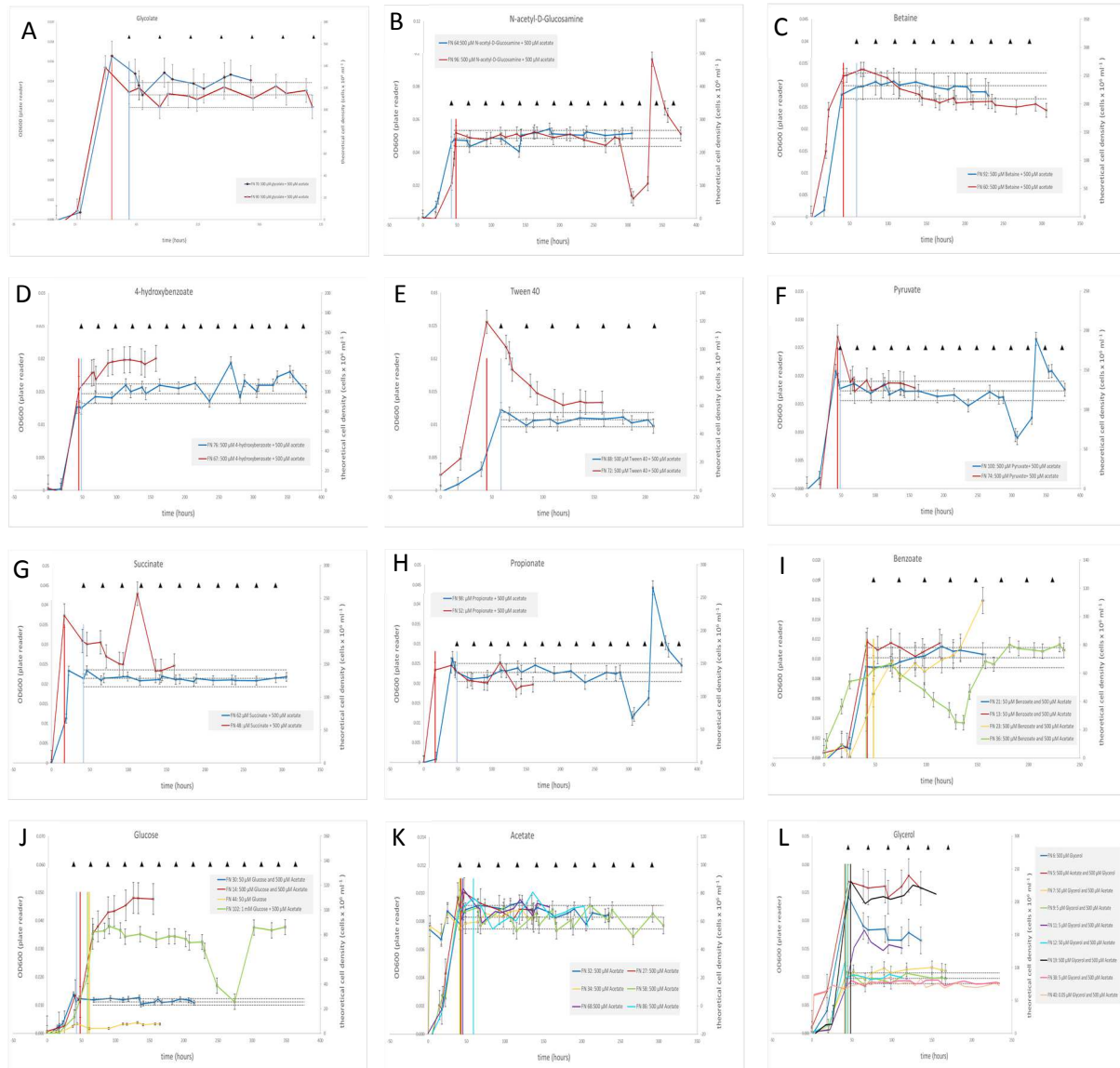
Figure 2. OD$_{600}$ time series for the chemostat experiments. Each triangle represents one turnover (chemostat turnover time (τ) represents the amount of time it takes for the media in a chemostat culture vessel to be completely replaced by new media). The vertical lines represent the time at which the pump was turned on (cells were allowed to grow after inoculation to prevent washout). Dotted lines show 10% variance from the mean OD after the pump was turned on for one replicate. NOTE: substantial drops and spikes in OD in figures B, F, and H around the 300-hour mark are due to a stir plate malfunction, which caused cells to collect at the bottom of the vessel before becoming resuspended after the stir plate was fixed.
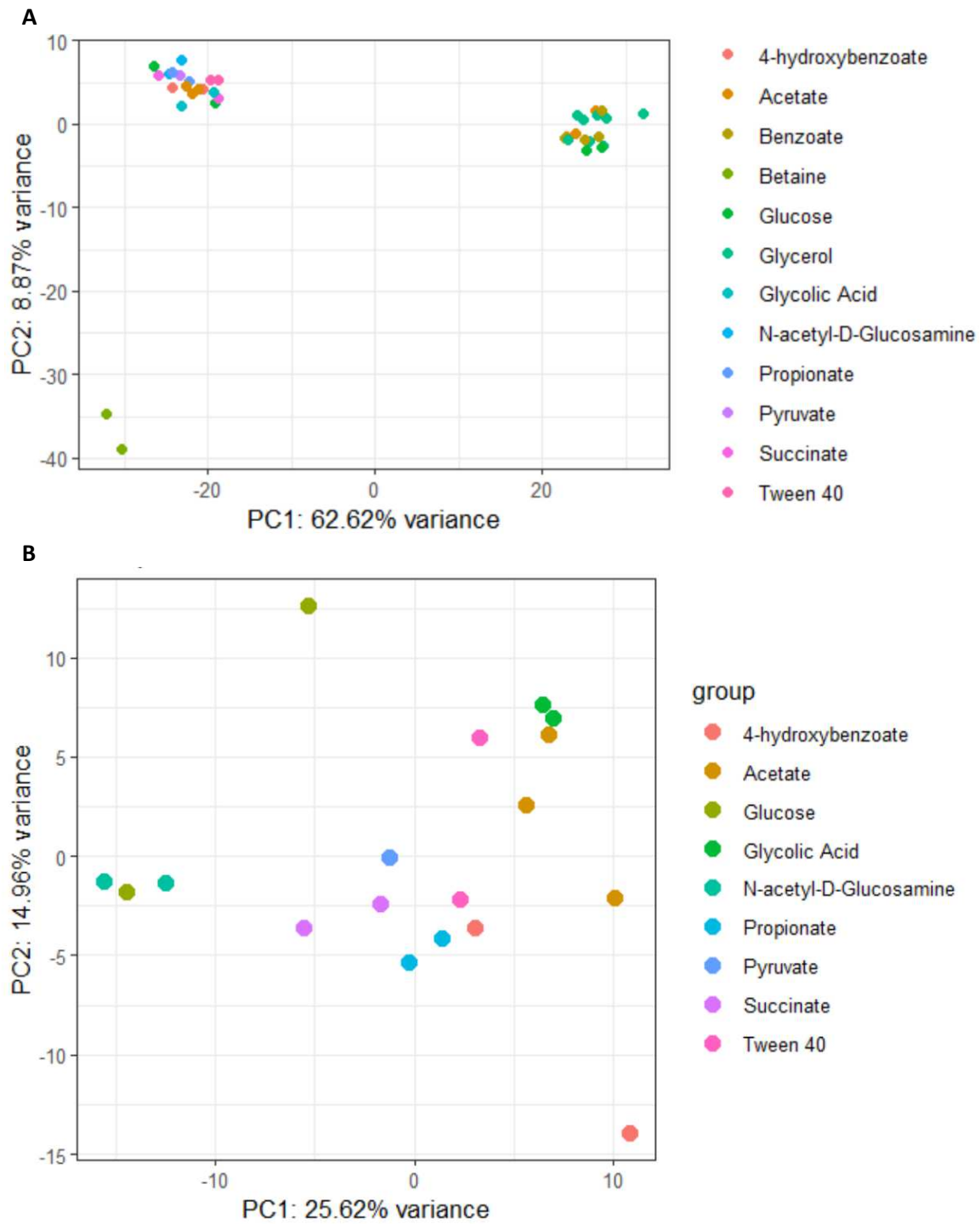
Figure 3. Principal Components Analysis of Gene Expression. Figure 3A shows a PCA for all samples and sample groups. Figure 3B shows a PCA for sample group 2, excluding betaine treatments.
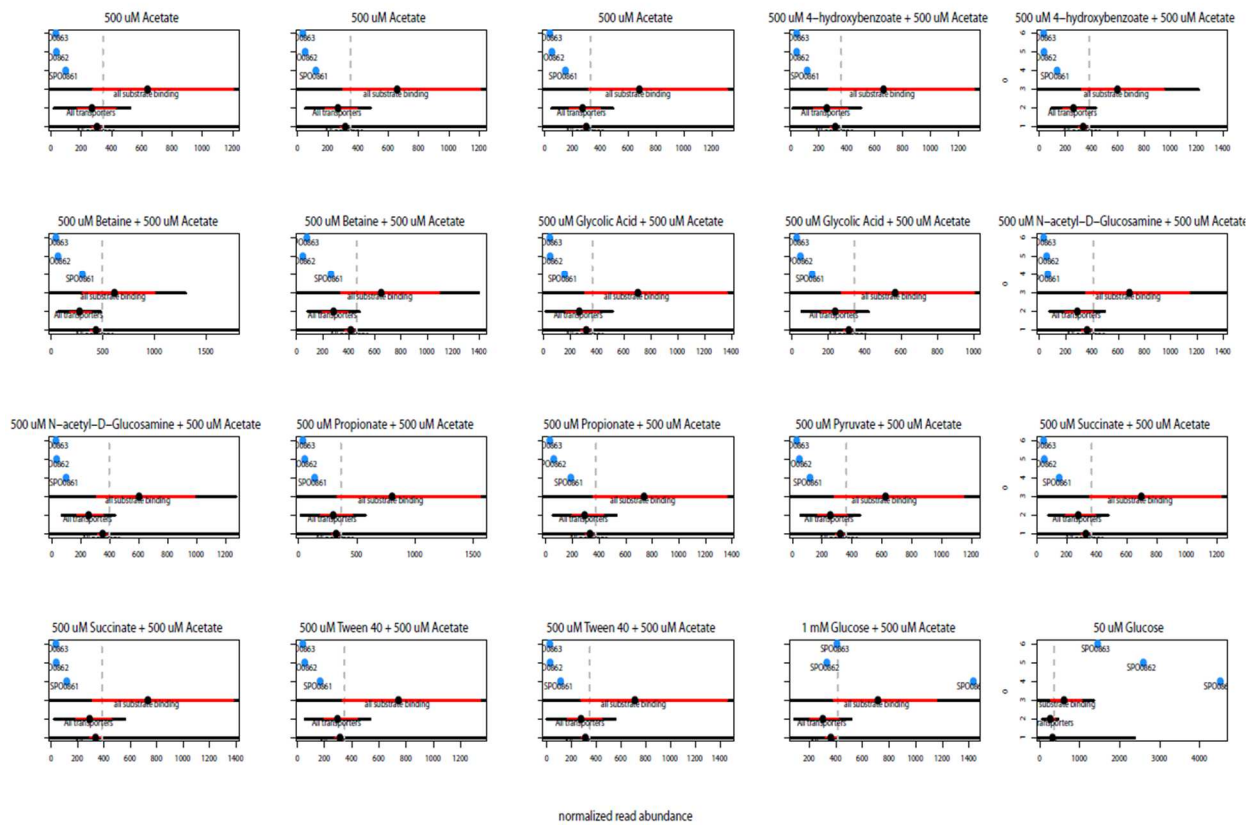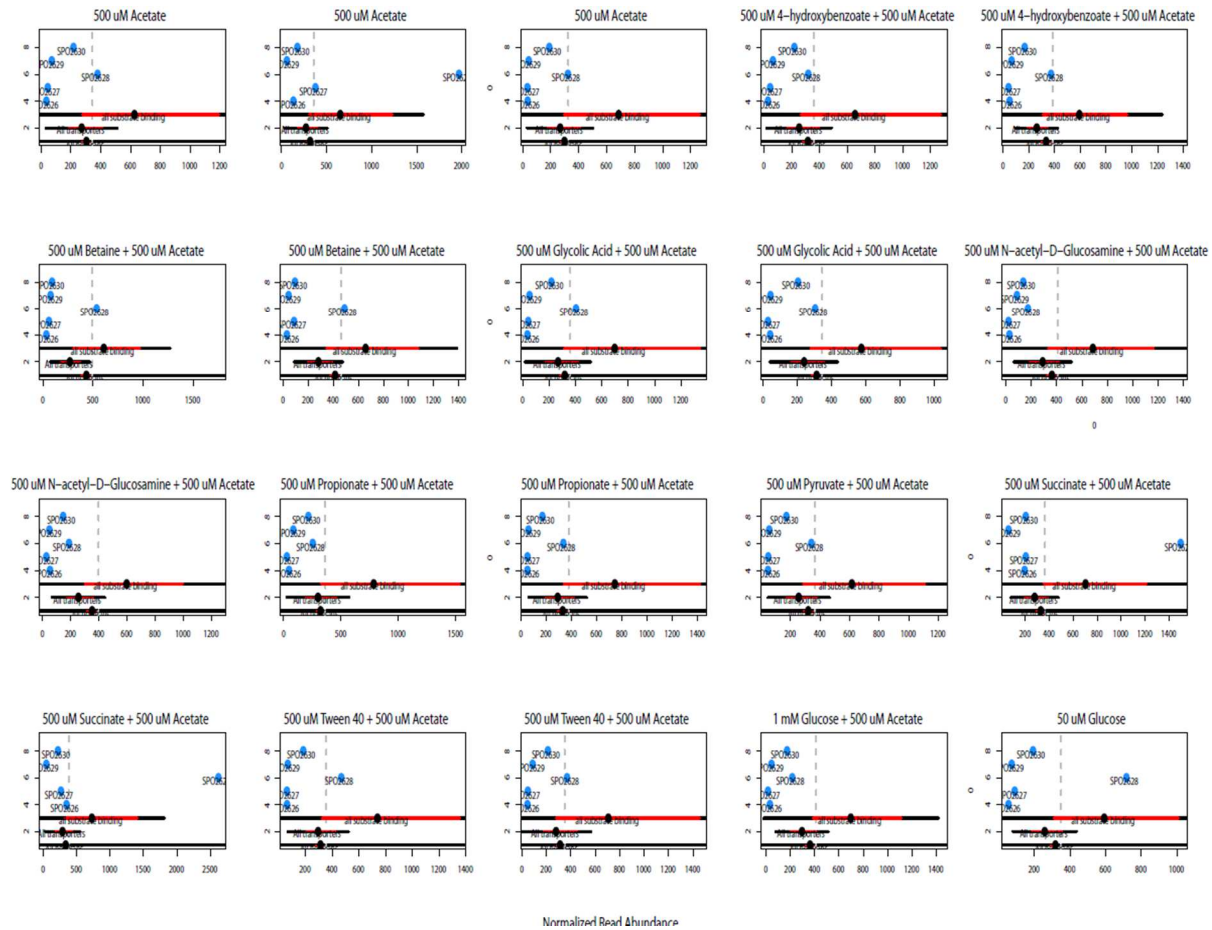
Figure 4. Transcript abundances of glucose ABC transporter genes (SPO0861-0863) compared to the average of all *R. pomeroyi* genes, transporter genes, and substrate binding transporter components. Each plot is a carbon treatment (with the substrate listed on top). The x-axis shows the normalized transcript abundances. At the bottom of the y-axis the average transcript abundance for all *R. pomeroyi* genes in the treatment as marked by a black dot. The 95% confidence interval for the mean is shown by the red line. The range of two standard deviations from the mean is shown by the black line. This same marking of mean (filled black circle), 95% CI of mean (red line), and two standard deviations (black line) is shown for all *R. pomeroyi* transporter gene transcript abundances on line 2, and all substrate binding protein genes on line 3. Transcript abundances for the genes of interest are shown by the blue dots are labeled with their locus tag number (SPO####). The grey dashed line is the upper limit of the 95% CI for all genes average.
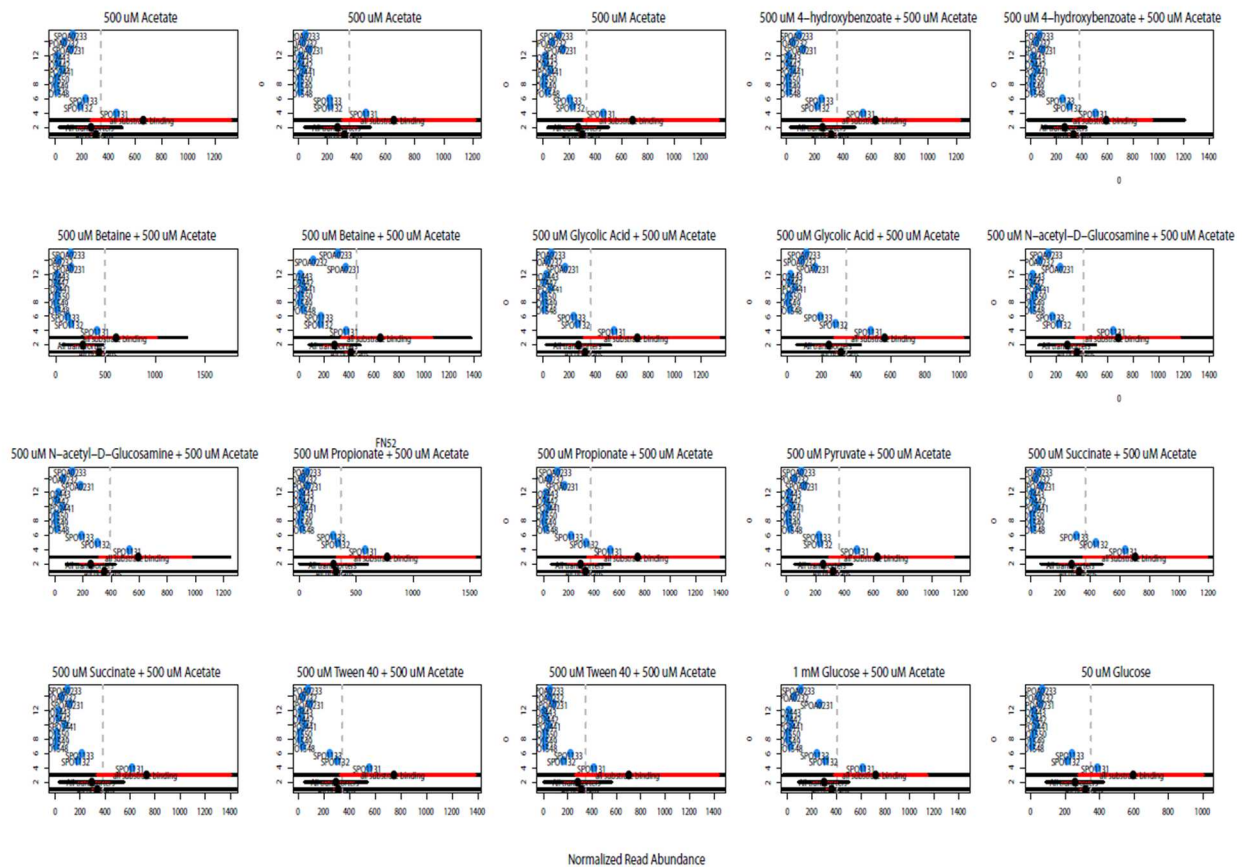
Figure 5. Transcript abundances of succinate TRAP transporter genes (SPO2626-2630) compared to the average of all *R. pomeroyi* genes, transporter genes, and substrate binding transporter components. Each plot is a carbon treatment (with the substrate listed on top). The x-axis shows the normalized transcript abundances. At the bottom of the y-axis the average transcript abundance for all *R. pomeroyi* genes in the treatment as marked by a black dot. The 95% confidence interval for the mean is shown by the red line. The range of two standard deviations from the mean is shown by the black line. This same marking of mean (filled black circle), 95% CI of mean (red line), and two standard deviations (black line) is shown for all *R. pomeroyi* transporter gene transcript abundances on line 2, and all substrate binding protein genes on line 3. Transcript abundances for the genes of interest are shown by the blue dots are labeled with their locus tag number (SPO####). The grey dashed line is the upper limit of the 95% CI for all genes average.
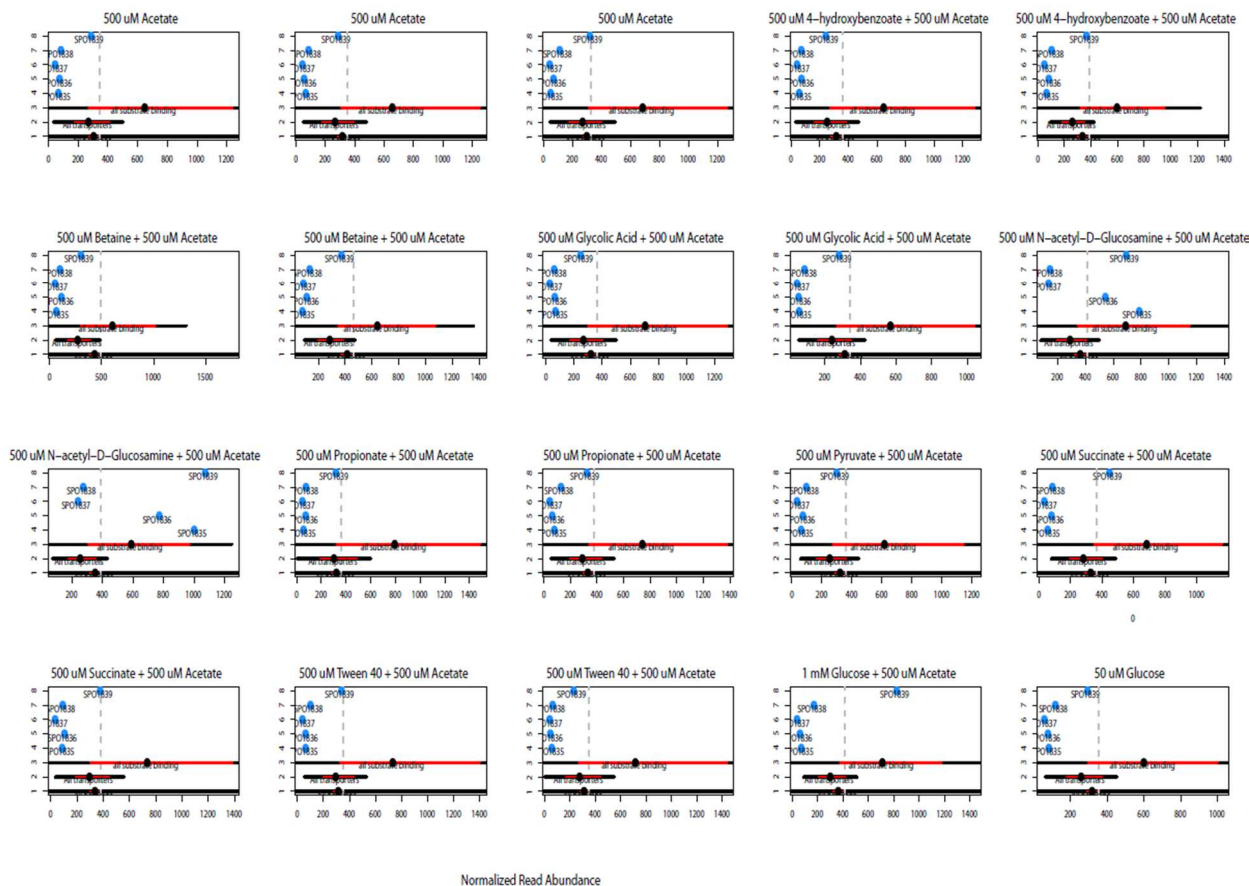
48

Figure 6. Transcript abundances of glycine betaine transporter genes compared to the average of all *R. pomeroyi* genes, transporter genes, and substrate binding transporter components. Four different transporter operons potentially involved with glycine betaine uptake are shown (SPO1131-1133; SPO1548-1550; SPO2441-2443; and SPOA0231-A0233). Each plot is a carbon treatment (with the substrate listed on top). The x-axis shows the normalized transcript abundances. At the bottom of the y-axis the average transcript abundance for all *R. pomeroyi* genes in the treatment as marked by a black dot. The 95% confidence interval for the mean is shown by the red line. The range of two standard deviations from the mean is shown by the black line. This same marking of mean (filled black circle), 95% CI of mean (red line), and two standard deviations (black line) is shown for all *R. pomeroyi* transporter gene transcript abundances on line 2, and all substrate binding protein genes on line 3. Transcript abundances for the genes of interest are shown by the blue dots are labeled with their locus tag number (SPO####). The grey dashed line is the upper limit of the 95% CI for all genes average.
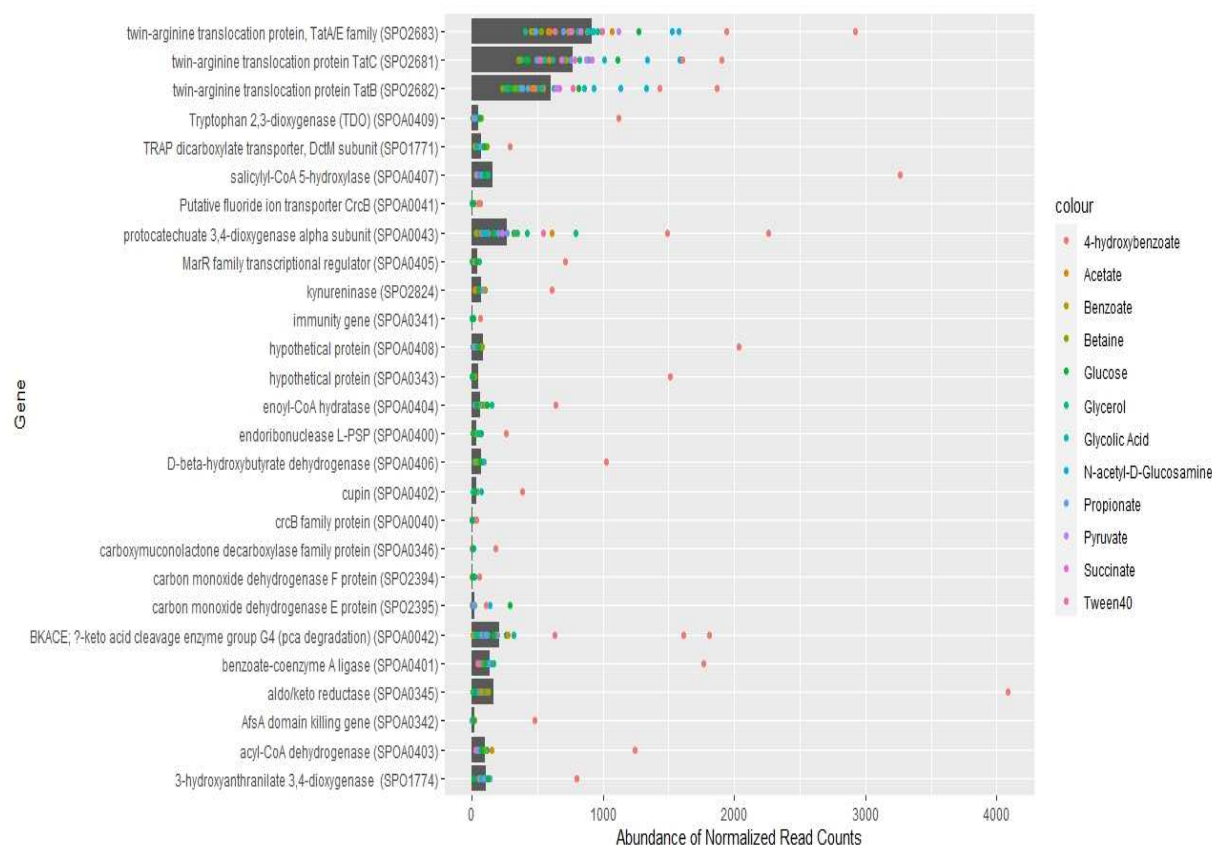
Figure 7. Transcript abundances of N-Acetylglucosamine transporter genes (SPO1835-1839) compared to the average of all *R. pomeroyi* genes, transporter genes, and substrate binding transporter components. Each plot is a carbon treatment (with the substrate listed on top). The x-axis shows the normalized transcript abundances. At the bottom of the y-axis the average transcript abundance for all *R. pomeroyi* genes in the treatment as marked by a black dot. The 95% confidence interval for the mean is shown by the red line. The range of two standard deviations from the mean is shown by the black line. This same marking of mean (filled black circle), 95% CI of mean (red line), and two standard deviations (black line) is shown for all *R. pomeroyi* transporter gene transcript abundances on line 2, and all substrate binding protein genes on line 3. Transcript abundances for the genes of interest are shown by the blue dots are labeled with their locus tag number (SPO####). The grey dashed line is the upper limit of the 95% CI for all genes average.

Figure 8. Relative abundance of normalized reads for all differentially expressed genes upregulated in 4-hydroxybenzoate. The grey bar is the average of all normalized read abundances for that gene. Each dot represents a single chemostat experiment, color represents carbon substrate treatment.
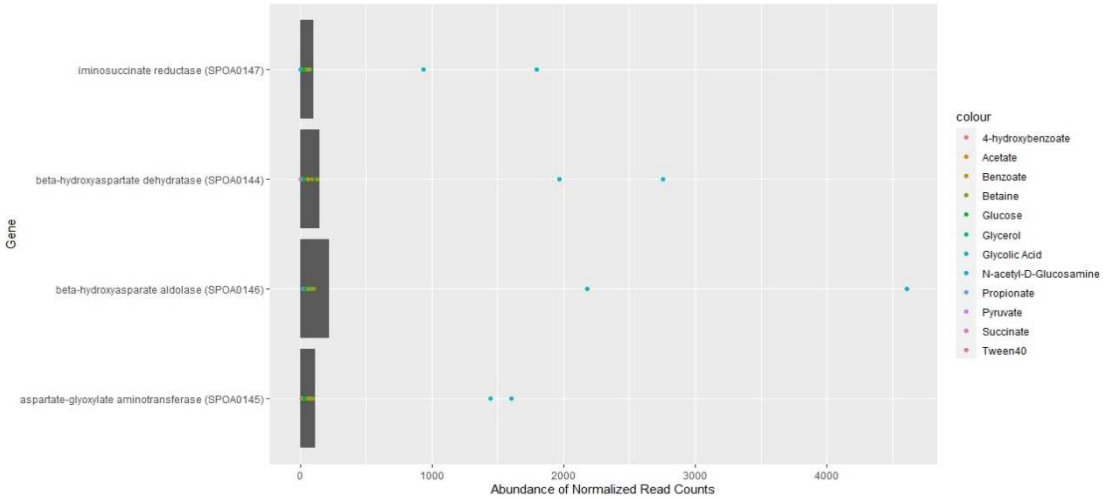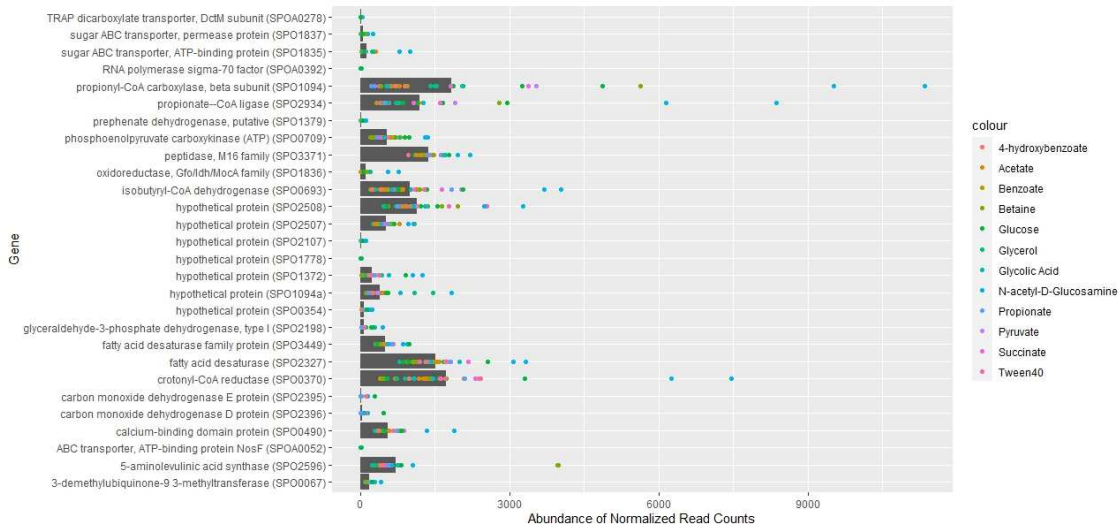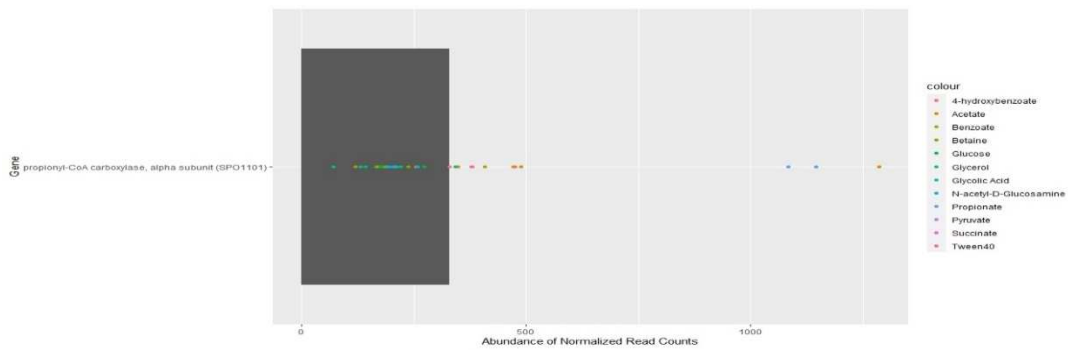
Figure 9. Relative abundance of normalized reads for all differentially expressed genes upregulated in glycolic acid treatments. The grey bar is the average of all normalized read abundances for that gene. Each dot represents a single chemostat experiment, color represents carbon substrate treatment.

Figure 10. Relative abundance of normalized reads for all differentially expressed genes upregulated in N-acetyl-D-glucosamine treatments. The grey bar is the average of all normalized read abundances for that gene. Each dot represents a single chemostat experiment, color represents carbon substrate treatment.
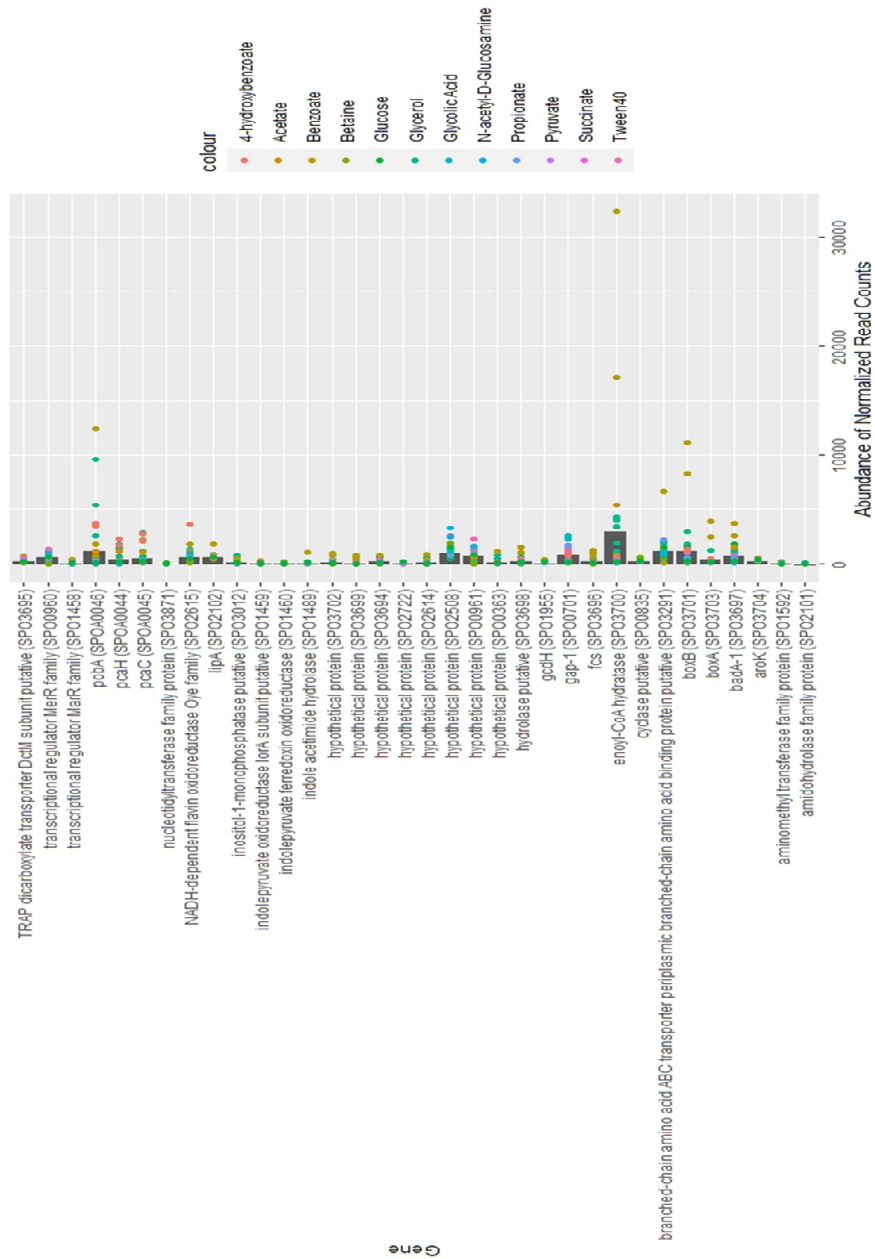
Figure 11. Relative abundance of normalized reads for all differentially expressed genes upregulated in propionate treatments. The grey bar is the average of all normalized read abundances for that gene. Each dot represents a single chemostat experiment, color represents carbon substrate treatment.
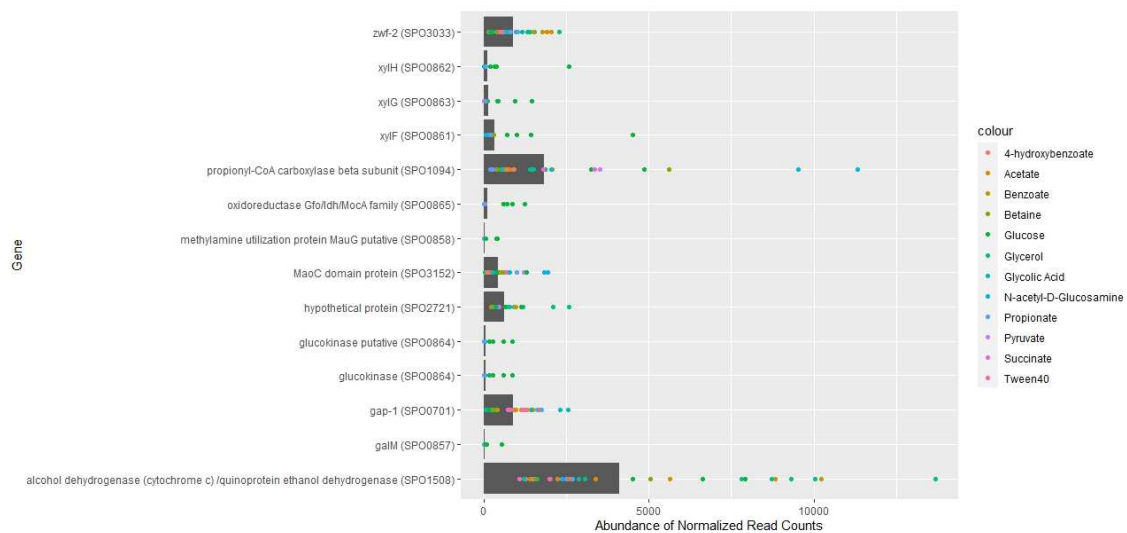
Figure 12. Relative abundance of normalized reads for all differentially expressed genes upregulated in benzoate treatments. The grey bar is the average of all normalized read abundances for that gene. Each dot represents a single chemostat experiment, color represents carbon substrate treatment.

Figure 13. Relative abundance of normalized reads for all differentially expressed genes upregulated in glucose treatments. The grey bar is the average of all normalized read abundances for that gene. Each dot represents a single chemostat experiment, color represents carbon substrate treatment.
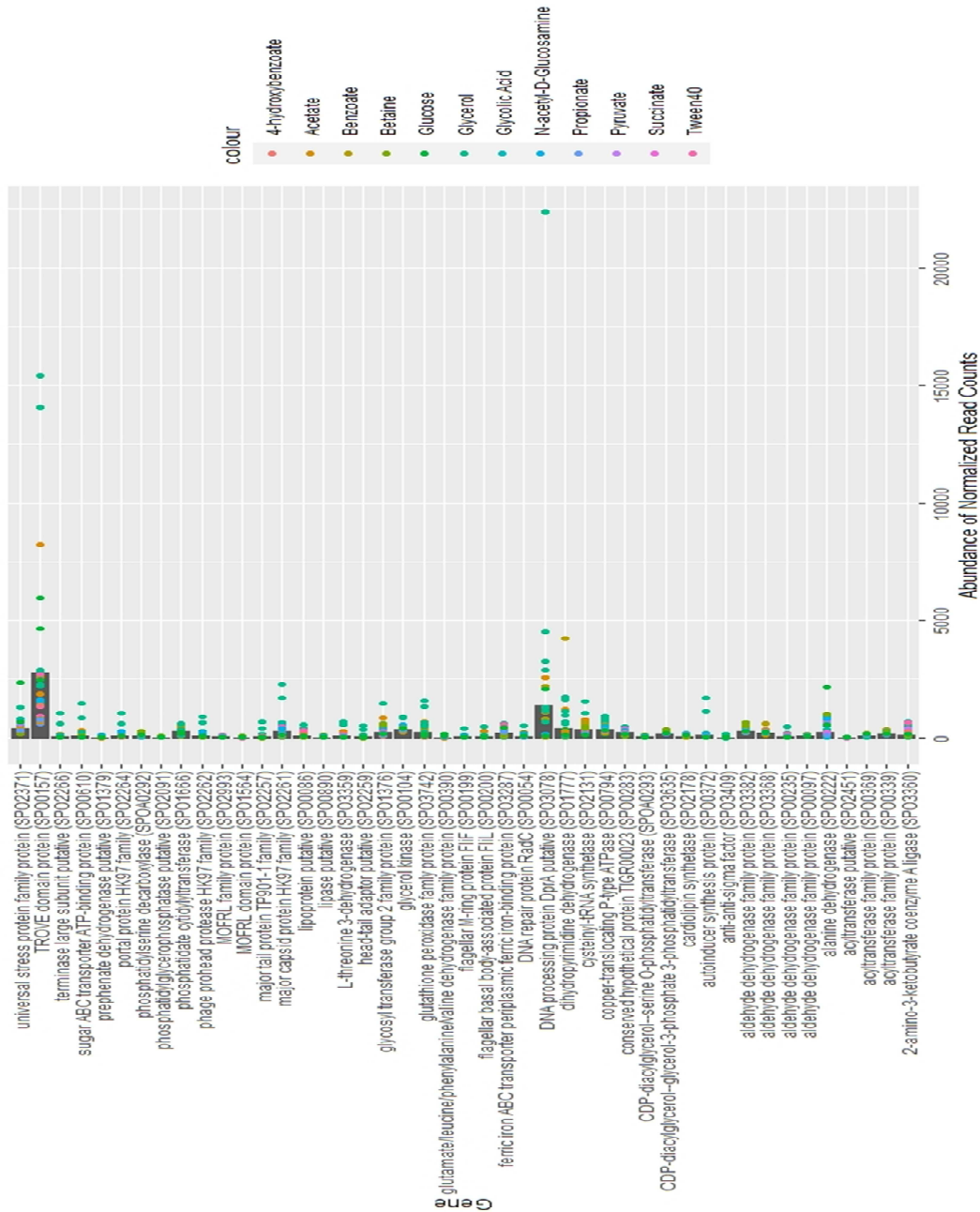
Figure 14.: Relative abundance of normalized reads for all differentially expressed genes upregulated in glycerol treatments. The grey bar is the average of all normalized read abundances for that gene. Each dot represents a single chemostat experiment, color represents carbon substrate treatment.
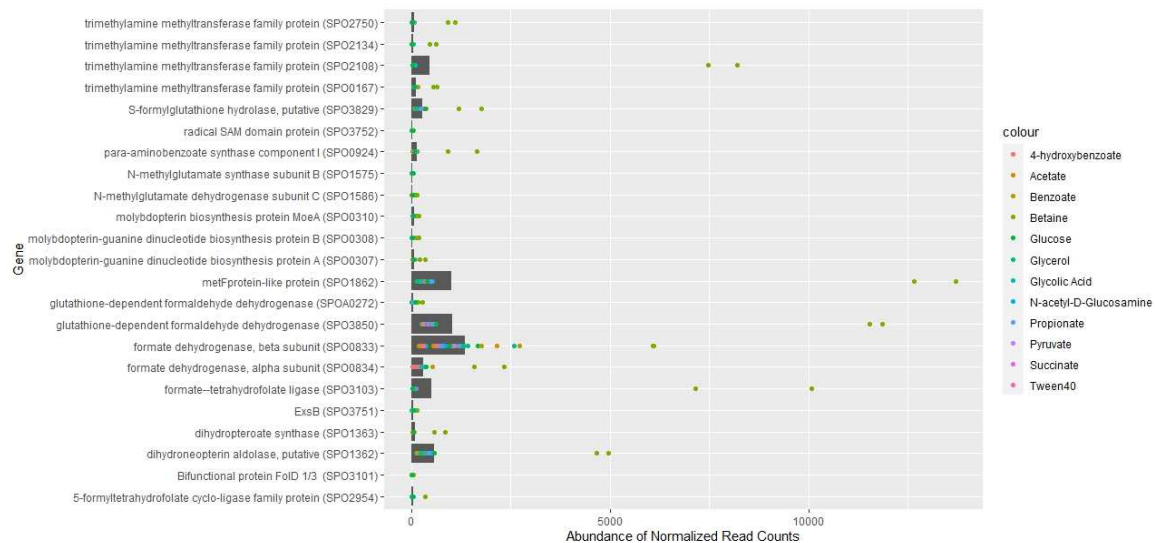
Figure 15. Relative abundance of normalized reads for a subset of differentially expressed genes upregulated in betaine treatments. The grey bar is the average of all normalized read abundances for that gene. Each dot represents a single chemostat experiment, color represents carbon substrate treatment.