8-2015

Relationships between Growth Rate and Gene Expression in *Ruegeria pomeroyi* DSS-3, a Model Marine Alphaproteobacterium

Nattasha Vinas
*Clemson University*, nattasha.vinas@gmail.com

Follow this and additional works at: https://tigerprints.clemson.edu/all_theses

Part of the Microbiology Commons

Recommended Citation

https://tigerprints.clemson.edu/all_theses/2184

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.
RELATIONSHIPS BETWEEN GROWTH RATE AND GENE EXPRESSION IN
*Ruegeria pomeroyi* DSS-3, A MODEL MARINE ALPHAPROTEOBACTERIUM

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Microbiology

by
Nattasha Viñas
August 2015

Accepted by:
Dr. Barbara J. Campbell, Committee Chair
Dr. J. Michael Henson
Dr. Harry D. Kurtz, Jr.
ABSTRACT

Microbes are important contributors to ecosystem processes such as biogeochemical cycling. Their activities vary, depending on environmental parameters such as carbon type and concentration, temperature, and salinity. Current estimates of in situ microbial growth is limited to gross estimates or rely on incubation based methods that disturb the natural state of the community. The goal of this master’s thesis is to develop molecular methods to directly assess microbial growth rates in the environment at the level of a taxonomic group. Here we grew *Ruegeria pomeroyi* DSS-3 under different temperatures (15 °C or 30 °C) and different carbon sources (yeast extract/tryptone, glucose, or acetate). We then characterized differences in growth rates and gene expression either with select growth-related genes (*rpoD, rpoB, rpoS, rplB*, and *ftsI*) or of the whole transcriptome. Ratios of *rpoB* and *rplB* mRNA:mRNA genes were significantly upregulated in log versus stationary phase as measured by qPCR for all three experiments utilizing a minimal media. Acetate-grown cells exhibited significant differences between growth phases for all five genes. These results indicate expression of the selected genes depend on growth phase as well carbon source availability. Conversely, a negative correlation between growth-related mRNA per cell and specific growth rate was observed. We also examined changes in other genes in log vs. stationary phases of growth using a transcriptomics approach. The bulk of differentially expressed genes were involved in amino acid transport and metabolism as well as translation and ribosomal structure. Our cultivation-based results indicate that monitoring differences in specific growth-related transcripts levels is a viable option for determining growth-related
activity changes in microbial taxa in marine environments. Future experiments should include growth in a mixed culture and with other bacteria with different ecological strategies in order to generalize our results to the total marine microbial community.
DEDICATION

To my mother
ACKNOWLEDGMENTS

I would firstly like to thank my advisor, Dr. Barbara Campbell, and my committee members Dr. J. Michael Henson and Dr. Harry Kurtz for their support and guidance throughout this project. Thanks are also in order for Dr. Matthew Cottrell of the University of Delaware for performing the RNA-seq experiment.

Thank you to undergraduates Tara Brown, Tianna Gore, and Brenton Davis for undertaking the burden of cell counting. Thank you to Dr. Abhiney Jain for thought-provoking conversations about my science. Thanks also to my fellow graduate students and lab members for listening to me complain, sometimes loudly, about failed experiments and also for strange discourse in the graduate office. Special thanks to my labmate Marco Valera for driving me home during unsavory weather after late night time points. I would also like to acknowledge the NSF Grant OCE-1261359. Finally, I would like to thank the friends I have made over the years that have stuck by my side, supported me though my lowest points, and celebrated my accomplishments. Truly, I could not have done this without your love.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE PAGE</td>
<td>i</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>Background</td>
<td>1</td>
</tr>
<tr>
<td>Marine Microbial Inhabitants</td>
<td>2</td>
</tr>
<tr>
<td>Current Modes of Measuring Activity</td>
<td>4</td>
</tr>
<tr>
<td>RNA-seq</td>
<td>5</td>
</tr>
<tr>
<td>Cellular Components that Vary with Growth Rate</td>
<td>6</td>
</tr>
<tr>
<td>Growth-related Genes</td>
<td>8</td>
</tr>
<tr>
<td>II. RESEARCH OBJECTIVES</td>
<td>10</td>
</tr>
<tr>
<td>Project Motivation</td>
<td>10</td>
</tr>
<tr>
<td>Experimental Goals</td>
<td>11</td>
</tr>
<tr>
<td>III. EXPERIMENTATIONAL METHODS</td>
<td>13</td>
</tr>
<tr>
<td>Primer Design</td>
<td>13</td>
</tr>
<tr>
<td>Culture Conditions</td>
<td>13</td>
</tr>
<tr>
<td>DNA/RNA Extractions</td>
<td>14</td>
</tr>
<tr>
<td>cDNA Synthesis and qPCR</td>
<td>14</td>
</tr>
<tr>
<td>Cell Enumeration</td>
<td>15</td>
</tr>
<tr>
<td>Protein Extractions and Analysis</td>
<td>15</td>
</tr>
<tr>
<td>RNA-seq Sample Collection, Processing, and Analysis</td>
<td>16</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>16</td>
</tr>
</tbody>
</table>
Table of Contents (Continued)

<table>
<thead>
<tr>
<th>IV. RESULTS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio Changes Between Growth Phases</td>
<td>18</td>
</tr>
<tr>
<td>Expression Level Changes Between Growth Rates</td>
<td>22</td>
</tr>
<tr>
<td>RNA-seq Analysis Between Log and Stationary Phase</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>V. DISCUSSION</th>
<th>Page</th>
</tr>
</thead>
</table>

APPENDICES

<table>
<thead>
<tr>
<th>A: Supplementary Tables</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>REFERENCES</td>
<td>45</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <em>R. pomeroyi</em> DSS-3 Primers Designed for qPCR</td>
<td>13</td>
</tr>
<tr>
<td>2. qPCR and Melt Curve Conditions for All Primer Sets</td>
<td>15</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Simplified schematic of a marine microbial food web</td>
<td>10</td>
</tr>
<tr>
<td>2.</td>
<td>Representative growth curve of acetate-grown cells</td>
<td>19</td>
</tr>
<tr>
<td>3.</td>
<td>Ratios of cDNA/DNA in acetate-grown cells</td>
<td>20</td>
</tr>
<tr>
<td>4.</td>
<td>Relationship between protein and cell counts in acetate-grown cells</td>
<td>22</td>
</tr>
<tr>
<td>5.</td>
<td>Representative growth curves of glucose-grown cells</td>
<td>24</td>
</tr>
<tr>
<td>6.</td>
<td>Ratios of cDNA/DNA in glucose-grown cells</td>
<td>25</td>
</tr>
<tr>
<td>7.</td>
<td>Relationship between protein and cell counts in glucose-grown cells</td>
<td>27</td>
</tr>
<tr>
<td>8.</td>
<td>Comparisons of gene expression and protein concentration between specific growth rates at mid log</td>
<td>29</td>
</tr>
<tr>
<td>9.</td>
<td>Breakdown of differentially expressed genes (p &lt; 0.000001) by COG category</td>
<td>31</td>
</tr>
<tr>
<td>10.</td>
<td>Differential expression confidence of select growth-related genes as a means of comparison to qPCR</td>
<td>32</td>
</tr>
<tr>
<td>11.</td>
<td>Differential expression confidence of ribosomal proteins that have a p-value &lt; 0.000001</td>
<td>33</td>
</tr>
</tbody>
</table>
CHAPTER ONE
INTRODUCTION

BACKGROUND

Aquatic habitats, in particular marine ecosystems, make up over 70% of the Earth’s surface. Because aquatic regions are an important resource and habitat for humans, animals, and microbes, their exposure to multiple sources of stresses such as pollution, nutrient run-off, climate change, and more are of concern. Microbes are key players in biogeochemical cycling such as the carbon and nitrogen cycles in these habitats. Bacteria have been shown to be responsible for the bulk of abundance and activity in the ocean (Azam and Malfatti, 2007). Furthermore, aquatic bacterial communities are changing in response to the above mentioned environmental factors (Lafferty et al. 2004). For instance, top down factors such as viral grazing may shape the abundance of both rare and dominant bacterial groups in marine systems (Bouvier and del Giorgio, 2007). Other biotic features such as the presence of different predatory bacteria may also select for the survival of certain bacterial species (Pineiro et al., 2013). Abiotic factors such as salinity, temperature, and oxygen contribute to the level of predictability in reoccurring bacterial groups (Fuhrman et al., 2006). Tracking the changes in microbial composition and activity in marine environments is important as variability can change the flux of biogeochemical processes (Azam and Malfatti, 2007).

The overall goal of marine microbial ecologists is to study which microbes live in marine systems, how they vary in response to the variety of environmental fluxes and factors, and what they contribute to these systems. Determining which taxa are active in
various biogeochemical cycles is necessary to guide researchers in deciding which microbes to further characterize and monitor in depth. Furthermore, contributions of individual taxa to community processes are largely unknown. Developing a method to directly measure environmental growth rates of individual taxa would help elucidate their role in the marine environment.

MARINE MICROBIAL INHABITANTS

A diverse array of microbes live in various habitats within marine systems. Microbial community composition has repeatedly been shown to change across salinity gradients from freshwater rivers feeding estuaries down to the open ocean (Bouvier and del Giorgio, 2002; Campbell and Kirchman, 2013; Kirchman et al. 2005). The open ocean is oligotrophic with very little influx of nutrients. In turn, microbial abundances tend to be low, with about half a million to a million cells per milliliter in the upper ocean (Whitman et al. 1998). A cultured representative of SAR11, one of the most abundant bacterial clades in the oceans, has slow growth rates and a streamlined genome (Rappé et al., 2002). These two traits are indicative of specialists, which tend to be selected for in homogenous environments such as the open ocean (Kassen, 2002). In contrast, heterogeneous environments, consisting of a diverse range of bottom up factors such as resource availability, tend to favor ecological generalists, those bacteria whose genome has evolved to be highly adaptable to a variety of environments (Kassen, 2002). The coastal ocean is an example of a heterogeneous environment because it experiences fluctuating levels of nutrients which results in a more variable microbial community.
A bacterial group representing both an abundant marine resident and a generalist species is the Roseobacter clade.

Roseobacters are found in most marine environmental samples, ranging from polar sea ice (Brinkmeyer et al., 2003), coastal biofilms (Dang and Lovell, 2000), as symbionts with dinoflagetts and sponges (Miller and Belas, 2004; Taylor et al. 2003). Certain areas contain a higher abundance of Roseobacters than others. For example, using community DNA hybridization, a pure culture of Roseovarius tolerans (ISM) was shown to be representative of 20% of the coastal community it was isolated from, while composing less than 1% of other marine communities (Fuhrman et al. 1994). Roseobacter abundance peaks at ten meters below the surface and declines as depth increases (González et al., 2000). The clade itself contains a diverse range of physiologies including the types of metabolism (and thus biogeochemical processes) they perform. There are currently at least 54 Roseobacter isolates (see www.roseobase.org).

The isolate used in this study, Ruegeria pomeroyi DSS-3, has been shown to be involved in dimethylsulfiniopropionate (DMSP) degradation (J. M. Gonzalez, 2003) and carbon monoxide oxidation (Cunliffe, 2013; Moran et al., 2004). Another isolate, Roseovarius sp. TM 1035, is also involved in DMSP degradation (Miller and Belas, 2004) but is an aerobic anoxygenic phototroph (AAnP) (Newton et al., 2010). Genomic analysis has further confirmed the vast metabolic range of these microbes, and that trophic strategy is the best predictor for genomic content, though even that is not terribly strong (Newton et al., 2010). While studying pure cultures gives in depth analysis into the physiology and
genomic studies give us an idea of the metabolic potential, it is still crucial to determine the actual activity of these microbes in their natural environment.

**CURRENT MODES OF MEASURING ACTIVITY**

A few different techniques have been developed over the years to measure microbial activities, from general measurements such as community growth rates, respiration, and bacterial production to more specific measures such as determining specific substrate uptake, enzyme activity, and the use of microautoradioactivity (MAR) with fluorescence in situ hybridization (FISH, collectively known as MAR-FISH) (del Giorgio and Gasol, 2008; Staley and Konopka, 1985). These different techniques can be sorted into two groups, depending on whether an incubation period is needed for the experiment or if data is collected directly from the environment, such as with proteomic techniques.

Bacterial biomass production can be measured by thymidine/leucine incorporation (Fuhrman et al. 1982). Their assimilation is evidence of heterotrophic production. Production by specific bacterial groups (and thus inferred contribution to geochemical cycling) can be measured using MAR-FISH (Cottrell and Kirchman, 2003). In one such study, it was found that about 50% of the variation of activity as measured by MAR-FISH can be explained by abundance (Cottrell and Kirchman, 2003). While specific substrate incorporation is very useful in determining activity, disadvantages are present. For example, a bacterium may not take up thymidine but could still be metabolically active (Pollard and Moriarty, 1984). Also, whether or not bacteria take up leucine (and other compounds) may depend on the concentration of the substrate (Alonso and
Pernthaler, 2006). Another drawback to incubation based experiments is potentially inducing artificial spikes and declines in growth that may not reflect natural activity in the environment (LaRock et al. 1988).

**RNA-seq**

Transcriptomics is the study of all the RNA transcripts produced under a certain condition or in a certain bacterial taxa. Microarrays were initially used to quantify gene expression (Schena et al. 1995). Development of high throughput technologies now allows researchers to probe environmental systems to collect data that can be used to reveal patterns in microbial activity. For example, two strains of *Prochlorococcus* adapted to either high-light or low-light conditions show highly diverse transcriptomes when grown in the same conditions (Voigt et al., 2014).

RNA sequencing (also called RNA-seq) is a form of transcriptomic analysis, giving researchers a snapshot of RNA present at a particular time without need of an incubation. In addition, the technology provides quantifiable data, allowing comparisons of differential gene expression between experiments (Creecy and Conway, 2015). RNA-seq can also give researchers valuable insight into how bacteria react to stressed conditions (Harke and Gobler, 2013; Pinto et al., 2014). This potentially allows researchers to use those differentially expressed genes as stress indicators in the environment. Similarly, transcriptomics can be used to determine genes expressed during utilization of specific substrates (Bullerjahn and Green, 2013). These particular RNAs may then be used as general activity indicators.
CELLULAR COMPONENTS THAT VARY WITH GROWTH RATE

Determining how certain macromolecular components change with growth has been of interest since the 1950’s with early studies conducted on *Salmonella typhimurium* and later with *Escherichia coli* B/r (Churchward et al. 1982; Schaechter et al. 1958). Certain parameters change with growth, with some positively correlating (RNA/protein for example) and others negatively correlating (DNA/protein) with growth rate (for instance, Figure 2a and b in (Bremer and Dennis, 1987)). Each of these macromolecules provides information with DNA used to evaluate the functional potential of a community, while RNA and protein are used to evaluate the actual metabolic activity at a specific time and environmental condition. In a study where the proteome of *R. pomeroyi* DSS-3 was analyzed under 30 different environmental conditions, the protein profile grouping was mostly related to the growth phase the cells were harvested – exponential/early stationary versus mid-late stationary – because ribosomal proteins represented over 26% and 14% of the normalized spectral abundance factor of all proteins, respectively (Christie-Oleza, 2012). RNA is typically less stable than protein, therefore it allows us a snapshot of what is happening in the cell at a particular period without risk of contamination from a previous time point.

Measuring cell-specific RNA content has shown to be useful in determining whether cells are active or not. Cell counts measured by 16S rRNA probes correlate with cells that were found to uptake $^3$H-labeled amino acids via autoradiography, suggesting these methods count the same type of cells (Karner and Fuhrman, 1997). This is probably because cells must be at least minimally active in order to have enough rRNA
be detected by probes. Past studies exploring the use of 16S rRNA to rRNA gene ratios as indicators of growth in marine environments have uncovered interesting and somewhat unexpected results. For example, in coastal ocean samples it was found that while abundance follows activity for the majority of bacterial taxa, many rare bacteria exhibited higher activity at low abundance levels (Campbell et al., 2011). Differences in the relationship between 16S rRNA and rDNA in some marine taxa also reflect differences in light, nutrient concentrations, and other environmental factors (Campbell and Kirchman, 2013).

Issues encountered with the use of rRNA as an activity indicator include rRNA concentration not always linearly correlating with growth rate across taxa and dormant cells having a high number of ribosomes (Blazewicz, Barnard, Daly, and Firestone, 2013); more issues and studies cited can be found summarized in Box 1 of the Blazewicz review. That said, since many other studies have found a correlation with growth rate and RNA concentration (Fegatella, et al., 1998; Kemp et al. 1993; Kerkhof and Kemp, 1999), it is worth studying how other growth related genes (besides rRNA) and their corresponding transcripts change with growth rate and across taxa. In this project, six genes (rpoB, rpoD, rpoS, rplB, and ftsI) were studied and their transcript abundance measured across growth phases and growth rates of a Roseobacter strain.

**GROWTH-RELATED GENES**

RNA polymerase, which transcribes DNA into RNA in bacteria, is made up of several subunits. There are two α subunits and single β, β', and ω subunits in the core enzyme. This core enzyme has a weak affinity to DNA and needs the σ subunit to bind
specifically to DNA promoters; this complete structure is called the holoenzyme. The $rpoB$ gene encodes the $\beta$-subunit of RNA polymerase. $rpoD$ encodes the $\sigma^{70}$ subunit, which is responsible for the binding and transcriptional initiation of housekeeping genes. The genes that are turned on during the transition to stationary phase and during various stressed conditions are managed by $\sigma^{38}$, encoded by $rpoS$. $\sigma^{70}$ has been shown in *E. coli* to have constant concentrations between exponential and stationary phase, while $\sigma^{38}$ has been shown to increase from exponential to stationary phase (Piper, et al., 2009; Sharma and Chatterji, 2010). There is also some evidence that the concentration of constitutively expressed genes may change as growth rate increases (Klumpp et al. 2009). The number of RNA polymerases has also been shown to remain relatively constant between growth phases, yet the number of RNA polymerases per cell seem to increase with growth rate (Sharma and Chatterji, 2010; Bremer and Dennis, 1987).

Two other genes were selected because of their potential for varying with growth rate. The penicillin-binding protein encoded by $ftsI$ is involved in the septal peptidoglycan synthesis that occurs during cell division (Errington et al. 2003). Since ribosome abundance is growth-rate dependent, the ribosomal protein encoding gene $rplB$ was chosen to monitor as well (Klumpp et al., 2009). While some of the above mentioned proteins have remained constant during different physiological conditions, it still remains to be seen how their corresponding transcript levels change. These growth-related genes were chosen for this study because they or their close relatives may be found in nearly all bacterial and archaeal genomes (Gil et al., 2004). The five selected genes occur only once per genome, and $rpoB$, $rpoD$, and $rplB$ are conserved between taxa, which may be
beneficial when applying the results of this work to other microorganisms (Gil et al., 2004; Raes et al., 2007).
CHAPTER TWO

RESEARCH OBJECTIVES

PROJECT MOTIVATION

Abundance of bacteria can only partially explain variations of activity in a given sample since some rare bacteria have been found to be more active than abundant taxa (Campbell and Kirchman, 2013; Cottrell and Kirchman, 2003). This discrepancy between abundance and perceived activity may be explained by the growth rate of individual taxa. We are interested in the growth rates of heterotrophic bacteria in particular because their uptake of organic matter can considerably change the overall flow of carbon in the ocean (Azam and Malfatti, 2007). The role of these heterotrophs in marine ecosystems can be seen in Figure 1, but in short, the growth of these bacteria is positively correlated to the amount of CO2 resired into the atmosphere.

Figure 1. Simplified schematic of a marine microbial food web. POM, particular organic matter; DOM, dissolved organic matter. Adapted from Figure 1 of Azam and Malfatti, 2007.
The model organism used in this study, *Ruegeria pomeroyi* DSS-3, is in the Roseobacter clade. One important attribute of this organism and clade is its ability to metabolize DMSP. This metabolism sometimes results in production of dimethyl sulfide (DMS), a compound which comprises the bulk of the sulfur flux into the atmosphere (Andreae, 1997). This flux can subsequently affect the climate system through the formation of aerosols and alterations in how much solar energy is reflected back into space (Charlson et al. 1987). In one study, the abundance of Roseobacter cells was found to be highly correlated to dissolved DMSP consumption and bacterioplankton production, while this correlation was not seen with the abundance of other taxa (Zubkov et al., 2001). This observation suggests a few bacterial species may at times dominate the production of compounds important to global climate change. The objectives of this study revolve around developing a method that utilizes transcriptional changes associated with different levels of growth in order to gauge environmental growth rates of marine bacterial taxa.

**EXPERIMENTAL GOALS**

We hypothesized ratios of specific mRNA:mRNA gene copy number can be used as an indicator of marine microbial activity. The overarching goal of this master’s thesis was to determine how expression levels of select genes vary between growth phases and specific growth rates in *R. pomeroyi* DSS-3. To address this goal, three objectives were created.

*Objective 1:* Determine whether expression levels change between growth phases in batch culture-grown cells. The chosen conditions for this experiment consisted of
cultures grown at 30 °C with minimal marine media supplemented with acetate and a vitamin solution. Samples from early and mid log, stationary, and death phase were analyzed. cDNA counts of specific growth-related genes were normalized to their respective DNA counterpart in a given sample. We hypothesized ratios of the transcripts to genes would correlate positively with log phase growth and negatively with stationary and death phase.

**Objective 2:** Determine if expression levels of the selected genes would differ when growth conditions (and thus rates) were changed. This is important to address since microbial communities are subjected to varying environmental conditions. We expected to see a positive correlation between ratios and growth rates. Samples were taken from mid log and stationary phase. Transcripts and corresponding genes were quantified via qPCR for the first two goals.

**Objective 3:** Analyze transcriptome data for differentially expressed genes between mid log and stationary phase in *R. pomeroyi* DSS-3. This allows comparisons between RNAseq and qPCR analysis and also provides a general view of which genes are associated with changes in growth in this organism.
CHAPTER THREE
EXPERIMENTAL METHODS

PRIMER DESIGN

Primers used in all qPCR experiments for this project were designed either with the NCBI primer designing tool or in Geneious with Primer3 and checked for the correct product amplification using BLAST (Table 1). Optimized annealing temperature was determined using a temperature gradient with four different temperatures in a qPCR test.

Table 1. R. pomeroyi DSS-3 Primers Designed for qPCR

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward</th>
<th>Reverse</th>
<th>Optimized Annealing Temp (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rpoB</td>
<td>CGGATGAACGTCGGTCAGAT</td>
<td>CGTCCATGCCCAGAGATACCC</td>
<td>60</td>
</tr>
<tr>
<td>rpoD</td>
<td>GCGGTGGACAAGTTTCGAGTA</td>
<td>CAGCGGCATCTGCGAGTTTTT</td>
<td>56.7</td>
</tr>
<tr>
<td>rpoS</td>
<td>CGAAAGCCTGACCCATTTGCG</td>
<td>TTTAGACCTCGCCCAGTTG</td>
<td>59</td>
</tr>
<tr>
<td>ftsI</td>
<td>CACATTTGCCAGCCCTGTTCC</td>
<td>CGAGGTGAGCGTAGCCAG</td>
<td>57</td>
</tr>
<tr>
<td>rplb</td>
<td>ATGGCGGCTATGTCCAGATC</td>
<td>TTGCCGCTGCTGCTTCGCTGCTGCT</td>
<td>57</td>
</tr>
<tr>
<td>16S rRNA</td>
<td>(1369F and 1492R)</td>
<td>CCGGTGAATACGTTTCYG</td>
<td>GWTACCTTGTTACGACTT</td>
</tr>
</tbody>
</table>

CULTURE CONDITIONS

Strains were prepared for growth curve experiments by recovering from half-strength YTSS glycerol stocks into half-strength YTSS broth (8 g yeast extract, 3 g tryptone, and 20 grams sea salt per liter for full strength). The sea salt concentration was the same in full and half-strength YTSS. Subsequent passages were performed in marine minimal basal media (250 mL Basal Medium, 50 mL FeEDTA Stock, and 699 mL DI H₂O Sea Salt Solution per 1 L mixture). The ingredients for each component of the minimal media are as follows: 699 mL DI H₂O and 20 g Sigma Sea Salts made up the sea
salt solution, 150 ml 1M Tris HCl pH 7.5, 87 mg K$_2$HPO$_4$, 1.5 g NH$_4$Cl, and 375 ml DI H$_2$O made up the basal medium stock, and 50 mg FeEDTA (ethylenediamine tetraacetic acid; ferric-sodium salt) plus 100 ml DI H$_2$O made up the FeEDTA stock. Each component was mixed together after autoclaving and carbon substrates were added at a final concentration of 10 mM. The minimal media was also supplemented with a vitamin solution (0.1% final volume) (Gonzalez et al., 1997). Biological replicates were created during the third passage in three separate 125 mL flasks. Growth curves were conducted on the third passage. Unless otherwise noted, strains were grown in the dark, at 30 °C, and in a shaker at 250 RPM. Other culture conditions (intended to change specific growth rate) include 1) growth in minimal media supplemented with either acetate or glucose and 2) an incubation temperature of 15 °C.

**DNA/RNA EXTRACTIONS**

Two milliliters of culture per time point were taken and spun down at 16,000 x g for 10 minutes. The supernatant was discarded and 600 µL of RLT buffer (provided by QIAGEN) and 6 µL of beta-mercaptoethanol was added and allowed to incubate on a rotator at 60 °C for 10 minutes. The protocol for the QIAGEN AllPrep DNA/RNA Mini Kit was then followed for both DNA and RNA extraction and purification. There was an extra elution step at the end of each RNA/DNA extraction with the same elute to increase the respective nucleic acid yield.

**cDNA SYNTHESIS AND qPCR**

RNA samples were treated with Turbo DNase from Ambion according to the manufacturer’s instructions. Samples were checked for complete DNA digestion by
performing a PCR test for 16S rRNA genes. RNA was diluted to concentrations demonstrated to be in the linear range of qPCR (data not shown). The diluted RNA was then reverse transcribed into cDNA using the High-Capacity cDNA Reverse Transcription Kit from Applied Biosystems. The manufacturer’s provided protocol was followed. Samples were amplified in triplicate with qPCR. A melt curve was performed after each reaction to determine if there was any contaminating DNA or nonspecific amplification. Conditions for qPCR are listed in Table 2 with the variations in annealing temperature dependent on the primer pair as indicated in Table 1.

Table 2. qPCR and Melt Curve Conditions for All Primer Sets

<table>
<thead>
<tr>
<th>Step</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Repeat from Step 2</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Repeat Step 7 150x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp. (°C)</td>
<td>95</td>
<td>95</td>
<td>57-60</td>
<td>72</td>
<td></td>
<td>72</td>
<td>65</td>
<td>65 + 0.2/cycle</td>
<td></td>
</tr>
<tr>
<td>Time (min)</td>
<td>5:00</td>
<td>00:15</td>
<td>00:30</td>
<td>00:30</td>
<td>Plateread</td>
<td>5:00</td>
<td>0:30</td>
<td>0:05</td>
<td>Plateread</td>
</tr>
</tbody>
</table>

**CELL ENUMERATION**

Samples for cell counts were diluted with paraformaldehyde (2% final concentration) buffered with sodium phosphate buffer (pH 7.2). After filtering and staining with 4’,6-diamidino-2-phenylindole (DAPI) for 5-10 minutes, cells were imaged with an epifluorescence microscope in at least ten fields per filter and counted with the aid of ImageJ software.

For repeat growth curves of the same conditions, cell counts from previous experiments were correlated with OD and fitted with an exponential trendline from lag to late-log time points. The resulting equation was used for enumeration. Samples taken
for OD$_{580}$ were diluted to or below 0.50 when necessary. Cell counts were always taken during stationary and death phases.

**PROTEIN EXTRACTIONS AND ANALYSIS**

One milliliter of culture collected for protein analysis was spun down at 15,800 x g at 4 °C for 10 minutes. After discarding the supernatant, the pellets were frozen at -80 °C until further analysis. For the extraction, protein pellets were thawed on ice and lysis buffer (20 mM Tris, 100 mM NaCl, 1 mM EDTA, and 0.5% Triton X-100) was added. After suspending cells, they were boiled in buffer for 5 minutes. Samples were then cooled on ice, and the instructions for the Bio-Rad Protein Assay Standard Procedure for Microtiter Plates was followed with the use of BSA standards.

**RNA-seq SAMPLE COLLECTION, PROCESSING, AND ANALYSIS**

RNA-seq samples were collected by Dr. Matt Cottrell of the University of Delaware from biological triplicates of cultures grown in a modified YTSS medium (0.4 grams yeast extract, 0.25 grams tryptone, and 20 grams sea salts per liter) at 18 °C. rRNA was partially removed prior to sequencing on an Illumina HiSeq machine at the University of Delaware. Sequences were processed and analyzed in Geneious version R8 (http://www.geneious.com, Kearse et al., 2012). Leftover rRNA reads were removed by mapping reads to a FASTA file containing the 5S, 16S, and 23S sequences of the respective reference genome. Sequences were trimmed from the 3’ end, resulting in an average read length of 75 bp. Reads were then mapped to the reference genome and megaplasmid using TopHat. The Bowtie2 preset was used at the lowest sensitivity (which is the same as -D 5 -R 1 -N 0 -L 22 -i S,0,2.50 in --end-to-end mode). Expression
levels were calculated with ambiguously mapped reads counted as partial matches. Transcript expression levels were compared and normalized by the median of gene expression ratios. Expression levels were calculated and compared between each of the biological replicates and also between pooled triplicates. Any resulting differentially expressed genes between replicates were treated as cellular noise and ignored when analyzing pooled replicates.

**DATA ANALYSIS**

After data collection, the original concentrations (in copies/µL) of each transcript and gene in extracted RNA and DNA were calculated. The ratios of mRNA:DNA were created and compared to see if there were any significant differences between growth phases. The mRNA:DNA ratios from mid log were compared to corresponding ratios at different growth rates to determine if there were any significant differences. The Student’s t-test was used when comparing means of ratios between growth phases. Linear regression analysis was used to compare mid log ratios with specific growth rate. Specific mRNA/cell and protein/cell were also calculated.
CHAPTER FOUR

RESULTS

RATIO CHANGES BETWEEN GROWTH PHASES

Two types of growth rates were analyzed in this study: specific growth rate and changes in growth from one time point to another. Growth in different culture conditions addressed specific growth rates changes and comparing different growth phases addressed the growth rate changes cultures experience during a single growth curve experiment. For the latter analysis, readings for optical density at 580 nm were taken as a general indicator of the progression of cells through the four growth phases: lag, logarithmic, stationary, and death. Cell counts were later processed and their relationship with OD was calculated. The growth rate for each growth curve conducted in this study was calculated by determining the slope of the line of best fit through three points in exponential phase using cells counts. This process is demonstrated in Figure 2, where the slope corresponds to a growth rate of 0.402 hour$^{-1}$ or 9.648 day$^{-1}$ for cells grown at 30 °C in minimal media supplemented with acetate as a sole carbon source.
Figure 2. Representative growth curve of acetate-grown cells. OD$_{580}$ is included for comparison. Arrows indicate where samples were taken for qPCR and protein analysis. Error bars indicate standard error (n=3).

Since the primary objective of this project was to identify which, if any, transcripts correspond to an increased activity level, expression levels of the selected growth-related genes from early log, mid log, stationary, and death phase were analyzed. Lag samples were not included because it didn’t seem there was enough cell mass for accurate transcript quantification, though it may be useful to study in future projects. Ratios of transcript per gene in copies/µL of extracted RNA and DNA for rpoB, rpoD, rpoS, rplB, and ftsI as determined by qPCR are shown in Figure 3. rpoB had significant differences (p < 0.05) in expression between samples taken at log, stationary, and death phases. Differences in expression of rpoB seemed to be most pronounced between mid log and stationary (p = 0.00643) and mid log and death (p = 0.00619). Similarly, rplB expression was significantly different (p < 0.01) in the two log phase samples versus death and stationary. rpoD expression was different between early log and stationary,
early log and death, and mid log and stationary. The difference in \textit{rpoD} expression between mid log and death was close to significant (p = 0.0522). \textit{rpoS} ratios showed significant differences (p <0.05) in early log and mid log versus death, mid log versus stationary, and stationary versus death. \textit{ftsI} differences were significant (p < 0.05) in mid log versus stationary and death. There were also differences (p <0.01) in early log versus stationary and death. There were no significant differences between early log and mid log, though the lowest p-values were associated with \textit{rpoB} and \textit{rplB} expression, p = 0.0966 and 0.0852, respectively. In all cases with significant differences for all five genes, each were upregulated in log phase growth when compared to stationary or death.

![Figure 3](image.png)

Figure 3. Ratios of cDNA/DNA in acetate-grown cells. Ratios are specific transcripts per gene copy number as determined by qPCR. Error bars indicate standard deviation.

We were interested in how protein per cell changed between growth phases and growth rates, since protein per cell has been shown to positively correlate with growth rate in \textit{E. coli} B/r (Bremer and Dennis, 1987). After collecting and centrifuging the cells, the resulting pellet did not undergo a wash step before lysis in buffer. Any residual
proteins in the media and from the extracellular matrix are thus included in the analysis. Since each protein pellet was treated the same way, we believed it would still be valuable to perform comparisons of protein levels. We first focused on comparing the protein content between growth phases. The general trend of the soluble protein quantified from mid log, stationary, and death phase followed cell counts, as shown in Figure 4a. There seemed to be a bit more protein in relation to cell counts in death phase when compared to mid log or stationary, and this discrepancy is made more apparent in Figure 4b when protein was normalized by cell counts. Though the differences between mid log and death (p = 0.0781) and stationary and death (p = 0.0950) do not quite reach significance, it may indicate an issue with protein quantification in death phase. Also, the protein quantified in mid log was outside of the standard curve, so this may have skewed results from that phase of growth.
Figure 4. Relationship between protein and cell counts in acetate-grown cells.  a) Protein concentration compared to cell concentration in three growth phases and b) amount of protein per cell. Error bars indicate standard deviation.

**EXPRESSION LEVEL CHANGES BETWEEN GROWTH RATES**

The second objective of this project addressed how expression levels of growth-related genes change with specific growth rate. Specific growth rates were manipulated by growing cells in glucose at 30 °C or at 15 °C (Figure 5). Specific growth rates were
11.225 and 0.4193 day$^{-1}$ for glucose-grown cells at 30 °C and 15 °C, respectively. The order of fastest to slowest growing cells were those grown in minimal media plus glucose at 30 °C, then acetate at 30 °C, and finally glucose at 15 °C. This order is expected since glucose is a richer carbon source than acetate and cells typically grow slower in colder temperatures. Previous growth rates for *Ruegeria pomeroyi* DSS-3 ranged from 0.42 day$^{-1}$ to 2.78 day$^{-1}$, though culture conditions varied from what was used in this study (Cunliffe, 2013; J. M. Gonzalez, 2003). The inclusion of a vitamin solution in the media and rigorous shaking at 250 RPM may have contributed to the higher growth rates measured in this study.
Figure 5. Representative growth curves of glucose-grown cells. a) Growth at 30 °C and b) at 15 °C. Arrows indicate where samples were taken for qPCR and protein analysis. Error bars indicate standard error where n=3, except for days 10 and 12 where n=2.

Along with comparing ratios of cDNA/DNA of various genes between growth curve experiments, we also wanted to determine whether expression ratios had consistent differences between growth phases as with the acetate-grown cells. Mid log and stationary phase ratios were chosen to compare because we assumed cells were in the
peak of growth in mid log while cells in stationary phase were not actively growing. It should be noted that samples taken from cells grown in glucose at 15 °C had two replicates instead of the typical three for day 10 and 12. Interestingly, for both growth curve experiments with cells grown in glucose, *rpoB* and *rplB* were the only genes with ratios that were significantly different (p < 0.01) between growth phases (Figure 6). *rpoD* was the only other gene whose expression to gene copy number ratios seemed to approach significance at p = 0.0543 and p = 0.0939 for the 30 °C and 15 °C experiments, respectively.
Figure 6. Ratios of cDNA/DNA in glucose-grown cells. Ratios are specific transcripts per gene copy number as determined by qPCR in cells grown a) at 30 °C and b) at 15 °C. Error bars indicate standard deviation. Note differences in y-axis scale.

Despite missing a wash step during protein processing, we were still interested in determining if protein concentration in the cell varied between growth phases of differing culture conditions. Cells grown in glucose at 30 °C exhibited a highly significant protein difference per cell ($p < 0.001$) between mid log and stationary (Figure 7b). This result is what was originally expected for all growth curve experiments. Conversely, cells grown
in glucose at 15 °C showed no difference between mid log and stationary (Figure 7c). A hypothesis for the unexpected protein levels in stationary and death (see Figure 4b and 7c) has two parts: 1) DNA of recently lysed cells had already degraded and cells were thus not counted after DAPI staining, but 2) the cell wall and protein content was still intact enough to be picked up in the protein assay. To test this, a linear line of best fit, seen in Figure 7b, was created to estimate the hypothetical amount of cells that may have contributed to the total measured protein in stationary phase. Figure 7c includes the following hypothetical protein/cell average if this hypothesis is true. The difference between mid log and this hypothetical average is not significant (p = 0.3124), but it does exhibit how residual proteins from nonviable cells could contribute to total measured protein.
Figure 7. Relationship between protein and cell counts in glucose-grown cells. a) Amount of protein per cell for cells grown at 30 °C. b) Linear trendline to determine hypothetical cell counts for day 10 for cells grown at 15°C. c) Amount of protein per cells grown at 15 °C, including hypothetical amount. Error bars indicate standard deviation. Three stars represent \( p < 0.001 \).
While cDNA/DNA ratios were consistently higher in log phase versus stationary or death phase, indicating higher ratios at higher growth rates, most ratios from mid log had no correlation to specific growth rates from other growth curve experiments (Figure 8a). The only gene that showed a positive correlation, with an $R^2$ of 0.9938, is $rpoB$ ($p = 0.05$). The rest of the ratios had $R^2$ values ranging from 0.03587 to 0.5164. Conflictingly, there was a negative correlation of copies of specific transcripts per cell in mid log phase to specific growth rates (Figure 8b). $R^2$ values ranged from 0.9982 for $rpoB$ and 0.9944 for $rpoS$ to the lowest value of 0.9569 for $rplB$. Only $rpoB$ and $rpoS$ had a significant correlation ($p < 0.05$). There was not a strong correlation between protein concentration per cell in mid log and specific growth rate (Figure 8c).
Figure 8. Comparisons of gene expression and protein concentration between specific growth rates at mid log. a) Ratios of cDNA/DNA and b) copies of transcripts per cell and c) amount of protein per cell. Error bars represent standard deviation. A single star represents p < 0.05.
RNAseq ANALYSIS BETWEEN LOG AND STATIONARY PHASE

In order to compare our qPCR results to relative mRNA levels between growth phases, we utilized a transcriptomic approach to identify differentially expressed genes between log and stationary phase. Cells were grown in a rich medium containing yeast extract and tryptone and incubated at 18 °C. The Geneious software program was used to map reads to the *R. pomeroyi* DSS-3 genome and megaplasmid. Differentially expressed genes between log and stationary phase time points were then identified. To be confident the genes analyzed in detail were differentially expressed, a p-value filter of 0.000001 was chosen. A total of 441 differentially expressed genes were identified and 306 had a Clusters of Orthologous Groups (COG) ID associated with them. The most abundant functional categories assigned to these 306 genes are those associated with Amino Acid Transport and Metabolism and also Translation, Ribosomal structure, and Biogenesis (Figure 9). The combined categories Function Unknown and General Functional Prediction Only make up 20% of the assigned genes. Gene product names of those genes without an assigned COG ID can be seen in Table A-1.
Figure 9. Breakdown of differentially expressed genes (p < 0.000001) by COG category.
In order to determine if the transcripts analyzed by qPCR showed similar differential expression patterns using RNA-seq analysis, we compiled the differential expression confidences of \textit{rpoB}, \textit{rpoD}, and \textit{rplB} (Figure 10). The differential expression confidence was calculated and defined in the Geneious software program as the negative base 10 log of the p-value. Values were adjusted to be negative for genes that were underexpressed in stationary compared to log and positive for overexpressed genes. Results for \textit{rpoA}, \textit{rpoN}, \textit{rpoH-2}, and \textit{Rne/Rng} were also included as qPCR primers were designed for these genes and could be useful in future experiments. DSS-3 genes \textit{rpoA}, \textit{rpoN}, and \textit{Rne/Rge} showed significant downregulation in stationary phase compared to log with p-values of 0.00027, $2.40 \times 10^{-13}$, and 0.0067, respectively. \textit{rpoD} was downregulated at stationary with a p-value of 0.0026. \textit{rplB} was upregulated in stationary with a p-value of 0.0012.

![Figure 10](image.png)

Figure 10. Differential expression confidence of select growth-related genes as a means of comparison to qPCR. One star represents \( p < 0.05 \), two stars represent \( p < 0.01 \), and three stars represent \( p < 0.001 \).
In the interest of identifying genes that could be used in similar qPCR experiments in the future, 16 differentially expressed ribosomal proteins (p < 0.000001) were identified (Figure 11). 87.5% of these genes were downregulated in stationary compared to mid log phase.

Figure 11. Differential expression confidence of ribosomal proteins that have a p-value < 0.000001.
CHAPTER FIVE
DISCUSSION

Microbes are essential to biogeochemical cycling. Variations in the activity of marine heterotrophic bacteria can have a global effect on the patterns of carbon flux (Azam and Malfatti, 2007). Many Roseobacters, including the marine model organism *Ruegeria pomeroyi* DSS-3, have the ability to metabolize DMSP (Buchan et al. 2005; Gonzalez, 2003). The metabolism of this compound can sometimes lead to the production of DMS which is responsible for the bulk of sulfur influx into the atmosphere (Andreae, 1997). Tracking the microbial activities responsible for regulating various greenhouse gases may provide insight to climate change models and could potentially aid in managing those microbial processes (Singh et al. 2010).

Many studies have found RNA content to be positively correlated with growth rate (Bremer and Dennis, 1987; Fegatella et al., 1998; Kemp et al., 1993; Kerkhof and Kemp, 1999). Researchers have also had success with using 16S rRNA:rDNA gene ratios as an indicator of growth (Campbell et al., 2011; Muttray and Mohn, 1999). However, there are substantial limitations to using these ratios as an activity indicator, as reviewed in Blazewicz et al., 2013. Because of these limitations, developing alternative measures of microbial activity in the environment is of great interest and importance. This study sought to determine if using ratios of growth-related mRNA and their corresponding genes are a viable index for estimating growth rates in the environment. Our data indicate select growth-related gene expression is different between growth phases when normalized by the corresponding gene. In acetate-grown cells, expression
of all five genes was significantly different when comparing at least one log phase growth point to a point in stationary or death phase. The lack of a significant difference between early log and mid log time points supports the hypothesis that expression of these growth-related genes is correlated with actively growing cells. Differences in gene expression may also depend on available carbon sources. Only \textit{rpoB} and \textit{rplB} were significantly different between log and stationary phase in glucose-grown cells, despite having very different growth rates at 15 °C and 30 °C. The gene expression patterns observed in this study follow the substrate metabolism patterns found in other bacteria (Kovárová-Kovar and Egli, 1998). Our results are also similar to oxygenase mRNA copies per cell obtained with \textit{Pseudomonas putida G7} grown in rich media that was extracted with a similar QIAGEN RNA/DNA extraction kit (Kong and Nakatsu, 2010).

In a study of nine marine Proteobacteria, seven of the nine strains exhibited a peak in 16S rRNA/cell in mid to late exponential phase (Kerkhof and Kemp, 1999). Because of this positive correlation, it was hypothesized all genes except \textit{rpoS} would be upregulated in log versus stationary phase. \textit{rpoS} was the exception as transcription has been shown to increase five- to ten-fold when entering stationary phase in \textit{E. coli} in rich medium (Khmel’, 2005). Previous studies have found an increase in \textit{rpoS} expression and abundance in exponential phase when comparing minimal versus rich media (Dong and Schellhorn, 2009; Tao et al. 1999). Since cells grown in glucose showed no significant differences in \textit{rpoS} expression, there must be another explanation besides the lack of a rich medium for the upregulation of \textit{rpoS} in cells grown in acetate. A simple explanation for the elevated levels of \textit{rpoS} in log compared to stationary for acetate-grown cells is
that cells were highly stressed due to the low carbon and energy availability, prompting
the “turning on” of genes regulated by \textit{rpoS}. It has been shown previously in \textit{E. coli} that
acetate induces expression of \textit{rpoS} and plays a role in acetate-induced acid tolerance
(Arnold et al. 2001; Schellhorn and Stones, 1992). These higher stress levels may also be
at least partly responsible for the differences in gene expression for \textit{ftsI} and \textit{rpoD} in
acetate-grown cells.

Comparing gene expression ratios between specific growth rates is less
straightforward than comparing between growth phases. \textit{rpoB} was the only gene whose
expression showed a positive correlation with specific growth rate when normalized by
gene copy number. When transcripts were normalized by cell counts, the expression of
all genes exhibited a negative correlation with growth rate, but only two of the
correlations were significant. When transcripts were normalized by cell counts and
compared between growth phases, similar patterns of differential expression were seen as
when normalized by gene copy number (data not shown). It is possible there was an
error in the quantification of cDNA in the 15 °C cells (either due to an error in the qPCR
standards or nonspecific amplification) and it was compensated for by the normalization
of the similarly overamplified gene copies. An alternative explanation is these genes are
actually negatively correlated with specific growth rate. Because of the previous
literature supporting the opposite result and because of our own results comparing gene
expression between growth phases, this seems the less likely option and needs further
support. This result sparks an interesting question – which of the two scenarios, specific
growth rates or the changes between growth phases, most resembles what is actually
occurring in the environment? When we compare between specific growth rates, we are making the assumption that bacteria in the environment are behaving similarly to cells in exponential phase growth. More likely, cells in the environment go through miniature growth phases – starting from dormancy when nutrient availability is poor and adjusting to nutrient influx (as from lag to log phase) or quickly switching gene expression to utilize a different substrate or combat a stressor (as from exponential to stationary). If cells in the environment indeed behave more closely to the phases in batch culture, perhaps it would be prudent to focus on the changes in expression between growth phases rather than specific growth rate.

The results of our protein analysis did not meet our expectations. We hypothesized there would be higher protein levels per cell in log phase growth when compared to stationary. Only glucose-grown cells incubated at 30 °C exhibited this difference. As already addressed, the protocol for protein analysis was missing a wash step before cell lysis, which resulted in including any contaminating protein from the extracellular matrix and media. However, I do not believe this would account for the high protein concentrations found in stationary phase for acetate-grown cells and cells grown at 15 °C. More likely, recently lysed cells that were not counted because of their loss of DNA had intact protein that contributed to the total protein quantified in stationary. It is also possible cell counts were incorrect for that time point. Additionally, there were no correlation between protein per cell and specific growth rate. In the past, protein content has been shown to increase with growth rate (Bremer and Dennis, 1987). There is also recent evidence that any increase in the abundance of some proteins results in an equal
decrease in others, resulting in a relatively constant protein mass per cell (Schmidt et al., 2011). Future experiments should be performed with an optimized protein extraction protocol.

Our transcriptome data provided additional insight into the differential gene expression of *R. pomeroyi* DSS-3 that could be helpful in selecting which genes to monitor via qPCR in the future. Amino acid transport and metabolism as well as translation and ribosomal structure made up the bulk of genes that were differentially expressed in DSS-3. Both of these functional groups are involved in changes in growth and growth rates in some bacteria. Under low nitrogen conditions, significant increases in transcripts for amino acid transporters were found in a cyanobacterium using RNA-seq analysis (Harke and Gobler, 2013). In a proteomic study performed on DSS-3 ribosomal proteins comprised 26% of the total proteins in exponential phase but only 14% in stationary (Christie-Oleza et al., 2012). Additionally, increases of rRNA and tRNA transcripts were correlated with an increase of growth rate in *E. coli* (Bremer and Dennis, 1987).

It is interesting to note which of the specific growth-related genes were up or downregulated in log or stationary phase according to RNA-seq analysis. While most highly significant ribosomal protein genes were downregulated in stationary phase, *rpsR* and *rpsQ* were upregulated. *rpsR* encodes protein S18 in the ribosomal 30S subunit in *E. coli* (Isono and Kitakawa, 1978). *rpsQ* has been shown to be nonessential in *E. coli* since *rpsQ*-knockout mutants were viable (Bubunenko et al. 2007). There is not much information on these genes that would suggest why these two genes would have high expression levels in stationary. *rpoD* only showed significant differences in acetate-
grown cells, the p-value was close to significance in glucose-grown cells at 30 °C, so to see a downregulation of \textit{rpoD} in stationary phase is not unexpected. It was surprising to see \textit{rplB} upregulated in stationary as we saw \textit{rplB} upregulated in log phase compared to stationary for three specific growth rates. \textit{rpoB} showed no differential expression. We perceive three possibilities for the discrepancy between qPCR and RNA-seq results. Firstly, the type of normalization of any transcriptome data will affect the subsequent differential expression values. Data in this transcriptome study was normalized using the median of gene expression ratios, recommended by the Geneious software and specifically Dillies et al., 2013. It is also the same type of normalization utilized in the popular R package DeSeq (Anders and Huber, 2010). It thus seems unlikely the normalization type is cause for variations in \textit{rplB} and \textit{rpoB} gene differentiation as compared to qPCR analysis.

Another possible explanation for the discrepancy between the qPCR and RNA-seq results is an issue with the RNA-seq experiment itself. After removing leftover rRNA reads in Geneious, an average of about 42\% and 26\% of reads from the log and stationary samples, respectively, mapped to the genome and megaplasmid indicating an issue in either the processing or sequencing of the reads or with the culture. Lastly, it is possible gene expression of \textit{rpoB} and \textit{rplB} changes in a rich medium. We have seen in this study that gene expression between growth phases can change depending on the carbon source. It is not unreasonable to then hypothesize gene expression may not be differentiated between growth phases for \textit{rpoB} in a rich medium. It seems less likely that \textit{rplB} would have opposite overexpression in minimal versus rich media. Regardless of which
explanation is favored, it is clear more supporting information is needed, either in the form of performing a qPCR analysis from cultures grown in a rich media, optimizing RNA-seq processing, and/or performing dual qPCR and RNA-seq analysis on the same growth curve conditions.

Although the cost-reduction of sequencing leads to cheaper projects yielding much data directly from environmental sources, these results indicate the importance of culture-dependent studies. More culture and isolation studies should be performed to elucidate the functions of the many differentially expressed hypothetical proteins found in this study. Along with the hypothetical proteins are genes that have been annotated but have unknown functions. Once these hypothetical proteins have assigned functions, we will have a clearer picture of what is happening in the cell between various conditions of growth and stress.

We found that rpoB and rplB are the top candidates for use in future qPCR experiments as they were significantly upregulated in conditions of high versus low activity across the three growth curves conducted with minimal media. In order to verify if there is a similar pattern of gene expression rich media experiments, qPCR analysis should be performed on these genes after cells have grown in half-strength YTSS media. Another possible growth curve experiment that could provide environmentally relevant information would be growing cells using DMSP as a sole or secondary carbon source. It would also be interesting to perform experiments with continuous cultures. This would eliminate any variations due to sampling at slightly different phases of growth between growth curves. Performing experiments with a mixed culture will eventually be
necessary. An ideal experimental set-up would be to perform RNA-seq analysis on a few different growth conditions, identify differentially expressed genes, and verify differentiation with qPCR analysis. An alternative to this would be to design and test primers on those transcripts we already found differentially expressed through RNA-seq. Since we ultimately aim to measure growth rates of individual taxa, it would be prudent to perform similar experiments with other microorganisms to see if 1) similar genes are differentially expressed and 2) if the amount of differentially expressed genes is dependent on ecological strategy.
### Appendix A

**Supplementary Tables**

Table A-1. Breakdown of Differentially Expressed Genes

<table>
<thead>
<tr>
<th>Differentially Expressed Genes, p &lt; 10^{-7}</th>
<th>Total Genes Identified by Locus Tag</th>
<th>Total Genes with COG ID Assignment</th>
<th>Total Genes Unassigned and Gene Product Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>441</td>
<td>439</td>
<td>306</td>
<td>133</td>
</tr>
</tbody>
</table>

- acetyltransferase, GNAT family
- alkane 1-monooxygenase (EC 1.14.15.3)
- ATP synthase F0 subcomplex C subunit
- CAAX amino terminal protease family protein
- coenzyme PQQ biosynthesis protein A
- cytochrome c550, putative
- dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133)
- helicase, ATP-dependent, putative
- hemolysin, putative
- Hpt domain protein
- ISSpo2, transposase
- 4 lipoproteins, putative
- MAPEG family protein
- monofunctional biosynthetic peptidoglycan transglycosylase
- nitrile hydratase beta subunit
- nitrile hydratase subunit alpha
- PaxA, putative
- R body protein RebB homolog response regulator
- SapC protein, putative
- serine protease, subtilase family
- tellurite resistance protein
- TPR domain protein
- 2 transcriptional regulators, LuxR family
- twin-arginine translocation pathway signal sequence domain protein
- 2 twin-arginine translocation pathway signal sequence domain proteins, putative
- universal stress family protein
- universal stress protein family protein
- 3 YeeE/YedE family proteins
- 97 hypothetical proteins

*Identified using Find Genes function in IMG/ER at http://img.jgi.doe.gov/*
REFERENCES


Schaechter, B., Malloe, O., and Kjeldgaard, N. O. (1958). Dependency on medium and
temperature of cell size and chemical composition during balanced growth of
Salmonella typhimurium. *Journal of General Microbiology*.

Schellhorn, H. E. and Stones, V. L. (1992). Regulation of *katF* and *katE* in *Escherichia
coli* K-12 by weak acids. *Journal of Bacteriology, 174*(14), 4769–4776.

of gene expression patterns with a complementary DNA microarray. *Science (New
York, N.Y.), 270*(5235), 467–470. doi:10.1126/science.270.5235.467

Schmidt, A., Beck, M., Malmström, J., Lam, H., Claassen, M., Campbell, D., and
Aebersold, R. (2011). Absolute quantification of microbial proteomes at different
doi:10.1038/msb.2011.37

during stress and starvation by modulation of sigma activity. *FEMS Microbiology
Reviews, 34*(5), 646–57. doi:10.1111/j.1574-6976.2010.00223.x

Singh, B. K., Bardgett, R. D., Smith, P., and Reay, D. S. (2010). Microorganisms and
Microbiology, 8*(11), 779–90. doi:10.1038/nrmicro2439

nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review
of Microbiology, 39*, 321 – 346.

Suzuki, M. T., Preston, C. M., Béjà, O., de la Torre, J. R., Steward, G. F., and DeLong, E.
Bacterial Artificial Chromosome (BAC) libraries from different depths in Monterey

genomics: expression analysis of *Escherichia coli* growing on minimal and rich

specificity in marine sponge-associated bacteria, and potential implications for
doi:10.1046/j.1462-2920.2003.00545.x

