# EFFECT OF ENERGY LIMITATION ON MICROBIAL EPIGENETIC MODIFICATIONS

by

Christina Baughan

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Marine Studies

Summer 2021

**EFFECT OF ENERGY LIMITATION ON MICROBIAL**

**EPIGENETIC MODIFICATIONS**

by

Christina Baughan

Approved: _____

Jennifer F. Biddle, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____

Mark Moline, Ph.D.
Director of the School of Marine Science and Policy

Approved: _____

Estella Atekwana, Ph.D.
Dean of the College of Earth, Ocean, and Environment

Approved: _____

Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

## ACKNOWLEDGMENTS

I do not know even where to begin to thank all those who have helped me throughout my graduate school career. First, I want to thank my advisor Jennifer Biddle for not only her guidance and expertise, but also for being an advocate when I needed her. I will forever be thankful for my community college professor Virginia Balke for encouraging me to apply for an internship with the Center for Dark Energy Biosphere Investigations which connected me to my advisor. I am so thankful that she warmly accepted me into her lab as an intern and technician to finally her graduate student. Thank you to all the lab members for all your help and support along the way. Thank you, Sabrina Beckmann, Andrew Franca, Malique Bowen, Rui Zhao, Ibrahim Farag, Kristin Yoshimura, Kaliopi Bousses, Leaundra Schopflin, Glen Christman, Nikki Dumigan, and Sydney Messick. I would also like to thank my committee members, Dr. Thomas Hanson and Dr. Adam Marsh. Thank you to all my professors for all your help. I've made some great friends at the college along the way thank you so much for being there. A big thanks to my Air Force family for your support. To my friends that I've made here thank you so much for your support and being my biggest cheerleaders through this. To my Mom and Dad, gracias por tu apoyo y tu cariño. Many thanks to my big sister for all your support miles away. My husband, Chris, has been my biggest supporter always taking a load off of me in the house when I needed him to, knowing exactly what I need before I even know myself.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## ABSTRACT

Epigenetic modifications are often induced in eukaryotic organisms in a response to stress. These modifications can alter gene expression by methylating the genome which can turn genes on or off. These modifications have not been as widely investigated in prokaryotic organisms. Microbial populations in the deep sea are able to sustain standing populations despite no new known energy inputs. While studies have attempted to investigate their capability of surviving in these low nutrient environments it is often attempted by inducing nutrient limited stress instead of energy limited stress. Energy limited stress would keep the organism metabolically active while nutrient limited stress would reduce an essential nutrient required for biomass. Here we induce energy limited stress to two organisms, *Pseudomonas aeruginosa* and *Methanosarcina mazei* and sequenced their genomes using long-read sequencing that includes kinetic information to detect modified bases. *Pseudomonas aeruginosa* PAO1 was energy starved using various amino acids as the sole carbon and energy source. Indication of stress in PAO1 include a slowed growth to the production of a secondary metabolite that is often present in virulent PAO1. *M. mazei* was energy starved by reducing acetate, its preferred electron source based on its methanogenesis pathway. DNA analysis of PAO1 revealed a hypomethylated genome during the stationary phase of the starved condition with the majority of the modifications occurring in the rRNA region of the genome. Further analysis reveals a difference in OMPK of the starved condition to the normal condition in which genes that were hypermethylated were associated with proton transport and protein metabolic processes.

# Chapter 1

# ENERGY LIMITATION ON *PSEUDOMONAS AERUGINOSA* AND *METHANOSARCINA MAZEI*

## 1.1 Introduction

Once thought to have harbored no life due to its extreme environments, we now know active microbial communities have been observed in sediments hundreds of meters beneath the seafloor. Even in the very low nutrient area of ocean sediment of the South Pacific gyre microbial communities are able to persist (D'Hondt et al., 2015). The deep biosphere maintains populations of up to 90% of Earth's total prokaryotic biomass (Bar-On, Phillips, & Milo, 2018). Geochemical measurements from these areas reveal that microbes are metabolically active but have not received any new energy inputs. Microbial communities in these areas are somehow able to survive despite the low amount of energy to sustain their populations. Our understanding of microbial energy metabolism largely comes from cultures that grow rapidly with high metabolic rates and high densities. Growth characteristics from cultures that are grown in the laboratory are quite different from microorganisms in nature.

Only about 1% of bacteria are able to be cultured in the laboratory, so the limited information we have on growth is focused on these typically fast growers. It is difficult to determine the growth of bacteria that could be living an oligotrophic lifestyle in the deep sea. It is thought these communities either exist in a state of

dormancy or starvation with no growth (Schut et al., 1993). To further explore the idea many studies have been carried out while organisms are placed under nutrient limitation stress to mimic their natural environment. A population of *E. coli* has been maintained in the stationary phase without the addition of nutrients for long periods of time. A type of mutation labeled the "growth advantage in stationary phase", or GASP, has been shown to increase fitness mutations (Finkel & Kolter, 1999). These mutations allow for the cells that carry the mutation to outcompete those that do not carry the mutation. However, when looking at the bioenergetics of necromass as an energy source it does not seem favorable to sustain growth in a large population (Bradley, Amend, & LaRowe, 2018). While the energy outputs from necromass do provide enough energy to a cell for growth it is not enough based on the populations on the seafloor.

Other bacterial responses to stress include the formation of spores, created by specific lineages of bacteria such as *Bacillus*. It has been discovered that non-sporulating *Bacillus subtilis* can survive months of deep starvation. Once thought to have been dormant they have been shown to be actively growing and have a distinct transcriptome from that of stationary cells (Gray et al., 2019). It was not clear if this type of growth is transcriptionally regulated. Considering the available models for the low energy subsurface, we find that no models actually describe the environment in question.

Studies that attempt to mimic the natural environment of deep sea tend to only focus on reducing the amount of available nutrients instead of focusing on the energetics of the organism. Nutrient limitation is the limitation of an essential nutrient that is required for biomass synthesis. The essential nutrient could be a compound or

element, but not the entire pool of nutrients (Lever et al., 2015). An organism's source of electrons could be nitrate or acetate, but these energy sources can also serve as nutrients. An example of nutrient limited stress would be an organism's inability to biosynthesize cellular material. The term nutrient limitation is often used interchangeably with energy limitation, but they are actually two different things. Energy limitation can be defined as reducing the energy inputs a microbe would need to be metabolically active such as reducing nitrate. Energy limitation is different across organisms, as the energy limitation for a phototrophic organism would be light. An organism can be both nutrient limited and energy limited and also one and not the other. Growth in the laboratory is often under optimum conditions and are not indicative of growth in nature. In order to gauge how microbes grow in nature, we have to look at growth when the energy supply is reduced. Even though there is organic matter, or food, available to these organisms it does not mean they are able to utilize it as energy (Morita, 1990). Their energy source could technically not be available, and these organisms would remain in a starvation or energy limited state.

To examine energy starvation, I will consider two drastically different microbe groups, methanogens and *Pseudomonas aeruginosa.* Methanogens are obligate anaerobes and have three pathways for substrate degradation and methane formation, acetoclastic, hydrogenotrophic, and methylotrophic. The pathways are named after their electron donors. Methanogens derive their metabolic energy by the degradation of substrate to methane (Deppenmeier, Müller, & Gottschalk, 1996). Limiting energy from methanogens would be to limit the substrate they utilize in their methanogenesis pathway. *Pseudomonas aeruginosa* PAO1 is an opportunistic pathogen that has been associated with infection in the lung in patients with cystic fibrosis. Previous studies

have revealed aromatic amino acids to enhance the *Pseudomonas* quinolone signal which are important for production of virulence factors (Palmer, Palmer, Jorth, & Whiteley, 2010). We are also going to take in consideration the difference in cells in exponential phase and in stationary phase. Physiological changes occurs within a cell once in stationary phase such as the thickening of the peptidoglycan layer and changes in gene expression (Payne, McCarthy, Johnson, North, & Blum, 2018).

It has not been determined as to how relatively "normal" bacteria are able to thrive in low energy environments, but a clue to how they survive energy stress may be held in their genomes. In eukaryotes, genes are often regulated by DNA methylation which can alter gene expression (Willyard, 2017). Environmental stress can induce epigenetic changes in eukaryotes. An example would be the introduction of sensitive corals to ocean acidification which revealed a decline in growth and an increase of methylated DNA (Aranda et al., 2018; Dixon, Liao, Bay, & Matz, 2018). The thought is that over time these epigenetic modifications can acclimatize these organisms to the fast-growing changes due to climate change.

We can view the impact of methylation on *Alphaproteobacteria* where the interaction between GcrA/CcrM-mediated regulation has been linked to a starvation response (Mouammine & Collier, 2018). We can also observe methylation within metagenomes. A metagenomic analysis of shallow estuarine sediments revealed the presence of methylation states in CpG sites (Rambo, Marsh, and Biddle, 2017). As such, we hypothesize that methylation may play a role in how microbes are able to regulate their machinery under low energy. Long-read sequencing by PacBio is able to directly detect DNA modifications by using kinetic information from each nucleotide addition to call bases. SMRT sequencing will be used to uncover the patterns of

modifications across different organisms under energy limited stress. Modification sites will be tracked and used to generate a methylome for pure cultures in order to determine the genes which are being regulated by DNA methylation and their relationship to energy limitation stress. We will also track energy stress via changes to the lipid pool of the organisms, as the lipids are typically modified under periods of stress in microbes.

## 1.2 Materials and Methods

### 1.2.1 Cell Culture

#### 1.2.1.1 PAO1 Growth Conditions and Cultivation

*Pseudomonas aeruginosa* PAO1 cultures in triplicate were inoculated in LB and grown over night in a shaking incubator at 37°C. PAO1 was then inoculated in a minimal media of a MOPS buffered salts base (50 mM MOPS, pH 7.2, 93 mM NH4Cl, 43 mM NaCl, 3.7 mM KH2PO4, 1 mM MgSO4, 3.5 mM FeSO4·7H2O) supplemented with 20 mM concentrations of either phenylalanine, tyrosine, alanine, or glucose as the sole carbon source. Growth was tracked by measuring cell density with OD600 readings on a spectrophotometer. Growth rates for each growing condition were plotted and the slowest growing method was chosen to carry out further studies.

#### 1.2.1.2 Methanogen Growth Condition and Cultivation

Archaeal strains *Methanosarcina baltica* GS-1A and *Methanogenium frigidum* Ace-2 were purchased from DSMZ. Archaeal strains *Methanosarcina mazei*, *Methanosarcina burtonii*, and *Methanosarcina barkeri* were provided by a collaborator. Vitamin solution and trace element solution were made according to

protocol provided by DSMZ. Medium 141 Methanogenium medium ($H_2/CO_2$) for *M. frigidum* was amended from the DSMZ protocol by excluding L-Cysteine-HCl x H2O from solution. Medium 141c Methanococcoides for *M. baltica* medium ($N_2/CO_2$) was amended from the DSMZ protocol by excluding L-Cysteine-HCl x H2O and trimethylammonium chloride. Medium 120 Methanosarcina medium for *M. mazei, M. burtonii, and M. barkeri,* was amended from the DSMZ by excluding L-Cysteine-HCl x H2O from solution. Methanogens were grown anaerobically in 250mL serum bottles with rubber stoppers. All media was sparged with $N_2$ gas for 20 minutes in the solution then 10 minutes in the headspace. It was then sparged with $N_2/CO_2$ gas for 15 minutes in the solution and 5 minutes in the headspace. The methanogenium medium was sparged with $H_2$ gas for 30 seconds in the solution. All media was then autoclaved for 30 minutes. Incubation conditions for *M. frigidum* were 15°C in the dark and for *M. baltica* 20°C in the dark. Growth was monitored by sampling 0.5mL of culture, staining with SYBR green, and direct cell counting under the microscope. Methane production was also quantified by GC analysis of a 1mL sample of the headspace using a gas tight syringe. Starvation conditions for each methanogen restricts the amount of electron donor availability in the media based on the energetic pathway of that methanogen. *Methanosarcina baltica* was inoculated in triplicates on varying concentrations of acetate from 0 mM to 30 mM for one condition and in varying concentration of methanol for another condition. *Methanogenium frigidum* was inoculated in varying concentrations of acetate from 0 mM to 30 mM. Growth rates for all methanogens were very slow growing that further starvation studies were carried out solely on *Methanosarcina mazei*. *M. mazei* was inoculated in triplicates on varying concentrations of acetate 0 to 30 mM in increments of 5.

### 1.2.2 DNA Extraction

Samples were collected from the exponential phase and stationary phase by centrifuging cell cultures at 10k rpm and removing the supernatant. Samples that were not processed immediately were stored at -80°C. DNA was extracted from samples using the Qiagen DNeasy kit. Modifications for the lysis step were made for the *M. mazei* samples which include a proteinase K treatment and an overnight hot water bath at 56°C. Vortexing was also done mildly as to prevent shearing of the DNA of the *M. mazei* samples. Concentrations of DNA were quantified using Qubit dsDNA HR Assay Kit and Qubit 3.0 Fluorometer. Samples were submitted to the University of Delaware DNA Sequencing & Genotyping Center for fragment size analysis and PacBio DNA sequencing. The library was prepared with SMRTbell Template Prep Kit 1.0, Sequel Binding Kit 2.1, Sequel DNA Internal Control 2.1, Sequel Sequencing Plate 2.1 (4 rxn), Sequel OS Enzyme, and Sequel SMRT Cell Oil, and run on a single PacBio Sequel SMRT cell for sequencing.

### 1.2.3 Sequence Analysis

The PacBio .bam files for PAO1( FE, FS, GE, and GS) were uploaded to the University of Delaware Biomix Cluster. Data was analyzed within the cluster using command-line tools. Coverage was plotted and compared for each condition. High confidence modifications were filtered with the Bacterial_Basemod_Analysis openly available at https://github.com/PacificBiosciences/Bioinformatics-Training/tree/master/basemods using the BaseModFunctions.v2.1.R and Mbovis_basemods.v2.1.R scripts with a slope of 60/150, x-intercept of 0 and greater than 30x coverage. Modifications in each condition were visualized in R. The bam files were aligned to the PAO1 reference genome obtained from NCBI using pbmm2.

Modifications types were clarified by either modified bases or m6A or m4c methylation using the ipdSummary tools from SMRTlink. The PAO1 total genome annotation obtained from NCBI was used to determine the location of unique modifications using BedTools. Gene load scores were compiled using a python script that parsed the gff file and summated the total number of modification type that was found within the positional boundaries of each gene. The gene load score was normalized with the following equation to obtain the observed modifications by kilobase(OMPK).

$$OMPK = \text{Gene load score}/(\text{Gene length}/1000)$$

The fold change ratio was calculated for FE and GE using OMPK and the equation.

$$(FE-GE)/GE$$

Genes with a fold change less than -1 and greater than 1 were chosen for further analysis. The list of genes was passed into DAVID, an online bioinformatics resource for functional annotation (Huang, Sherman, & Lempicki, 2009). The GO biological process database was used to curate a list of GO terms along with their p-value and counts.

### 1.2.4 Archaeal Lipid Extraction

The modified method for total lipid extraction was adapted by Bligh and Dyer, 1959. One Teflon tube was prepared for each sample by washing with 10 mL DCM:20 mL MeOH: 8 mL P-buffer, sonicating for 10 minutes right-side up followed by 10 minutes upside-down. Tubes were then rinsed with DCM and MeOH. Tubes were pre-filled with a B&D solvent mix of 8 mL TCA< 5 mL DCM and 10 mL MeOH. Cell pellets from the exponential and stationary phase were thawed and dislodged from the microcentrifuge tube with DCM and moved into prepared 50mL tubes. Solvent was

added to the tube and inverted to mix. Samples were sonicated for 20 minutes in a water bath, rested for 10 minutes in an ice water bath, and sonicated for an additional 20 minutes. Contents of extract were transferred to a separatory funnel with 10 mL of DCM and 10 mL of Gen Pure water and inverted to mix. The organic phase was drained off and collected. After draining off the organic phase 20 mL of DCM was added, and the organic phase was drained off and collected again. This process was repeated 3 times. The samples were dried down and concentrated to 1 mL volume. Samples were stored in -80°C until processed. Samples were run on an Agilent QTOF mass spectrophotometer to observe lipid structure.

## 1.3 Results

### 1.3.1 PAO1

#### 1.3.1.1 Starvation

Growth varied across all substrates as seen in Figure 1.1. Growth is stunted on all amino acids while growth on the same minimal media but supplemented with glucose shows more favorable growth. Growth on LB while initially is at a higher cell density shows a steeper slope and is therefore growing the quickest (Figure 1.1). Another growth experiment was carried out with varying concentrations of phenylalanine to observe if the growth rate could be further restricted. Varying the concentrations of phenylalanine did little to alter the growth rate further and 20 mM of phenylalanine remained as the starvation condition with 20 mM glucose as the normal condition.

#### 1.3.1.2 DNA Extraction

DNA was extracted from the exponential phase of the glucose starvation at 10 hours at stationary phase at 24 hours. DNA was extracted from phenylalanine at exponential phase at 40 hours and at stationary phase at 72 hours. A total of 14µg.of DNA was extracted for sample GE, 5 µg for sample GS,7.5 µg for sample FE, and 6.16 for sample FS.

#### 1.3.1.3 DNA Analysis

The total number of modifications was extracted from the bam files using command line tools as seen in Figure 1.3. Glucose exponential, GE, contained a total of 319,852 modifications. Glucose stationary, GS, contained a total of 319,551. Phenylalanine exponential, FE, contained a total of 326,848 modifications. Phenylalanine stationary, FS, contained a total of 3,568 modifications. FS has a significantly lower number of modifications compared to other three conditions (Figure 1.3). The sequence files obtained from the sequencing center as a bam file were all similar sizes of about 7-8 gb. The SMRT link ipd summary tool analyzed the kinetics of the sequencing files to obtain gff and csv files containing the modifications. Modifications that were called were m6a, m4c, and modified bases. Modified bases are not necessarily methylated, but the ipd ratio indicated a pause at that location. Modifications sorted by type reveal similar trends across three of the conditions (Figure 1.3). The gff file obtained from the SMRT tools was much smaller for FS than the other three conditions. It was reanalyzed to verify the SMRT tool was working correctly.

The modifications annotated to a previously assembled PAO1 genome reveal similarities within each condition (Figures 1.4-1.8). In the tRNA regions FE contained

a larger amount of m6a (Figure 1.4). Modifications across the protein coding regions are similar across all conditions (Figure 1.5). Modifications were much less on FE, but of those modifications they were localized to rRNA regions (Figure 1.6). Modifications on the intergenic regions show no detectable data for most types of modifications for FS (Figure 1.7). Modifications on the noncoding regions are similar for FE, GE, & GS, but FS does not contain as many (Figure 1.8).

OMPK values are displayed in a heatmap (Figure 1.2) to show the similarity between each condition. GE and GS are mapped close to each other which indicated that they have a similar number of modifications on the same genes. FE and FS have branched off on their own. Most of the variation comes from both glucose conditions and phenylalanine exponential. The histogram of the fold change between FE & GS shows the frequency to determine the cut-off score for futher analysis. After removing genes that have no score an any condition and a fold change of either ≥1 or ≤-1. Fold change between FE& GE generated a list of 550 upregulated genes and 41 downregulated genes. Table (1.2 & 1.3) show the results of the functional annotation of the enrichment analysis. Filtering out results based on p-value > 0.05 leaves us with a list of GO terms.

### 1.3.2 Methanogens

#### 1.3.2.1 Starvation Experiments

Starvation experiments started with three psychrophilic methanogens, *M. frigidum, M. baltica, and M. burtonii*. Growth was only observed occurring in *M. frigidum* after 4 weeks. The two other methanogens were no longer considered for further experiments. Growth with *M. frigidum* occurred very slowly, without

producing a discernible exponential phase to use for further analysis. Starvation experiments moved forward with the more easily cultivatable methanogen *Methanosarcina mazei.* Figure 1.11 reveals slow growth with acetate as the sole energy source. 30mM was used as the baseline condition and 15 mM was used for the starvation condition.

#### 1.3.2.2 DNA Extraction and Analysis

DNA was extracted from both the exponential and stationary phases of both conditions. DNA concentrations were low for both conditions despite calculations from direct cell counts that suggest more biomass should have been available. Demultiplexed samples files were received from the sequencing center and the data is currently being processed.

#### 1.3.2.3 Lipid Analysis

Lipid analysis failed to produce any results on the first run. Results are pending on a successful second run.

Figure 1.1: PAO1 growth curve on 20mM concentrations of glucose, alanine, phenylalanine, and tyrosine. Growth on LB for comparison.

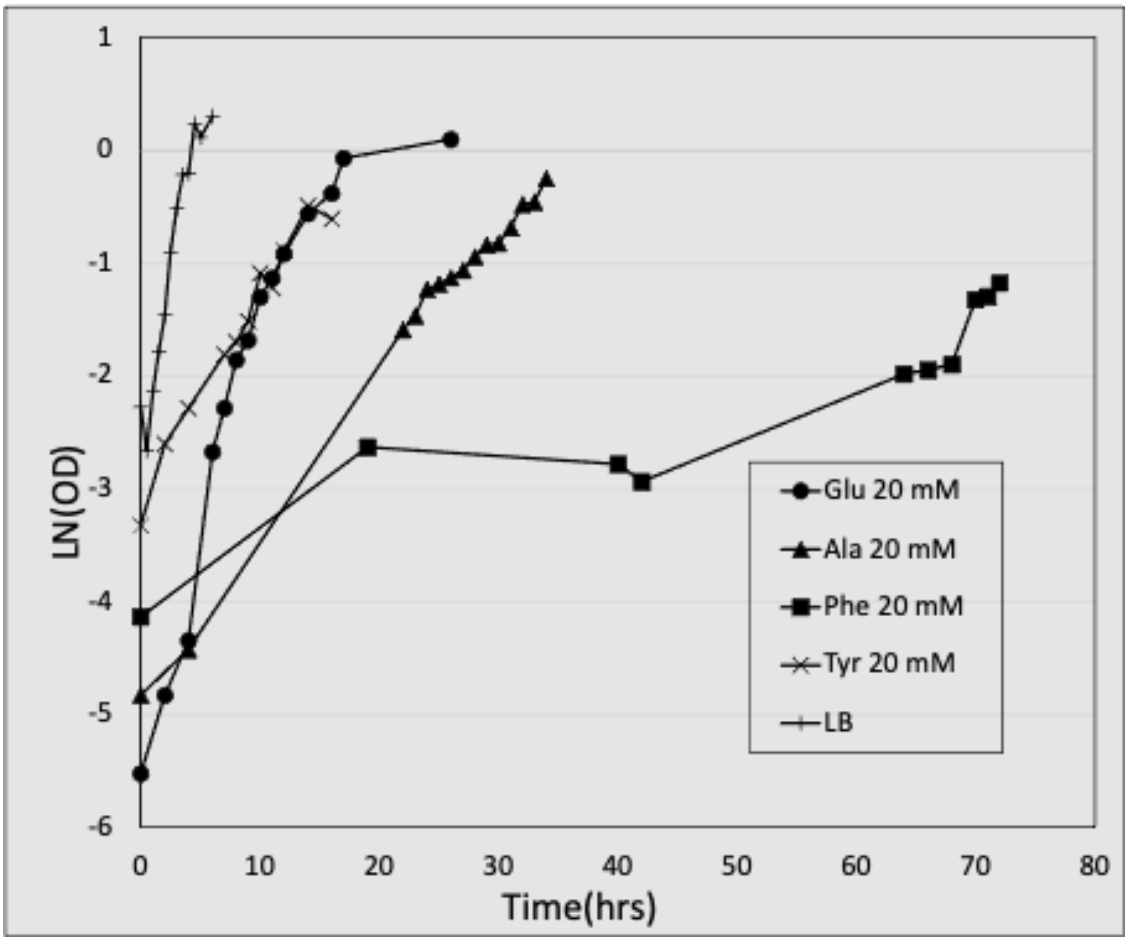


Figure 1.2: PAO1 growth curve on varying concentrations of phenylalanine.

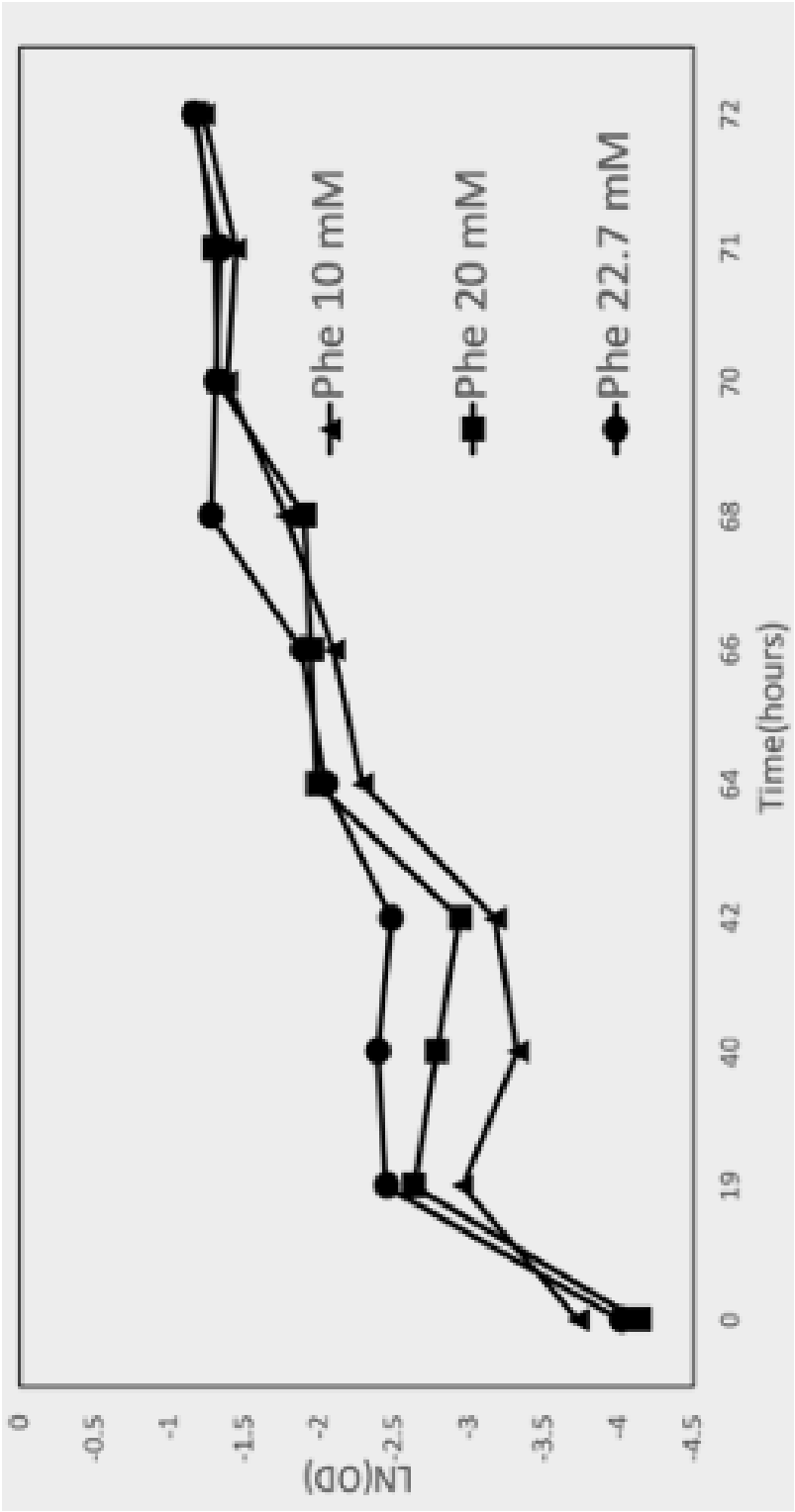


Figure 1.3: Bar plot of modifications by type in each condition. Modification type is either m4C, m6a, or a modified base. FE, FS, GE, GS are phenylalanine exponential, phenylalanine stationary, glucose exponential, and glucose stationary.

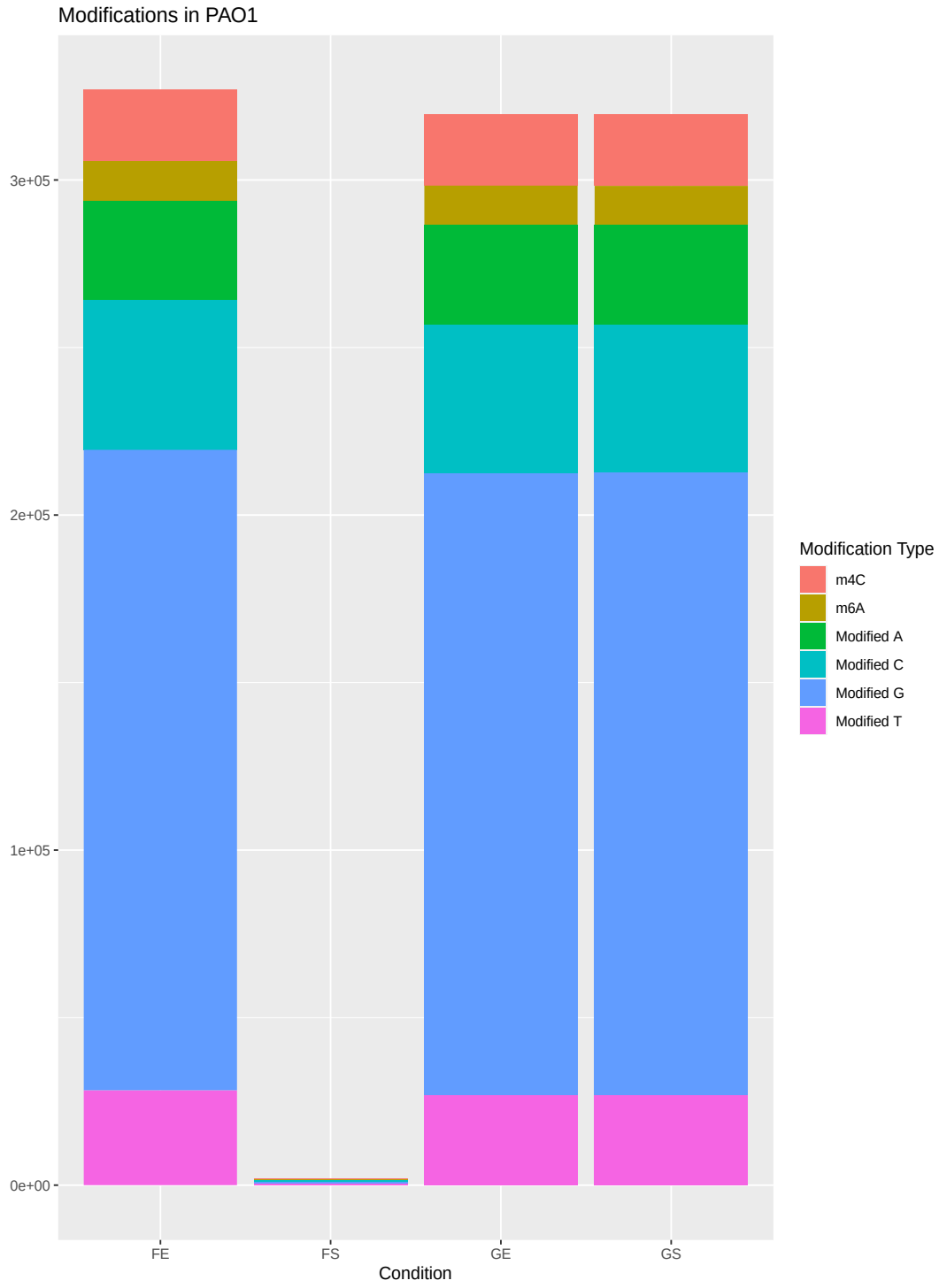


Table 1.1: Sequence quality data obtained from the PacBio SMRT Portal.

| Sample Name | Polymerase Reads | Subreads | Bases | Mean Read Length | Longest Subread Length | Mean Barcode Quality |
|---|---|---|---|---|---|---|
| FE | 123,438 | 651,770 | 4,468,408,904 | 36,536 | 72,969 | 66 |
| FS | 76,628 | 356,406 | 2,749,919,404 | 36,213 | 73,642 | 67 |
| GE | 120,639 | 617,430 | 4,342,831,010 | 36,375 | 70,756 | 66 |
| GS | 107,756 | 580,119 | 4,008,986,537 | 37,567 | 83,221 | 68 |

Figure 1.4: Modifications by type in the tRNA region of the PAO1 genome.

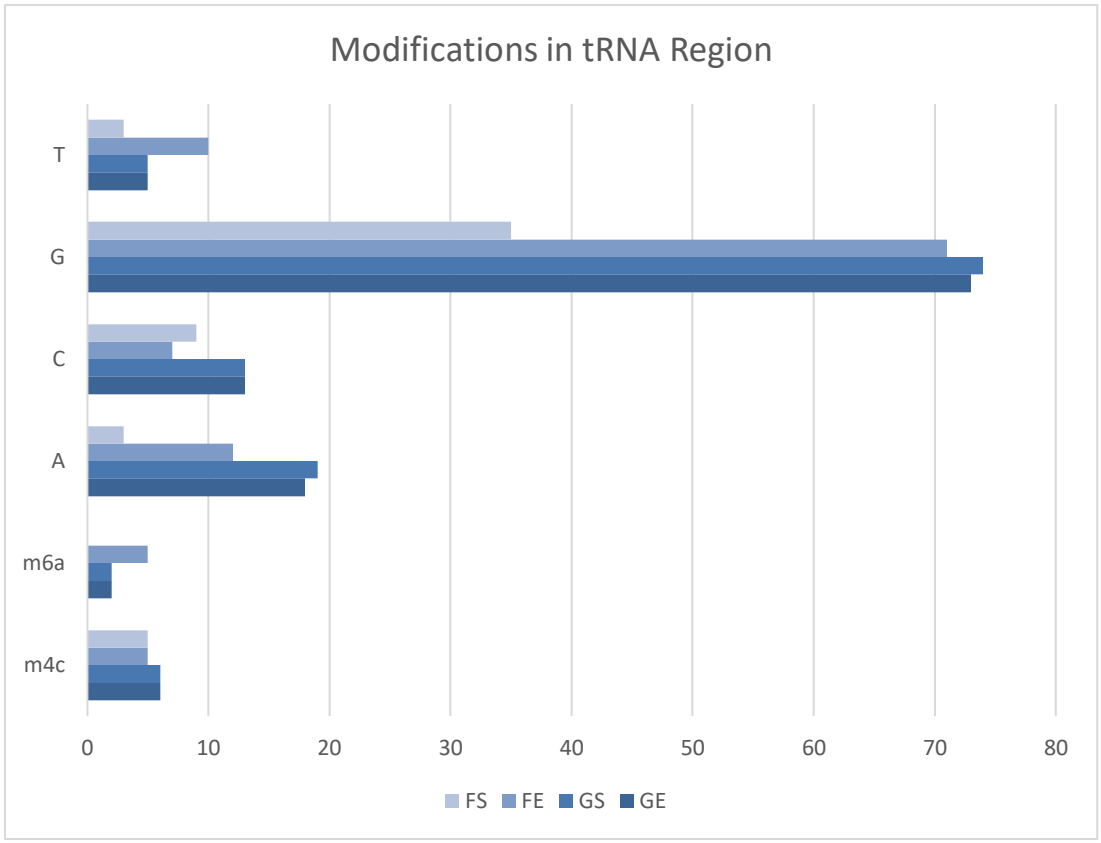


Figure 1.5: Modifications by type in the protein coding region of the PAO1 genome.

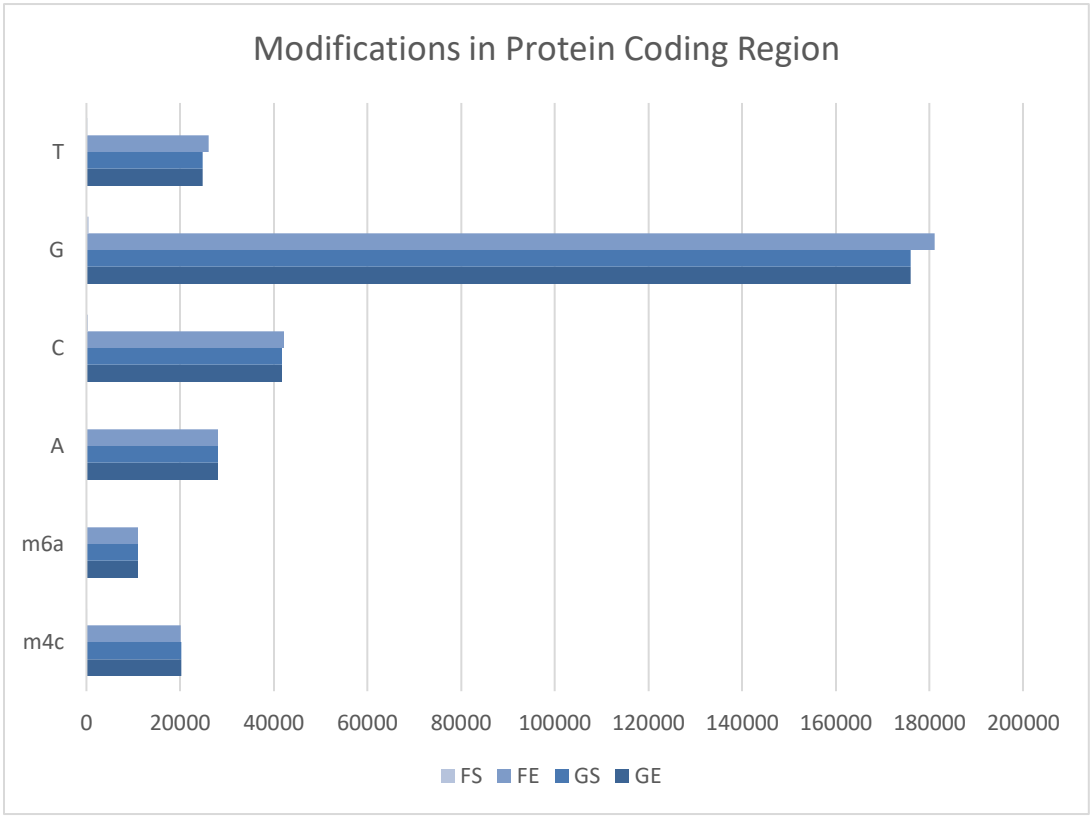


Figure 1.6: Modifications by type in the rRNA region of the PAO1 genome.

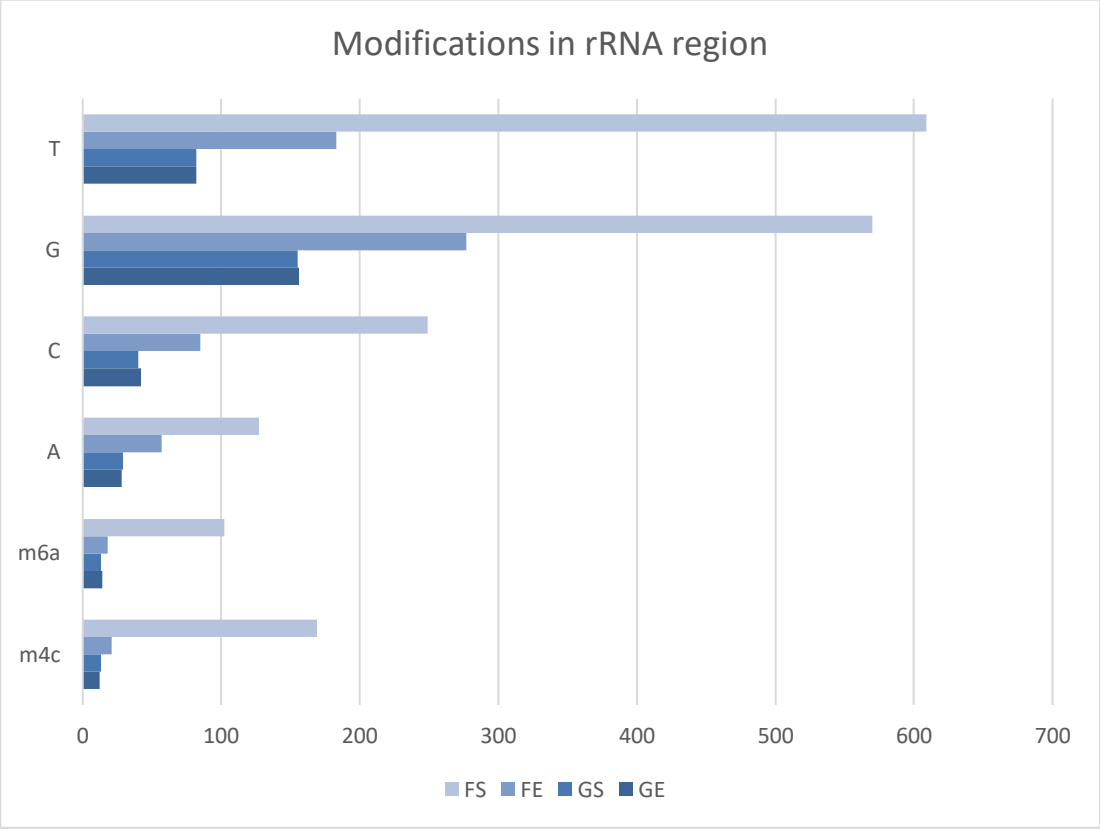


Figure 1.7: Modifications by type in the noncoding region of the PAO1 genome. FS did not detect any modifications in this region.

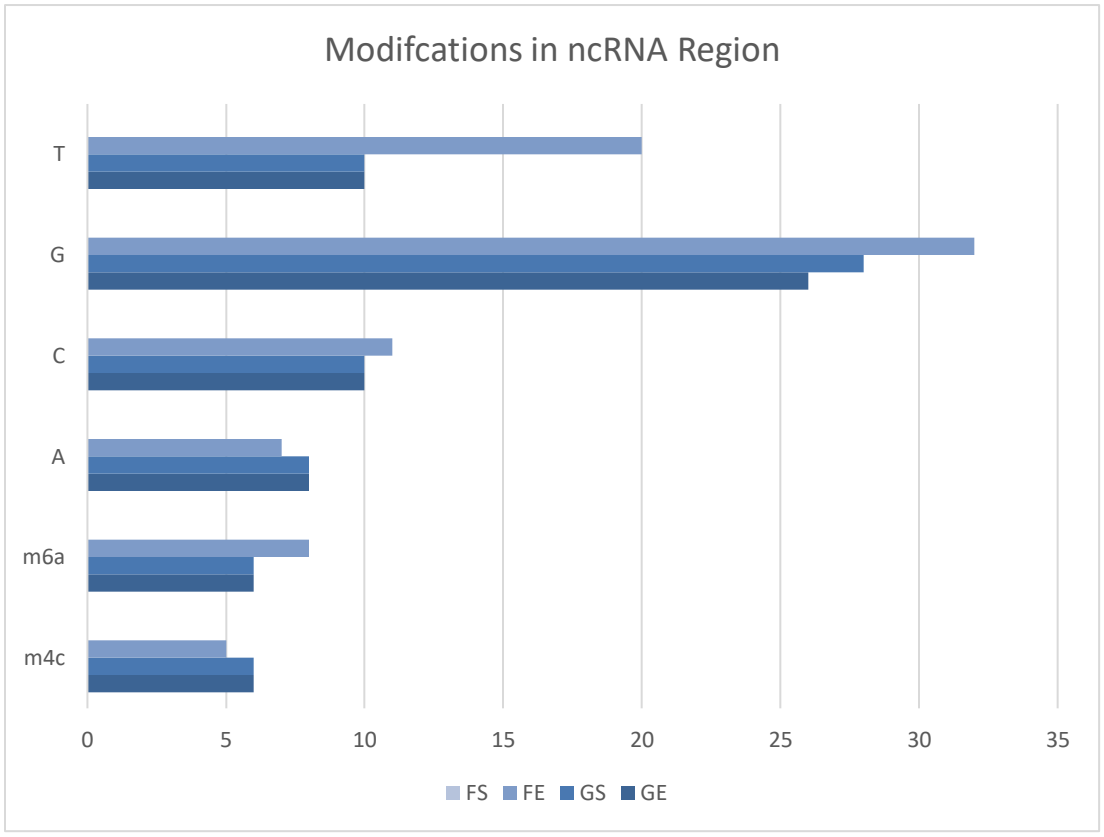


Figure 1.8: Modifications by type in the intergenic region of the PAO1 genome.

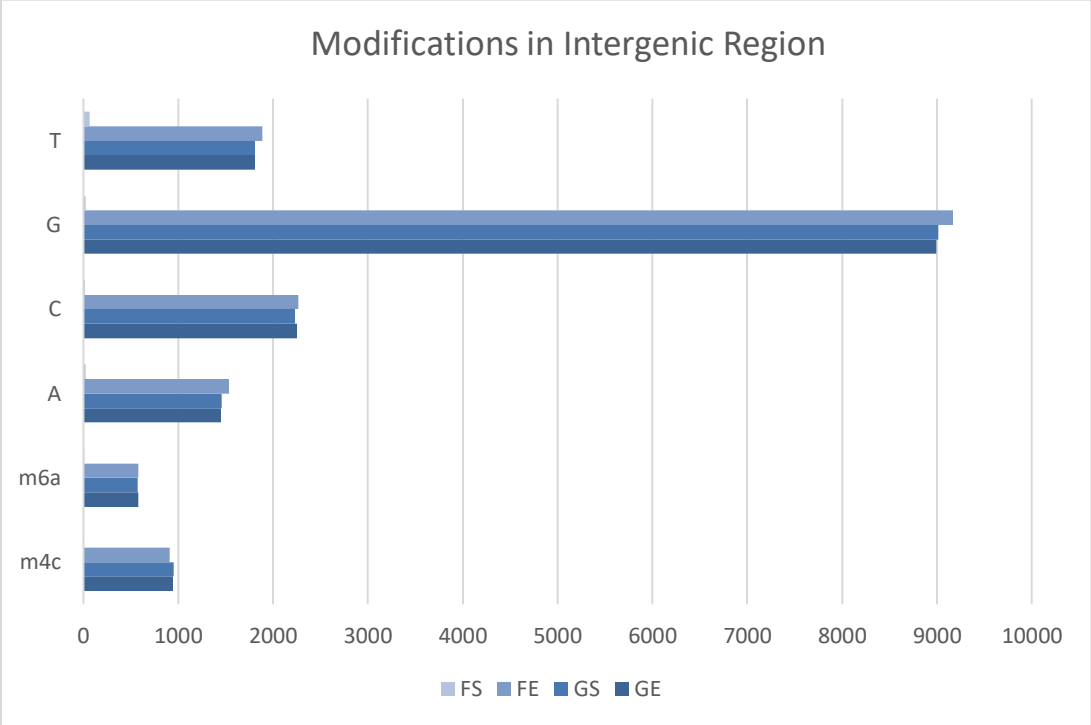


Figure 1.9: Per strand coverage plots of each condition. A) Glucose exponential phase B) Glucose stationary phase C) Phenylalanine exponential phase D) Phenylalanine stationary phase. Pink is modified As, green is modified Cs, blue is modified Gs, and purple are modified Ts.

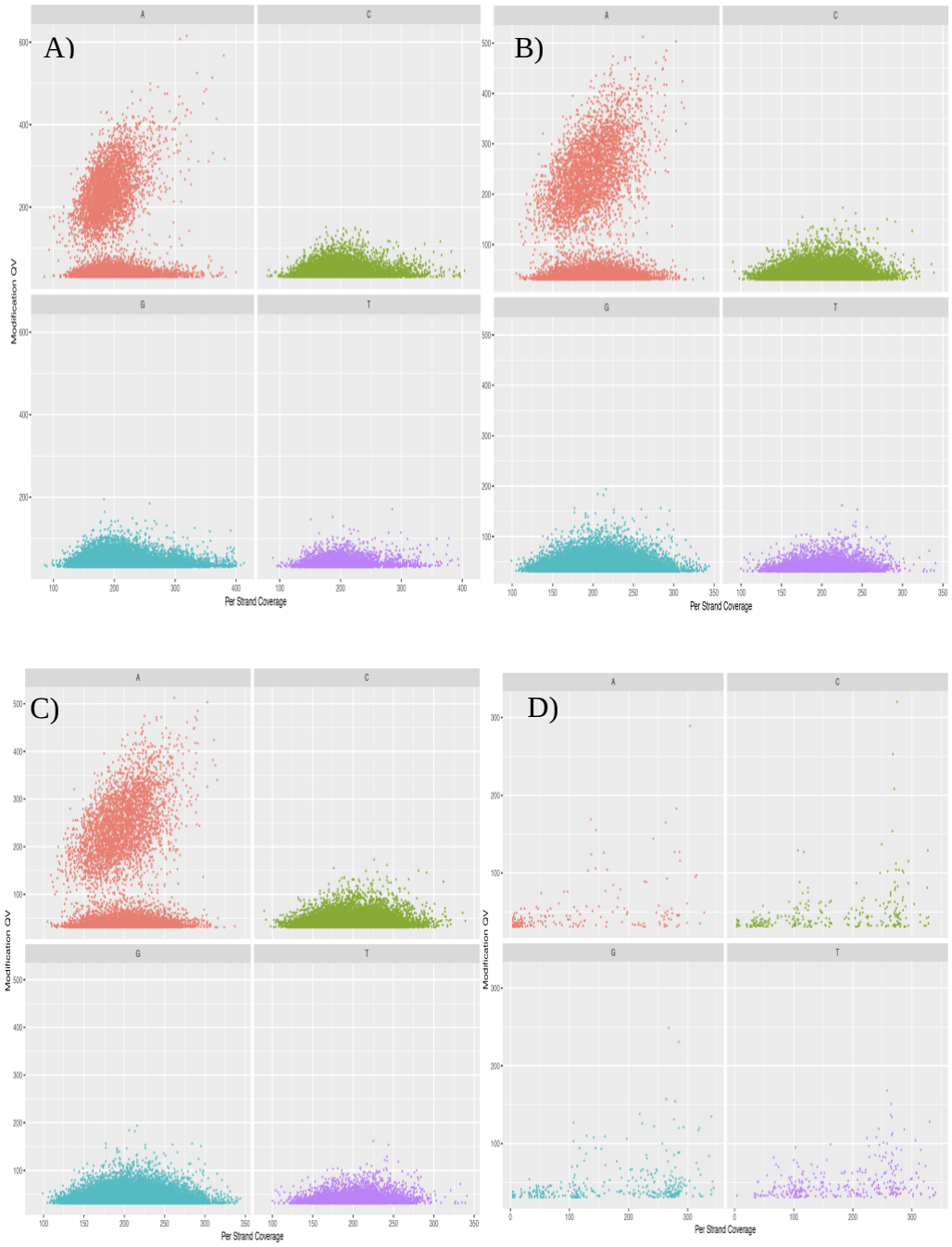


Figure 1.10: Heatmap of observed modifications per kilobase (OMPK) for each condition showing similarity between conditions. Heatmap was generated with scaling of rows. List of genes are on the x-axis.

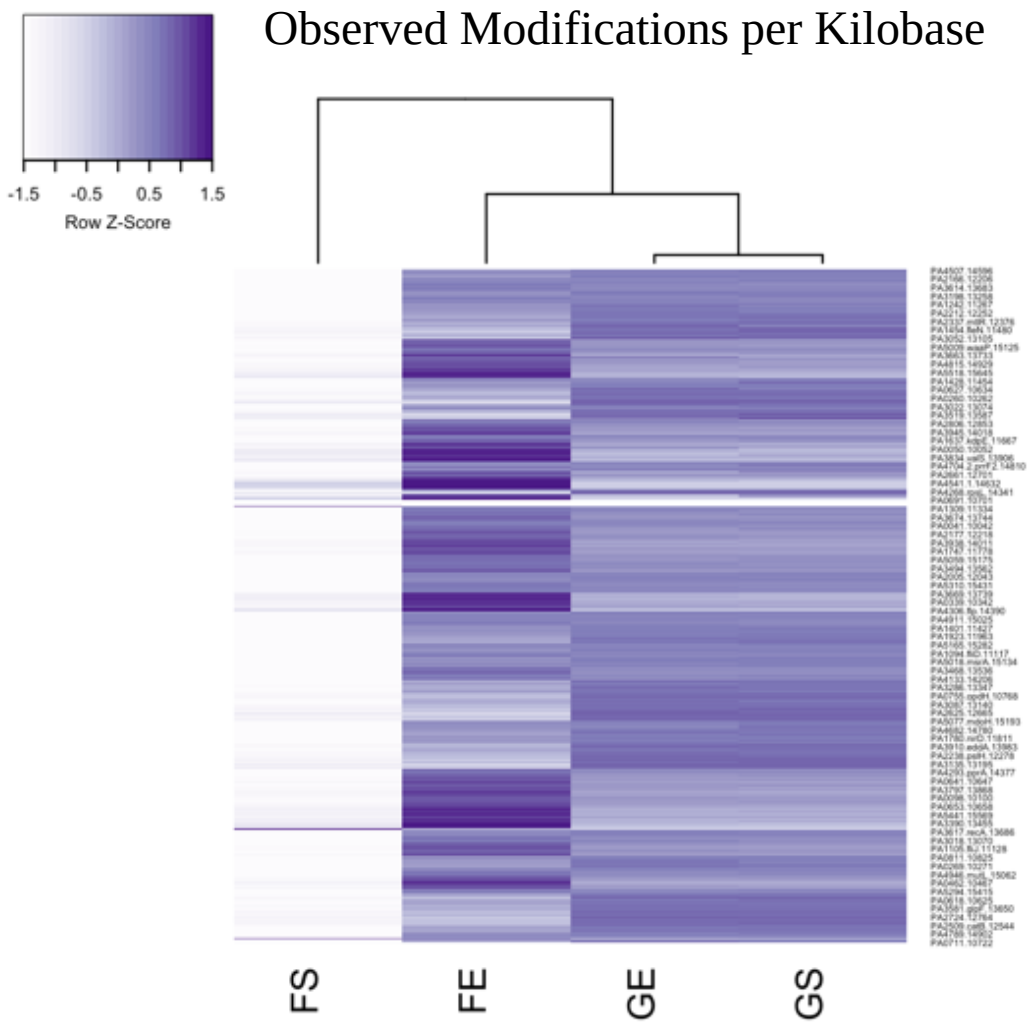


Figure 1.11: Histogram with the fold change between conditions FE & GE. Fold change represents the ratio for differentially modified genes. Fold change greater than 1 and less than -1 represent differentially modified genes.

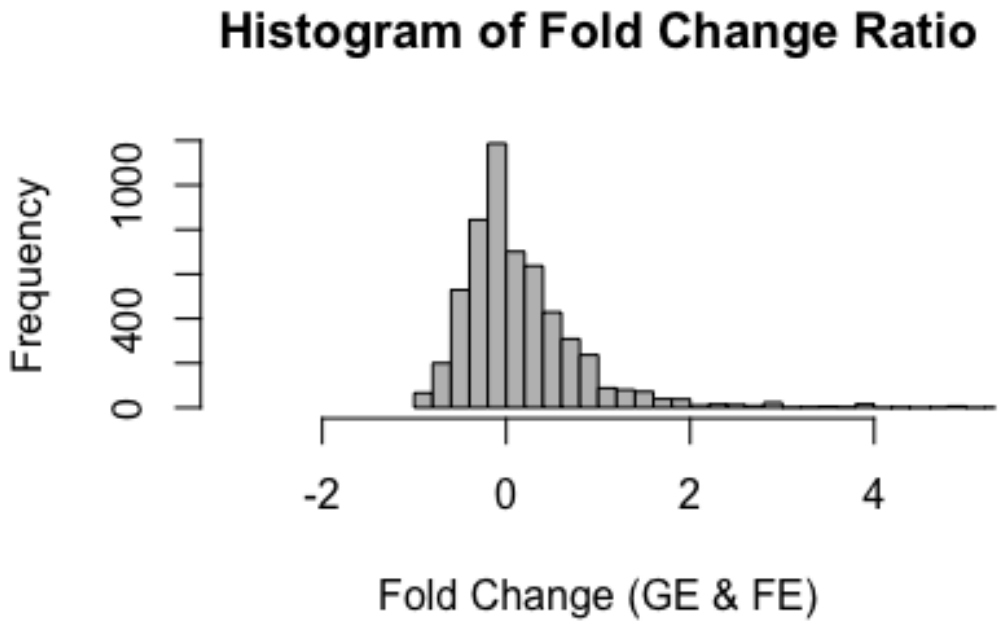


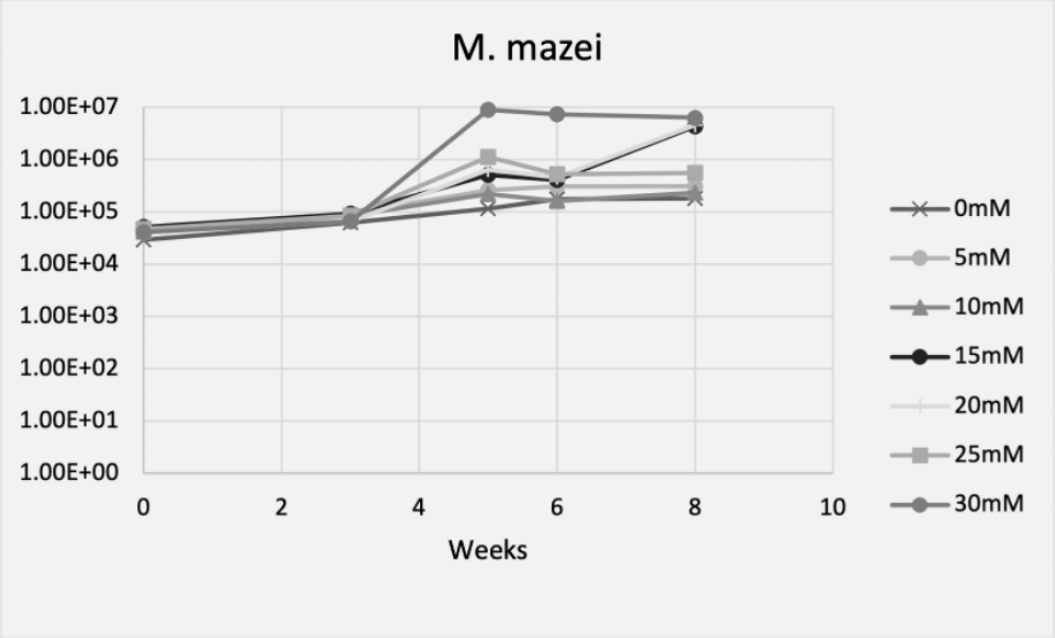
Table 1.2: DAVID functional annotation results curated from a list of 550 upregulated genes. Only significantly enriched biological themes are shown.

| GO ID | GO Term | Count | PValue |
|---|---|---|---|
| GO:0006818 | hydrogen transport | 34 | 4.10E-04 |
| GO:0015992 | proton transport | 34 | 4.10E-04 |
| GO:0044267 | cellular protein metabolic process | 24 | 5.14E-04 |
| GO:0019538 | protein metabolic process | 20 | 0.002183491 |
| GO:0015672 | monovalent inorganic cation transport | 16 | 0.006662614 |
| GO:0051704 | multi-organism process | 11 | 0.01671263 |
| GO:0010035 | response to inorganic substance | 9 | 0.020466769 |
| GO:0043455 | regulation of secondary metabolic process | 8 | 0.021017997 |
| GO:1900376 | regulation of secondary metabolite biosynthetic process | 8 | 0.021017997 |
| GO:2000021 | regulation of ion homeostasis | 8 | 0.021218446 |
| GO:0061088 | regulation of sequestering of zinc ion | 7 | 0.021218446 |
| GO:0032844 | regulation of homeostatic process | 7 | 0.021218446 |
| GO:0034220 | ion transmembrane transport | 7 | 0.023926473 |
| GO:0015986 | ATP synthesis coupled proton transport | 7 | 0.026035802 |
| GO:0015985 | energy coupled proton transport, down electrochemical gradient | 6 | 0.026035802 |
| GO:0030162 | regulation of proteolysis | 6 | 0.027612303 |
| GO:0098655 | cation transmembrane transport | 5 | 0.030294381 |
| GO:0043603 | cellular amide metabolic process | 5 | 0.035024972 |
| GO:0009201 | ribonucleoside triphosphate biosynthetic process | 5 | 0.035154266 |
| GO:0050801 | ion homeostasis | 4 | 0.036517312 |
| GO:1902600 | hydrogen ion transmembrane transport | 4 | 0.03659533 |
| GO:0006812 | cation transport | 4 | 0.038418685 |
| GO:0009405 | pathogenesis | 4 | 0.041409595 |
| GO:0098662 | inorganic cation transmembrane transport | 3 | 0.043772275 |
| GO:0006518 | peptide metabolic process | 3 | 0.047749615 |
| GO:0006754 | ATP biosynthetic process | 3 | 0.049003942 |

Table 1.3: DAVID functional annotation results curated from a list of 41 downregulated genes. Only significantly enriched biological themes are shown.

| GO ID | GO Term | Count | PValue |
|---|---|---|---|
| GO:0044271 | cellular nitrogen compound biosynthetic process | 11 | 0.01205193 |
| GO:1901576 | organic substance biosynthetic process | 13 | 0.02307872 |
| GO:0044249 | cellular biosynthetic process | 13 | 0.02442621 |
| GO:0010467 | gene expression | 9 | 0.0339309 |
| GO:0009058 | biosynthetic process | 13 | 0.03596719 |

Figure 1.12: Growth curve of *M. mazei* growing on varying amounts of acetate to identify starvation conditions.



## 1.4 Discussion

*Pseudomonas aeruginosa* is an opportunistic pathogen that has been associated with lung infections in patients diagnosed with cystic fibrosis. It has been widely used as a model organism because of its pathogenicity. Growth on a minimal media with restricted amino acids has been shown to slow the growth rate (Palmer et al., 2010). The amino acids act not only as the sole carbon source, but also as the sole energy source. The growth rate (Figure 1.1) is showing that despite having a non-preferential energy source, PAO1 was still able to grow. Cultures that were grown on phenylalanine turned the media into a blue-green color after a few days of growth. The color change is caused by the presence of a secondary metabolite, pyocyanin. Pyocyanin is indicative of infection in patients with cystic fibrosis (Lau, Hassett, Ran, & Kong, 2004). It is has been observed in studies that induce oxidative stress in PAO1 in which pyocyanin acts as a protective mechanism (Orlandi, Bolognese, Chiodaroli, Tolker-Nielsen, & Barbieri, 2015). Furthermore, changes in methylation status in PAO1 are dependent on growth conditions and it can affect its pathogenicity (Doberenz et al., 2017). The appearance of pyocyanin and the slow growth rate indicate that the conditions to which PAO1 was subjected to were sufficient to elicit a stress response.

Having seen a stress response in the FS condition we would expect to see a higher number of modifications in that condition compared to the other conditions. Modifications from FS in Figure 1.3 do not appear as expected. They are at a drastically lower number than the other conditions that share very similar number of modifications. It needed to be determined if the data was real. Coverage can be a cause for one condition to have a smaller dataset. The coverage per base was visualized and observed to have had similar coverage across all conditions (Figure 1.9). Sequence

quality could also have an effect on the dataset. Summary statistics from the PacBio SMRT portal show no significant difference in quality among the samples (Table 1.1). DNA methylation has long thought to repress transcription. In some instances, methylation of genes can show a reduction of spurious transcription. Concomitantly, the loss of DNA methylation can cause spurious transcription which has been seen in eukaryotic cells (Neri et al., 2017; Wade & Grainger, 2018a). Spurious transcription can also have an impact on cell function by prematurely terminating transcription (Wade & Grainger, 2018). A cell would likely develop a mechanism to prevent spurious transcription from occurring and wasting resources. A stressed cell might not be able to keep its methylation in order to prevent spurious transcription. The result of spurious transcription can result in the slow growth of an organism as seen in PAO1 grown on phenylalanine. It has long been assumed methylation and modifications repress transcription, but that might not always be the case.

The modification profiles for the conditions did not show a change from GE and GS to FE as expected. Since the modifications profile from FS was so different, we would expect to see a similarity with FE. FE did not produce pyocyanin to indicate it was stressed it was still growing very slowly compared to GE and GS. The annotated files still had many similarities between GE, GS, and FE across the regions. It could be that the exponential phase of a starvation condition may not be stressful enough to show a response which would have been missed had we not looked at both phases of growth. Interestingly, even though FS had a small number of modifications in its dataset, it contained more modifications than any other condition on the rRNA region of the genome (Figure 1.6). The production of rRNA is the rate-limiting step in the creation of a ribosome in prokaryotes.

RNA synthesis is often the first response to the change in  a nutrient medium (Grummt & Ladurner, 2008). To meet the demand for protein synthesis in *E. coli* the number of ribosomes in the cell is proportional to the growth rate (Nomura, Gourse, & Baughman, 1984). The transcription rate of "active" rDNA is altered as a short-term regulation in response to stress in eukaryotic organisms, but for more stable rDNA transcription it is epigenetically regulated (Grummt & Ladurner, 2008). Active rRNA genes are often hypomethylated at the promoter region and hypermethylated at promoters of silent rRNA genes. The silencing of genes is seen in eukaryotes at CpG residues and regulated by recruitment of a DNA methyltransferase (Santoro & Grummt, 2001). The transcription rate of "active" rDNA is altered as a short-term regulation in response to stress in eukaryotic organisms, but for more stable rDNA transcription it is epigenetically regulated (Grummt & Ladurner, 2008). The growth rate of FS corresponds to a hypermethylated rRNA region. During this energy limited state, the modifications could be restoring energy balance by limiting RNA transcription. It could explain why there are more modifications in that region compared to the other conditions. Further analysis will need to be carried out to verify.

No determinable difference has been seen in comparing GE & GS to FE & FS based on their global modification profile. Even though we are not seeing any trends across the two condition it could be that there are fewer major changes overall and more smaller changes that would be missed. Observed modifications per kilobase (OMPK) scores give us the ability to observe which genes are being differentially modified. Figure 1.10 shows the similarity between GE and GS still remains true when observing the number of modifications on each gene. FE shares some similarities with GE and GS, but it still branches separately indicating that there are some genes of

interest to isolate. Despite the very few modifications observed in FS the data suggests the gain in modifications in the starvation condition and not retention. Not much information can be obtained from the FS dataset due to the low number of modifications it is difficult to see any similarities. GO terms that are enriched include proton transport, metabolic processes, and ATP synthesis. Functional profiling of differentially regulated genes as a function of the presence of a specific MTase revealed an enrichment of genes involved in iron metabolism (Doberenz et al., 2017). Genes that were expressed, but not associated with a motif were also involved in iron metabolism. None of the GO terms that were enriched were associated with iron metabolism, but they are involved in other metabolic processes. Most of the fold change came from FE genes being differentially more modified that GE. Modifications are occurring more frequently on genes involved with proton transport and metabolism in upregulated genes (Table 1.2). The downregulated genes are involved in biosynthetic processes (Table 1.3), which is typically the energy requiring part of metabolism. Fewer modifications are occurring on these genes which could indicate the organism conserving energy in the starvation state.

The next steps include analysis of the *M. mazei* samples to determine if the same trends occur in an archaeal system. A pipeline will be generated to allow a straightforward method of analyzing epigenetic data from long-read sequencing. qPCR of target genes or RNAseq to quantify gene expression of the differentially modified genes will determine if they are being transcriptionally regulated. There are four rRNA operons in PAO1 (Stover et al., 2000), so the rRNA increase seen in the FS condition could possibly be positional change of operons used. We will look into specific operons to view the patterns. Once it has been established that epigenetics

plays a role in energy regulation in an organism we can then apply the analysis to metagenomes from the deep sea. This will hopefully solve the mystery of their survival in the deep sea.

## REFERENCES


Aranda, M., Michell, C. T., Voolstra, C. R., Li, Y., Zahran, N., Cui, G., … Liew, Y. J. (2018). DNA methylation regulates transcriptional homeostasis of algal endosymbiosis in the coral model Aiptasia. *Science Advances*, *4*(8), eaat2142. https://doi.org/10.1126/sciadv.aat2142

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), 6506–6511. https://doi.org/10.1073/pnas.1711842115

Bradley, J. A., Amend, J. P., & LaRowe, D. E. (2018). Survival of the fewest: Microbial dormancy and maintenance in marine sediments through deep time. *Geobiology*, (July), 1–17. https://doi.org/10.1111/gbi.12313

D'Hondt, S., Inagaki, F., Zarikian, C. A., Abrams, L. J., Dubois, N., Engelhardt, T., … Ziebis, W. (2015). Presence of oxygen and aerobic communities from sea floor to basement in deep-sea sediments. *Nature Geoscience*, *8*(4), 299–304. https://doi.org/10.1038/ngeo2387

Deppenmeier, U., Müller, V., & Gottschalk, G. (1996). Pathways of energy conservation in methanogenic archaea. *Archives of Microbiology*, *165*(3), 149–163. https://doi.org/10.1007/BF01692856

Dixon, G., Liao, Y., Bay, L. K., & Matz, M. V. (2018). Role of gene body methylation in acclimatization and adaptation in a basal metazoan. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(52), 13342–13346. https://doi.org/10.1073/pnas.1813749115

Doberenz, S., Eckweiler, D., Reichert, O., Jensen, V., Bunk, B., Spröer, C., … Häussler, S. (2017). Identification of a Pseudomonas aeruginosa PAO1 DNA Methyltransferase, Its Targets, and Physiological Roles. *MBio*, *8*(1), e02312-16. https://doi.org/10.1128/mBio.02312-16

Finkel, S. E., & Kolter, R. (1999). Evolution of microbial diversity during prolonged starvation. *Proceedings of the National Academy of Sciences*, *96*(7), 4023–4027. https://doi.org/10.1073/pnas.96.7.4023

Grummt, I., & Ladurner, A. G. (2008, May 16). A Metabolic Throttle Regulates the Epigenetic State of rDNA. *Cell*. Elsevier B.V. https://doi.org/10.1016/j.cell.2008.04.026

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature*

*Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Lau, G. W., Hassett, D. J., Ran, H., & Kong, F. (2004). The role of pyocyanin in Pseudomonas aeruginosa infection. *Trends in Molecular Medicine*, *10*(12), 599–606. https://doi.org/10.1016/J.MOLMED.2004.10.002

Lever, M. A., Rogers, K. L., Lloyd, K. G., Overmann, J., Schink, B., Thauer, R. K., … Jørgensen, B. B. (2015). Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations. *FEMS Microbiology Reviews*, *39*(5), 688–728. https://doi.org/10.1093/femsre/fuv020

Mouammine, A., & Collier, J. (2018). The impact of DNA methylation in *Alphaproteobacteria*. *Molecular Microbiology*, 0–3. https://doi.org/10.1111/mmi.14079

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., … Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, *543*(7643), 72–77. https://doi.org/10.1038/nature21373

Nomura, M., Gourse, R., & Baughman, G. (1984). REGULATION OF THE SYNTHESIS OF RIBOSOMES AND RIBOSOMAL COMPONENTS, *53*, 75–117. Retrieved from www.annualreviews.org

Orlandi, V. T., Bolognese, F., Chiodaroli, L., Tolker-Nielsen, T., & Barbieri, P. (2015). Pigments influence the tolerance of pseudomonas aeruginosa PAO1 to photodynamically induced oxidative stress. *Microbiology (United Kingdom)*, *161*(12), 2298–2309. https://doi.org/10.1099/mic.0.000193

Palmer, G. C., Palmer, K. L., Jorth, P. A., & Whiteley, M. (2010). Characterization of the Pseudomonas aeruginosa Transcriptional Response to Phenylalanine and Tyrosine □, *192*(11), 2722–2728. https://doi.org/10.1128/JB.00112-10

Payne, S., McCarthy, S., Johnson, T., North, E., & Blum, P. (2018). Nonmutational mechanism of inheritance in the Archaeon Sulfolobus solfataricus. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(48), 12271–12276. https://doi.org/10.1073/pnas.1808221115

Santoro, R., & Grummt, I. (2001). Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. *Molecular Cell*, *8*(3), 719–725. https://doi.org/10.1016/S1097-2765(01)00317-3

Schut, F., De Vries,', E. J., Gotitschal, J. C., Robertson, B. R., Harder, W., Prins,' And, R. A., & Button2, D. K. (1993). *Isolation of Typical Marine Bacteria by Dilution Culture: Growth, Maintenance, and Characteristics of Isolates under*

*Laboratory Conditions*. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* (Vol. 59). Retrieved from http://aem.asm.org/

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., … Olson, M. V. (2000). Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature 2000 406:6799*, *406*(6799), 959–964. https://doi.org/10.1038/35023079

Wade, J. T., & Grainger, D. C. (2018a). Spurious transcription and its impact on cell function. *Transcription*, *9*(3), 182–189. https://doi.org/10.1080/21541264.2017.1381794

Wade, J. T., & Grainger, D. C. (2018b, May 27). Spurious transcription and its impact on cell function. *Transcription*. Taylor and Francis Inc. https://doi.org/10.1080/21541264.2017.1381794

Willyard, C. (2017, February 22). An epigenetics gold rush: New controls for gene expression. *Nature*. Nature Publishing Group. https://doi.org/10.1038/542406a

# Appendix A

# ANALYSIS SCRIPTS

## A.1 Modification detection and motif identification

```
#!/bin/bash
#SBATCH --job-name=index_ref
#SBATCH --ntasks=16
#SBATCH --mem=128000
SMRT_ROOT=/usr/local/smrtlink/
PATH=$SMRT_ROOT/smrtcmds/bin:$PATH
##pbmm2 replaces pbalign
pbmm2 index PAO1.fa PAO1.mmi
pbmm2 align PAO1.mmi FE.bam FE_aln.bam
ipdSummary FS_aln.bam --reference PAO1.fa --identify m6A,m4C --methylFraction
# --gff FS_mods.gff --csv FS_motifs.csv
MotifMaker find reprocess -g FS_mods.gff -f PAO1.fa -m FS_motifs.csv
 -o FS_mods_mots.gff
```

## A.2 Visualize base modifications in R

```
        # This script is for use with SMRTPortal 2.1 modifications.csv and
modifications.gff output files.
        # It reads gff3 compliant files.
        # Meredith Ashby, Khai Luong, Jonas Korlach 02/2014
        library(Hmisc)
        source("BaseModFunctions.v2.1.R")

        #### USER ENTERED VARIABLES HERE
        # NOTE: The modification detection should be done on the reference you wish
to use for circos, or the
        # coordinates will not match up.

        refPath <- 'PAO1.fa'
        gff <- 'GS_mods.gff'
        csv <- 'GS_mods.csv'
        outputPath <- 'BaseModFS300x.'

        #### READ IN THE GFF FILE
        hits_GS<- readModificationsGff(gff)
        attach(hits)

        #### IDENTIFY AND EXAMINE THE HIGH CONFIDENCE HITS
```

```
hits$score
###PAGE 1
# Plotting the base modification coverage vs scores on separate axie by base let's us consider applying different cutoffs in different channels.
pdf(paste(outputPath, '1.ModQV_vs_Coverage_scatter_plot_Phe_SP.pdf', sep=""), width=11, height=8.5, onefile=T)
p <- qplot(coverage, score, colour=CognateBase, size=I(0.5), data=subset(hits, CognateBase %in% c('A','T','G','C') & score > 30)) +
  facet_wrap(~CognateBase) +
  labs(title="Modification QV vs. Coverage Phe SP", x="Per Strand Coverage", y="Modification QV", colour="Cognate Base") +
  theme(legend.position =
        "top")
show(p)
dev.off()


###################### OPTIONAL ADDITIONAL FILTERING OF HITS FROM THE GFF FILE

###PAGE 2
# Iteratively adjust the slope for each base, determining which detections to include in motif finding depending on your level of coveraqe.
# Your selection of of the minimum QV can affect the % detected for each motif.  In SMRT Portal or with Motif Maker, you
# can selected hits using just the minumim QV; in R you can set a cutoff that takes into account the coverage dependence of the modificaiton QV
# score.

cutoffs <- data.frame(CognateBase=c('A','C','G','T'), slope=c(80/150, 50/150, 50/150, 50/150), intercept=c(0,0,0,0), x=c(60, 60, 60, 60))
cutoffs

# change the values for slope, intercept, and x for each base. X is a vertical line on the plot and
# specifies a minimum coverage, independent of score. 'Slope' and 'intercept' define an additional

# minimum score which depends on coverage. For a straight line cut independent of coverage, use
# slope = 0 and intercept = desired minimum score.

# This plot shows how the cutoff choices made above will filter hits.
```

```
pdf(paste(outputPath, '2.DataFiltering.pdf', sep=""), width=11, height=8.5,
onefile=T)
equations <- data.frame(CognateBase=c('A','C','G','T'),
equation=formatEquationText(cutoffs)) # print formatting, for the plot.
p <- qplot(data=subset(hits, CognateBase %in% c('A','T','G','C')), x=coverage,
y=score, colour=CognateBase, size=I(0.5)) +
  facet_wrap(~CognateBase, scales="free_y") +
  labs(title="Subsetting the Potential Hits to Limit False Positives", x="Per
Strand Coverage", y="Modification QV", colour="Cognate Base") +
  theme(legend.position = c(0.85,0.3)) +
  geom_abline(data=cutoffs, aes(slope=slope, intercept=intercept)) +
  geom_vline(data=cutoffs, aes(xintercept=x)) +
  geom_text(data=equations, aes(label=equation, x=30, y=200), vjust=0,
hjust=0)
show(p)
dev.off()
```

### A.3 Annotation of modifications

```
#!/bin/bash
#SBATCH --job-name=multibed_PAO1_TLS1_v2_feature_type_2
#SBATCH --ntasks=4
#SBATCH --mem=16000
# look at methylations in feature types
bedtools intersect -a PAO1-Common_FE_FS_GE_GS.gff -b
PAO1_annot_protein_coding.gff3 > PAO1-FE_FS_GE_GS-comm_protein_coding.gff
bedtools intersect -a PAO1-Common_FE_FS_GE_GS.gff -b
PAO1_annot_ncRNAs.gff3 > PAO1-FE_FS_GE_GS-comm_ncRNAs.gff
bedtools intersect -a PAO1-Common_FE_FS_GE_GS.gff -b
PAO1_annot_rRNAs.gff3 > PAO1-FE_FS_GE_GS-comm_rRNAs.gff
bedtools intersect -a PAO1-Common_FE_FS_GE_GS.gff -b
PAO1_annot_tRNAs.gff3 > PAO1-FE_FS_GE_GS-comm_tRNAs.gff
bedtools intersect -a PAO1-Common_FE_FS_GE_GS.gff -b
PAO1_annot_pseudogenes.gff3 > PAO1-FE_FS_GE_GS-comm_pseudogenes.gff
bedtools intersect -a PAO1-Common_FE_FS_GE_GS.gff -b
PAO1_annot_allGenes.gff3 -v > PAO1FE_FS_GE_GS-comm_intergenic.gff
```