IDENTIFICATION OF SMALL RNAS AND DIFFERENTIAL GENE EXPRESSION

IN RHODOBACTER SPHAEROIDES UNDER GOLD CHLORIDE STRESS

_____

A Thesis

Presented to

The Faculty of the Department of Biological Sciences

Sam Houston State University

_____

In Partial Fulfillment

of the Requirements for the Degree of

Master of Science

_____

by

Jovinna Laryssa Mendel

December, 2021

IDENTIFICATION OF SMALL RNAS AND DIFFERENTIAL GENE EXPRESSION

IN RHODOBACTER SPHAEROIDES UNDER GOLD CHLORIDE STRESS

by

Jovinna Laryssa Mendel

_____

APPROVED:

Madhusudan Choudhary, PhD
Committee Director

Anne Gaillard, PhD
Committee Member

Todd Primm, PhD
Committee Member

Hyuk Cho, PhD
Committee Member

John Pascarella, PhD
Dean, College of Science and Engineering
Technology

DEDICATION

I would like to dedicate this thesis to all the people who have believed in me, supported me, and prayed for me through this very challenging season. To my family, my husband, and my close friends: thank you for always encouraging me when things were difficult.

I would also, most importantly, like to dedicate this thesis to the Lord. Without my faith I would not be where I am today. To God be the glory forever.

ABSTRACT

Mendel, Jovinna Laryssa, *Identification of small RNAs and differential gene expression in Rhodobacter sphaeroides under gold chloride stress.* Master of Science (Biology), December, 2021, Sam Houston State University, Huntsville, Texas.

Small, regulatory RNAs (sRNA) play an important role in mediating transcriptional and translational processes within bacterial organisms. Understanding how these sRNAs play a role in heavy metal stress is of importance for bacteria involved in bioremediation. The following study aims to (i) identify novel sRNA sequences within *Rhodobacter sphaeroides* using RNAspace, a bioinformatic approach, (ii) validate a set of sRNAs expressed when the bacterium is grown under an aerobic and/or gold chloride stress condition, and (iii) analyze the gene expression profiles to identify specific target genes involved in the gold chloride stress condition. A total of 712 sRNAs were predicted within the genome of *R. sphaeroides* using the bioinformatic approach. *R. sphaeroides* growth characteristics were observed under different concentrations of gold chloride and were found to withstand up to a 1.0 μM concentration. Total RNA isolated from the untreated control group and the 1.0 μM $AuCl_3$ treated group were selected for small RNA and total RNA sequencing. A total of three differentially expressed sRNA sequences were detected in the 1.0μM $AuCl_3$ group, thus implying the role of these sRNAs in gold chloride stress. Additionally, targets were predicted for each sRNA utilizing the CopraRNA prediction program. A transcriptomic analysis was performed to identify differentially expressed genes between the control and 1.0 μM $AuCl_3$ groups at lag/early-log and late-log/stationary growth phases. A total of 121 genes representing a wide variety of gene functions exhibited up- or down- gene regulation at the lag/early-log phase, while 604 genes were up-/down-regulated at the late-log/stationary phase. A

majority of commonly differentially expressed genes were observed to be involved in membrane alteration, chemotactic response, energy production, and intracellular/extracellular transport across the membrane. Small RNAs that were detected by sRNA sequencing were predicted to additionally target differentially expressed genes observed within this comparison. A compiled list of identified sRNAs and their corresponding target genes were used to further elucidate the regulatory roles of these sRNAs under gold chloride stress.

KEY WORDS: sRNA; Gene regulation; Heavy metal tolerance; Gold chloride; Bioremediation; *Rhodobacter sphaeroides*

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Dr. Madhusudan Choudhary, for his constant support throughout my undergraduate and graduate academic career. Thank you for always having a passion to teach young scientists like myself about all that the scientific world has to offer. Thank you for all the time and effort that you have spent with me to complete this thesis, and for the encouragement you have given to me along the way.

I would like to thank my thesis committee: Dr. Hyuk Cho, Dr. Todd Primm, and Dr. Anne Gaillard, for your support and guidance in completing this project. Thank you for sharing your ideas with me and for pushing me closer to becoming a scientist. To Dr. Hyuk Cho, thank you for your willingness to help me with coding, even when you were very busy. I appreciate the guidance you have provided me and thank you for this newfound passion I now have for learning how to code.

I would like to thank my colleagues, Marco, Dennis, and Jessica who have helped me with the data collection for this project. I would also like to thank my officemates who were always willing to give input into some of the ideas or questions I had about this project. It is truly a gift to think alongside other scientists.

Lastly, I would like to thank my husband, my family, and my friends for all their support during this season of life. To my husband, I could never thank you enough for everything you have done for me. You have constantly shown me your support through your actions and your words, and I am forever grateful. To my family, thank you for always believing in me and for constantly showing your appreciation for my work. To my friends, both "the girls" and "the guys", thank you for not only supporting me, but for

helping me relax during this stressful season. Thank you for all the hangouts, fun

conversations, game nights, dinners, and so much more. I appreciate every single one of

you and I know that I can always lean on you for encouragement and fun. Thank you to

all those who have made this experience one to remember. Here's to all the future

endeavors we will have in this new, upcoming season.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Bioinformatic Prediction of Small RNAs in *Rhodobacter sphaeroides*

Gene regulation represents the process by which coding genes in any given organism are controlled under different conditions. Gene regulation is an important phenomenon that exists across all domains of life. It is responsible for maintaining interactions that occur within and between all living organisms in a range of environments. Gene regulation can occur at each of the different levels involved with gene expression, from pre-transcriptional processes to post-translational modifications. In bacteria, the processes of transcription and translation occur simultaneously due to the uncompartmentalized nature of a bacterial cell. Consequently, transcriptional and translational machineries exist in the same locations in the cell. DNA-dependent RNA polymerases are directed to promoter regions along the DNA strands by sigma (σ) factors. Once the polymerase is fully docked onto the strand, RNA synthesis begins taking place to generate a new, continuous strand of RNA that is complementary to the noncoding DNA template strand. This RNA strand, commonly known as messenger RNA (mRNA), gets anchored onto the ribosome and is translated into a polypeptide. Translation is initiated when the ribosomal subunits bind to an AG rich sequence, commonly referred to as the Shine-Dalgarno sequence, located around 10 nucleotides upstream from the start codon (Anders, 2004). Additionally, the ribosome will bind and interact with the translation start codon, AUG. This process can occur immediately after a portion of mRNA is synthesized and while transcription is still continued, thus increasing the efficiency of protein synthesis in the cell (Rathoure & Srivastava, 2016). However, since these processes of transcription and translation occur rapidly and concurrently, a

regulatory system utilizing a series of mechanisms involving noncoding RNA molecules and proteins is needed to produce an accurate yield of gene products. The regulation of gene expression in prokaryotic organisms involves a variety of factors such as RNA-binding proteins, small metabolites, and even various RNA molecules (Bervoets & Charlier, 2019). Such factors impact gene regulation at different levels, from pre-transcription to post-translation. Prokaryotes use these mechanisms to adapt and survive within their environment and respond rapidly to changes that occur within the cell.

Classes of Regulatory RNAs

There are different classes of regulatory RNAs found in prokaryotic organisms which are involved in regulating transcription and translation. These RNAs, such as riboswitches, protein-binding RNAs, and cis/trans-encoded base-pairing RNAs, are all vital components in the bacterial response to an internal or external stimulus (Waters & Storz, 2009). Each class consists of distinguishing characteristics, both in sequence composition and in interaction with other RNA molecules. Moreover, each class works to regulate in a specific way. The incorporation of all different types of regulatory RNAs provides an efficient means of surviving a stressful and changing environment.

Riboswitches

Riboswitches are one type of regulatory RNA molecule found within prokaryotic species. These RNAs are connected to the messenger RNA (mRNA) in which they regulate and are generally characterized as a secondary structure found in the 5' untranslated region of the mRNA (Waters & Storz, 2009). Riboswitches are further characterized by two distinct regions: an aptamer region which is responsible for interacting with an internal/external stimulus, and an expression platform which consists

of a sequence that undergoes a conformational change to induce a downstream effect (Roth & Breaker, 2009). Various conditions such as changes in temperature, interaction with stalled ribosomes, and presence of metabolites are responsible for inducing a conformational shift in the secondary structure of the riboswitch (Grundy & Henkin, 2006) . This phenomenon can lead to repression or activation of transcription or translation of the mRNA, as well as mRNA processing (Blouin et al., 2009). A total of 28 classes of riboswitches have been identified across bacterial species, with unique aptamer features observed for each class. Well-studied and highly common riboswitches include the TPP, cobalamin, FMN, glycine, and SAM I riboswitch. Riboswitches have been observed to be highly present in pathogenic strains, thus expanding the possibility of riboswitch regulation as pharmaceutical targets for pathogens (Pavlova et al., 2019).

Protein-binding RNAs

Protein-binding RNAs are responsible for altering the functionality of a protein target and play a major role in gene regulation at the post-transcriptional level. Types of protein binding RNAs include ribonuclease (RNase) P, 4.5S RNA, 6S RNA, and bifunctional transfer-messenger RNA (tmRNA) (Waters & Storz, 2009). RNase P is a structurally characterized ribozyme involved in the 5' end processing of different transfer RNA (tRNA) molecules within prokaryotic and eukaryotic organisms (Feltens et al., 2003). In bacteria, RNase P consists of a single, large catalytic RNA segment paired with a small protein cofactor responsible for tRNA alignment with the active site of the catalytic RNA (Reiter et al., 2010). While RNase P plays a major role in regular housekeeping of tRNA molecules, the ribonucleoprotein has also been shown to bind and overlap with a downstream gene in *Thermus thermophilus,* thus affecting the expression

of that gene (Feltens et al., 2003). 4.5S RNA represents another catalytic RNA which forms a complex with the Ffh protein in *Escherichia coli* to produce a signal recognition particle (Peluso et al., 2000). This ribonucleoprotein complex is responsible for regulating translation by interacting with an actively translated signal peptide emerging from a ribosome and translocating it towards the plasma membrane for secretion or insertion into the membrane (Nagai et al., 2003). 6S RNA plays an important role in regulating transcription by interacting with the σ^{70} unit of RNA polymerase (Trotochaud & Wassarman, 2005). 6S RNA was first identified in *Escherichia coli* (Hindley, 1967) and has since been discovered in a wide range of bacterial species (Cavanagh & Wassarman, 2014). Two small protein binding RNAs, CsrB and CsrC, work in tandem to interfere with the CsrA protein in *E. coli* (Liu et al., 1997). CsrA plays a major role in the microbial transition from exponential to stationary phase via the glycogen synthesis pathway, as well as playing a role in bacterial pathogenesis for the plant pathogen *Erwinia carotovora* (Romeo, 1998). Interaction with CsrB blocks the RNA binding site of CsrA, thus inhibiting the protein activity and its influence on regulating transcription of genes (Liu et al., 1997).

Clustered Regulatory Interspersed Small Palindromic Repeat RNAs (crRNAs)

The clustered regulatory interspersed small palindromic repeat (CRISPR) and Cas-9 associated enzyme system acts as a defense mechanism in bacteria against foreign DNA transferred via conjugation by other bacteria or via transduction mediated by bacteriophages (Thomas & Nielsen, 2005). The CRISPR/Cas9 system consists of a set of DNA palindromic repeats with spacers in between that contain sequences identical to foreign DNA elements. This set of repeats and spacers, known as a CRISPR cassette, is

flanked by an AT-rich leader sequence and *cas* genes. Upon injection of foreign DNA, the cassette is transcribed to produce an RNA molecule which gets cleaved into short RNA sequences (crRNAs). Additionally, the *cas* genes are transcribed and translated into Cas9 and associated proteins. The interaction of the crRNAs and proteins works to recognize and cleave foreign DNA/RNA that has entered the bacterium (Pougach et al., 2012).

Small RNAs (sRNAs)

Small, noncoding RNAs (sRNAs) are endogenous, base-pairing molecules in bacteria which range between 50 to 400 nucleotides in length. These sRNAs display a wide range of mechanisms in regulating expression of their target genes at both transcriptional and translational levels (Dutta & Srivastava, 2018). While sRNAs are classified in different ways, they can be easily divided into two main classes: cis-encoded sRNAs and trans-encoded sRNAs. Cis-encoded sRNAs are generally transcribed in regions that overlap with their corresponding target genes, such as the open reading frame (ORF) of a gene or between ORFs in a bacterial operon. These sRNAs display complete base-pairing over an extensive region of their target genes and function to down-regulate or terminate expression at both transcriptional and translational levels. Trans-encoded sRNAs are transcribed in locations which are separate from their corresponding targets. Based on this phenomenon, these sRNAs bind to their targets with limited complementarity and work to up-regulate or down-regulate translation (Waters & Storz, 2009). Most trans-encoded sRNAs act in a negative manner by repressing transcription or translation of their mRNA targets. A large sum of trans-encoded sRNAs require the use of Hfq, an RNA chaperone which facilitates RNA-RNA interactions (Jørgensen et al.,

2020). Both cis- and trans-encoded sRNAs are found to be highly expressed under stress-inducing conditions, such as pH, temperature, and nutrient deficiency stress (Hoe et al., 2013). Additionally, sRNAs are responsible for regulating genes involved in virulence mechanisms of pathogenic bacteria (Papenfort & Vogel, 2010). The diversity exhibited by sRNA regulation makes these molecules attractive candidates for studying prokaryotic regulatory networks.

Additionally, while sRNAs can be characterized by their method of action, they can also be characterized by their association with RNA-binding proteins. Hfq and ProQ are two of the most prominent RNA chaperones that have been shown to facilitate in sRNA-mRNA binding in prokaryotic organisms. These proteins help stabilize RNA-RNA interactions and provide accuracy to base pairing between the sRNA and mRNA target sequences, especially with sRNAs encoded in *trans*. While sRNAs can influence their respective targets without the presence of RNA chaperones, the proteins help to accelerate the regulatory function of the sRNA (Quendera et al., 2020). This mechanism of action has made it easier for scientists to detect novel sRNAs by performing co-immunoprecipitations with Hfq, as well as comparing transcriptomic and proteomic data correlations in Hfq mutants (Faner & Feig, 2013).

Small RNAs in Bacterial Model Organisms

Small RNA molecules have been a topic of interest within the field of RNA biology and have been extensively studied in various bacterial model organisms. The very first mRNA-sRNA interaction was discovered in the well-known model organism *Escherichia coli*. This sRNA, known as micF, was coincidentally observed through studying the genetic structure of *ompC* and *ompF* genes responsible for osmoregulation

in the organism. The micF sRNA was shown to hybridize with the *ompF* mRNA, a transcript responsible for encoding the outer membrane protein, resulting in a decrease of *ompF* expression under high osmolarity conditions (Mizuno et al., 1983). This novel discovery of a small regulatory molecule being responsible for influencing gene expression led to the expansion of sRNA discovery in a wide variety of prokaryotic organisms. Since then, scientists have expanded upon the list of sRNAs in *E. coli* and other microorganisms such as *Salmonella enterica* (Kröger et al., 2012)*, Staphylococcus aureus* (Guillet et al., 2013), and *Pseudomonas aeruginosa* (Dutta & Srivastava, 2018). In *E. coli*, over 100 sRNAs have been extensively studied and characterized, ranging from protein-binding regulatory sRNAs to specialized RNAs involved in toxin-antitoxin systems (Brantl & Jahn, 2015; Gottesman & Storz, 2011). In *S. enterica*, over 871 novel sRNAs have been identified, with some characterized to play roles in carbon starvation and host-pathogen interaction via outer membrane vesicles (Houserova et al., 2021; Malabirade et al., 2018) This new understanding of gene regulation has led to ideas of how sRNAs might be involved in pathogenicity and bacterial symbiosis.

New discovery of sRNAs in highly studied pathogenic bacteria has given new significance to these molecules in mediating virulence, specifically by regulating genes involved in colonization, tissue tropism, and overall bacterial fitness under stress (Caldelari et al., 2013). In *Streptococcus pneumoniae*, multiple sRNAs were observed to play a role in bacterial pathogenesis at different stages, such as nasopharynx colonization, lung infection, and bacterial sepsis. The putative identification of targets for these sRNAs indicates that each sRNA exhibits a pleiotropic effect in the cell, however further investigation is needed to determine if the interactions are direct or indirect (Mann et al.,

2012). In *Vibrio cholerae*, a genome-wide search detected 18 potential sRNAs transcribed from ToxT, a transcription factor responsible for regulating genes involved in virulence. Of the 18 sRNAs detected, two sRNAs known as TarA and TarB, were discovered on a pathogenicity island and were found to impact virulency by targeting genes involved in bacterial fitness during colonization of a host (Bradley et al., 2011; Richard et al., 2010). Additionally*, S. enterica* has been shown to contain an abundance of sRNAs encoded on pathogenicity islands that play a role in bacterial survival under stress-inducing conditions (Padalon-Brauch et al., 2008).

While sRNAs are highly studied in most pathogenic bacteria, the field is moving towards unveiling new sRNAs and their roles in non-pathogenic, environmental microbes. *Bacillus subtilis* is a highly studied, Gram-positive model microorganism found in soil environments. This microbe's ability to produce endospores, form biofilms, and efficiently secrete signaling molecules has made it a large contributor for industrial uses (Errington & Aart, 2020). Over 100 sRNAs have been detected in *B. subtilis*, and very few have been assigned a characterized function involving cellular heterogeneity, arginine catabolism, and iron homeostasis (Ul Haq et al., 2020). In *Streptomyces coelicolor*, another soil-dwelling Gram-positive bacterium, over 50 sRNAs were experimentally validated and only a few have been classified to regulate cell growth and metabolism (Heueis et al., 2014). Additionally, photosynthetic cyanobacteria such as *Synechocystis* sp. PCC6803 have been observed to encode hundreds of sRNA molecules, with some characterized to regulate genes involved in adaptation to nutrient availability and environmental stressors (Kopf & Hess, 2015).

Small RNAs in *Rhodobacter sphaeroides*

Rhodobacter sphaeroides is a Gram-negative, non-sulfur bacterium found within the Rhodobacteraceae family. There are many characteristics that make this bacterium an ideal organism for genetic analysis. For example, it represents a group of bacterial species which are comprised of a multipartite genome and can be found in a variety of environments and ecological niches. Its genome codes for a variety of gene functions that allow it to survive and thrive with expanded metabolic and genetic regulatory networks (Mackenzie et al., 2007).

Previous studies have identified differentially expressed sRNAs within *R. sphaeroides* (Table 1), yet only a few have been experimentally characterized. Of the small RNAs found within *R. sphaeroides*, most play a role in mediating photooxidative stress within the organism (Berghoff et al., 2009). *R. sphaeroides* has a multifunctional metabolic system that allows it to generate ATP through a series of mechanisms: aerobic respiration, anaerobic respiration, and anoxygenic photosynthesis in the presence of light (Zannoni et al., 2009). During anoxygenic photosynthesis, the slight increase in oxygen tension can result in toxic effects to the cell through the formation of singlet oxygen by the bacteriochlorophylls present in photosynthetic complexes (Glaeser et al., 2011). To combat the stress-induced situation, *R. sphaeroides* has developed a means of regulatory networks involving proteins and sRNAs to reduce the presence of singlet oxygen in the cell by balancing expression of photosynthetic genes.

Two sRNAs, known as PcrZ and PcrX, are important in regulating the formation of photosynthetic complexes in the bacterium. PcrZ (RSs2430) is a *trans*-acting, 136-nucleotide long, intergenic sRNA that is located between RSP_0819 and RSP_6134

(Mank et al., 2012). This sRNA has its own promoter that is induced by the response regulator PrrA, a component of the redox-responsive Prr system that is responsible for inducing transcription of photosynthetic genes under little to no oxygen tension (Eraso & Kaplan, 1994; Zeilstra-Ryalls et al., 1998). Major targets for PcrZ include pigment binding proteins and bacteriochlorophyll synthesis enzymes, along with genes involved in carotenoid synthesis and cyclic photosynthetic electron transport. PcrZ was shown to directly down-regulate expression of *bchN* and *puc2A*, both of which are involved in reactions regarding photosynthetic complexes in the cell (Reinbothe et al., 2010; Zeng et al., 2003). Additionally, the low oxygen tension conditions within the organism generates a 50-nucleotide byproduct from the 5' end of PcrZ. This small segment of PcrZ is not responsible for influencing the interaction with mRNA targets, and may be a result of RNA decay (Mank et al., 2012). Separate from PcrZ, PcrX is an sRNA that is derived from the 3' UTR of the *puf* operon in *R. sphaeroides* (Eisenhardt et al., 2018). The *puf* operon consists of five genes all involved in the formation of the reaction-center light-harvesting complex 1 (RC-LH1) (Donohue et al., 1988). When low oxygen tension is present, the response regulator PrrA is employed to activate transcription of the operon, and RNase E-mediated cleavage is used for maturation of the transcript (Eraso & Kaplan, 1994). During this maturation process, the 3' UTR is cleaved to produce PcrX which targets the *pufX* mRNA transcript, thus preventing further processing (Eisenhardt et al., 2018). PufX represents a scaffolding protein necessary for the assembly of the of RC-LH1 complexes, and the absence of this protein results in a decrease of photosynthetic complex organization (Francia et al., 2002). Additionally, the increase of PcrX leads to a reduction in the half-life of the *puf-BALMX* mRNA transcript, resulting in a decrease of

photosynthetic complex formation in the cell (Eisenhardt et al., 2018). Regulation of photosynthetic complex formation by PcrZ and PcrX helps to minimize the generation of singlet oxygen in the bacterium.

Along with regulating photosynthetic complexes, *R. sphaeroides* harbors an additional regulatory system of proteins and sRNAs to minimize singlet oxygen through interaction with cellular transporters and enzymes involved in carbon metabolism. These sRNAs, such as SorX (RSs2461), SorY (RSs1543), CcsR1-4 (RSs0680a-d), and Pos19 (RSs0019) were detected under singlet oxygen-mediated stress (Berghoff et al., 2009) and were later characterized by independent studies (Adnan et al., 2015; Billenkamp et al., 2015; Müller et al., 2016; Peng et al., 2016). The *trans*-acting ability of these sRNAs helps to mitigate downstream oxidative damage on the cell.

SorX is an sRNA found in the 3' UTR of RSP_0847, a gene encoding a OmpR-type transcriptional regulator (Peng et al., 2016). Much like previously mentioned sRNAs, SorX is co-transcribed under a RpoHI/HII promoter which is activated under singlet oxygen stress and heat shock (Nuss et al., 2009). Upon formation of the transcript, RNase E-mediated cleavage is used to generate a pre-SorX transcript of 116 nucleotides long, and then a further processing product of 75 nucleotides long that is cleaved from the 3' end of the pre-SorX sRNA. The 75-nt segment contains the functional regulatory component of the sRNA and is found to be highly conserved within the Rhodobacteraceae family (Peng et al., 2016). SorX impacts *R. sphaeroides'* resistance to singlet oxygen damage by increasing the bacterium's resistance to organic hydroperoxides, such as *tert*-butyl hydroperoxide (tBOOH), which are produced because of secondary damage from singlet oxygen stress (Vatansever et al., 2013). The increased

resistance is a result of the sRNA interaction with the transcript of *potA*, a gene encoding a polyamine transporter responsible for cellular uptake of spermidine (Igarashi & Kashiwagi, 1999). In the presence of spermidine, *R. sphaeroides* becomes increasingly sensitive to damage caused by organic hydroperoxides. The interaction of SorX and *potA* mRNA, which occurs at the Shine-Dalgarno sequence of the mRNA transcript and requires the Hfq chaperone for stable interaction, results in down-regulated expression of *potA* leading to a reduced presence of spermidine influx through the transporter. Additionally, the overexpression of SorX was observed to impact expression of two sRNAs also involved in singlet oxygen stress: CcsR1-4 and SorY. Although mechanisms of this sRNA-sRNA mediated influence are unknown, it is speculated that SorX plays a role in the transcription of the two sRNAs (Peng et al., 2016).

SorY is an sRNA which contains its own RpoHI/HII promoter and a transcriptional terminator sequence (Adnan et al., 2015). Since RpoHI/HII-dependent genes are known for being activated under a variety of stressors (de Lucena et al., 2010; Dufour et al., 2012; Martínez-Salazar et al., 2009), SorY has been observed to be induced by stressors such as singlet oxygen, hydrogen peroxide, heat, cadmium chloride ($CdCl_2$), and paraquat (Adnan et al., 2015). Upon singlet oxygen stress, SorY increases *R. sphaeroides'* resistance to singlet oxygen damage by targeting *takP* mRNA, a transcript encoding the extra-cytoplasmic soluble receptor subunit of a TRAP-T transporter (Gonin et al., 2007). This sRNA-mRNA interaction is highly dependent on the presence of the Hfq chaperone, and results in decreased stability and translation of the *takP* mRNA. The downregulation of *takP* reduces the transportation of malate into the cell, and thus reduces the flow of malate being incorporated into the tricarboxylic acid (TCA) cycle

(Adnan et al., 2015). The direct interaction between SorY and *takP* mRNA provides a way for *R. sphaeroides* to shift its metabolic system from the TCA cycle to the pentose phosphate pathway and the Entner-Doudoroff (ED) pathway, two major pathways involved in oxidative stress response (Chavarría et al., 2013; Rui et al., 2010).

While SorX and SorY regulate oxidative stress via cellular transporters, a set of four homologous sRNAs, known as CcsR1-4 (previously RSs0680a-d), are responsible for regulation of genes involved in C_1 metabolism and genes involved in the formation of the pyruvate dehydrogenase complex. These four sRNAs are co-transcribed with RSP_6037 through a RpoHI/HII promoter and are terminated via a Rho-independent terminator structure (Billenkamp et al., 2015). The four sRNA transcripts share a conserved CCUCCUCCC motif found within two hairpin loop structures and are therefore classified within the "cuckoo" RNA family (Reinkensmeier & Giegerich, 2015). Upon overexpression of CcsR1-4 in *R. sphaeroides*, genes involved in C_1 metabolism (*pqqA. xoxJ, xoxF, cycB, coxS, coxL*) and genes encoding subunits of the pyruvate dehydrogenase complex (*pdhD, pdhAb, pdhB)* were indirectly downregulated. Additionally, a direct interaction between CcsR1-4 and *flhR*, a gene encoding a transcriptional activator of glutathione (GSH)-dependent methanol/formaldehyde metabolism, was observed under photooxidative stress. This sRNA-mRNA interaction led to an increased resistance to oxidative stress in *R. sphaeroides* through an increased concentration of GSH within the cell. The higher GSH concentration paired with inactivation of genes regulated by FlhR resulted in re-allocation of GSH for repairing proteins with oxidative-stress induced damage. Additionally, the redundancy of sRNA-

mRNA interactions between each of the four sRNAs indicates an enhanced efficiency of sRNA regulation through this mechanism (Billenkamp et al., 2015).

Along with the previously mentioned sRNAs, Pos19 is induced under various stresses such as singlet oxygen, hydrogen peroxide, and iron limitation, and is preceded by a RpoE promoter. This sRNA is 219 nucleotides in length, contains a Rho-independent terminator, and contains a small open reading frame (sORF) within its sequence that generates a small peptide. While a functional role of the small peptide has not been determined, it was observed to have no interference on the regulatory effect of the sRNA. Pos19 was observed to have an indirect, negative impact on genes involved in serine and sulfur metabolism. Additionally, the sRNA was shown to directly regulate RSP_0557 (a hypothetical protein) and *cysH* (a thioredoxin involved in sulfur metabolism). Both interactions were shown to be reliant on the presence of Hfq. The regulatory function of Pos19 is speculated to be involved in glutathione (GSH) biosynthesis by its impact on genes involved in sulfur metabolism (Müller et al., 2016). Since sulfur metabolism results in an increase of GSH levels in the cell to prevent oxidative damage (Li et al., 2004), the reduction of GSH by overexpression of Pos19 indicates that the sRNA functions to regulate sulfur metabolism occurring in the cell by preventing overabundance of products generated from the process. This mechanism ultimately leads to a decrease in the presence of reactive oxygen species within the bacterium (Müller et al., 2016).

Lastly, UpsM is an sRNA previously identified under photooxidative stress conditions and is observed to play a possible role in regulating cellular growth and division (Weber et al., 2016). This sRNA was observed to be located within an extended

5' UTR region upstream the *mraZ* gene, the first gene located in the division and cell wall (*dcw*) gene cluster in *R. sphaeroides*. This cluster consists of genes responsible for regulating cellular growth and are highly conserved amongst rod-shaped, gram-negative bacteria (Mingorance et al., 2004). The extended 5' region of this gene cluster was observed to be unique to the Rhodobacteraceae family, indicating a diversified role of *dcw* gene cluster transcription in these bacteria. UpsM was found to contain a Rho-independent terminator and was shown to require RNase E, Hfq, and RpoHI/HII-dependent target mRNAs to undergo further processing into a 130-nucleotide sequence. UpsM was also predicted to have riboswitch capability due to the aptamer/terminator configuration seen in the secondary structure of the sRNA, however a functional role of UpsM as a riboswitch has not been determined. Additionally, a transcriptomic analysis when UpsM was overexpressed in *R. sphaeroides* showed a negative effect on genes involved in cellular growth, however no further classification of interaction has been made between UpsM and its seemingly corresponding targets (Weber et al., 2016).

While a handful of sRNAs were observed and characterized to be involved in regulating processes under (photo-)oxidative stress response, novel sRNAs and their mechanisms in gene regulation have yet to be determined under alternative stress-inducing systems. Promoter systems such as RpoE and RpoHI/HII show promising results for sRNA activation under various stresses. Therefore, it is interesting to know whether these sRNAs, along with novel sRNAs, function to regulate targets under different mechanisms of stress.

Table 1

Previously Discovered Noncoding RNA in Rhodobacter sphaeroides

Classification[a]	Name[b]	Size[c]	Coordinates[d]		Strand[e]	Location[f]	Database[g]	Articles[h]	Validation Method[i]
			Start	Stop					
cis-reg	Rhodo-rpoB	93	293504	293597	+	1	Rfam	n/a	Similarity (80.4 bit score)
cis-reg; riboswitch	SAM-alpha riboswitch	76	444939	445015	+	1	Rfam, RNAcentral	n/a	Similarity (62.8 bit score)
ncRNA	LR-PK1	251	536887	536637	-	1	BSRD, Rfam, RNAcentral	n/a	Similarity (75.77 bit score)
Gene; sRNA	CcsR1 (RSs0680a)	82	692456	692374	-	1	Rfam	(Billenkamp et al., 2015)	454-sequencing, Northern blot (PMID:19906181)
Gene; sRNA	CcsR2 (RSs0680b)	81	692350	692270	-	1	Rfam	(Billenkamp et al., 2015)	454-sequencing, Northern blot (PMID:19906181)
Gene; sRNA	CcsR3 (RSs0680c)	77	692235	692159	-	1	Rfam	(Billenkamp et al., 2015)	454-sequencing, Northern blot (PMID:19906181)
Gene; sRNA	CcsR4 (RSs0680d)	77	691778	691702	-	1	Rfam	(Billenkamp et al., 2015)	454-sequencing, Northern blot (PMID:19906181)

Classification[a]	Name[b]	Size[c]	Coordinates[d]		Strand[e]	Location[f]	Database[g]	Articles[h]	Validation Method[i]
			Start	Stop					
ncRNA	RSs0682 (UpsM)	206	694636	694841	+	1	BSRD	(Berghoff et al., 2009; Weber et al., 2016)	454-sequencing, Northern blot (PMID:19906181)
Gene	alpha-tmRNA	355	741215	740861	-	1	BSRD	n/a	Similarity (296.13 bit score)
cis-reg; riboswitch	Glycine riboswitch	88	804587	804674	+	1	BSRD, Rfam, RNAcentral	n/a	Similarity (48.94 bit score)
cis-reg; riboswitch	Glycine riboswitch	77	804675	804751	+	1	BSRD	n/a	Similarity (35.39 bit score)
Gene; ribozyme	RNaseP_bact_a	398	890713	891110	+	1	BSRD, Rfam, RNAcentral	n/a	Similarity (254.24 bit score)
ncRNA	RSs0940	68	976062	976129	+	1	BSRD	(Berghoff et al., 2009)	454-sequencing, Northern blot (PMID:19906181)
cis-reg; riboswitch	Cobalamin riboswitch	201	1030432	1030632	+	1	BSRD, Rfam, RNAcentral	n/a	Similarity (129.12 bit score)
ncRNA	RSas2750	124	1400946	1401069	+	1	BSRD	(Berghoff et al., 2009)	454-sequencing, Northern blot (PMID:19906181)

(continued)

| Classification[a] | Name[b] | Size[c] | Coordinates[d] | | Strand[e] | Location[f] | Database[g] | Articles[h] | Validation Method[i] |
			Start	Stop					
ncRNA	RSs1368	71	1426531	1426601	+	1	BSRD	(Berghoff et al., 2009)	454-sequencing, Northern blot (PMID:19906181)
ncRNA	RSs1386	74	1443998	1444071	+	1	BSRD	(Berghoff et al., 2009)	454-sequencing, Northern blot (PMID:19906181)
cis-reg; riboswitch	Cobalamin riboswitch	195	1481651	1481845	+	1	BSRD, Rfam, RNAcentral	n/a	Similarity (86.9 bit score)
Gene; sRNA	SorY	82	1632645	1632727	+	1	Rfam, RNAcentral	(Adnan et al., 2015)	Microarray, Northern blot (PMID: 25833751)
ncRNA	RSs1624	108	1724393	1724500	+	1	BSRD	(Berghoff et al., 2009)	454-sequencing, Northern blot (PMID:19906181)
ncRNA	RSs1740	74	1849171	1849244	+	1	BSRD	(Berghoff et al., 2009)	454-sequencing, Northern blot (PMID:19906181)
ncRNA	PcrX	108	1980346	1980239	-	1	BSRD	(Eisenhardt et al., 2018)	RNA-seq, Northern blot (PMID: 29995316)
Gene; sRNA	5_ureB	288	2036299	2036587	+	1	Rfam, RNAcentral	n/a	Similarity (120.9 bit score)

Classification[a]	Name[b]	Size[c]	Coordinates[d]		Strand[e]	Location[f]	Database[g]	Articles[h]	Validation Method[i]
			Start	Stop					
cis-reg	COG3680	55	2053168	2053222	+	1	Rfam	n/a	Similarity (70.5 bit score)
cis-reg	terC	60	2191904	2191963	+	1	Rfam	n/a	Similarity (50 bit score)
cis-reg; riboswitch	TPP riboswitch	98	2388603	2388506	-	1	BSRD, Rfam, RNAcentral	n/a	Similarity (54.33 bit score)
gene	bacteria small SRP	99	2455877	2455975	+	1	Rfam, RNAcentral	n/a	Similarity (58.1 bit score)
ncRNA	RSs2363	163	2495719	2495881	+	1	BSRD	(Berghoff et al., 2009)	454-sequencing, Northern blot (PMID:19906181)
ncRNA	6S RNA (SsrS)	156	2495722	2495877	+	1	BSRD, RNAcentral	(Elkina et al., 2017)	RNA-seq, Northern blot (PMID: 28692405)
cis-reg; riboswitch	Cobalamin riboswitch	208	2531207	2531414	+	1	BSRD, Rfam, RNAcentral	n/a	Similarity (107.7 bit score)
ncRNA	PcrZ (RSs2430)	136	2565819	2565954	+	1	BSRD	(Berghoff et al., 2009; Mank et al., 2012)	454-sequencing, Northern blot (PMID:19906181)

(continued)

Classification[a]	Name[b]	Size[c]	Coordinates[d]		Strand[e]	Location[f]	Database[g]	Articles[h]	Validation Method[i]
			Start	Stop					
Gene; sRNA	SorX (RSs2461)	74	2598225	2598299	+	1	BSRD, Rfam	(Berghoff et al., 2009; Peng et al., 2016)	454-sequencing, Northern blot (PMID:19906181)
Gene; sRNA	Ffh	51	2811690	2811741	+	1	Rfam, RNAcentral	n/a	Similarity (63.4 bit score)
ncRNA	RSs2778	70	2924382	2924451	+	1	BSRD	(Berghoff et al., 2009)	454-sequencing, Northern blot (PMID:19906181)
ncRNA	RSas1198	299	2969111	2969409	+	1	BSRD	(Berghoff et al., 2009)	454-sequencing, Northern blot (PMID:19906181)
cis-reg; leader	SerC	52	3120985	3121036	+	1	BSRD, Rfam	n/a	Similarity (69.8 bit score)
ncRNA	RSs2978	187	3144046	3144232	+	1	BSRD	(Berghoff et al., 2009)	454-sequencing, Northern blot (PMID:19906181)
cis-reg; riboswitch	TPP riboswitch	100	3168810	3168909	+	1	BSRD, Rfam, RNAcentral	n/a	Similarity (67.5 bit score)

Classification[a]	Name[b]	Size[c]	Coordinates[d]		Strand[e]	Location[f]	Database[g]	Articles[h]	Validation Method[i]
			Start	Stop					
ncRNA	Pos19 (RSs0019)	219	28936	29153	+	2	BSRD	(Berghoff et al., 2009; Müller et al., 2016)	454-sequencing, Northern blot (PMID:19906181)
cis-reg; riboswitch	sul1 riboswitch	55	101101	101155	+	2	Rfam	n/a	Similarity (74.2 bit score)
cis-reg; riboswitch	Cobalamin riboswitch	214	239721	239508	-	2	BSRD, Rfam, RNAcentral	n/a	Similarity (109.1 bit score)
cis-reg; riboswitch	Cobalamin riboswitch	180	274276	274455	+	2	BSRD, Rfam, RNAcentral	n/a	Similarity (92.21 bit score)
ncRNA	RSs0245	124	417288	417411	+	2	BSRD	(Berghoff et al., 2009)	454-sequencing, Northern blot (PMID:19906181)
cis-reg; riboswitch	Cobalamin riboswitch	201	458214	458014	-	2	BSRD, Rfam, RNAcentral	n/a	Similarity (100.61 bit score)
ncRNA	RSs0252	105	519442	519546	+	2	BSRD	(Berghoff et al., 2009)	454-sequencing, Northern blot (PMID:19906181)

Classification[a]	Name[b]	Size[c]	Coordinates[d]		Strand[e]	Location[f]	Database[g]	Articles[h]	Validation Method[i]
			Start	Stop					
cis-reg; riboswitch	SAM-SAH	50	688628	688579	-	2	BSRD, Rfam, RNAcentral	n/a	Similarity (57.46 bit score)
cis-reg; riboswitch	sul1 riboswitch	58	47536	47593	+	pD	Rfam	n/a	Similarity (71.8 bit score)
									(continued)
cis-reg; riboswitch	sul1 riboswitch	57	54799	54855	+	pD	Rfam	n/a	Similarity (60.4 bit score)
cis-reg; riboswitch	sul1 riboswitch	58	1944	2001	+	pE	Rfam	n/a	Similarity (73.5 bit score)
									(continued)

Note. All sRNAs in this table are found in *Rhodobacter sphaeroides* 2.4.1. [a]Classification is based on common small noncoding RNA categories. Noncoding RNA is a generalized category for sRNAs not yet identified in function. [b]The name of each sRNA as published in their respective articles. [c]Represents the chromosome which contains the sRNA sequence. [d]The strand where the sRNA is located; (+) indicates the forward strand while (-) indicates the reverse strand. [e]The location of each sRNA in the Rhodobacter sphaeroides 2.4.1 genome; coordinates for an sRNA located on the reverse (-) strand are written in reverse order. [f]The approximate length of each sRNA found in the published database or article. [g]BSRD: bacterial small RNA repository database, Rfam: RNA family database version 13.0, RNAcentral: noncoding RNA sequence database. [h]The following articles have experimentally validated a sRNA sequence by means of RNA sequencing, Northern blotting, or RT-PCR. [i]Methods of experimental validation or identification by homology are stated, along with the following PubMed identification number (PMID).

Heavy Metal Contamination

Rhodobacter sphaeroides is known to possess tolerance to heavy metal environments, and has been discovered to utilize such metals in its own metabolic processes (Johnson et al., 2017). The identification of sRNA under this stress condition can provide new insight to understand naturally occurring, regulatory sRNA mechanisms in purple, non-sulfur bacteria.

Heavy metal contamination is a major environmental issue currently impacting the present human population and the environment. Heavy metals consist of a density larger than 5 g/cm^3 and are commonly found in molecular compounds throughout the environment. However, when heavy metals become solubilized and highly concentrated, they can become toxic to cells. Common examples of heavy metals include zinc (Zn), lead (Pb), gold (Au), cadmium (Cd), and arsenic (As) (Nies, 1999). Heavy metal contamination can lead to many serious health problems, such as bone loss, neurological damage, and different types of cancers (Järup, 2003). The severity of heavy metal contamination has led to various means of bioremediation tactics to minimize toxic concentrations of heavy metals in the environment. One method of bioremediation is the use of bacteria in contaminated freshwater ecosystems. Different bacterial species have been shown to tolerate heavy metal environments through various resistance mechanisms. For example, efflux pumps are commonly used to remove toxic concentrations of heavy metals that enter the cell via transport channel proteins. Additionally, bacterial species can use enzymes to detoxify the heavy metals and make them more inert (Silver & Phung, 1996). These resistance mechanisms allow bacteria to be promising candidates for heavy metal bioremediation.

Rhodobacter sphaeroides poses as the ideal model organism for the following study of analyzing sRNAs involved in heavy metal tolerance, especially due to its ability to tolerate a heavy metal contaminated system. Previous experimentation has shown that *R. sphaeroides* can interact with gold particles in a toxic concentrated environment. The bacterium has been shown to uptake gold chloride and localize gold particles in its plasma membrane (Johnson et al., 2017). Additionally, *R. sphaeroides* has been shown to withstand high concentrations of gold chloride, as well as metabolize the gold chloride $(AuCl_3)$ particles into elemental gold (Au^0) by extracellular formation of bio-nanoparticles (Italiano et al., 2018). However, the regulatory mechanisms of gene expression that are involved with *R. sphaeroides'* metabolic capability are not understood. The following proposed study will aim to understand how sRNA regulation is involved with the above concept, and whether it can be manipulated to further advance the bioremediation quality of *R. sphaeroides* in gold-contaminated environments.

Objective

The purpose of this research is to identify novel sRNAs involved in the process of gold metal tolerance in *R. sphaeroides*, as well as identify alterations in gene expression when the bacterium is exposed to this heavy metal stress. The presence of sRNAs paired with gene expression data can be used to envision the role of sRNAs involved in heavy metal stress. By understanding how sRNA regulates under the stress-inducing environment, further research can be conducted to manipulate the regulatory system of sRNA for enhancement of *R. sphaeroides* as a bioremediating agent. A diagram can be observed in Figure 1 describing the overall workflow developed for the following study. It is first hypothesized that a large number of small regulatory RNA (sRNA) exists within

the *R. sphaeroides* genome. This hypothesis will be tested by using RNAspace, a

bioinformatic tool used to predict noncoding sequences within a given genome.

Figure 1

Diagram of Overall Workflow for Identification of sRNAs and Differential Gene Expression in this Study
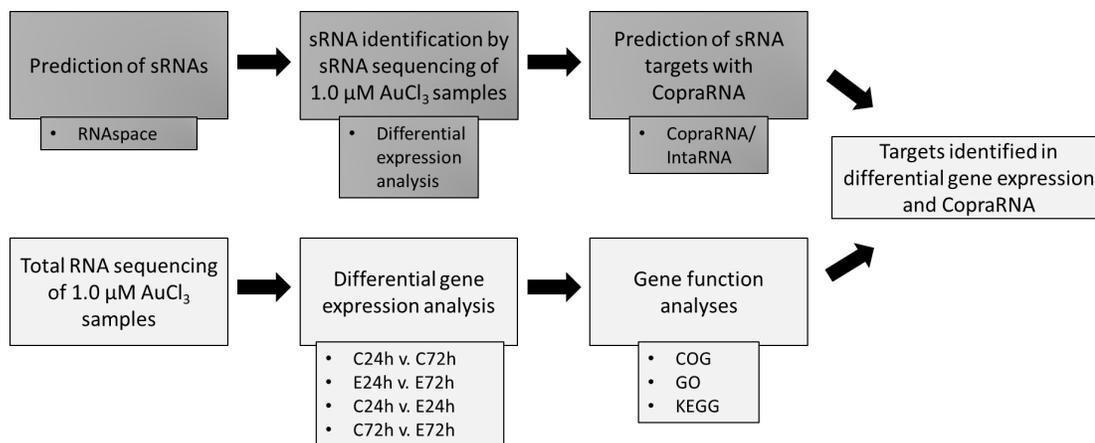
Methods

Identification of sRNAs in Rhodobacter sphaeroides using RNAspace

The growing interest in understanding mechanisms by which novel sRNAs

regulate their targets has created a demand for robust approaches to observe and predict

sRNA-mRNA interactions. Recent innovation of bioinformatics programs directed

towards predicting such interactions has allowed scientists to 1) develop streamlined

experimental tactics for validation of novel sRNA-mRNA interactions, 2) identify the

functionality of a novel sRNA in the genome, and 3) utilize a fast and cost-effective

technique for identifying sRNAs and their targets. To identify the presence of regulatory

sRNAs in *Rhodobacter sphaeroides* 2.4.1, an integrative and web-accessible platform

known as RNAspace.org is utilized to compute sRNA predictions within the genome. A

diagram of the workflow using RNAspace can be seen in Figure 2.

Figure 2

Diagram of sRNA Prediction Workflow using RNAspace Interface
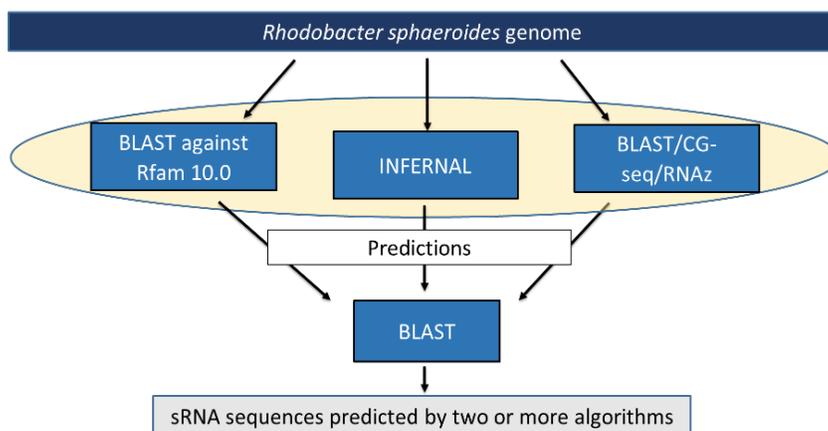
RNAspace for sRNA Prediction in Escherichia coli K12 substrate MG1655 Genome

An openly available and integrative platform known as RNAspace (Cros et al., 2011) was used to identify potential sRNA sequences in the *Escherichia coli* strain K12 substrate MG1655 genome. The *E. coli* K12 substr. MG1655 genome served as the control for sRNA prediction using RNAspace. This strain of bacteria is a well-studied organism in the field of regulatory RNA, with over 100 regulatory RNA sequences collected in literature and in RNA databases (Gottesman & Storz, 2011; Rau et al., 2015). Therefore, *E. coli* K12 substr. MG1655 was used to determine the effectiveness of RNAspace at detecting regulatory sRNAs. The FASTA file (RefSeqID: NC_000913.3) was downloaded from the National Center for Biotechnology Information (NCBI) database to be used in the RNAspace interface. Within the interface, a homology search was performed by selecting the BLAST (Altschul et al., 1990) and INFERNAL (Nawrocki et al., 2009) programs to compare the genome against the Rfam 10.0 database (Gardner et al., 2011). BLAST is used to detect regions of local similarity between the

query sequence and a set of representative sequences for an Rfam family. INFERNAL uses covariance models to identify homologous sequences between the query and RNA families identified in Rfam on a sequence and secondary structure level (Cros et al., 2011). Both gene finders were run under default parameters provided by the RNAspace webserver. In addition to the homology search, a comparative analysis was also performed within RNAspace using BLAST for sequence alignment, CG-seq for sequence aggregation, and RNAz for structural inference. The annotated genomes of *Escherichia coli* O157H7, *Escherichia coli* ATCC 8739, *Escherichia coli* K12 substr. MG1655, and *Paracoccus denitrificans* PD1222 were selected for use in a BLAST homology search against the query to identify conserved regions across closely related species. Only the intergenic regions of the genome were processed during the analysis. Conserved regions identified in the pairwise alignment were clustered together through CG-seq, and clustered regions were examined by RNAz to detect a highly conserved and thermodynamically stable secondary structure. All algorithms were used with their respective default parameters provided in the RNAspace webserver. The option to combine results was not chosen to allow individual collection of sRNA predictions between the three different gene finders (BLAST, INFERNAL, and comparative analysis with BLAST/CG-seq/RNAz). For the BLAST gene finder, scores were allocated to each prediction to depict the lowest E-value generated by pairwise alignments of the query against an Rfam RNA family. The default E-value threshold determined by RNAspace was set to 0.001, meaning that 0.001% matches were expected to happen by chance (Altschul et al., 1990). Sequences with an E-value greater than the 0.001 threshold were removed. Additionally, predictions generated by the homology search that were classified

into an RNA family not representative of regulatory RNA (ex: tRNA, rRNA, etc.) were removed. Sequences predicted by INFERNAL were also assigned with an E-value. The default E-value inclusion threshold for the multiple sequence alignment performed within INFERNAL was set to 0.01, and the reporting threshold set to 10.0 (Nawrocki et al., 2009). All final sequences that exhibited an E-value greater than 0.01 were removed. Within the comparative analysis gene finder, a BLAST similarity search was performed with the query genome and the genomes of four chosen species. The CG-seq algorithm was used to gather matching sequences into clusters of conserved sections, with focus on only intergenic regions of the genome (Grenier-Boley et al., 2010). Lastly, each cluster was subjected to the RNAz algorithm for secondary structure analysis. A consensus secondary structure was generated based on the conservation of the sequence amongst the different species. The thermodynamic stability of the consensus structure was determined, and a probability value was assigned to a given prediction to indicate the likelihood of the sequence being a functional, noncoding RNA. Predictions with a probability value higher than the default cut-off value of 0.7 were kept and recorded in the RNAspace webserver results (Cros et al., 2011).

The results generated by each of the three algorithms were compared to one another using the alignment of two sequences function in nucleotide BLAST. Additionally, all RNAspace predictions were compared to a list of *E. coli* K12 sRNA sequences previously identified in major regulatory RNA databases (Appendix A). Sequences that matched with the RNAspace predictions were isolated and observed for each algorithm.

RNAspace for sRNA Prediction in Rhodobacter sphaeroides Genome

RNAspace was used to predict sRNA sequences in the *Rhodobacter sphaeroides* 2.4.1 genome. The *R. sphaeroides* 2.4.1 FASTA files (RefSeqID: NC_007493.2, NC_007494.2, NC_009007.1, NC_007488.2, NC_007489.1, NC_007490.2, NC_009008.1) were downloaded from the National Center for Biotechnology Information (NCBI) database to be used in the RNAspace interface. A homology search was performed using the BLAST and INFERNAL gene finders to compare against the Rfam 10.0 database. Both algorithms were run under default parameters. The comparative analysis was performed using BLAST for sequence alignment, CG-seq for sequence aggregation, and RNAz for structural inference. The annotated genomes of *Rhodobacter sphaeroides* ATCC 17025, *Rhodobacter sphaeroides* ATCC 17029, *Rhodobacter sphaeroides* KD131, and *Paracoccus denitrificans* PD1222 were selected for use in a BLAST homology search. Only intergenic regions of the genome were processed during the analysis. Conserved regions identified in the pairwise alignment were clustered together through CG-seq, and clustered regions were examined by RNAz to detect a highly conserved and thermodynamically stable secondary structure. All algorithms were used with their respective default parameters provided in the RNAspace webserver. The option to combine results was not chosen to allow individual collection of sRNA predictions between the three different processes (BLAST, INFERNAL, and comparative analysis with BLAST/CG-seq/RNAz). Predictions generated by the homology search that were classified into an RNA family not representative of regulatory RNA (e.g.: tRNA, rRNA, etc.) were removed. Additionally, sequences identified by BLAST and INFERNAL with an E-value greater than the threshold of 0.001 and 0.01

were removed. The results generated by each of the three algorithms were compared to one another using the alignment of two sequences function in nucleotide BLAST. Additionally, all RNAspace predictions were compared to a list of *R. sphaeroides* sRNA sequences previously identified in major regulatory RNA databases (Table 1). Sequences that matched with the RNAspace predictions were isolated and observed for each algorithm.

Results and Discussion

sRNA Predictions in E. coli K12 substr. MG1655 Genome

A total of 1,893 sRNA predictions were made by the BLAST, INFERNAL, and comparative analysis programs within RNAspace. The following distribution of predictions can be observed in Figure 3. BLAST predicted a total of 51 unique sequences, INFERNAL predicted a total of 123 unique sequences, and the comparative analysis using BLAST, CG-seq, and RNAz predicted a total of 1,414 unique sequences. A total of 137 predictions were made by two or more algorithms, with 31 sequences being predicted by all three. The overlap of predictions made by the algorithms signifies a higher likelihood of the prediction being observed *in vivo*, as seen previously in a study that utilized RNAspace to identify novel sRNAs in *Actinobacillus pleuropneumoniae* (Rossi et al., 2016).

Figure 3

Venn Diagram of All sRNA Predictions for E. coli K12 substr. MG1655 Genome
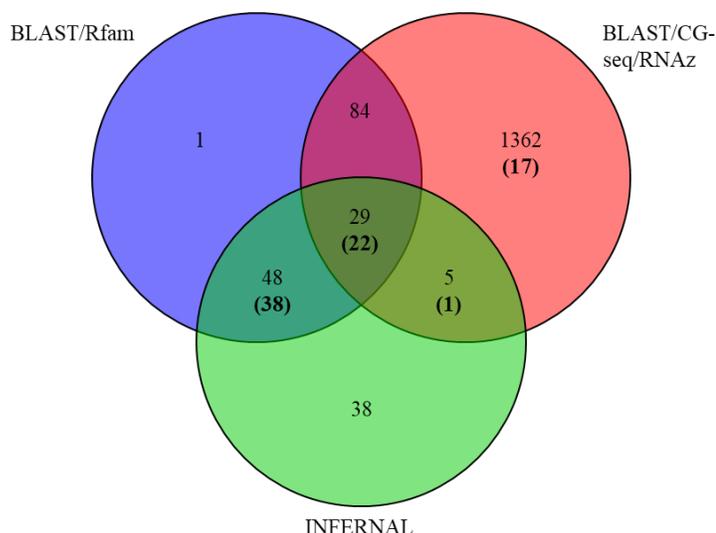
Note. The following image depicts the total sRNA predictions generated by the three algorithms used in the RNAspace webserver. BLAST represents the homology search of BLAST against the Rfam 10.0 database. INFERNAL represents the RNA structural motif homology search against the Rfam 10.0 database. RNAz represents the comparative analysis consisting of BLAST, CG-seq, and RNAz. Numbers represented in parentheses indicate total number of previously identified sRNAs detected by RNAspace.

To evaluate the predictive capabilities of RNAspace, we compared the total sRNAs predicted in *E. coli* to a list of 108 previously discovered sRNAs. This list was generated through collection of sRNAs from published articles and various noncoding RNA databases such as Rfam (Griffiths-Jones et al., 2003), RegulonDB (Santos-Zavaleta et al., 2019), and Ecocyc (Keseler et al., 2017). The following list can be observed in Appendix A. Out of the 108 previously identified sequences, RNAspace was able to capture 80 sequences, with 61 of those sequences being detected by two or more programs. Therefore, RNAspace was able to detect previously identified sRNAs with 74% success. With this knowledge, RNAspace was used for sRNA prediction within the *Rhodobacter sphaeroides* genome.

sRNA Predictions in Rhodobacter sphaeroides Genome

A total of 712 sRNA predictions were made by the BLAST, INFERNAL, and comparative analysis gene finders in RNAspace. The distribution of predictions across the three groups can be observed in Figure 4. BLAST predicted a total of 16 unique sequences, INFERNAL predicted a total of 3 unique sequences, and the comparative analysis performed by BLAST/CG-seq/RNAz predicted a total of 646 unique sequences. A total of 26 predictions were made by two or more algorithms, with 1 sequence being predicted by all three.

Figure 4

Venn Diagram of All sRNA Predictions for R. sphaeroides Genome
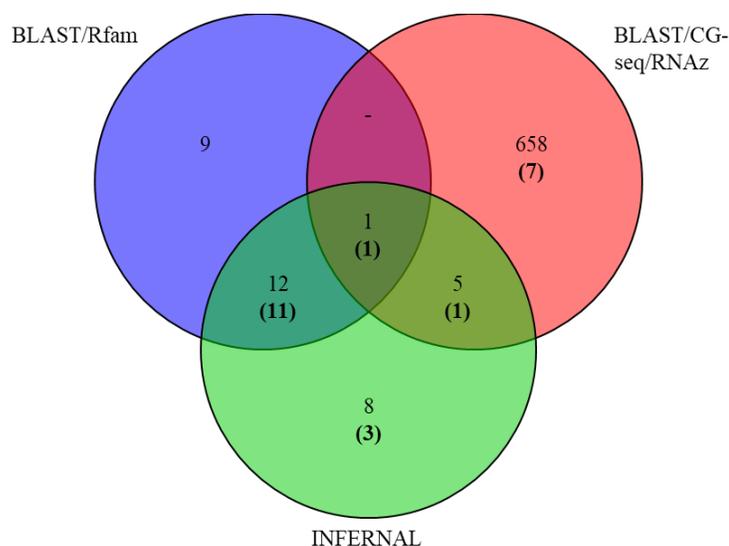
Note. The following image depicts the total sRNA predictions generated by the three algorithms used in the RNAspace webserver. BLAST is depicted in blue and represents the homology search of BLAST against the Rfam 10.0 database. INFERNAL is depicted in green and represents the RNA structural motif homology search against the Rfam 10.0 database. RNAz is depicted in red represents the comparative analysis consisting of BLAST, CG-seq, and RNAz. Numbers represented in parentheses indicate total number of previously identified sRNAs detected by RNAspace.

Appendix B provides a breakdown of all predicted sequences in *R. sphaeroides*. Similar to the *E. coli* RNAspace analysis, predictions generated by RNAspace were compared to a list of known sRNA sequences collected from different web-accessible databases such as Rfam (Griffiths-Jones et al., 2003), BSRD (Li et al., 2013), and RNAcentral (The Rnacentral Consortium, 2019). The following sequences can be previously observed in Table 1. A total of 23 out of the 46 sequences were matched with the RNAspace predictions, thus indicating that RNAspace was able to predict previously discovered sRNAs in *R. sphaeroides* with 50% success. Additionally, out of the 23 sequences, a total of 13 were predicted by two or more programs.

The total 712 sRNAs predicted in *R. sphaeroides* provide a basis for identification of these sequences *in vivo*. While a bioinformatic approach does not distinguish between different growth conditions where sRNA may be most prevalent, it is possible to determine whether these predictions are considered as true sRNA sequences through various wet lab experimentation. Methods such as microarray analysis, small RNA sequencing, reverse transcriptase PCR (RT-PCR), and Northern blot analysis can be utilized to verify these predictions within *R. sphaeroides* under different growth conditions of interest (Pichon & Felden, 2008).

Future Work

While a total of 712 sRNA predictions were generated for *R. sphaeroides*, the amount pales in comparison to the 1,893 predictions generated for *E. coli*. One reason for this occurrence may be due to a bias of sRNA prediction programs towards enteric and pathogenic related bacteria. The bioinformatic approach to identify sRNAs has merely relied on using organisms such as *E. coli* or *Salmonella* strains as the basis for creating

the predictive algorithm (Li et al., 2012). RNAspace was chosen due to the variety of predictive programs within the platform in an attempt increase the capturing ability of sRNA sequences. However, it is possible that some of these programs have an inherent bias towards bacterial organisms like *E. coli* or *S. enterica*.

In addition to a possible inherent bias, the sRNA predictions generated by each individual gene finder were not evenly distributed for both the *R. sphaeroides* and *E. coli* genomes, indicating uneven predictive capabilities when comparing the programs. This phenomenon may exist due to the nature of each prediction program. For example, the BLAST gene finder compares the input genome to annotated RNA families within the Rfam database and detects highly similar sequences as predictions. However, the comparative analysis gene finder performs very differently in that it focuses on identifying sequences within the intergenic regions of the genome which are conserved across a group of species and exhibit a thermodynamically stable RNA secondary structure. While it was originally believed that identifying sequences predicted by two or more algorithms were deemed robust predictions, the results generated very little overlap between the three gene finders used within RNAspace. Therefore, a simple comparative analysis of the predictions generated by each program may not suffice for robust sRNA prediction. In the future, a different approach should be used to mitigate the two issues mentioned above. One way scientists are working to resolve these two problems is through use of machine learning models. One study found that utilizing sequence derived features unique to sRNAs within the machine learning algorithm increased the performance of detecting sRNAs when compared to a comparative analysis approach

(Tang et al., 2018). Therefore, the use of machine learning models may provide robust

sRNA predictions for a wide range of bacterial organisms.

CHAPTER II

Molecular Analysis of sRNAs and their Corresponding Target Genes in *Rhodobacter*

***sphaeroides* under Gold Chloride Stress**

Recent innovation of bioinformatic tools directed towards predicting sRNAs and

their respective mRNA targets have allowed scientists to 1) develop streamlined

experimental approaches for validation, 2) classify regulatory functions of the sRNA, and

3) utilize a quick and cost-effective technique for sRNA and sRNA target identification

(Mendel, 2019). Small RNA prediction programs can provide users with a candidate list

of sRNAs that exist within their respective organism. This candidate list can then be

surveyed, and sRNA sequences of interest can be chosen for experimental validation and

examination. Common validation methods for sRNAs include microarray analysis, RNA

sequencing, and Northern blot analysis (Sharma & Vogel, 2009; Wassarman et al., 2001).

Additional methods utilizing co-immunoprecipitation with Hfq, the RNA chaperone

commonly associated with sRNA stability, have also been used to aid in identification of

sRNAs in an organism (Zhang et al., 2003). Moreover, the growing increase in

technology has allowed scientists to advance sRNA detection methods with better

accuracy and efficiency. For example, the use of a chimeric deoxyuracil (dU) stem-loop

primer in quantitative reverse transcription (RT-qPCR) can increase the sensitivity and

specificity of the sRNAs being detected in a bacterium (Wu et al., 2017). A new and

comprehensive sequencing approach, deemed RIL-seq for RNA interaction by ligation

and sequencing, uses co-immunoprecipitation with Hfq and RNA-ligation to identify

novel sRNAs and their interacting mRNA targets *in vivo* (Melamed et al., 2018). Using

wet-lab approaches can not only identify sRNAs present in each condition but can also validate the prediction accuracy of any given sRNA prediction program.

Small RNA sequences were previously predicted in *Rhodobacter sphaeroides* using RNAspace, a web-accessible bioinformatic tool for noncoding RNA prediction (Cros et al., 2011). The 712 predicted sRNA sequences generated by RNAspace provides a list of key candidates to observe in wet-lab experimentation. While sRNA prediction cannot specify certain environmental conditions in which an sRNA may be present, it can provide an overall snapshot of sequences that may exist throughout the entire genome. For the verification of sRNAs in *R. sphaeroides* under gold chloride contamination, sRNA sequencing paired with a size selection of 50-200 nucleotides will be used in the following experiment. The sRNAs that are identified in this growth condition will be matched to the previously generated list of predictions to determine the prediction accuracy of RNAspace. Additionally, these sRNA sequences will be further used for identification of target genes through use of another bioinformatic prediction program: CopraRNA. By identifying which target genes have a likelihood to interact with the sRNAs present in the cell, we can predict the regulatory functions of each sRNA sequence.

Identification of sRNA Target Genes with CopraRNA

With identification of small, noncoding RNAs comes a rising demand for distinguishing mechanisms by which these sRNAs regulate their respective targets. Bioinformatic approaches to detect sRNA targets have become more prominent in scientific research schemes. Each target prediction program harbors a unique algorithm for detecting sRNA-mRNA interaction predictions, yet all center around four major

characteristic approaches. These approaches aim to observe the structural integrity and

sequence conservation of the sRNA, the structural integrity and sequence conservation of

the mRNA transcript, and the predicted interacting sites of the sRNA-mRNA hybrid

(Ahmed et al., 2018). Additionally, some programs provide the option to pinpoint

predicted interaction sites near the 5' UTRs or translational start sites of the mRNA

transcripts; regions which are highly responsible for successful initiation of translation

(Kery et al., 2014; Mann et al., 2017). Out of the few sRNA target predictors readily

available to researchers, data has shown that CopraRNA provides the most accuracy in

detecting novel sRNA-mRNA interactions for any given organism (Pain et al., 2015;

Wright et al., 2013).

CopraRNA, an acronym short for Comparative Prediction Algorithm for sRNA

Targets, is a web-accessible program which uses phylogenetic information and sequence

conservation to predict sRNA targets within an organism of interest. By focusing on

conservation of target genes, the CopraRNA algorithm can accurately predict areas where

sRNAs recognize and interact with their targets. CopraRNA requires at least three sRNA

sequences that share similarity with the sRNA of interest for the input. The results consist

of a list of the top 200 putative targets which are organized by the corresponding

CopraRNA p-value (Wright et al., 2013).

CopraRNA can also be paired with the IntaRNA algorithm, a separate Freiburg

RNA tool which exists as an individual program as well as an algorithm built into

CopraRNA. IntaRNA, which stands for Interacting RNAs, is used to measure potential

RNA-RNA interactions between a query sRNA sequence and mRNA targets. The

algorithm calculates the free energy values of the sRNA and mRNA sequences to

determine target site accessibility. Additionally, the algorithm locates an interacting seed region, or a short-base pairing region, where the sRNA and mRNA are most likely to interact (Busch et al., 2008). Results generated by IntaRNA consist of the sRNA and mRNA sequence, the base-pairing region and hybridization energy of the site, and a generated heatmap of the interaction site for visualization (Mann et al., 2017).

The CopraRNA/IntaRNA hybrid system was tested on datasets from 18 enteric bacterial species for validation. The results generated by the program showed a 74% success rate of prediction after the sRNA-mRNA target interactions were verified with compensatory mutation assays (Wright et al., 2014). Additional validation of target prediction has been performed using this hybrid approach in non-enteric organisms such as *Sinorhizobium meliloti* (Baumgardt et al., 2015). A recent study comparing a new target prediction program, sRNARFTarget, against the standalone IntaRNA platform and CopraRNA with built in IntaRNA parameters showed that CopraRNA was able to outperform this program based on accuracy and running time (Naskulwar & Peña-Castillo, 2021). The combined use of CopraRNA and IntaRNA for target identification provides a robust approach to accurately identify sRNA targets for further experimentation.

It is hypothesized that a set of sRNA sequences will be expressed under gold chloride-contaminated stress conditions and will be predicted to influence expression of corresponding target gene(s). The use of experimental identification of previously predicted sRNAs with wet lab sequencing, and the pairing of sRNA target prediction through bioinformatic means can help shed light on the regulatory functions associated with each sRNA under this stress-induced growth condition.

Methods

The following experiment was adopted by a former graduate student who studied growth characteristics of *Rhodobacter sphaeroides* 2.4.1 under different concentrations of gold chloride solution (Johnson et al., 2017).

Bacterial Strain and Media Preparation

Rhodobacter sphaeroides 2.4.1 was used to conduct the following experiment. Sistrom's minimal media (SIS) was used for growing the bacterial cells, since it is a selective minimal media (Sistrom, 1960). *R. sphaeroides* was taken from a -80°C frozen glycerol stock and plated on a SIS plate where it grew aerobically at 30°C to achieve colony formation. Culture tubes containing 5 mL of liquid SIS media were inoculated with individual colonies and grown aerobically in a shaking incubator at 30°C and at 140 rotations per minute (rpm). The bacterial culture stocks were grown until the log phase was reached, which is represented by an optical density reading at the 600nm wavelength (OD_{600}) between 0.6 – 0.8. Growing bacterial stocks up to the logarithmic phase has been observed as the optimal growth phase for inoculating heavy metals to study their effects on bacteria (Daughney et al., 2001).

Gold Chloride Solutions

A total of 500 mg of gold chloride ($AuCl_3$) was purchased from Acros Organics in powdered form. Due to the hygroscopic nature of this chemical, a heavily concentrated stock of 29.97 mM was made using MilliQ-filtered water as the solvent, and the remaining chemical was stored in a vacuum-sealed desiccator. This concentrated stock was then used to make a 1.0 mM working stock solution for the following experiment.

The working stock solution was used to inoculate a series of culture tubes with the following concentrations of gold chloride: 0.0 µM, 0.5 µM, 1.0 µM, 2.5 µM, and 5.0 µM.

Growth Characteristics

A series of analyses to identify the growth kinetics of *R. sphaeroides* in gold chloride contaminated media was performed under aerobic growth conditions. As mentioned previously, culture stocks of *R. sphaeroides* were grown in liquid SIS media until an OD_{600} reading of 0.6-0.8 was reached. At this point, three culture stocks with similar optical density readings were chosen to represent the three biological replicates for the experiment. The bacterial stocks were used to inoculate their respective experimental tubes containing liquid SIS media mixed with a given concentration of gold chloride. A control group of SIS media without gold chloride was also inoculated. Within each gold chloride concentration group, individual tubes were created for each time point to not disrupt the cellular growth during sample analysis. A schematic diagram of this experimental setup can be observed in Figure 5.

Figure 5

Schematic Diagram of Growth Curve Experiment Setup
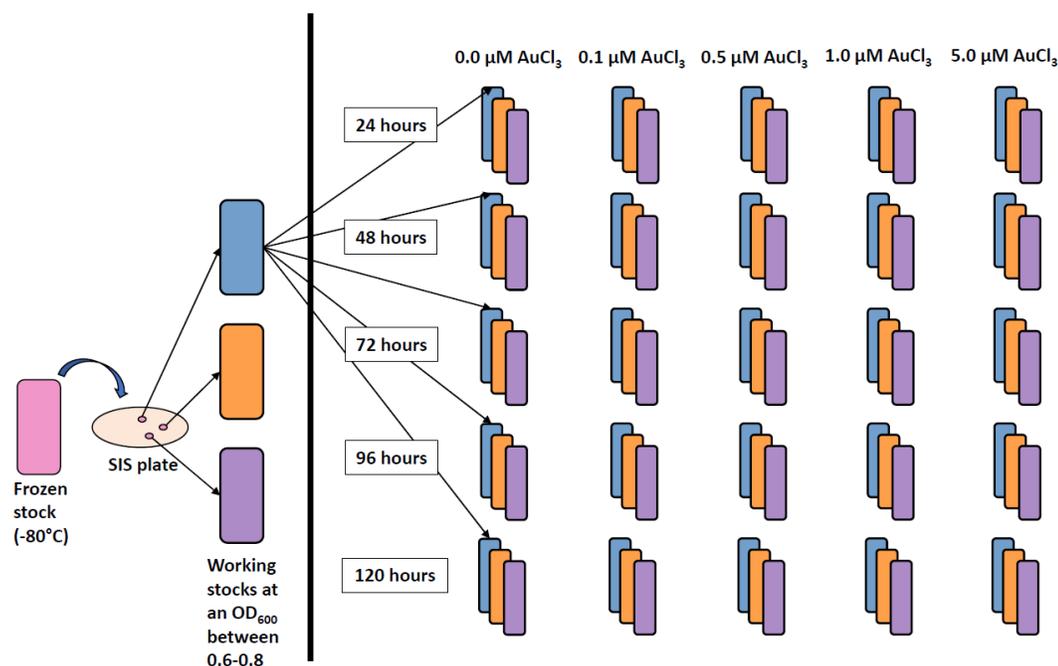
Note. The following image depicts the method of inoculation for the growth curve experiment of *Rhodobacter sphaeroides* under varying gold chloride concentrations. Three culture stocks were generated from individual colonies to represent individual biological replicates. Each stock was used to inoculate its respective tubes for each gold chloride concentration and at each time point where data was going to be collected.

The optical density measurements were recorded using a UV-vis spectrophotometer set to the 600nm wavelength. Readings were recorded for each sample at 24-hour intervals up until the 120-hour timepoint. A total of 1 mL of each sample was used to obtain the reading with a spectrophotometer. Blanks were created for the control group and for each of the four gold chloride concentrations. The data collected from each sample was plotted to represent the growth curve, as seen in Figure 8.

Cell viability was also performed by plating dilutions of the samples at each 24-hour interval. A serial dilution was performed on 1 mL of each sample until a dilution

factor of 10^{-6} was achieved. A 100 µL aliquot was plated onto SIS plates containing the appropriate gold chloride concentration using glass bead plating. The plates were incubated at 30°C for a span of four days. On the fourth day, the colony counts were recorded for each sample and plotted, as seen in Figure 9.

Statistical Analysis

A one-way ANOVA was conducted to determine a difference in optical density and CFU counts between the groups at each timepoint. A p-value of less than or equal to 0.05 indicated that the compared groups were significantly different. A post-hoc analysis using the Tukey's test was performed on timepoints that were deemed statistically significant to identify which groups were different from one another.

RNA Isolation of Rhodobacter sphaeroides

According to statistical analysis of the growth kinetics, the control and 1.0 µM gold chloride group showed a statistically significant difference at the 24-hour timepoint. All other timepoints were not statistically different. Therefore, cells were collected for all three replicates at the 24- and 72-hour time points for the control and 1.0 µM gold chloride groups to undergo RNA analysis. A total of 7 mL of cells were spun down in a centrifuge (8,000 rpm, 5 min., 4°C) to form a pellet and were flash frozen in a dry ice/ethanol bath. The frozen bacterial pellets were stored at -80°C until further processing. The Norgen Biotek Total RNA Isolation Kit was used to isolate RNA from the frozen samples. An on-column DNase I treatment (Norgen Biotek DNase I Kit) was used to remove contaminating DNA. Total RNA of each sample was examined on the Nanodrop One machine for quality and quantity, and then were stored at -80°C. The

samples were packaged with dry ice and sent to LC Sciences in Houston, TX for total

RNA sequencing and small RNA sequencing.

Micro/Small RNA Sequencing

The total RNA isolated from *R. sphaeroides* was used for small RNA sequencing. The following library construction (Figure 6) and sequencing procedure were performed by LC Sciences (Houston, TX). In summary, a quality control check was performed on the total RNA samples using a Bioanalyzer 2100 (Agilent, CA, USA). The library prep was performed using the TruSeq Small RNA Sample Prep Kit (Illumina, San Diego, USA). Single end 50bp sequencing was performed on an Illumina Hiseq 2500 at LC Sciences (Hangzhou, China). Raw reads were obtained and run through an analysis platform generated by LC Sciences. The raw data files and an analysis report were obtained from LC Sciences.

Figure 6

Library Construction of Total RNA for Small/MicroRNA Sequencing
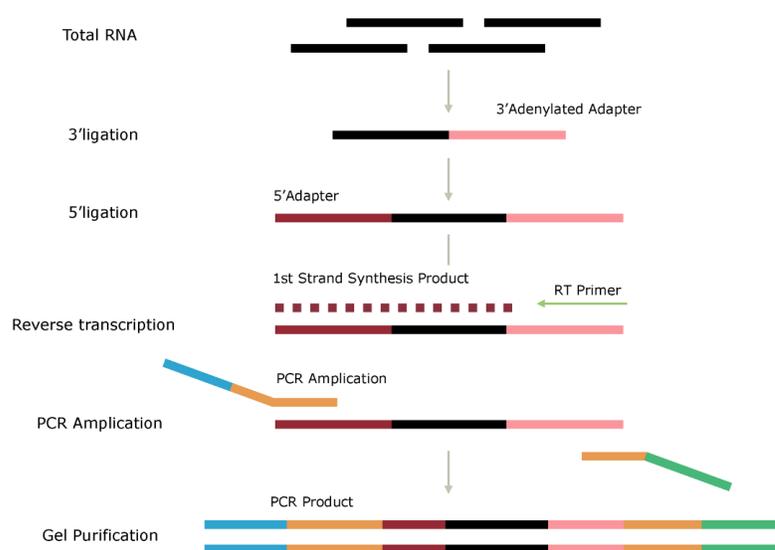
Data Analysis of Sequencing Data

The following bioinformatic pipeline for data analysis can be observed in Figure 7. The raw reads generated by sequencing were analyzed using an in-house program ACGT101-miR (LC Sciences, Houston, TX) to remove adapter sequences, low quality reads, and contaminating sequences. Remaining sequences greater than or equal to 18 nucleotides in length were annotated using the Rfam database (Griffiths-Jones et al., 2003) to remove noncoding RNAs such as rRNA, tRNA, snRNA, snoRNA, and degraded fragments of mRNA sequences. Sequences that were not mapped to Rfam were aligned to miRbase, a miRNA database, and perfectly matched sequences were considered conserved miRNAs in *R. sphaeroides*. Since *R. sphaeroides* is not found to contain miRNA sequences, reads were not annotated in the miRNA database. Instead, reads were mapped to the *R. sphaeroides* genome and hairpin identification of sequences were determined based on the following criteria: number of nucleotides in one bulge in stem $\leq$ 12, number of base pairs in the stem region of the predicted hairpin $\geq$ 16, cutoff of free energy kCal/mol $\leq$ -15, length of hairpin up and down stems and terminal loop $\geq$ 50, number of biased bulges in mature region $\leq$ 2, number of base pairs in the mature region of the predicted hairpin $\geq$ 12, percent of mature in stem $\geq$ 80. The final sequences detected from the hairpin prediction were determined as small RNAs identified within the sequencing. The normalization of sequence counts mentioned in Li et. al. was also performed on each data set (2016). Differential expression of sRNAs was observed in each comparison group as seen in Figure 9. Student's t-tests were performed on normalized sequencing counts, and p-values of $\leq$ 0.05 were deemed significant.

Figure 7

Bioinformatic Pipeline of Small/MicroRNA Sequencing Data Analysis
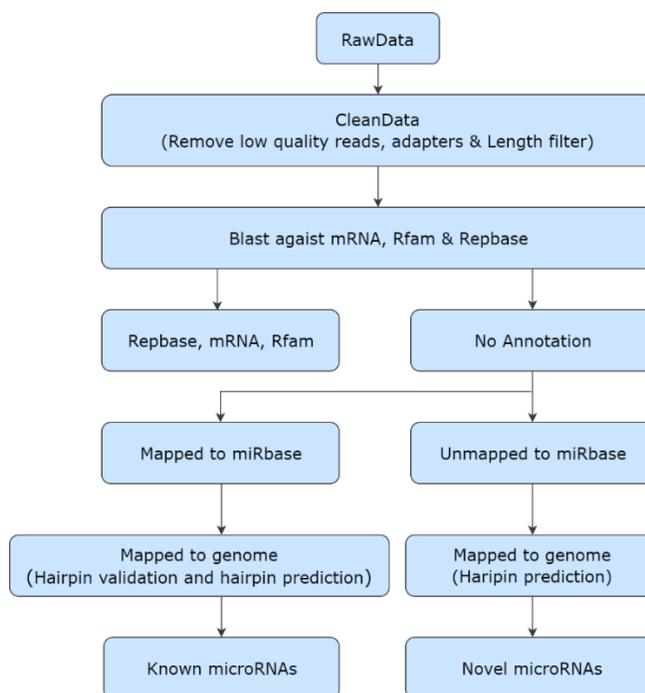
Identification of sRNA Homologs using GLASSgo

The web-accessible program GLASSgo (http://rna.informatik.uni-freiburg.de/GLASSgo/Input.jsp) (Lott et al., 2018) provided by the Freiburg RNA Tools webserver was used to identify sRNA homologs for the differentially expressed sRNAs identified in sequencing. Each sRNA sequence was inserted into the query and searched against the Alphaproteobacteria taxon selection category under default conditions. The results were collected, and each sequence was labelled with the corresponding Reference Sequence ID provided by the National Center for Biotechnology Information (NCBI). The generated list of each sRNA and its respective homologs were used as input for CopraRNA.

Identification of sRNA Targets using CopraRNA

The web-accessible program CopraRNA (http://rna.informatik.uni-freiburg.de/CopraRNA/Input.jsp) was utilized on the Freiburg RNA Tools platform to identify putative target genes for sRNAs found in *R. sphaeroides* sequencing data (Mattheis et al., 2018). Each sRNA and its respective list of homologs generated by GLASSgo were inserted into the query of the CopraRNA website. The CopraRNA/IntaRNA integrated platform was run under default conditions. CopraRNA default consists of sequences extracted around the start codon (200 nucleotides upstream, 100 nucleotides downstream), a dynamic p-value setting, and no consensus prediction (Wright et al., 2013). IntaRNA default consists of a seed region minimum of 7 base pairs, interaction overlap in the query only, and a 0°C maximum absolute energy for an interaction (Mann et al., 2017). A list of the top 200 targets were generated as a default for each sRNA and were compared to the up and down regulated genes identified in the total RNA sequencing data that were deemed significant.

GLASSgo and CopraRNA Program Validation

To identify the feasibility of GLASSgo and CopraRNA in identifying sRNA target sequences, the programs were first used to identify targets of previously studied sRNAs in *R. sphaeroides*. The following list of sRNAs and their experimentally validated targets were obtained for analysis and can be seen in Table 3. Each sRNA was put through GLASSgo and CopraRNA following the parameters and methods mentioned above. The predicted targets of each sRNA were compared to the list of experimentally validated targets, and information such as CopraRNA p-value, IntaRNA p-value, hybridization energy, and rank out of 200 were obtained.

Results and Discussion

Growth Characteristics under Gold Chloride Contamination

The optical density readings (shown in Figure 8) and colony forming units (shown in Figure 9) were recorded when *R. sphaeroides* was grown in different concentrations of gold chloride solution. Bacterial growth was not observed at the 2.5 µM and 5.0 µM concentrations, suggesting that these concentrations are lethal for *R. sphaeroides* survival. Due to the absence of bacterial growth, these two groups were removed from statistical analysis. As seen in Figure 8, the increasing concentration of gold chloride significantly impacted *R. sphaeroides'* growth. The lag phase, which can be observed at the 24-hour interval, was prolonged when *R. sphaeroides* was grown in 1.0 µM of gold chloride. The one-way ANOVA for this timepoint calculated a p-value of <0.001, indicating the growth of the groups at this timepoint were significantly different. The Tukey's test confirmed that the bacterial growth of the control group and 1.0 µM treated group showed a difference with a p-value of 0.001. However, at the 72-hour timepoint, the bacterial growth between these groups was not significantly different. This phenomenon indicates an adaptation mechanism being exhibited by *R. sphaeroides*. When exposed to the gold chloride solution, the bacteria are needing to re-adjust to the environment by altering certain cellular and molecular functions within the cell. Therefore, we can see an elongated lag phase as the bacteria grow to adapt within the toxic environment. Once they reach the late log/stationary, the bacteria are then able to fully adapt and grow in a fashion similar to the control group as observed at the 72-hour timepoint.

Figure 8

Growth Curves of R. sphaeroides with Different Gold Chloride Concentrations
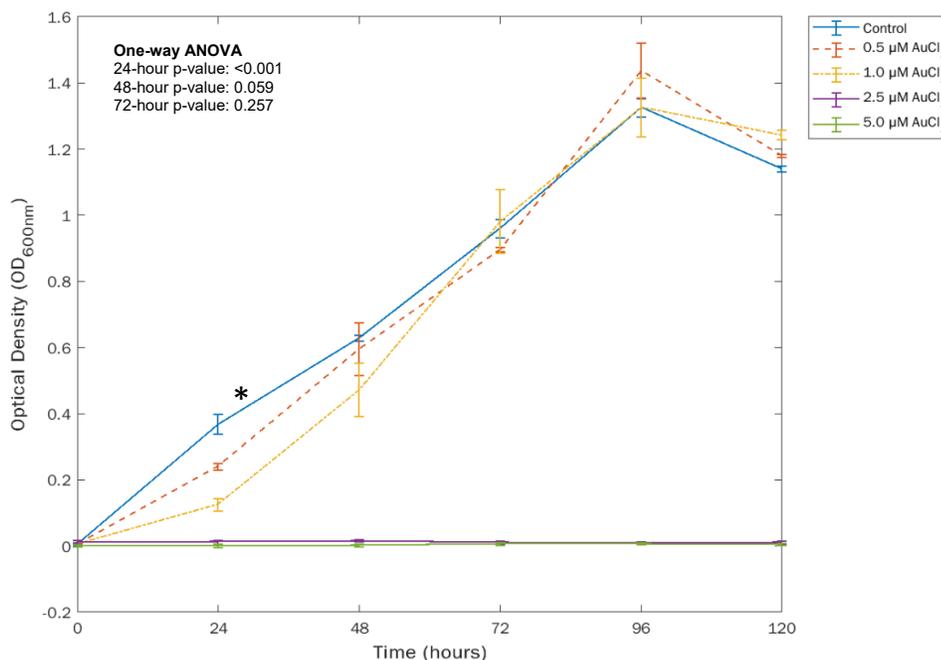
Note. The following graph depicts the growth curve analysis of *Rhodobacter sphaeroides* under varying gold chloride concentrations. Optical density readings were taken at 24-hour intervals up until 120 hours. A wavelength of 600 nm was used to measure absorbance of each sample. Each point represents the average of three biological replicates, with the standard deviation shown by the error bars. Asterisk represents p-value less than 0.05 for one-way ANOVA analysis.

Interestingly, a one-way ANOVA performed on the CFU data at the 24-hour timepoint showed no significant difference. One possibility for this occurrence could be the adaptation mechanism being observed by the cells. Since the samples were seemingly adapting to the growth environment at the 24-hour timepoint, the plating of these bacteria on an agar plate containing the same gold chloride concentration had no impact on the overall cell survival. The bacterial cells were already adapted and therefore able to grow optimally. Additionally, bacterial cells were not synchronized during the collection of optical density and CFU data points and could thus influence the differences observed in the one-way ANOVA analysis between the two data sets. Future replication of this

experiment will help to understand the observed difference. More repeated experiments with synchronized cell growth will be needed to identify if this phenomenon exists over multiple testing.

Figure 9

Colony Forming Units (CFU) of R. sphaeroides with Different Gold Chloride Concentrations
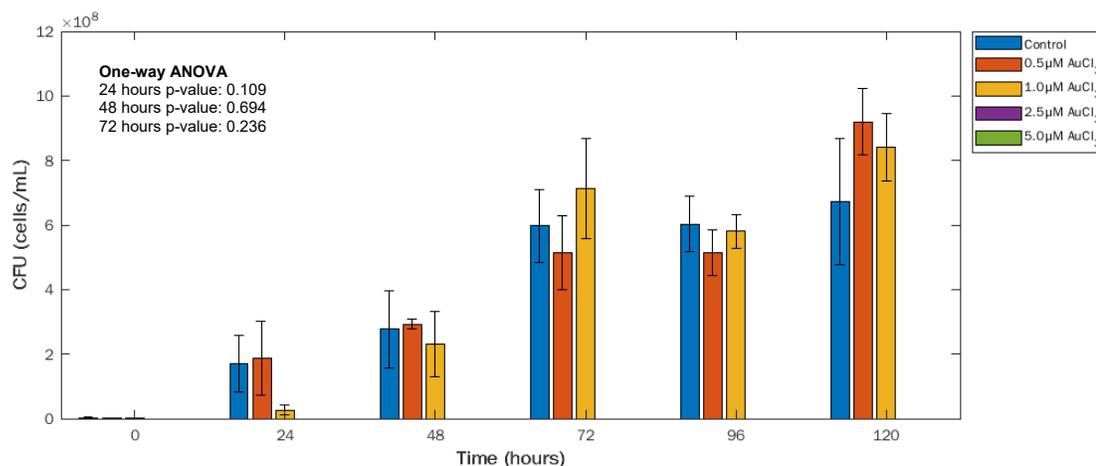
Note. The following graph depicts the cell viability of *Rhodobacter sphaeroides* under varying gold chloride concentrations. Colony forming units (CFUs) were collected for each 24-hour interval up until 120 hours. Each point represents the average of three biological replicates, with the standard deviation shown by the error bars.

Since the 1.0 μM gold chloride treatment group showed the most difference when compared to the control group at the 24-hour timepoint, these two groups were chosen for further RNA sequencing analysis at the 24-hour and 72-hour timepoints.

Small RNA Sequencing Analysis

To assess the quality of the sRNA sequencing results, a Pearson correlation and a principle component analysis (PCA) were generated to determine the clustering of biological replicates within each sample group and can be observed in Figures 10 and 11.

A strong Pearson correlation is represented by an R^2 greater than 0.9 for any two replicates existing within the same sample group (Conesa et al., 2016).

Figure 10

Pearson Correlation Between Biological Replicates in sRNA Sequencing Data

Figure 11

Principle Component Analysis of Biological Replicates in sRNA Sequencing Data
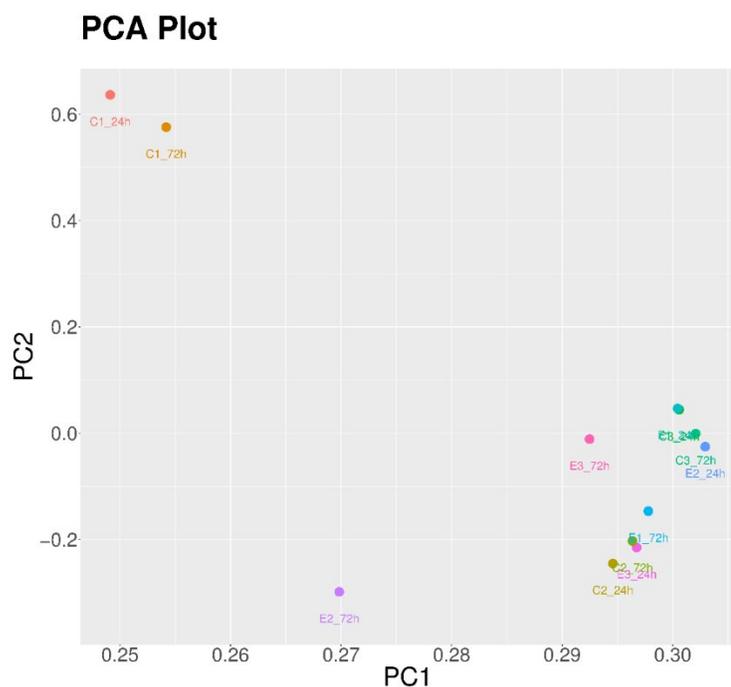
It can be observed in both Figure 10 and Figure 11 that the correlation between biological replicates within the control groups are low, with an R^2 of around 0.6-0.7. This could be due to the lack of sample quality from the original total RNA sample, or sample contamination introduced during the RNA isolation or library construction step. To account for this discrepancy, sRNAs with minimal representation presented in the sequencing (less than 2 of the 3 replicates exhibiting reads for the sequence where present) were removed from further analysis. Additionally, sequences which exhibited low expression levels were also removed.

After removal of poorly correlated sRNA sequences, A total of 24 sRNA sequences were detected across all sample groups, with sequence lengths varying between 50-300 nucleotides. The distribution of sequence lengths is consistent with previously published data of different bacterial species. A differential expression analysis comparing sample groups in four different ways identified seven unique sRNAs which were observed to be significantly differentially expressed (Student's t-test p-value less than 0.05). The following data can be observed in Figure 12.

Figure 12

Differential Expression of sRNAs Observed Across Four Comparison Groups
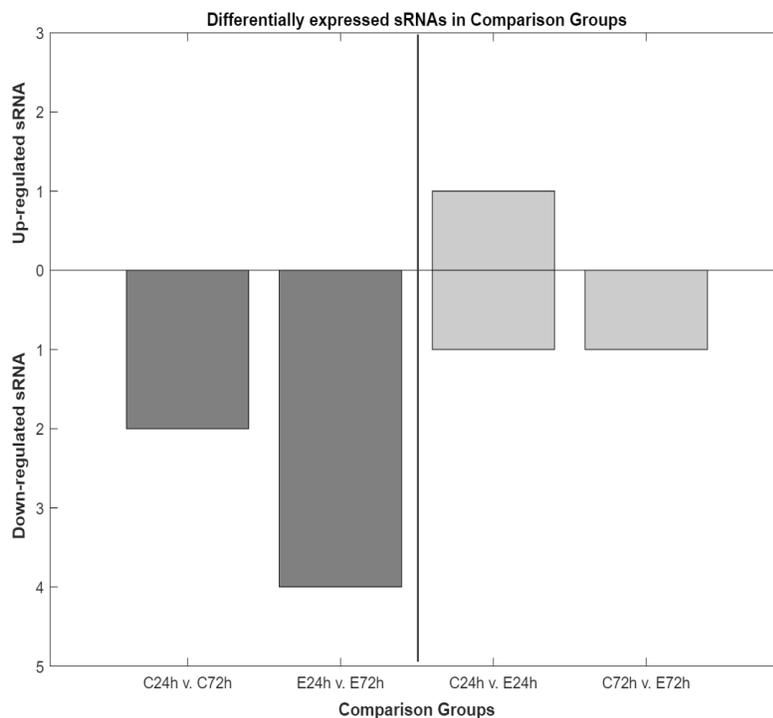
Note. Differentially expressed sRNAs identified in each comparison group. The representation of groups are as follows: control group at 24 hours vs. the control group at 72 hours (C24h v. C72h), 1.0 μM AuCl₃ group at 24 hours versus the 1.0 μM AuCl₃ group at 72 hours (E24h v. E72h), control group at 24 hours versus the 1.0 μM AuCl₃ group at 24 hours (C24h v. E24h), control group at 72 hours versus the 1.0 μM AuCl₃ group at 72 hours (C72h v. E72h). The first two groups were compared to one another (dark gray), and the last two groups were compared (light gray), and no sRNAs were found similar between the comparisons.

Out of the seven differentially expressed sRNAs, two were found to be down-regulated in the first group comparing the control 24-hour samples to the control 72-hour samples. Additionally, four sRNAs were observed to be down-regulated within the 1.0 μM AuCl₃ (or experimental) group when between the 24-hour and 72-hour timepoints. The comparison between the control group at 24 hours and the experimental group at 24 hours identified one up-regulated and one down-regulated sRNA. Lastly, the comparison

between the control and experimental group at the 72-hour timepoint exhibited one down-regulated sRNA. To identify potential targets, each of these seven sRNA sequences was subjected to CopraRNA.

Sequenced sRNAs Predicted by RNAspace

To determine if RNAspace was able to predict any of the 24 identified sRNA sequences, the predictions were compared to each sequence using the align two or more sequences function in BLASTn. Upon analysis, a total of three out of the 24 sRNAs were found to match with RNAspace predictions and can be observed in Table 2.

Table 2

Sequenced sRNAs Detected by RNAspace

sRNA sequence ID	RNAspace Prediction ID	Chromosome	RNAspace Program	RNAspace score
PC-3p-999_2009	000276	1	BLAST/CG-seq/RNAz	0.756037
PC-3p-79238_5	000273	1	BLAST/CG-seq/RNAz	0.999977
PC-5p-38477_14	000085	1	BLAST/CG-seq/RNAz	0.99444

Since only three out of the 24 sRNAs were able to be detected by RNAspace, it reveals that RNAspace may not suffice for sRNA prediction, particularly when observing the bacterium under gold chloride stress. RNAspace provides an overall identification of sRNA sequences and does not discriminate between sRNAs that exist in a particular condition. Therefore, it is likely that some of the predictions generated by RNAspace are bonafide sRNA sequences, however they do not appear within our tested condition. Additionally, it is possible that RNAspace was not able to predict the remaining 21

sRNAs due to discrepancies in the methods used to identify the sRNA predictions. Therefore, more testing with a different, more robust sRNA prediction method like deep-learning and machine learning will be needed to improve upon the bioinformatic detection of sRNAs within *R. sphaeroides.*

CopraRNA Target Prediction

To evaluate the capability of CopraRNA at detecting targets within *R. sphaeroides*, a set of sRNAs with experimentally identified targets were chosen for analysis through the prediction platform. The following list of sRNAs, their targets, and the capability of CopraRNA to detect each target can be observed in Table 3.

Table 3

R. sphaeroides Experimentally Identified sRNA Targets Detected by CopraRNA

sRNA sequence	Experimentally identified targets	Ranking in CopraRNA (out of 200)	CopraRNA p-value	IntaRNA p-value
CcsR1-4	flhR (RSP_2591)	4	0.01334	0.012671
	pqqA (RSP_6132)	15	0.02252	0.017618
	pdhB (RSP_4050)	n/a	n/a	n/a
	coxL (RSP_2877)	47	0.0388	0.034870
PcrX	pufX (RSP_0255)	122	0.01957	0.125152
PcrZ	bchN (RSP_0285)	n/a	n/a	n/a
	puc2A (RSP_6158)	n/a	n/a	n/a
Pos19	RSP_0557	n/a	n/a	n/a
SorX	potA (RSP_1882)	135	0.0138	0.107237
SorY	takP (RSP_0097)	12	0.0008648	0.007858

For the following six sRNAs, CopraRNA was able to detect six out of the 10 targets within the *R. sphaeroides* genome, with CopraRNA p-values ranging from less than 0.001 to 0.05. Two sRNAs, PcrZ and Pos19, did not have their respective targets detected by CopraRNA. This lack of result may be due to the sRNA-mRNA target interaction being a species-specific occurrence. Since CopraRNA focuses on the conservation of the sRNA and its corresponding target sequences, it does not detect targets that are specific to a single organism (Wright et al., 2013). It is possible that these sRNA sequences and/or their interactions with other targets are species-specific to *R. sphaeroides* and share little commonality with other related species. Additionally, information covering sRNAs and their detected targets *in vivo* is limited for *R. sphaeroides*. It is highly likely that the sRNA sequences contain far more targets under different growth conditions, which may very well be detected by the CopraRNA program. However, to determine if this is real and not artifact, further experimentation is needed with *R. sphaeroides* to broaden the regulatory role of each sRNA. Since CopraRNA was able to detect more than half of the validated targets for the sRNAs identified in *R. sphaeroides*, it was used for further identification of target genes for the seven differentially expressed sRNAs detected by sRNA sequencing.

A total of 200 putative targets were generated for each sRNA sequence, and the distribution of sRNAs according to various p-value cutoffs can be observed in Table 4.

Table 4

P-value Distribution of CopraRNA Targets for Seven Differentially Expressed sRNAs

Comparison	sRNA	Total number of targets predicted	Number of targets with Copra p-value cutoff			
			≤0.05	≤0.01	≤0.005	≤0.001
Control 24 hours v. Control 72 hours	PC-5p-21399_35	200	200	81	55	19
Experimental 24 hours v. Experimental 72 hours	PC-5p-14897_58	200	125	11	5	2
	PC-3p-999_2009	200	200	44	23	5
Control 24 hours v. Experimental 24 hours	PC-5p-25298_28	200	200	59	29	9
	PC-3p-14954_58	200	197	39	20	4
	PC-5p-2842_438	200	200	200	200	59
	PC-5p-21399_35	200	200	81	55	19
Control 72 hours v. Experimental 72 hours	PC-3p-14954_58	200	197	39	20	4
	PC-3p-999_2009	200	200	44	23	5

According to the benchmark testing of CopraRNA, a p-value of 0.01 was considered a stringent cutoff, since it was able to predict 50 out of 101 experimentally verified sRNAs in *Escherichia coli* (Wright et al., 2013). However, when comparing the experimentally verified targets found in *R. sphaeroides* which were detected by CopraRNA (Table 2), it can be observed that some of the targets were predicted with p-values greater than the 0.01 cutoff value. It is highly possible that gene targets found within *R. sphaeroides* may not be detected at highly stringent p-values within

CopraRNA. Therefore, to determine the potential role of each sRNA in regulating its respective targets, the CopraRNA predictions generated for each sequence will be compared to differentially expressed genes which are observed within the same comparison group. By doing so, this will give a better indication of the regulatory mechanisms imposed for each sRNA sequence.

Future Work

The identification of novel sRNAs within the 1.0 µM gold chloride treated group provides promising information about the usage of sRNAs in mediating heavy metal stress. To adequately identify the presence of these sRNAs in *R. sphaeroides*, a second method of experimental identification is needed to increase confidence. Therefore, the following sRNAs will be further detected using RT-qPCR with primers specific to the sequence identified in the RNA sequencing. The validation of these sRNAs using a second method can help confirm the presence within the organism and remove any false discovery that may have been detected through sequencing. Any sRNAs that are found by both sRNA-seq and RT-qPCR will be chosen for further experimental analysis. Specifically, a series of overexpression analyses for each sRNA will be conducted to determine the sRNA-specific effect in mediating heavy metal stress, as well as determine targets that are differentially regulated upon overexpression. Additionally, sRNA-mRNA target identification can be observed by genetically modifying bases involved in the sRNA-mRNA hybridization to determine the sRNA effect on the respective target.

A second method that will be used to further enhance the work done in this study will be to explore a novel target identification program, sRNARFTarget, which utilizes machine learning modelling to detect robust sRNA target predictions (Naskulwar &

Peña-Castillo, 2021). While CopraRNA performs better than sRNARFTarget, there is still an inherent limitation with CopraRNA by requiring highly conserved sRNA and target sequences (Wright et al., 2013). In sRNARFTarget, the conservation of sequences is not necessary, therefore allowing potential organism-specific target interactions to be detected (Naskulwar & Peña-Castillo, 2021). Therefore, combining results generated by both CopraRNA and sRNARFTarget may enhance the detection of true gene targets associated with each sRNA sequence.

CHAPTER III

Differential Gene Expression Profiles of *Rhodobacter sphaeroides* in Gold Chloride

Stress

The understanding of heavy metal tolerance and heavy metal resistance in bacterial organisms has been of interest for utilizing these microorganisms in bioremediation. Understanding the mechanisms by which bacteria adapt and survive in toxic metal environments can help researchers identify new ways to mitigate damage caused by heavy metals within the environment. The analysis of gene expression, a phenomenon commonly associated with detecting adaptive strategies of bacteria, can provide insight to the genes and metabolic pathways used for heavy metal tolerance. Additionally, the understanding of gene regulation which occurs during heavy metal stress can provide researchers with a useful gene set to further manipulate for more efficient and effective means of heavy metal removal from the environment.

Heavy Metal Stress in Bacteria

Many bacterial species is known to tolerate heavy metal stress within their environment. Microorganisms common to soil environments where heavy metals are most found have been isolated and identified for further study involving bioremediatory properties (Abdu et al., 2017; Dhanwal et al., 2018). Such bacteria have been observed to respond to heavy metal stress in a variety of ways, such as biosorption, sequestration, metallic oxidation and reduction, and efflux (Ramasamy et al., 2007). For example, the copper resistant bacterium *Pseudomonas syringae* utilizes three copper-associated proteins (CopA, CopB, and CopC) located within the periplasmic and outer membrane regions of the cell to transport copper out of the cytoplasm (Cha & Cooksey, 1991).

Furthermore, *Bacillus cereus* has been observed to reduce toxic hexavalent chromium (CrVI) into less toxic trivalent chromium (CrIII) through use of a chromium reductase enzyme (Zhao et al., 2012). Whole transcriptomic analyses have been conducted on different organisms such as *Caulobacter crescentus* and *Sinorhizobium meliloti* to identify genes responsible for mediating the heavy metal tolerance within the bacterium (Hu et al., 2005; Lu et al., 2017). In *C. crescentus*, differentially expressed genes were analyzed with four different types of heavy metals, and each metal exhibited a different response. However, common expression of efflux pumps, membrane proteins, and enzymes involved in mediating oxidative stress were identified for all metals (Hu et al., 2005). In *S. meliloti*, exposure to copper and zinc resulted in induction of an oxidase, outer membrane protein, and multiple sulfite oxidoreductases (Lu et al., 2017). Moreover, transcriptomic analysis of Pseudomonas aeruginosa when exposed to copper showed induction of active transport enzymes, iron-associated proteins, and enzymes involved in oxidative stress (Teitzel et al., 2006).

Heavy Metal Stress in *Rhodobacter sphaeroides*

Previous studies have identified the suitability of *Rhodobacter sphaeroides* as a bioremediation agent due to its ability to adapt to various heavy metal ions and oxyanions within the environment (Buccolieri et al., 2006; Moore & Kaplan, 1992). For example, *R. sphaeroides* was identified to withstand varying concentrations of heavy metals such as mercury, copper, iron, nickel, and cobalt under photosynthetic growth. Additionally, the response of *R. sphaeroides* growth differentiated between the different metals observed, thus indicating individual mechanisms of adaptation for each heavy metal (Giotta et al., 2006). A different study observed the capabilities *R. sphaeroides* to withstand high

concentrations of chromium and furthermore exhibited the capability of *R. sphaeroides* to reduce toxic hexavalent chromium into the less toxic trivalent form (Nepple et al., 2000).

While studies have observed the impact of heavy metals on *R. sphaeroides* growth, little is known about the mechanisms by which the bacterium can adapt to each heavy metal upon exposure. One study aimed to identify the adaptive capabilities of *R. sphaeroides* when the organism was placed under cobalt ion stress. *R. sphaeroides* was grown under aerobic and photosynthetic conditions with 5 mM of cobalt ions (Co^{2+}), and the impact of toxicity on the bacterium was shown to be more powerful under aerobic growth than photosynthetic growth. Therefore, it was implicated that the energy producing pathways involved in photosynthetic growth help lessen the toxic effect of the metal on the bacterium. Additionally, the adaptive capabilities of *R. sphaeroides* to cobalt ions was observed to rely on the shifting of energetic metabolisms within the cell under aerobic and photosynthetic growth, particularly through the use of an ABC sugar transporter system (Volpicella et al., 2014). Mechanisms of heavy metal resistance with manganese ions were also observed in *R. sphaeroides* under photosynthetic and aerobic growth. The ions were observed to negatively influence the expression of genes involved with photosynthetic complexes, such as the *puc* operon. While it was imposed that the heavy metal ions were influencing these genes through a direct impact with PpsR, a known repressor of photosynthetic genes, the exact mechanism of action was still undefined (Horne et al., 1998).

Impact of Gold Chloride Contamination on *Rhodobacter sphaeroides*

Previous research has observed the impact of gold chloride contamination on the heavy metal tolerance of *R. sphaeroides* and Rhodobacter-related species (Feng et al.,

2007; Johnson et al., 2017). *R. sphaeroides* cells were observed to tolerate up to 1.0 µM

of trivalent gold (Au^{3+}) ions and were found to have accumulated gold nanoparticles

within the cytoplasmic and membranous fractions of the cell (Johnson et al., 2017). A

similar observation was made in an alternate study observing the influence of gold

nanoparticle generation by *R. sphaeroides*. In this study, spherical gold nanoparticles

were seen to have accumulated on the surface of the cells, thus indicating a mechanism of

gold bioaccumulation and reduction into a less toxic form occurring between the

extracellular trivalent gold and the outer membrane (Italiano et al., 2018). Additionally,

R. sphaeroides was observed to display cellular elongation upon exposure to 10 µM of

trivalent gold, thus revealing a bacterial defense mechanism similar to that observed

when *R. sphaeroides* was placed under chromate stress (Italiano et al., 2012).

While research has identified the tolerance capability of *R. sphaeroides* under

gold stress, little is known about the exact mechanisms by which *R. sphaeroides* is able to

mediate gold toxicity and generate gold bio-nanoparticles. To gain insight into the

mechanisms utilized by *R. sphaeroides* to mediate gold chloride stress, transcriptomic

analysis of the bacteria grown under aerobic conditions at the previously determined gold

chloride concentration (1.0 µM) will be observed. It is hypothesized that a specific set of

genes will be up- or down-regulated in the gold-chloride stress condition which influence

the bacterium's ability to tolerate the toxic metal.

Methods

RNA Isolation of Rhodobacter sphaeroides

As previously mentioned, bacteria from the control and 1.0 µM gold chloride

groups were chosen for sequencing analysis. Cells were collected for all three replicates

at the 24- and 72-hour time points. A total of 7 mL of cells were spun down in a

centrifuge (8,000 rpm, 5 min., 4°C) to form a pellet and were flash frozen in a dry

ice/ethanol bath. The frozen bacterial pellets were stored at -80°C until further

processing. The Norgen Biotek Total RNA Isolation Kit was used to isolate RNA from

the frozen samples. An on-column DNase I treatment (Norgen Biotek DNase I Kit) was

used to remove contaminating DNA. Total RNA of each sample was examined on the

Nanodrop One (Thermofisher Scientific) for quality and quantity, and then stored at -

80ºC. The samples were packaged with dry ice and sent to LC Sciences in Houston, TX

for total RNA sequencing and small RNA sequencing.

Total RNA Sequencing

The total RNA isolated from *R. sphaeroides* was used for total RNA sequencing.

The following sequencing procedure, which can be observed in Figure 13, was performed

by LC Sciences (Houston, TX). In summary, a quality control check was performed on

the total RNA samples using an Agilent Technologies 2100 Bioanalyzer. The ribosomal

RNA was removed, and the samples were then fragmented with divalent cation buffers

and elevated temperatures. The Illumina TruSeq stranded total RNA library prep kit was

used for library prep. A second quality control check was done on the prepared libraries

using an Agilent Technologies 2100 Bioanalyzer High Sensitivity DNA chip. Paired-end

sequencing was then performed using the Illumina NovaSeq 6000. Raw reads were

obtained and run through an analysis platform. The trimmed fastq files and an analysis

report were obtained from LC Sciences.

Figure 13

Flow Chart of Library Construction for Total RNA Sequencing
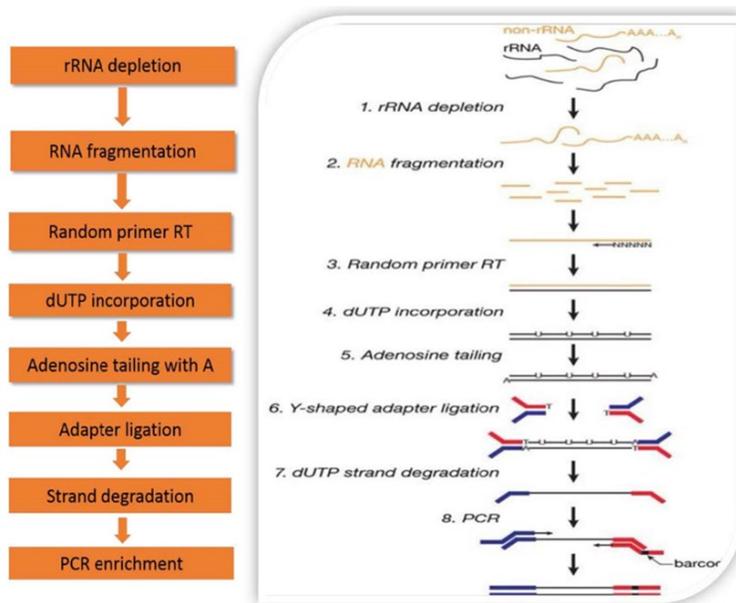
Bioinformatic Analysis of Sequencing Data

The following bioinformatic analysis, which can be observed in Figure 14, was performed by LC Sciences (Houston, TX). Sequencing reads with low quality bases, undetermined bases, and adaptor contamination were removed using Cutadapt and in-house perl scripts (Martin, 2011). FastQC was used to determine sequencing quality (Andrews, 2010). Reads were mapped to the *Rhodobacter sphaeroides* 2.4.1 genome (NCBI RefSeq IDs: NZ_CP030271.1, NZ_CP030272.1, NZ_CP030273.1, NZ_CP030274.1, NZ_CP030275.1, NZ_CP030276.1) using Bowtie2 (Langmead & Salzberg, 2012) and HISAT (Kim et al., 2015) and were assembled using StringTie (Pertea et al., 2015). A comprehensive transcriptome was generated using reads from all 12 samples through perl scripts and gffcompare (https://github.com/gpertea/gffcompare/). Once the final transcriptome was generated, StringTie and edgeR (Robinson et al., 2010) were used to determine expression levels of all transcripts. StringTie was used to

calculate FPKM (fragments per kilobase per million). Differentially expressed genes were determined by edgeR using the following parameters: a log2 fold change greater than 1 or less than -1 and a p-value of less than 0.05 generated by a parametric F-test.

Gene expression analysis was conducted by comparing the following groups: control 24 hours versus control 72 hours, 1.0µM AuCl₃ 24 hours versus 1.0µM AuCl₃ 72 hours, control 24 hours versus 1.0µM AuCl₃ 24 hours, and control 72 hours versus 1.0µM AuCl₃ 72 hours.

Figure 14

Bioinformatic Pipeline of Total RNA Sequencing Data
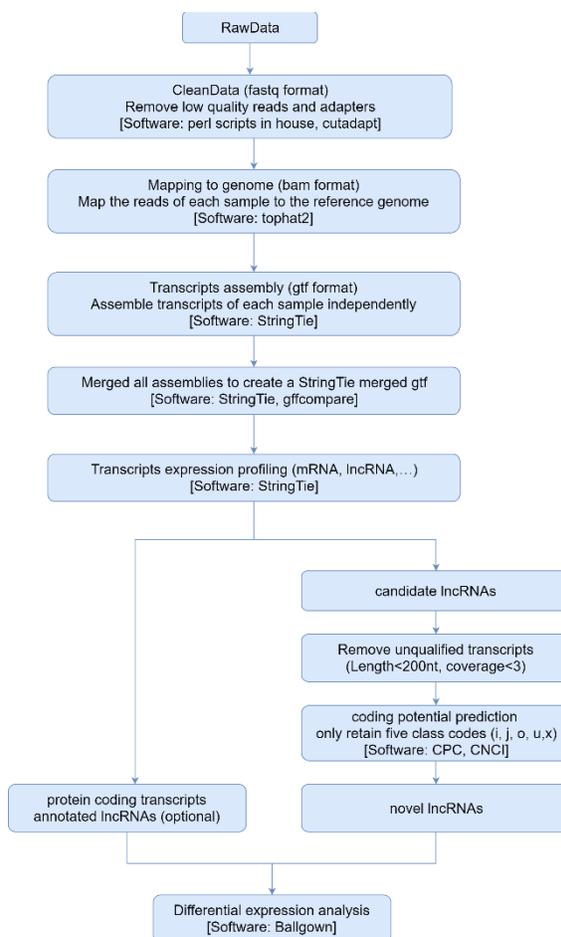
Data Mining and Enrichment Analysis

The data generated from each comparison group was observed and used for further graphical analysis. Genes were re-annotated using MATLAB scripts (*version R2021a*) to provide the common RSP_#### identifiers. Genes with a log2 fold change of less than -1 and greater than 1, and a q-value of less than 0.05 were chosen for graphical representation and enrichment analysis. The q-value indicates an adjusted p-value with the false discovery rate being taken into consideration. Q-values are considered more robust since it removes the multiple hypothesis testing problem observed by RNA-sequencing data (Menyhart et al., 2021) . Genes across all comparison groups which met these criteria were used to generate heatmaps for all samples. The total amount of up- and down-regulated genes within each comparison group was observed, and common genes were observed between the first two comparison groups and the last two comparison groups, as seen in Figure 19.

Differentially expressed genes within the control 24 hours versus 1.0 µM AuCl$_3$ 24 hours and control 72 hours versus 1.0 µM AuCl$_3$ 72 hours groups were further classified using the DAVID annotation web tool (Huang da et al., 2009) and the clusters of orthologous groups (COG) database (Tatusov et al., 2000). For the DAVID annotation tool, a list of the up-regulated and down-regulated genes were analyzed from each comparison group to identify gene enrichment. For gene enrichment, a statistical test is performed to identify whether genes within a dataset assigned to a given term are over-represented when compared to the total number of genes found within the organism that are assigned to the same term (Subramanian et al., 2005). The gene ontology (GO) terms which were found within the GO_direct categories were collected, since the terms within

this category were directly annotated to the annotation source (*R. sphaeroides* genome).

Additionally, KEGG pathways represented in the gene list were also captured.

Parameters used to identify terms represented within each gene list consisted of a

minimum gene count of two and an EASE score of 1.0. The EASE score represents the p-

value generated by a modified Fisher's Exact test and is the method used by DAVID to

identify enriched gene terms (Huang da et al., 2009). An EASE score/p-value of less than

or equal to 0.05 represents a significantly enriched term found within the dataset.

However, by increasing the EASE score to 1.0, all the terms that were annotated

according to the gene list were captured and detected.

For the COG analysis, the up-regulated and down-regulated genes from each

comparison group were annotated with the corresponding COG identification using

previously captured .ptt files for *R. sphaeroides* and MATLAB scripts. The genes were

grouped into the four major categories observed in the COG database, as well as

classified into the 27 subcategories to identify general gene functions.

All heatmaps and bar graphs generated for each individual analysis were created

using MATLAB (*version R2021a*).

Results and Discussion

To assess the quality of the generated sequencing results, a quality control

analysis was performed on the raw data generated for each replicate. The following

results can be observed in Table 5. The Q20% and Q30% represent the Phred scores

generated from the sequencing results. The Phred score represents the probability that a

nucleotide base is called correctly within the sequencing, and therefore a higher

generated Phred score of 30% is considered a good result (Richterich, 1998).

Table 5

Quality Control Statistics of Total RNA Sequencing Data

Sample	Raw Data		Valid Data		Valid Ratio (reads)	Q20%	Q30%	GC content (%)
	Reads	Base	Reads	Base				
C1_24h	66706494	10.01G	57238546	8.59G	85.81	99.89	97.67	65
C1_72h	66593316	9.99G	60020626	9.00G	90.13	99.94	97.78	65
C2_24h	61337420	9.20G	54206714	8.13G	88.37	99.95	97.91	64
C2_72h	63548900	9.53G	56900622	8.54G	89.54	99.94	97.82	64
C3_24h	61456496	9.22G	53735078	8.06G	87.44	99.96	98.10	63
C3_72h	66703756	10.01G	59378726	8.91G	89.02	99.95	97.80	63
E1_24h	63986472	9.60G	58641390	8.80G	91.65	99.95	97.89	65
E1_72h	68479624	10.27G	60350342	9.05G	88.13	99.92	97.54	65
E2_24h	69388202	10.41G	63615216	9.54G	91.68	99.92	97.86	65
E2_72h	66355536	9.95G	57719462	8.66G	86.99	99.95	97.87	63
E3_24h	63351104	9.50G	54988384	8.25G	86.80	99.67	96.30	65
E3_72h	64148476	9.62G	56152278	8.42G	87.53	99.95	97.90	64

Additionally, a Pearson Correlation and a Principle Component Analysis (PCA) were generated to determine the clustering of biological replicates within each group and can be observed in Figures 15 and 16. As previously mentioned, a strong Pearson correlation is represented by an R^2 greater than 0.9 for any two replicates existing within the same sample group (Conesa et al., 2016). As seen in the PCA plot, biological replicates are found to cluster within the same sample group. Additionally, variance is exhibited between the 24-hour samples and the 72-hour samples, which is expected for this study.

Figure 15

Pearson Correlation of Sequencing Data Among All Replicates
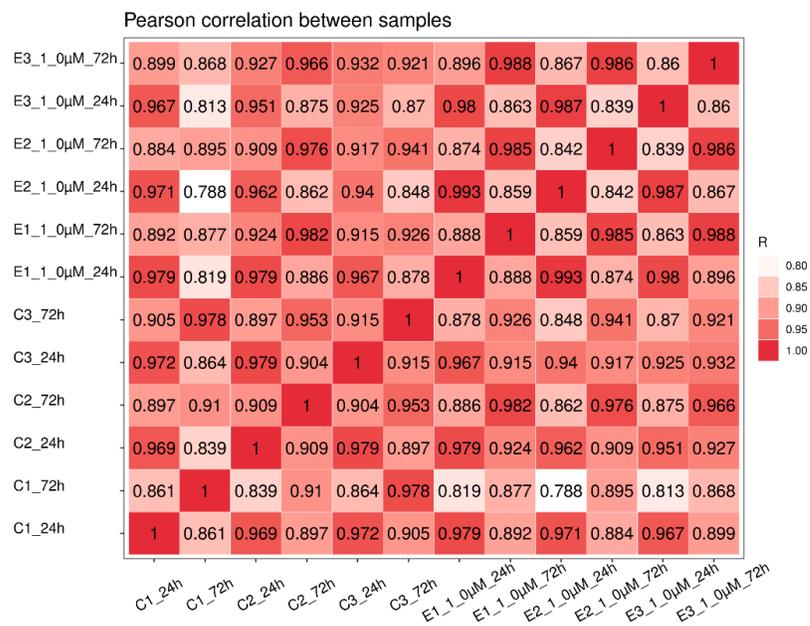
Figure 16

Principle Component Analysis of Sequenced Replicates
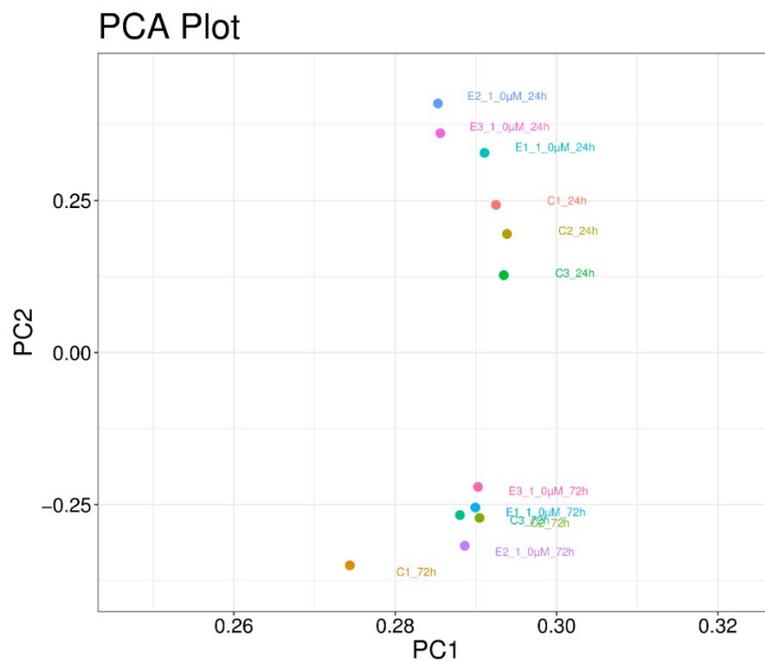
Total Gene Expression Differences Between Different Group Comparisons

An overview of the differentially expressed genes observed between the control and 1.0 µM AuCl₃ groups at 24 and 72 hours is shown in Figures 17 and 18. In Figure 17, all up-regulated genes (in red) and down-regulated genes (in blue) are displayed for each replicate to visualize the quality of the biological replicates within each group. While the replicates found within the control group at 24 and 72 hours have some visible anomalies, an overall pattern of expression is similar, and the expression patterns are correlated within each group.

Figure 17

Differentially Expressed Genes Across All Biological Replicates
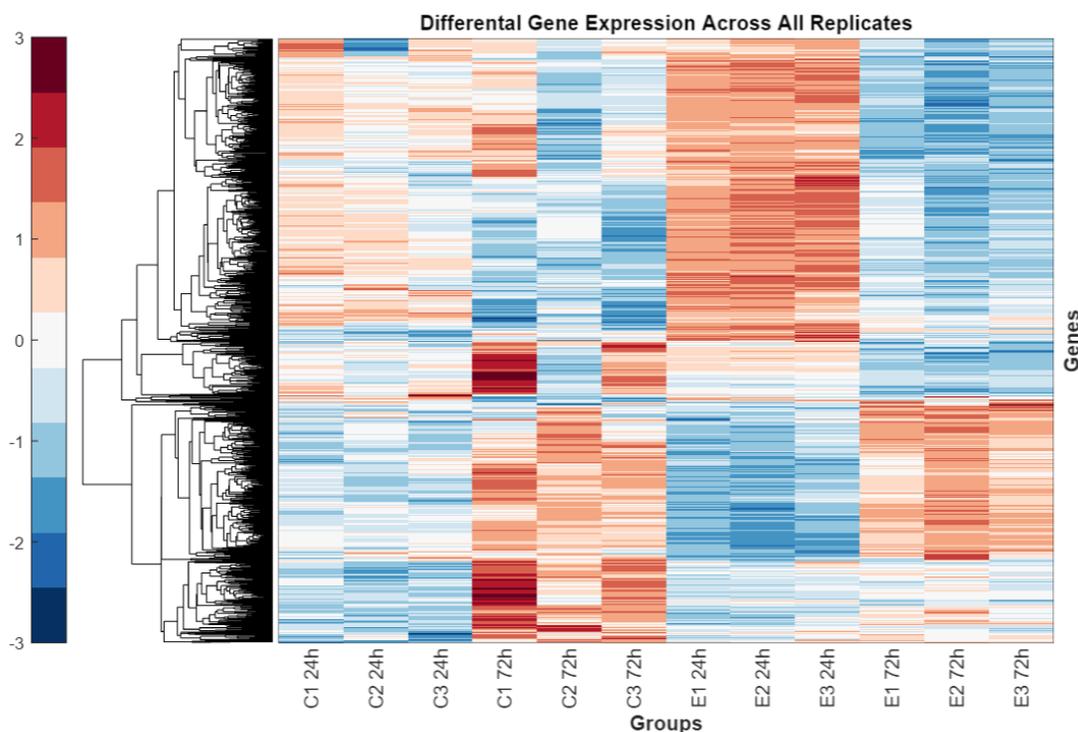
Note. Control groups are represented by the letter C and the 1.0 µM AuCl₃ treated group (or experimental group) is represented by the letter E. The color bar on the left represents a log scale of the expression data, with red indicating up-regulation and blue indicating down-regulation. Groups of genes with similar expression patterns are represented by the clustergram observed on the left of the heatmap.

To minimize the anomalies, the average of the three replicates for each group was shown in a heatmap in Figure 18. From this heatmap, it can be observed that a large group of genes were expressed in similar patterns when comparing the control group (without gold chloride treatment) at 24 and 72 hours to the experimental group (1.0 µM $AuCl_3$) at 24 and 72 hours. These genes which express a similar pattern are predicted to be involved in phase-dependent bacterial growth. However, it can be observed that most genes up-regulated within the control 24 hours group are also up-regulated at a higher magnitude in the 1.0 µM $AuCl_3$ group at 24 hours. This could indicate an effect of the heavy metal stress on the need to overexpress necessary growth-phase dependent genes. While there is a distinct difference pattern of expression between the control and experimental groups, there are some clusters of genes which exhibit unique expression patterns between the control and experimental groups over the observed timepoints. This phenomenon suggests that these genes are responsible for mediating the stress response that the gold chloride exhibits on the bacterial cells.

Figure 18

Averages of Differentially Expressed Genes Across Sample Groups
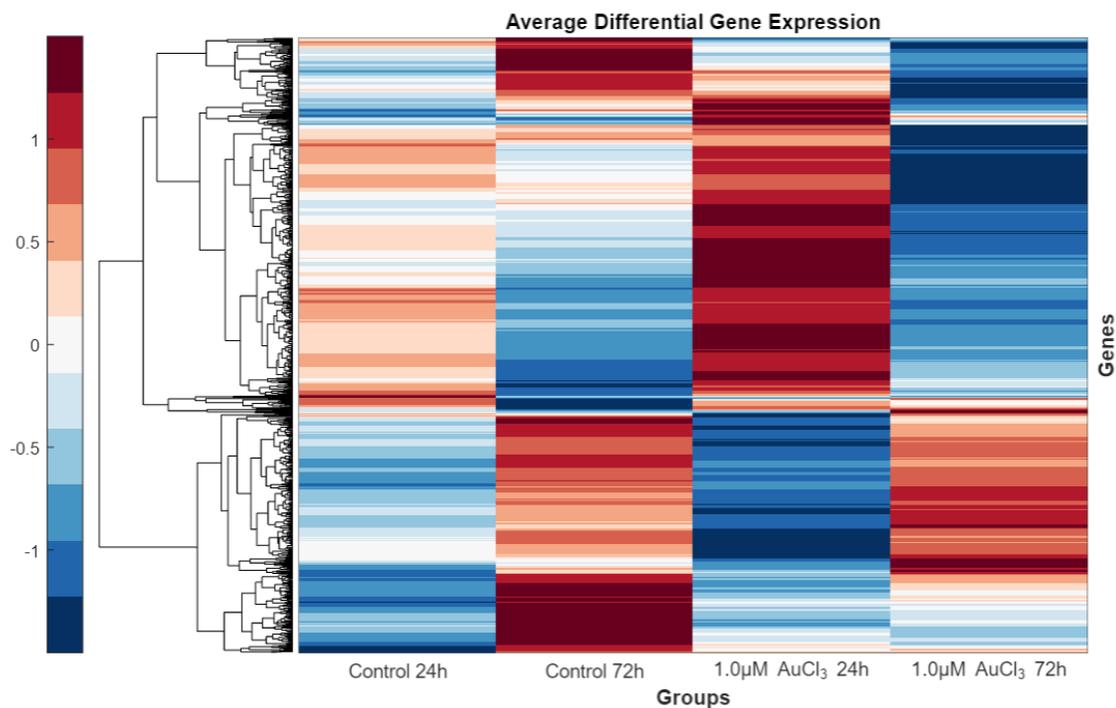
Note. The color bar on the left represents a log scale of the expression data, with red indicating up-regulation and blue indicating down-regulation. Groups of genes with similar expression patterns are represented by the clustergram observed on the left of the heatmap.

To further analyze the differential expression of genes involved in each sample group, a set of four comparisons were observed to identify gene expression differences between different group comparisons. The numbers of up- and down-regulated genes between these groups are shown in Figure 19.

The comparison between control groups at 24 hours and 72 hours reveals differential expression of 300 genes, of which 231 genes were up-regulated and 69 genes were down-regulated. Results suggest that the specific set of genes which are up- or down-regulated between control group at 24 hours (lag and early-log phase) or at 72

hours (late-log and stationary phase) are involved for the growth phase transition from lag/early-log phase to late-log and stationary phase (Zavala et al., 2019).

The second comparison was made between the experimental group (treated with 1.0 μM AuCl$_3$) at 24 hours and 72 hours. This comparison revealed the differential expression of 604 genes; 217 genes were up-regulated, and 387 genes were down-regulated. These genes represent both combined sets of genes which are involved in the growth phase and gold chloride stress responses together. Since both groups aimed to observe the changes in gene expression over the growth phase transition, the differentially expressed genes within each of the two groups were compared to one another, and genes that matched between the two groups were highlighted in green. A total of 162 genes were found to be the same between the two above comparison groups, with the control group exhibiting 121 up-regulated and 41 down-regulated genes, while the experimental (treated) group exhibited 119 up-regulated and 43 down-regulated genes. The identification of genes being expressed within these two groups signifies a possible set of common genes that are strictly growth-phase related. Additionally, it can be observed that the experimental group exhibited a large amount of down-regulation when compared to the control group. This phenomenon may be a result of the adaptation mechanism by *R. sphaeroides* to survive the gold chloride stress condition, however further analysis is needed to determine the types of genes being represented within the down-regulated segment.

Figure 19

Frequency of Differentially Expressed Genes Across Four Comparisons
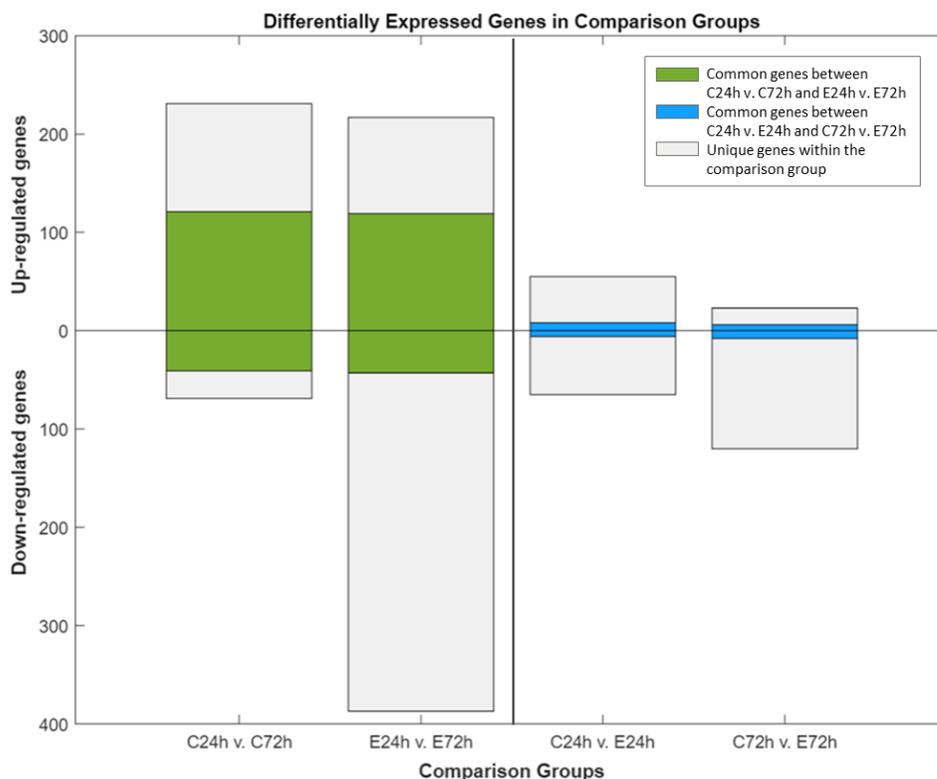
Note. Differentially expressed genes identified in each comparison group. The control group at 24 hours vs. the control group at 72 hours (C24h v. C72h) has a total of 300 differentially expressed genes, with 231 up-regulated genes and 69 down-regulated genes. The 1.0 µM $AuCl_3$ group at 24 hours versus the 1.0 µM $AuCl_3$ group at 72 hours (E24h v. E72h) has a total of 604 differentially expressed genes, with 217 being up-regulated and 377 being down-regulated. The control group at 24 hours versus the 1.0 µM $AuCl_3$ group at 24 hours (C24h v. E24h) has a total of 120 differentially expressed genes, with 55 up-regulated and 65 down-regulated. The control group at 72 hours versus the 1.0 µM $AuCl_3$ group at 72 hours (C72h v. E72h) has a total of 143 differentially expressed genes, with 23 up-regulated and 120 down-regulated. The first two groups were compared, and 162 genes were found common between the two (represented in green). The last two groups were compared, and 14 genes were found in common between the two (represented in blue).

The third comparison was made between the control (untreated) and experimental

(treated) group at the 24-hour timepoint. Results from the comparison demonstrated that

a total of 120 genes were differentially expressed; 59 genes were found to be up-regulated and 65 genes were down-regulated. This group was meant to represent the gene expression changes between the control and treated groups when the bacteria were within the early log-phase of growth at the initial 24-hour incubation period.

The fourth comparison was made between the control group and experimental group at the 72-hour time point. The comparison indicates a total of 143 genes, which are differentially expressed. Within this group, 23 genes were found to be up-regulated while 120 genes were down-regulated. The following comparison was made to represent the gene expression differences between the control and treated group at the later log phase/stationary phase of growth. Like the two groups, a comparison was made between the last two groups to detect the genes that were being maintained over both groups. A total of 14 genes were identified between the two groups (highlighted in blue), indicating these genes may play a major role in mediating bacterial survival under gold chloride stress over time.

Classification of Different COG Gene Functions

To identify types of gene functions of all up- and down-regulated genes for each of the four comparisons shown in Figure 19, a COG analysis was performed using the descriptions of major COG categories and subcategories listed in Table 6. The COG analysis of major categories is shown in Figure 20, while the COG subcategories are observed in Figure 21.

Table 6

COG Subcategories

COG Category	COG Subcategory	Description
Uncharacterized	- / #	Not matched to COG database
Information storage and processing	A	RNA processing and modification
	B	Chromatin structure and dynamics
	J	Translation, ribosomal structure, and biogenesis
	K	Transcription
	L	Replication, recombination, and repair
Cellular processes and signaling	D	Cell cycle control, cell division, chromosome partitioning
	M	Cell wall/membrane/envelope biogenesis
	N	Cell motility
	O	Post-translational modification, protein turnover, and chaperones
	T	Signal transduction mechanisms
	U	Intracellular trafficking, secretion, and vesicular transport
	V	Defense mechanisms
	W	Extracellular structures
	Y	Nuclear structure
	Z	Cytoskeleton
Metabolism	C	Energy production and conversion
	E	Amino acid transport and metabolism
	F	Nucleotide transport and metabolism
	G	Carbohydrate transport and metabolism
	H	Coenzyme transport and metabolism
	I	Lipid transport and metabolism
	P	Inorganic ion transport and metabolism
	Q	Secondary metabolites biosynthesis, transport, and catabolism
Poorly characterized	R	General function prediction only
	S	Function unknown

The first comparison group, which represents the gene expression changes within the control group from 24 hours to 72 hours indicated that an overall 75% of the total differentially expressed genes were up-regulated, and the same trend of increased up-regulation is also observed for all major COG categories and most of the COG subcategories, except for translation, ribosome structure and biogenesis (J), intracellular trafficking, secretion and vesicular transport (U), nucleotide transport and metabolism (F), lipid transport and metabolism (I), and secondary metabolites biosynthesis, transport and catabolism (Q).

The second comparison group, which represents the gene expression changes within the experimental group from 24 hours to 72 hours indicated approximately 70% of the total differentially expressed genes were downregulated, while the remaining 30% were upregulated. Each subcategory had a larger representation of down-regulated genes compared to their up-regulated counterparts except for post-translational modification, protein turnover, and chaperones (O), energy production and conversion (C), and coenzyme transport and metabolism (H). It is speculated that these gene functions may be important for the bacterial resistance and survival under the gold chloride stress. These results suggest that the bacterium selects genes of specific pathways to mitigate the heavy metal stress, and therefore only a few of the selected subcategories are exhibiting more up-regulation compared to the many subcategories exhibiting down-regulation.

Figure 20

COG Major Categories for All Comparison Groups
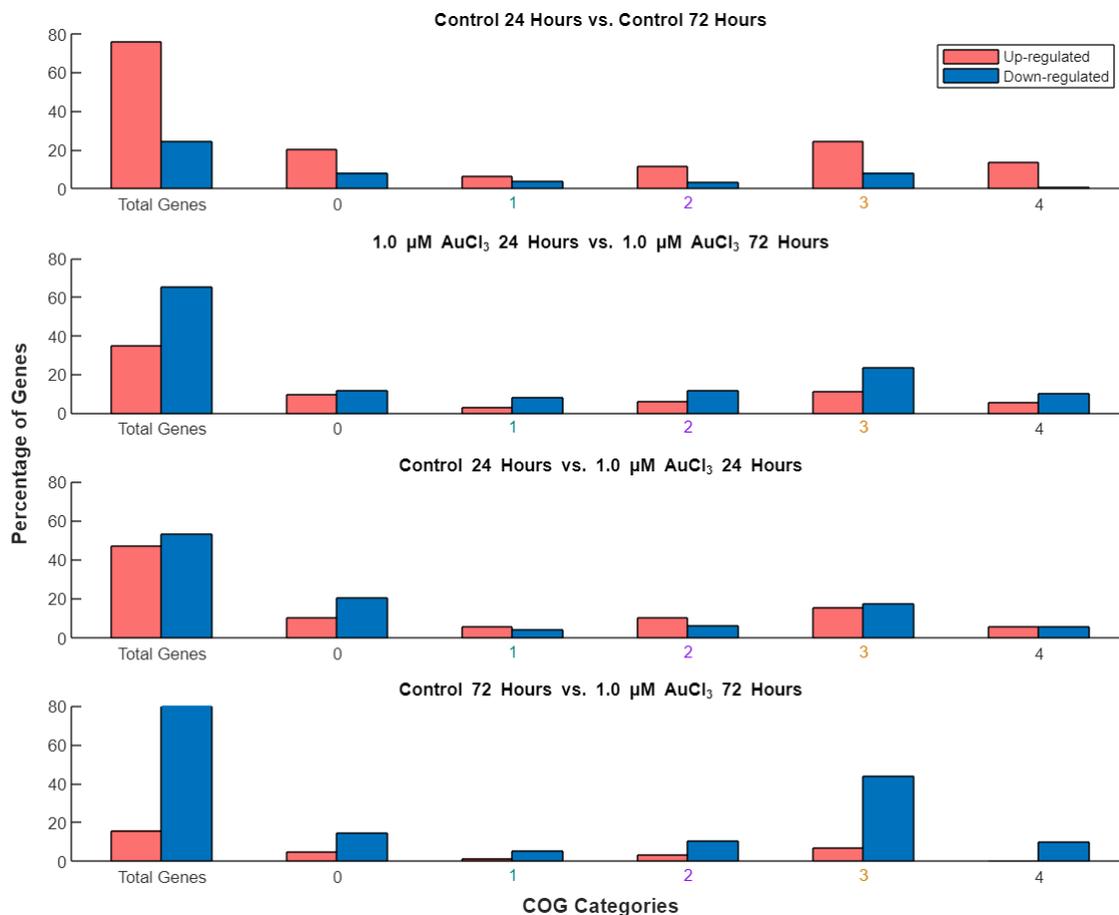
Note. The following graph represents the total percentage of up-regulated and down-regulated genes within each comparison group along with the four major COG categories depicted by each color-coded number. The following descriptions of each group are as follows: 0 – uncategorized, 1 – information storage and processing, 2 – cellular processes and signaling, 3 – metabolism, 4 – poorly characterized.

Figure 21

COG Subcategories for All Comparison Groups
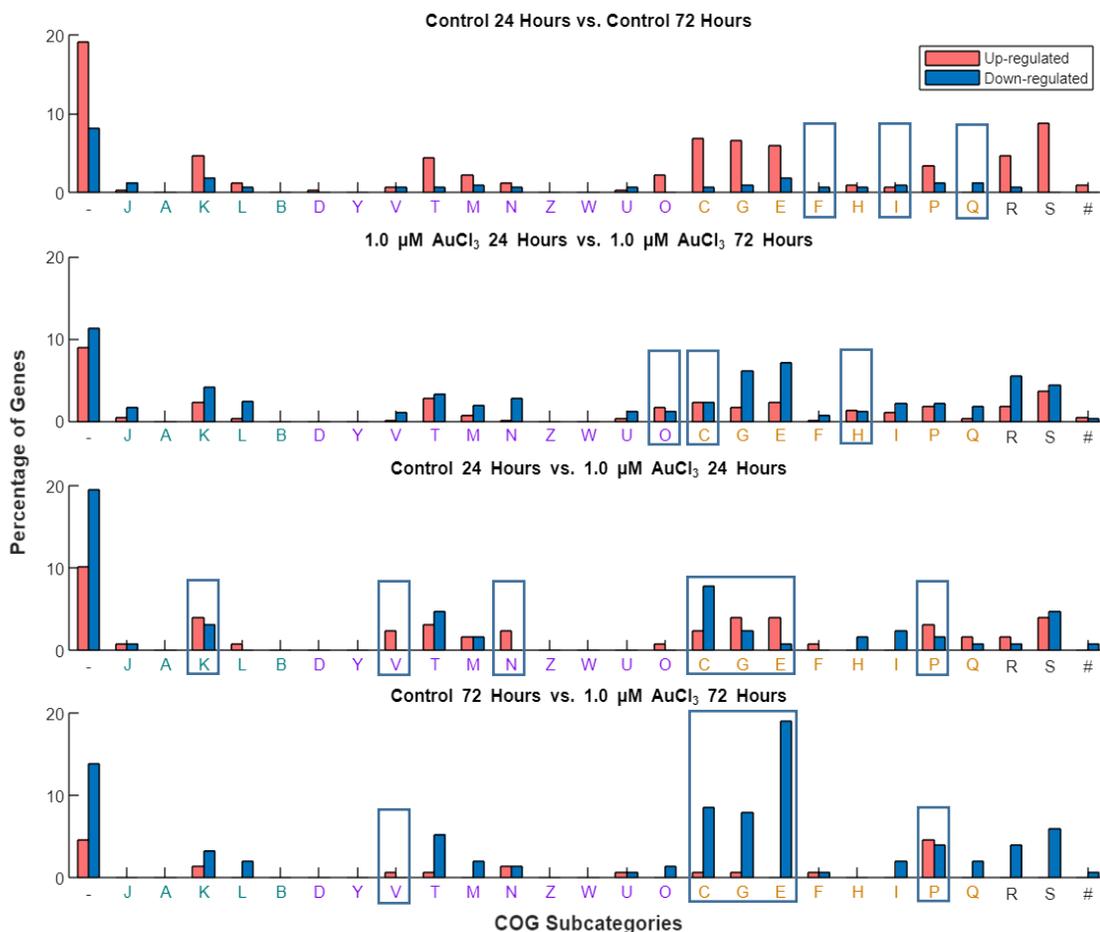
Note. The following graph represents the individual subcategories found within the four major categories observed in Figure 20. The subcategories are color-coded to represent their corresponding major group. The following descriptions of each subcategory can be found in Table 4. Boxes around each subcategory are of interest and discussed within the text.

The third comparison representing gene expression differences between the control 24-hour group and the experimental 24-hour group indicates that ~ 45% genes are up-regulated compared to ~55% genes which are down-regulated. There is a considerably higher number of genes which are up-regulated in two major COG categories:

information storage and processing (1) and cellular processes (2). The upregulation of genes under several subcategories, such as transcription (K), defense mechanisms (V), cell motility (N), carbohydrate transport and metabolism (G), amino acid transport and metabolism (E), and inorganic transport and metabolism (P) were observed. However, down-regulation of genes under subcategories such as signal transduction (T), energy production and conversion (C), coenzyme transport and metabolism (H), and lipid transport and metabolism (I) were observed. This phenomenon may be due to the bacterium's adaptive mechanism towards the stress condition. Since the organism is under stress, it is adjusting to the surrounding environment by utilizing specific types of metabolisms. Additionally, this adaptation process may be influencing the transcription mechanisms occurring within the cell, which can be observed by a slightly higher up-regulation in the transcription subcategory (K). The subcategory representing energy production and conversion (C) shows a large percentage of down-regulated genes, indicating the bacterium's need to conserve energy during the adaptation to the toxic environment. The defense mechanisms (V) and cell motility (N) subcategories consist of only up-regulated genes. A common phenomenon has been observed in previous bacteria where genes regarding defense mechanisms such as antibiotic resistance were also induced when the bacterium was exposed to varying heavy metals (Nguyen et al., 2019). Therefore, it is possible that genes involved in antibiotic defense play a potential role in mitigating heavy metal stress. Additionally, the induction of cellular motility was also found within this analysis, suggesting the gold metal stress as a possible repellent for *R. sphaeroides.* This mechanism of heavy metals as repellents has been previously observed in organisms such as *Escherichia coli* (Tso & Adler, 1974).

The fourth comparison representing gene expression difference between the control 72-hour group and the experimental 72-hour group exhibits overall down regulation of ~80% differentially expressed genes, and the trend remained in all major COG categories and subcategories with exceptions for defense mechanisms (V) and inorganic transport and metabolism (P). When looking at each subcategory, it can be observed that major metabolisms such as carbohydrate (G) and amino acid metabolisms (E) are severely repressed under this condition. Additionally, as seen in the 24-hour timepoint comparison, the energy production and conversion (C) subcategory exhibits a large amount of down-regulation. In contrast, the inorganic ion transport and metabolism (P) subcategory exhibits slightly higher up-regulation, thus indicating the potential role of this metabolism being important for maintaining the bacterium's survival. Additionally, there is still some up-regulation being observed for the defense mechanism (V) subcategory. Although not as prominent when compared to the 24-hour timepoint, this mechanism may be useful in protecting the bacteria from its surrounding environment.

Gene Ontology (GO) Enrichment Within Different Comparison Groups

For the comparison of the control 24-hour and experimental (gold chloride treated) 24-hour groups, a summary of GO terms was detected among the differentially expressed genes for both up-regulated and down-regulated gene sets. The GO terms associated with the up-regulated gene set can be observed in Figure 22A, while the down-regulated gene set is found in Figure 22B.

Figure 22

Gene Ontology (GO) Analysis of Control 24 Hours vs. Experimental 24 Hours – Up-regulated (A) and Down-regulated (B) Genes
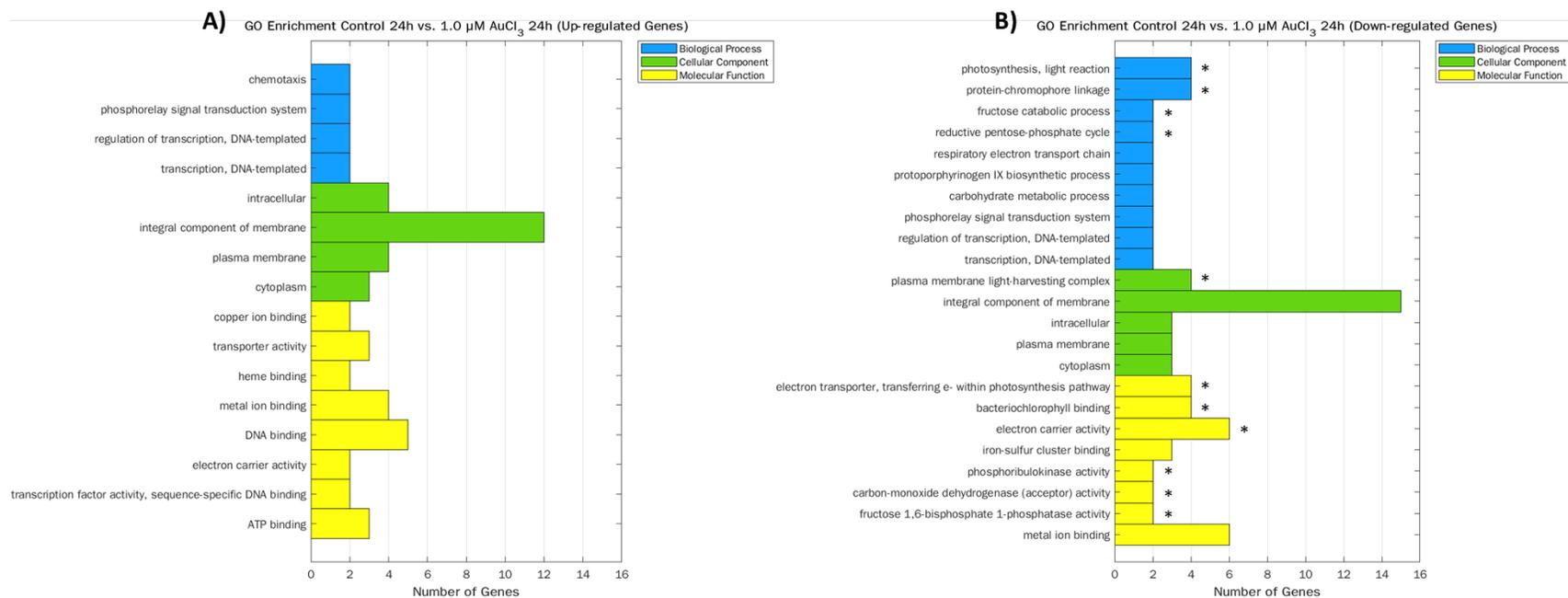
Note. Gene ontology (GO) terms identified in the up-regulated gene set (A) and down-regulated gene set (B) of the control group at 24 hours versus the 1.0 µM $AuCl_3$ group at 24 hours. Enriched terms are noted by an asterisk (*).

In summary, genes were classified into three major categories: biological process, cellular component, and molecular function. For the up-regulated gene set, there were no terms that were deemed enriched by the DAVID annotation tool. However, groups of genes matched to different terms involved in chemotaxis, cellular signaling, membrane components, and various forms of binding activity. The result of this finding may indicate the bacterial response to the initial effect of the gold chloride stress. It is possible that the stress influences the bacterial motility and membrane integrity. A similar mechanism has been observed in *Rhodobacter sphaeroides* R26, where exposure to cobalt and chromate stress significantly impacted the membrane lipidome (Calvano et al., 2014) . Additionally, the various binding activity being detected may be a result of the cellular involvement with the gold chloride ions within the cell (Azam et al., 2012). A chemotactic response was also detected, and thus supports the previous observation that a possible negative chemotaxis is occurring under gold chloride stress.

In the down-regulated gene set, genes which were deemed enriched by the DAVID annotation tool were noted with an asterisk (*) as observed in Figure 22B. Many genes belonging to photosynthetic-related GO terms and enzymes related to metabolism were down-regulated, implying a possible shift in metabolic pathways within the bacterium as it is exposed to the heavy metal stress. Additionally, this down-regulation may be due to the bacterium's need to conserve energy within the cell by repressing any basal level expression of photosynthetic genes and putting that energy towards identifying the best ways possible for surviving in this stress-induced condition.

A similar analysis was conducted on the comparison between the control and experimental group at the 72-hour timepoint to observe any noticeable shifts in gene

functions when compared to the 24-hour timepoint. To start, a summary of GO terms

detected among the differentially expressed genes can be observed in Figure 23. As seen

in the up-regulated gene set (Figure 23A), only a few GO terms were annotated to the

genes and only one term was identified as enriched. These GO terms can also be

observed in the GO analysis for the 24-hour timepoint, and therefore may indicate genes

which are still being utilized by the cell to withstand the heavy metal environment.

Figure 23

Gene Ontology (GO) Analysis of Control 72 Hours vs. Experimental 72 Hours – Up-regulated (A) and Down-regulated (B) Genes
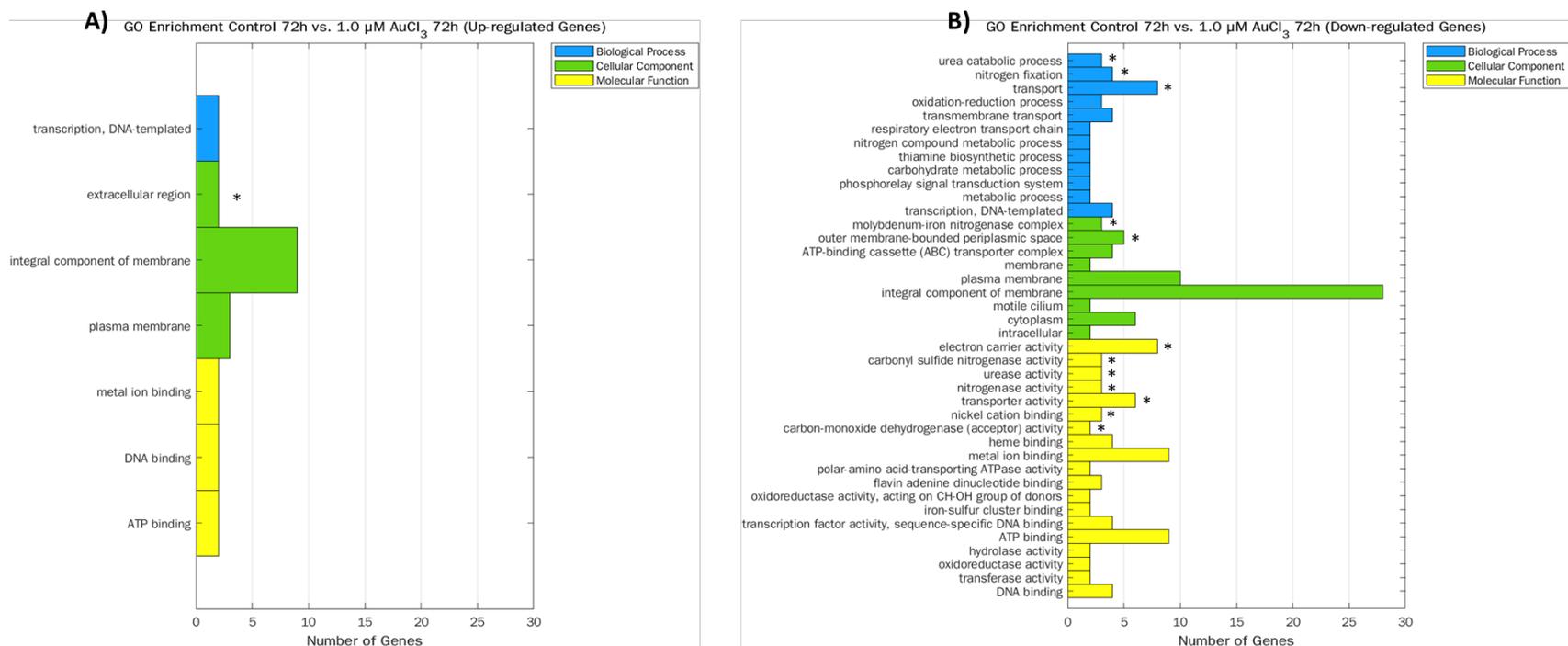
Note. Gene ontology (GO) terms identified in the up-regulated gene set (A) and down-regulated gene set (B) of the control group at 72 hours versus the 1.0 μM $AuCl_3$ group at 72 hours. Enriched terms are noted by an asterisk (*).

In the GO analysis of the down-regulated gene set, a variety of gene functions are observed with respect to various metabolic processes within the cell. Additionally, it can be noted that some of the same GO terms represented in the down-regulated gene set are also found to be represented in the up-regulated gene set. This may imply that only certain genes involved in the classifications like plasma membrane, metal ion binding, and ATP binding are responsible for mediating heavy metal stress compared to the overall list of genes present within this GO term. Nonetheless, a common theme of metabolic down-regulation can be observed within this specific condition, and therefore may support the idea that a selected metabolic pathway is being utilized to allow the bacterium to survive within this condition while other supporting metabolisms are suppressed.

Identification of KEGG Enrichment Annotation

In addition to the GO analysis described above, a KEGG enrichment analysis was performed using the DAVID annotation tool. The following pathways associated with the up-regulated and down-regulated gene sets of the 24-hour (A) and 72-hour (B) timepoints can be observed in Figure 24.

Figure 24

KEGG Analysis of Control 24 Hours vs. Experimental 24 Hours (A) and Control 72 Hours vs. Experimental 72 Hours (B)
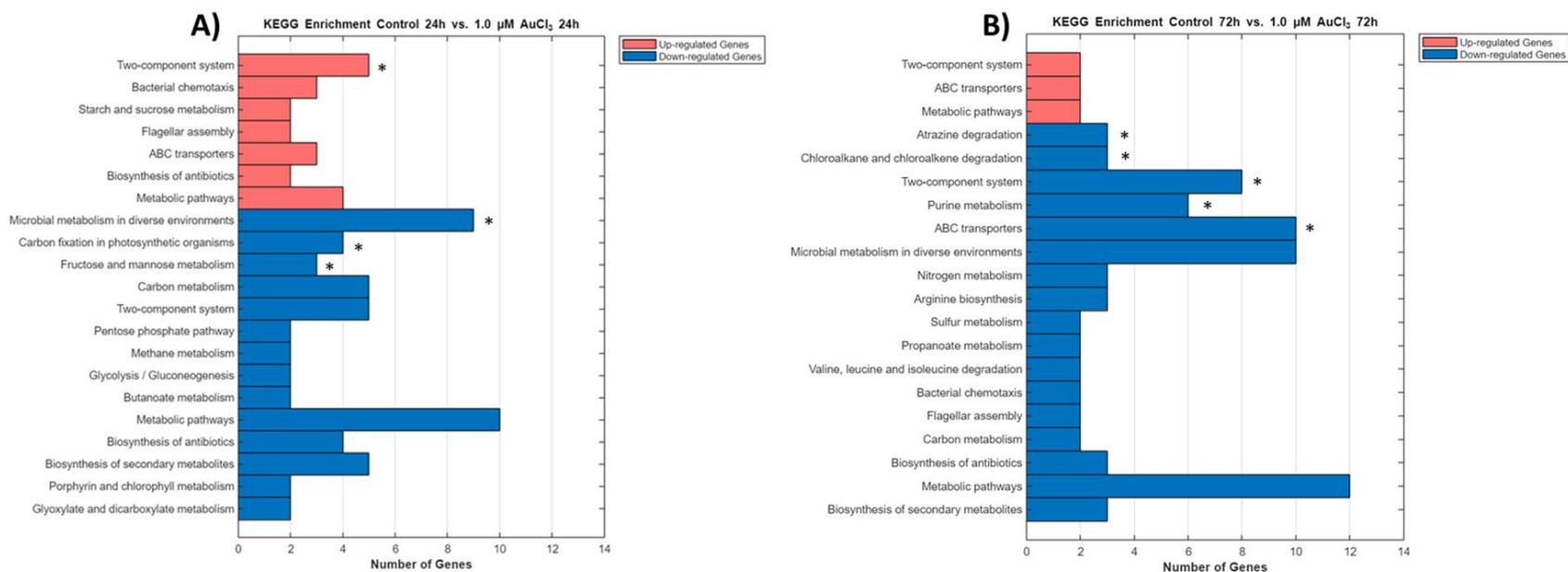
Note. KEGG pathways identified in the up- and down-regulated gene set of the control group at 24 hours versus the 1.0 μM AuCl₃ group at 24 hours (A) and control group at 72 hours versus the 1.0 μM AuCl₃ group at 72 hours (B). Enriched terms are noted by an asterisk (*).

Within the up-regulated gene set of the 24-hour comparison, pathways involving cellular chemotaxis, carbohydrate metabolism, and ABC transporters are represented, with the two-component system pathway being enriched. Additionally, various types of metabolisms including photosynthetic related pathways are found to be downregulated within this comparison group. These findings support the observations made in the GO analysis for this condition. Of particular interest within the KEGG pathways observed in Figure 24A is the presence of ABC transporter activity in the up-regulated genes of this condition. It has previously been observed in *R. sphaeroides* that ABC transporters play a role in the heavy metal tolerance of cobalt ions within the cell (Volpicella et al., 2014). A particular set of ABC transporters were found to be induced upon exposure of the cell to the heavy metal condition and were thus shown to play a role in heavy metal tolerance. Therefore, it is possible that a similar mechanism is occurring within the bacterium's response to the gold chloride stress.

For the 72-hour timepoint, it can be observed that very few pathways are being up-regulated, while a majority are being down-regulated. As previously observed with the corresponding GO analysis, there are some pathways being represented in both up-regulated and down-regulated gene sets which may indicate only a few selected genes from that pathway being induced under the stress condition compared to the total genes represented within the cell. Interestingly, there are still some genes being induced for ABC transporter systems at the 72-hour point. These genes have been previously identified in *R. sphaeroides* response to cobalt ion stress, and therefore may be of importance in this gold chloride stress condition.

Genes Identified in both Comparison Groups Potentially Responsible for Mitigating Heavy Metal Stress

To identify genes which may be heavily involved in heavy metal stress, the differentially expressed genes within the control 24 hours versus experimental 24 hours condition were compared to the control 72 hours versus experimental 72 hours condition. A total of 14 genes were identified in both datasets, thus indicating the possible necessity of these genes in the bacterium's response to surviving in gold chloride stress. The following list of genes can be observed in Table 7.

Within the list of the 14 common genes identified between the two comparison groups, a total of four genes are down-regulated across the 24-hour and 72-hour timepoints. The quinoprotein dehydrogenase associated-SoxYZ-like carrier (RSP_2590) has been extensively studied in the alphaproteobacterium *Paracoccus pantotrophus* and plays a major role in sulfur oxidation within the periplasm of the bacterial cell (Friedrich Cornelius et al., 2001). The three carbon monoxide dehydrogenase subunits (RSP_2876, RSP_2877, and RSP_2878) have been previously characterized as units involved in carbon monoxide metabolism, particularly in oxidizing carbon monoxide (CO) to carbon dioxide (CO_2) (Kerby et al., 1992) and have been shown to exist within a single operon in bacteria such as *Oligotropha carboxidovorans* and *Mycobacterium* sp. strain JC1 (Oh et al., 2010; Schübel et al., 1995). Additionally, the operon encoding the CO dehydrogenase in *Mycobacterium* was found to exhibit catabolite repression when glucose was present in high concentrations within the cell. Therefore, it is likely that a similar mechanism is occurring within *R. sphaeroides*. Since it is speculated that a shift in metabolism is taking place within the cell, it is possible that the bacterium is shutting down certain metabolic

pathways, in this case the carbon monoxide dehydrogenase operon, to utilize other metabolisms within the cell to produce energy.

Within Table 7, the list of genes which exhibit up-regulation at the 24-hour timepoint and down-regulation at the 72-hour timepoint are of interest in understanding the bacterium's response to the heavy metal stress during the early log phase and later in the late log to early stationary phase. The cytochrome b/diheme cytochrome c hybrid protein (RSP_2022) has not been extensively studied in *R. sphaeroides*. However, neighboring genes encoding a diheme cytochrome c protein (RSP_2020) and a sphaeroides heme protein (SHP) have been shown to be involved in an electron transfer pathway which results in the formation of stable oxygen complexes by SHP (Meyer et al., 2010). With an up-regulated occurrence of the RSP_2022 gene, it is possible that the bacterium is utilizing a particular redox pathway to mediate electron transfer within the cell.

The flagellar motor switch protein (FliG) is one of three major proteins involved in the "switch complex" and is responsible for motor rotation (Brown et al., 2007). Previous studies have identified a repression of flagellar and chemotactic signaling genes when a bacterium was placed in a heavy metal environment (Prabhakaran et al., 2016). However, the following is not the case presented in this study. Further analysis on the chemotactic response by *R. sphaeroides* within gold chloride stress is needed to identify whether the upregulation observed for the single protein is representing an increased flagellar activity.

MsrQ, a heme-binding subunit of the MsrPQ enzyme, has been observed to protect bacterial cells from oxidative damage and enhance cell envelope integrity by

utilizing respiratory chain electrons to repair damaged Met-O proteins within the

bacterial cell envelope (Gennaris et al., 2015). The up-regulation observed within the

experimental group at the 24-hour timepoint suggests the need for this enzyme to

mitigate harmful reactive oxygen species. The formation of reactive oxygen species has

been previously observed in microorganisms under different types of heavy metal stress

(Abskharon et al., 2010; Behera et al., 2014), and therefore may exist within *R.

sphaeroides* when under gold chloride stress. Further experimentation identifying

reactive oxygen species within the cell during this stress condition is needed to elucidate

this proposed mechanism.

The putative L,D-transpeptidase (RSP_3073) has been identified as an important

component in peptidoglycan synthesis. Particularly in *Escherichia coli*, L,D-

transpeptidase has been observed to create unusual 3'-3' crosslinks between

peptidoglycan layers and has proven useful in exhibiting beta-lactam resistance

(Hugonnet et al., 2016). Additionally, L,D-transpeptidase exhibits involvement in

strengthening cell wall integrity when the lipopolysaccharide (LPS) transport pathway is

compromised (Morè et al., 2019). Although not previously elucidated in *R. sphaeroides*,

it can be hypothesized that the up-regulation of L,D-transpeptidase is a result of a

potential compromised LPS transport system within the cell due to the presence of heavy

metal stress impacting enzymatic functions. Since most heavy metals are transported into

the cell via diffusion by membrane-associated importers (Tambosi et al., 2018), it is

highly likely that the effects of the heavy metal are influencing peptidoglycan synthesis

in an intracellular fashion.

The last, and perhaps most important set of genes identified in Table 5 are those which were observed to be upregulated across both conditions. These genes represent a selective advantage within the stress response and are of upmost priority for observing the bacterial response to gold chloride. Nucleotide diphosphate kinase (ndk) is a housekeeping enzyme conversed across various species of bacteria and is responsible for a variety of cellular processes. Ndks play a major role in regulating the nucleotide triphosphate (NTP) pool within bacterial cells, as well as exhibit histidine kinase activity and regulation of gene expression (Yu et al., 2017). As previously observed in *Bacillus lichenformis* under cadmium stress, the up-regulation of ndk might be due the bacterium's need to generate more energy in the form of various NTPs (including ATP) in response to the energy loss being exhibited by the stress-induced growth conditions (Sun et al., 2014). In addition to nucleoside diphosphate kinase, the TonB-dependent receptor, copper-translocating P-type ATPase, and siderophore-interacting protein have all been previously observed to play important roles in heavy metal tolerance within bacteria. TonB-dependent receptors are located in the outer membrane regions of bacterial cells and have been observed to transport various substances, including nickel and rare earth metals, into the cell through use of a proton motive force generated by the TonB complex (Ochsner et al., 2019; Schauer et al., 2008). In *Caulobacter cresentus*, the presence of cadmium, uranium, and chromium resulted in induced expression of TonB-dependent receptors but did not show an increase in the expression of the TonB complex. Therefore, it was speculated that the TonB receptors may act as extracellular sensors which interact with the metals in the environment (Hu et al., 2005). Since a similar phenomenon is observed in *R. sphaeroides*, it is possible that the up-regulation of the

TonB-dependent receptor plays a comparable role as an environmental sensor. However, further analysis is needed to determine if metal intake occurs through the interaction of the TonB receptor with the TonB complex under this stress. Siderophore interacting proteins play major roles in iron homeostasis within bacterial cells and are known to exhibit reductase activity on the ferric ions (Fe^{3+}) within the cytosol (Trindade et al., 2019). However, recent studies have shown that siderophores interact with a variety of heavy metals, particularly to mediate heavy metal toxicity and decrease or slow down the intake of toxic metals by sequestering the metals in the extracellular environment (Schalk et al., 2011). While siderophore-associated proteins have not been observed to reduce metals other than iron, it is possible that a similar mechanism may exist since siderophores are involved in cellular uptake of a variety of heavy metals. Lastly, copper-translocating P-type ATPases have been observed in mediating copper tolerance within different types of bacteria and are responsible for the active transport of copper ions from the cytosol to the periplasm (León-Torres et al., 2020; Teitzel et al., 2006). While the function of the up-regulated copper-translocating P-type ATPase has not been observed in *R. sphaeroides*, it is of interest to determine the potential role of this enzyme mediating gold chloride stress. It is important to note that additional genes involved in copper response, such as a copper responsive transcriptional regulator (RSP_2889) and a putative copper chaperone (RSP_2017) exhibited up-regulation within the 24-hour timepoint (data not shown). Therefore, it is highly possible that the mechanism for copper metal tolerance is involved in the tolerance of gold chloride stress. Further experimentation via observation of the bacterium's sensitivity to gold chloride when the

copper-associated genes are mutated will help identify the involvement of these genes in mediating gold chloride stress.

Table 7

Genes Commonly Observed Between the Control vs. Experimental Groups at 24- and 72-Hour Timepoints

Gene ID	Function	Up/Down Regulated in C24vE24	Up/Down Regulated in C72vE72	Targeted by sRNA	GO term	COG Group	References
RSP_2590	quinoprotein dehydrogenase-associated SoxYZ-like carrier	down	down	-	molecular function, cellular component, biological process	Function unknown (S)	(Friedrich Cornelius et al., 2001)
RSP_7577	hypothetical protein	down	up	-	n/a	n/a	n/a
RSP_2022	cytochrome b/diheme cytochrome c hybrid protein	up	down	-	electron transfer activity	Energy production and conversion (C)	(Meyer et al., 2010)
RSP_2220	flagellar motor switch protein (fliG)	up	down	-	cilium or flagellum-dependent cell motility, motor activity, structural molecule activity, chemotaxis, bacterial-type flagellum	Cell motility (N)	(Brown et al., 2007)
RSP_2876	Putative carbon monoxide dehydrogenase small chain	down	down	-	one-carbon metabolic process, carbon-monoxide dehydrogenase (acceptor) activity	Energy production and conversion (C)	(Kerby et al., 1992)

Gene ID	Function	Up/Down Regulated in C24vE24	Up/Down Regulated in C72vE72	Targeted by sRNA	GO term	COG Group	References
RSP_2877	Putative carbon-monoxide dehydrogenase large chain	down	down	-	one-carbon metabolic process, carbon-monoxide dehydrogenase (acceptor) activity	Energy production and conversion (C)	(Kerby et al., 1992)
RSP_2878	Putative carbon-monoxide dehydrogenase small chain	down	down	-	one-carbon metabolic process, carbon-monoxide dehydrogenase (acceptor) activity	Energy production and conversion (C)	(Kerby et al., 1992)
RSP_2894	nucleoside diphosphate kinase	up	up	-	nucleoside diphosphate kinase activity, nucleobase-containing small molecule interconversion	Nucleotide metabolism and transport (F)	(Yu et al., 2017)
RSP_6025	hypothetical protein	down	up	-	n/a	n/a	n/a
RSP_1411	protein-methionine-sulfoxide reductase heme-binding subunit MsrQ	up	down	-	molecular function, cellular component, biological process	Function unknown (S)	(Gennaris et al., 2015)
RSP_3073	Putative L,D-transpeptidase	up	down	-	molecular function, biological process	n/a	(Hugonnet et al., 2016)
RSP_3223	TonB-dependent receptor	up	up	-	cell outer membrane, siderophore uptake transmembrane transporter activity	n/a	(Schauer et al., 2008)

(continued)

Gene ID	Function	Up/Down Regulated in C24vE24	Up/Down Regulated in C72vE72	Targeted by sRNA	GO term	COG Group	References
RSP_2890	copper-translocating P-type ATPase	up	up	-	copper-exporting ATPase activity, copper ion transmembrane transporter activity, copper ion transport, membrane	Inorganic ion transport and metabolism (P)	(Teitzel et al., 2006)
RSP_3678	siderophore-interacting protein	up	up	-	iron assimilation, ferric-chelate reductase (NADPH) activity	Inorganic ion transport and metabolism (P)	(Trindade et al., 2019)

(continued)

When compared to the previously identified sRNAs which showed differential expression within the control versus experimental group comparisons at both 24- and 72-hour timepoints, the sRNAs present within each condition were not found to target genes related to those identified within the list in Table 7. A separate list of genes, which can be observed in Table 8, shows the targets identified by CopraRNA which also matched with differentially expressed genes found within the corresponding condition. This list of genes, along with the list previously mentioned in Table 7, will be used for further molecular analysis to identify each gene's effect on the heavy metal tolerance mechanisms observed by *R. sphaeroides*. While the 14 common genes identified between the two growth conditions are of main interest, it is possible that underlying mechanisms are occurring by the sRNAs regulation of the targeted genes exhibited in Table 8. One main example of this could be explained by the sRNA target RSP_2879. This target encodes a carbon monoxide dehydrogenase subunit (CoxG) and is found to be expressed from the same operon as the previously mentioned carbon monoxide dehydrogenase genes (RSP_2876, RSP_2877, and RSP_2878). Therefore, it is possible that the presence of the corresponding sRNA (PC-3p-14954-58) is directly impacting the expression of *coxG* and might be exhibiting a downstream regulatory effect on the other genes existing within the operon. The other genes listed as targets by sRNA do not have a direct implication in mediating heavy metal stress or influencing genes observed within both conditions. However, since the basis of these findings relies on correlation of bioinformatic target predictions and differentially expressed genes observed in RNA-sequencing, further experimental identification is required to accurately determine the regulatory roles of each sRNA in response to the heavy metal stress environment. It is

very likely that although these genes were found to be targets of each sRNA, the

interactions may not exist *in vivo.* Additionally, it is possible that targets of each sRNA

identified in the following conditions were missed by the CopraRNA prediction program.

Table 8

Differentially Expressed Genes as Targets of sRNA Predicted by CopraRNA

Condition	sRNAs Detected	sRNA up/down-regulated?	Differentially expressed genes in condition detected by CopraRNA		CopraRNA p-value	Gene up/down-regulated?
			Gene ID	Description		
Control 24 hours v. Experimental 24 hours	PC-5p-21399-35	Down	RSP_2507	outer membrane beta-barrel protein (*ompW*)	0.0214	Down
			RSP_3509	hemolysin-type calcium-binding region, RTX (*expE1*)	0.0134	Up
			RSP_2976	hypothetical protein/putative integral membrane protein	0.0235	Up
	PC-3p-14954-58	Up	RSP_2879	carbon monoxide dehydrogenase subunit G (*coxG*)	0.0299	Down
			RSP_2272	N-acetylmuramoyl-L-alanine amidase (*ampD*)	0.0499	Up
Control 72 hours v. Experimental 72 hours	PC-3p-999-2009	Down	RSP_0146	nitrogen regulatory protein P-II (*glnB*)	0.0066	Down
			RSP_0307	antifreeze protein, type I	0.0152	Down
			RSP_1807	DUF1223 domain-containing protein	0.0293	Down

Conclusions and Future Work

Understanding how bacteria regulate gene expression in heavy metal stress-inducing conditions is important to determine the candidacy of bacteria for bioremediatory processes. The use of small, noncoding RNAs (sRNAs) in regulating gene expression adds to the many molecular and cellular processes which exist to help bacteria adapt and survive in stressful situations. The present study aimed to identify sRNAs in the genome of *Rhodobacter sphaeroides* as well as detect the expression of sRNAs in the bacterium grown under gold chloride contaminated media. While over 700 sRNA sequences were predicted within the *R. sphaeroides* genome using a bioinformatic approach, only 24 sRNAs were experimentally detected in this study, with seven of the 24 sRNAs exhibiting differential expression. This is the first study identifying the presence of sRNAs in *R. sphaeroides* under a gold chloride stress condition, implying that sRNAs do play a role in regulating gene expression to help mitigate gold toxicity. Similar observations have been made in bacteria particularly with iron and copper stressors (Chareyre & Mandin, 2018; Maertens et al., 2020), inciting that sRNAs may be important for gene regulation of different heavy metal stressors and may be of further importance for bioremediation. Further investigation of each sRNA identified in this study is needed to determine the regulatory function of these sRNAs in gold chloride stress. Additionally, it is important to note that the identification of novel sRNA sequences is highly dependent on the type of stress-inducing condition being studied. While over 700 sRNAs were predicted in the genome, only a few were captured within this study of gold chloride stress. Therefore, to increase the identification of sRNAs which exists in *R. sphaeroides*, research involving exposure of the bacterium to a wide

range of heavy metal stressors is essential for broadening the list of sRNAs useful for future bioremediation studies.

In addition to the detection of sRNAs, the identification of potential genes involved in the bacterium's response to gold chloride were studied. When observing the impact of gold chloride stress on gene expression, all three gene function analyses (GO, KEGG, and COG) exhibited similar patterns when comparing the control group against the 1.0 µM AuCl$_3$ group at the 24-hour and 72-hour timepoints. At the 24-hour timepoint, genes involved in membrane composition and chemotactic signaling were found to be up-regulated in the gold treated group. These findings have been detected in previous studies, where certain bacteria utilize chemotactic responses to remove themselves from surrounding toxic concentrations of metals (Barrionuevo & Vullo, 2012). Additionally, alterations to the lipid composition of the bacterial membrane has been observed for various heavy metal stressors, indicating an important mechanism of cellular defense to the toxic environment (Markowicz et al., 2010). Only a few metabolic processes were found to be induced such as ABC transporter activity, which has been previously identified as important for transporting different metals in and out of bacterial cells (Ma et al., 2009). The remaining metabolic and energy producing processes were found to be repressed for the gold treated group at 24 hours, indicating a potential shift in energy production under gold chloride stress.

At the 72-hour timepoint, many cellular processes exhibited down-regulation, with only very few categories exhibiting upregulation. Shifts in the overall metabolic nature of the bacterium plays a large role in how the bacterium survives and adapts to the surrounding environment. The large amounts of downregulation exhibited at the 72-hour

timepoint give insight to the adaptive mechanism exhibited by *R. sphaeroides* under gold

chloride stress. Previous studies have observed this same phenomenon when examining

bacterial responses to cell envelope and nutrient deficiency stress (Gottesman, 2017;

Picard et al., 2013). Therefore, it is possible that bacterial adaptation relies on the

downregulation of genes to aid in shifting the metabolism for proper survival in the

stress-inducing condition.

When comparing the list of differentially expressed genes observed at the 24-hour

and 72-hour timepoints, a total of 14 genes were determined to be maintained throughout

the experimental condition and were therefore hypothesized to be involved in the

bacterium's response to gold chloride stress. While sRNAs were not found to target any

of these 14 genes, it is possible that the sRNAs may have an indirect effect on the

expression of the genes, as speculated for the sRNA PC-3p-14954-58. Potential targets of

the differentially expressed sRNAs within the 24-hour and 72-hour timepoints were

determined by comparison of the predicted CopraRNA targets of each sRNA to the list of

differentially expressed genes found within the same condition and may provide further

characterization of the overarching regulatory role of sRNAs in the gold chloride stress.

Therefore, future experimentation using an additional validation method and induced

overexpression is needed to confidently determine the sRNA's involvement in mediating

gold-chloride stress.

In future studies, each sRNA sequence will be subjected to being cloned in an

expression plasmid vector in *Escherichia coli*. The plasmids will then be transferred to *R.*

sphaeroides via conjugation. After the transfer of plasmids, the gene expression patterns

of the genes predicted as targets for the given sRNA sequence will be analyzed using

reverse transcriptase PCR (RT-PCR). By understanding how these sRNAs regulate gene expression, bacteria can be better utilized for bioremediatory processes to help reduce heavy metal toxicity from the environment.

REFERENCES

Abdu, N., Abdullahi, A. A., & Abdulkadir, A. (2017). Heavy metals and soil microbes.
Environmental Chemistry Letters, *15*(1), 65-84. https://doi.org/10.1007/s10311-016-0587-x

Abskharon, R. N. N., Hassan, S. H. A., Kabir, M. H., Qadir, S. A., Gad El-Rab, S. M. F.,
& Wang, M.-H. (2010). The role of antioxidants enzymes of E. coli ASU3, a
tolerant strain to heavy metals toxicity, in combating oxidative stress of copper.
World Journal of Microbiology and Biotechnology, *26*(2), 241-247.
https://doi.org/10.1007/s11274-009-0166-4

Adnan, F., Weber, L., & Klug, G. (2015). The sRNA SorY confers resistance during
photooxidative stress by affecting a metabolite transporter in Rhodobacter
sphaeroides. *RNA biology*, *12*(5), 569-577.
https://doi.org/10.1080/15476286.2015.1031948

Ahmed, W., Hafeez, M. A., & Mahmood, S. (2018). Identification and functional
characterization of bacterial small non-coding RNAs and their target: A review
[Review Article]. *Gene Reports*, *10*, 167-176.
https://doi.org/10.1016/j.genrep.2018.01.001

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local
alignment search tool. *J Mol Biol*, *215*(3), 403-410. https://doi.org/10.1016/s0022-2836(05)80360-2

Anders, L. (2004). *Structural Aspects Of Protein Synthesis* [Book]. World Scientific.

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence
Data [Online]*. In http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Azam, A., Ahmed, A. S., Oves, M., Khan, M. S., & Memic, A. (2012). Size-dependent antimicrobial properties of CuO nanoparticles against Gram-positive and -negative bacterial strains. *International journal of nanomedicine*, *7*, 3527-3535. https://doi.org/10.2147/IJN.S29020

Barrionuevo, M. R., & Vullo, D. L. (2012). Bacterial swimming, swarming and chemotactic response to heavy metal presence: which could be the influence on wastewater biotreatment efficiency? *World Journal of Microbiology and Biotechnology*, *28*(9), 2813-2825. https://doi.org/10.1007/s11274-012-1091-5

Baumgardt, K., Šmídová, K., Rahn, H., Lochnit, G., Robledo, M., & Evguenieva-Hackenberg, E. (2015). The stress-related, rhizobial small RNA RcsR1 destabilizes the autoinducer synthase encoding mRNA sinI in Sinorhizobium meliloti. *RNA biology*, *13*(5), 486-499. https://doi.org/10.1080/15476286.2015.1110673

Behera, M., Dandapat, J., & Rath, C. C. (2014). Effect of heavy metals on growth response and antioxidant defense protection in Bacillus cereus [https://doi.org/10.1002/jobm.201300805]. *Journal of Basic Microbiology*, *54*(11), 1201-1209. https://doi.org/https://doi.org/10.1002/jobm.201300805

Berghoff, B. A., Glaeser, J., Sharma, C. M., Vogel, J., & Klug, G. (2009). Photooxidative stress-induced and abundant small RNAs in Rhodobacter sphaeroides. *Molecular Microbiology*, *74*(6), 1497-1512. https://doi.org/10.1111/j.1365-2958.2009.06949.x

Bervoets, I., & Charlier, D. (2019). Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic

biology. *FEMS Microbiology Reviews*, *43*(3), 304-339. https://doi.org/10.1093/femsre/fuz001

Billenkamp, F., Peng, T., Berghoff, B. A., & Klug, G. (2015). A cluster of four homologous small RNAs modulates C1 metabolism and the pyruvate dehydrogenase complex in Rhodobacter sphaeroides under various stress conditions. *Journal of bacteriology*, *197*(10), 1839-1852. https://doi.org/10.1128/JB.02475-14

Blouin, S., Mulhbacher, J., Penedo, J. C., & Lafontaine, D. A. (2009). Riboswitches: Ancient and Promising Genetic Regulators. *ChemBioChem*, *10*(3), 400-416. https://doi.org/10.1002/cbic.200800593

Bradley, E. S., Bodi, K., Ismail, A. M., & Camilli, A. (2011). A genome-wide approach to discovery of small RNAs involved in regulation of virulence in Vibrio cholerae. *PLoS pathogens*, *7*(7), e1002126-e1002126. https://doi.org/10.1371/journal.ppat.1002126

Brantl, S., & Jahn, N. (2015). sRNAs in bacterial type I and type III toxin-antitoxin systems. *FEMS Microbiology Reviews*, *39*(3), 413-427. https://doi.org/10.1093/femsre/fuv003

Brown, P. N., Terrazas, M., Paul, K., & Blair, D. F. (2007). Mutational analysis of the flagellar protein FliG: sites of interaction with FliM and implications for organization of the switch complex. *Journal of bacteriology*, *189*(2), 305-312. https://doi.org/10.1128/JB.01281-06

Buccolieri, A., Italiano, F., Dell'Atti, A., Buccolieri, G., Giotta, L., Agostiano, A., . . . Trotta, M. (2006). Testing the photosynthetic bacterium Rhodobacter sphaeroides

as heavy metal removal tool. *Ann Chim*, *96*(3-4), 195-203.

https://doi.org/10.1002/adic.200690019

Busch, A., Richter, A. S., & Backofen, R. (2008). IntaRNA: efficient prediction of

bacterial sRNA targets incorporating target site accessibility and seed regions.

Bioinformatics (Oxford, England), *24*(24), 2849-2856.

https://doi.org/10.1093/bioinformatics/btn544

Caldelari, I., Chao, Y., Romby, P., & Vogel, J. (2013). RNA-mediated regulation in

pathogenic bacteria. *Cold Spring Harbor perspectives in medicine*, *3*(9), a010298-

a010298. https://doi.org/10.1101/cshperspect.a010298

Calvano, C. D., Italiano, F., Catucci, L., Agostiano, A., Cataldi, T. R. I., Palmisano, F., &

Trotta, M. (2014). The lipidome of the photosynthetic bacterium Rhodobacter

sphaeroides R26 is affected by cobalt and chromate ions stress. *BioMetals*, *27*(1),

65-73. https://doi.org/10.1007/s10534-013-9687-2

Cavanagh, A. T., & Wassarman, K. M. (2014). 6S RNA, a Global Regulator of

Transcription in Escherichia coli, Bacillus subtilis, and Beyond. *Annual Review of

Microbiology*, *68*, 45.

Cha, J. S., & Cooksey, D. A. (1991). Copper resistance in Pseudomonas syringae

mediated by periplasmic and outer membrane proteins. *Proceedings of the

National Academy of Sciences of the United States of America*, *88*(20), 8915-

8919. https://doi.org/10.1073/pnas.88.20.8915

Chareyre, S., & Mandin, P. (2018). Bacterial Iron Homeostasis Regulation by sRNAs.

Microbiol Spectr, *6*(2). https://doi.org/10.1128/microbiolspec.RWR-0010-2017

Chavarría, M., Nikel, P. I., Pérez-Pantoja, D., & de Lorenzo, V. (2013). The Entner–
Doudoroff pathway empowers Pseudomonas putida KT2440 with a high tolerance
to oxidative stress [https://doi.org/10.1111/1462-2920.12069]. *Environmental
Microbiology*, *15*(6), 1772-1785. https://doi.org/https://doi.org/10.1111/1462-
2920.12069

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A.,
. . . Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis.
Genome Biology, *17*(1), 13. https://doi.org/10.1186/s13059-016-0881-8

Cros, M.-J., de Monte, A., Mariette, J., Bardou, P., Grenier-Boley, B., Gautheret, D., . . .
Gaspin, C. (2011). RNAspace.org: An integrated environment for the prediction,
annotation, and analysis of ncRNA. *RNA (New York, N.Y.)*, *17*(11), 1947-1956.
https://doi.org/10.1261/rna.2844911

Daughney, C. J., Fowle, D. A., & Fortin, D. (2001). The effect of growth phase on proton
and metal adsorption by Bacillus subtilis. *Geochimica et Cosmochimica Acta*,
65(7), 1025-1035. https://doi.org/https://doi.org/10.1016/S0016-7037(00)00587-1

de Lucena, D. K. C., Pühler, A., & Weidner, S. (2010). The role of sigma factor RpoH1
in the pH stress response of Sinorhizobium meliloti. *BMC Microbiology*, *10*(1),
265. https://doi.org/10.1186/1471-2180-10-265

Dhanwal, P., Kumar, A., Dudeja, S., Badgujar, H., Chauhan, R., Kumar, A., . . . Beniwal,
V. (2018). Biosorption of Heavy Metals from Aqueous Solution by Bacteria
Isolated from Contaminated Soil
[https://doi.org/10.2175/106143017X15131012152979]. *Water Environment*

Research, *90*(5), 424-430.

https://doi.org/https://doi.org/10.2175/106143017X15131012152979

Donohue, T. J., Kiley, P. J., & Kaplan, S. (1988). The puf operon region of Rhodobacter

sphaeroides. *Photosynth Res*, *19*(1-2), 39-61. https://doi.org/10.1007/bf00114568

Dufour, Y. S., Imam, S., Koo, B. M., Green, H. A., & Donohue, T. J. (2012).

Convergence of the transcriptional responses to heat shock and singlet oxygen

stresses. *PLoS Genet*, *8*(9), e1002929.

https://doi.org/10.1371/journal.pgen.1002929

Dutta, T., & Srivastava, S. (2018). Small RNA-mediated regulation in bacteria: A

growing palette of diverse mechanisms. *Gene*, *656*, 60-72.

https://doi.org/https://doi.org/10.1016/j.gene.2018.02.068

Eisenhardt, K. M. H., Reuscher, C. M., & Klug, G. (2018). PcrX, an sRNA derived from

the 3′- UTR of the Rhodobacter sphaeroides puf operon modulates expression of

puf genes encoding proteins of the bacterial photosynthetic apparatus [Article].

Molecular Microbiology, *110*(3), 325-334. https://doi.org/10.1111/mmi.14076

Elkina, D., Weber, L., Lechner, M., Burenina, O., Weisert, A., Kubareva, E., . . . Klug, G.

(2017). 6S RNA in Rhodobacter sphaeroides: 6S RNA and pRNA transcript

levels peak in late exponential phase and gene deletion causes a high salt stress

phenotype. *RNA biology*, *14*(11), 1627-1637.

https://doi.org/10.1080/15476286.2017.1342933

Eraso, J. M., & Kaplan, S. (1994). prrA, a putative response regulator involved in oxygen

regulation of photosynthesis gene expression in Rhodobacter sphaeroides. *Journal

of Bacteriology*, *176*(1), 32. https://doi.org/10.1128/jb.176.1.32-43.1994

Errington, J., & Aart, L. T. v. d. (2020). Microbe Profile: Bacillus subtilis: model organism for cellular development, and industrial workhorse. *Microbiology (Reading, England)*, *166*(5), 425-427. https://doi.org/10.1099/mic.0.000922

Faner, M. A., & Feig, A. L. (2013). Identifying and characterizing Hfq-RNA interactions. *Methods (San Diego, Calif.)*, *63*(2), 144-159. https://doi.org/10.1016/j.ymeth.2013.04.023

Feltens, R., Gößringer, M., Willkomm, D. K., Urlaub, H., & Hartmann, R. K. (2003). An unusual mechanism of bacterial gene expression revealed for the RNase P protein of Thermus strains. *Proceedings of the National Academy of Sciences*, *100*(10), 5724. https://doi.org/10.1073/pnas.0931462100

Feng, Y., Yu, Y., Wang, Y., & Lin, X. (2007). Biosorption and Bioreduction of Trivalent Aurum by Photosynthetic Bacteria Rhodobacter capsulatus. *Current Microbiology*, *55*(5), 402-408. https://doi.org/10.1007/s00284-007-9007-6

Francia, F., Wang, J., Zischka, H., Venturoli, G., & Oesterhelt, D. (2002). Role of the N- and C-terminal regions of the PufX protein in the structural organization of the photosynthetic core complex of Rhodobacter sphaeroides. *Eur J Biochem*, *269*(7), 1877-1885. https://doi.org/10.1046/j.1432-1033.2002.02834.x

Friedrich Cornelius, G., Rother, D., Bardischewsky, F., Quentmeier, A., & Fischer, J. (2001). Oxidation of Reduced Inorganic Sulfur Compounds by Bacteria: Emergence of a Common Mechanism? *Applied and Environmental Microbiology*, *67*(7), 2873-2882. https://doi.org/10.1128/AEM.67.7.2873-2882.2001

Gardner, P. P., Daub, J., Tate, J., Moore, B. L., Osuch, I. H., Griffiths-Jones, S., . . .

    Bateman, A. (2011). Rfam: Wikipedia, clans and the "decimal" release. *Nucleic*

    Acids Research, *39*(suppl_1), D141-D145. https://doi.org/10.1093/nar/gkq1129

Gennaris, A., Ezraty, B., Henry, C., Agrebi, R., Vergnes, A., Oheix, E., . . . Barras, F.

    (2015). Repairing oxidized proteins in the bacterial envelope using respiratory

    chain electrons. *Nature*, *528*(7582), 409-412. https://doi.org/10.1038/nature15764

Giotta, L., Agostiano, A., Italiano, F., Milano, F., & Trotta, M. (2006). Heavy metal ion

    influence on the photosynthetic growth of Rhodobacter sphaeroides.

    Chemosphere, *62*(9), 1490-1499.

    https://doi.org/https://doi.org/10.1016/j.chemosphere.2005.06.014

Glaeser, J., Nuss, A. M., Berghoff, B. A., & Klug, G. (2011). Chapter 4 - Singlet Oxygen

    Stress in Microorganisms. In R. K. Poole (Ed.), *Advances in Microbial*

    Physiology (Vol. 58, pp. 141-173). Academic Press.

    https://doi.org/https://doi.org/10.1016/B978-0-12-381043-4.00004-0

Gonin, S., Arnoux, P., Pierru, B., Lavergne, J., Alonso, B., Sabaty, M., & Pignol, D.

    (2007). Crystal structures of an Extracytoplasmic Solute Receptor from a TRAP

    transporter in its open and closed forms reveal a helix-swapped dimer requiring a

    cation for alpha-keto acid binding. *BMC Struct Biol*, *7*, 11.

    https://doi.org/10.1186/1472-6807-7-11

Gottesman, S. (2017). Stress Reduction, Bacterial Style. *Journal of bacteriology*,

    199(20), e00433-00417. https://doi.org/10.1128/JB.00433-17

Gottesman, S., & Storz, G. (2011). Bacterial small RNA regulators: versatile roles and

   rapidly evolving variations. *Cold Spring Harbor perspectives in biology*, *3*(12),

   a003798. https://doi.org/10.1101/cshperspect.a003798

Grenier-Boley, B., Monte, A., & Touzet, H. (2010). CG-seq: a toolbox for automatic

   annotation of genomes by comparative analysis.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., & Eddy, S. R. (2003). Rfam:

   an RNA family database. *Nucleic Acids Research*, *31*(1), 439-441.

   https://doi.org/10.1093/nar/gkg006

Grundy, F. J., & Henkin, T. M. (2006). From Ribosome to Riboswitch: Control of Gene

   Expression in Bacteria by RNA Structural Rearrangements. (6), 329.

Guillet, J., Hallier, M., & Felden, B. (2013). Emerging Functions for the Staphylococcus

   aureus RNome. *PLOS Pathogens*, *9*(12), e1003767.

   https://doi.org/10.1371/journal.ppat.1003767

Heueis, N., Vockenhuber, M.-P., & Suess, B. (2014). Small non-coding RNAs in

   streptomycetes. *RNA biology*, *11*(5), 464-469. https://doi.org/10.4161/rna.28262

Hindley, J. (1967). Fractionation of 32P-labelled ribonucleic acids on polyacrylamide

   gels and their characterization by fingerprinting. *Journal of Molecular Biology*,

   30(1), 125-136. https://doi.org/https://doi.org/10.1016/0022-2836(67)90248-3

Hoe, C.-H., Raabe, C. A., Rozhdestvensky, T. S., & Tang, T.-H. (2013). Bacterial

   sRNAs: Regulation in stress. *International Journal of Medical Microbiology*,

   303(5), 217-229. https://doi.org/https://doi.org/10.1016/j.ijmm.2013.04.002

Horne, I. M., Pemberton, J. M., & McEwan, A. G. (1998). Manganous ions suppress photosynthesis gene expression in Rhodobacter sphaeroides. *Microbiology (Reading)*, *144 (Pt 8)*, 2255-2261. https://doi.org/10.1099/00221287-144-8-2255

Houserova, D., Dahmer, D. J., Amin, S. V., Crucello, A., King, V. M., Barnhill, E. C., . . . Borchert, G. M. (2021). Characterization of 475 novel, putative small RNAs (sRNAs) in Carbon-starved Salmonella enterica serovar typhimurium. *bioRxiv*, 2021.2001.2011.426214. https://doi.org/10.1101/2021.01.11.426214

Hu, P., Brodie, E. L., Suzuki, Y., McAdams, H. H., & Andersen, G. L. (2005). Whole-genome transcriptional analysis of heavy metal stresses in Caulobacter crescentus. *Journal of bacteriology*, *187*(24), 8437-8449. https://doi.org/10.1128/JB.187.24.8437-8449.2005

Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, *4*(1), 44-57. https://doi.org/10.1038/nprot.2008.211

Hugonnet, J.-E., Mengin-Lecreulx, D., Monton, A., den Blaauwen, T., Carbonnelle, E., Veckerlé, C., . . . Arthur, M. (2016). Factors essential for L,D-transpeptidase-mediated peptidoglycan cross-linking and β-lactam resistance in Escherichia coli. *eLife*, *5*, e19469. https://doi.org/10.7554/eLife.19469

Igarashi, K., & Kashiwagi, K. (1999). Polyamine transport in bacteria and yeast. *Biochem J*, *344 Pt 3*(Pt 3), 633-642.

Italiano, F., Agostiano, A., Belviso, B. D., Caliandro, R., Carrozzini, B., Comparelli, R., . . . Trotta, M. (2018). Interaction between the photosynthetic anoxygenic

microorganism Rhodobacter sphaeroides and soluble gold compounds. From toxicity to gold nanoparticle synthesis. *Colloids and Surfaces B: Biointerfaces*, *172*, 362-371. https://doi.org/https://doi.org/10.1016/j.colsurfb.2018.06.010

Italiano, F., Rinalducci, S., Agostiano, A., Zolla, L., De Leo, F., Ceci, L. R., & Trotta, M. (2012). Changes in morphology, cell wall composition and soluble proteome in Rhodobacter sphaeroides cells exposed to chromate. *BioMetals*, *25*(5), 939-949. https://doi.org/10.1007/s10534-012-9561-7

Johnson, H., Kafle, R. C., & Choudhary, M. (2017). Cellular Localization of Gold and Mechanisms of Gold Resistance in Rhodobacter sphaeroides. *Advances in Microbiology*, *7*, 602-616. https://doi.org/https://doi.org/10.4236/aim.2017.78047

Järup, L. (2003). Hazards of heavy metal contamination. *British Medical Bulletin*, *68*(1), 167-182. https://doi.org/10.1093/bmb/ldg032

Jørgensen, M. G., Pettersen, J. S., & Kallipolitis, B. H. (2020). sRNA-mediated control in bacteria: An increasing diversity of regulatory mechanisms. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1863*(5), 194504. https://doi.org/https://doi.org/10.1016/j.bbagrm.2020.194504

Kerby, R. L., Hong, S. S., Ensign, S. A., Coppoc, L. J., Ludden, P. W., & Roberts, G. P. (1992). Genetic and physiological characterization of the Rhodospirillum rubrum carbon monoxide dehydrogenase system. *Journal of bacteriology*, *174*(16), 5284-5294. https://doi.org/10.1128/jb.174.16.5284-5294.1992

Kery, M. B., Feldman, M., Livny, J., & Tjaden, B. (2014). TargetRNA2: identifying targets of small regulatory RNAs in bacteria. In (Vol. 42, pp. W124-W129).

Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C.,
Caspi, R., . . . Karp, P. D. (2017). The EcoCyc database: reflecting new
knowledge about Escherichia coli K-12. *Nucleic Acids Research*, *45*(D1), D543-
D550. https://doi.org/10.1093/nar/gkw1003

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low
memory requirements. *Nat Methods*, *12*(4), 357-360.
https://doi.org/10.1038/nmeth.3317

Kopf, M., & Hess, W. R. (2015). Regulatory RNAs in photosynthetic cyanobacteria.
FEMS microbiology reviews, *39*(3), 301-315.
https://doi.org/10.1093/femsre/fuv017

Kröger, C., Dillon, S. C., Cameron, A. D. S., Papenfort, K., Sivasankaran, S. K.,
Hokamp, K., . . . Hinton, J. C. D. (2012). The transcriptional landscape and small
RNAs of Salmonella enterica serovar Typhimurium. *Proceedings of
the National Academy of Sciences*, *109*(20), E1277.
https://doi.org/10.1073/pnas.1201061109

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat
Methods*, *9*(4), 357-359. https://doi.org/10.1038/nmeth.1923

León-Torres, A., Arango, E., Castillo, E., & Soto, C. Y. (2020). CtpB is a plasma
membrane copper (I) transporting P-type ATPase of Mycobacterium tuberculosis.
Biological Research, *53*(1), 6. https://doi.org/10.1186/s40659-020-00274-7

Li, K., Hein, S., Zou, W., & Klug, G. (2004). The glutathione-glutaredoxin system in
Rhodobacter capsulatus: part of a complex regulatory network controlling defense

against oxidative stress. *Journal of bacteriology*, *186*(20), 6800-6808. https://doi.org/10.1128/JB.186.20.6800-6808.2004

Li, L., Huang, D., Cheung, M. K., Nong, W., Huang, Q., & Kwan, H. S. (2013). BSRD: a repository for bacterial small regulatory RNA. *Nucleic acids research*, *41*(Database issue), D233-D238. https://doi.org/10.1093/nar/gks1264

Li, W., Ying, X., Lu, Q., & Chen, L. (2012). Predicting sRNAs and Their Targets in Bacteria. *Genomics, Proteomics & Bioinformatics*, *10*(5), 276-284. https://doi.org/https://doi.org/10.1016/j.gpb.2012.09.004

Li, X., Shahid, M. Q., Wu, J., Wang, L., Liu, X., & Lu, Y. (2016). Comparative Small RNA Analysis of Pollen Development in Autotetraploid and Diploid Rice. *Int J Mol Sci*, *17*(4), 499. https://doi.org/10.3390/ijms17040499

Liu, M. Y., Gui, G., Wei, B., Preston, J. F., 3rd, Oakford, L., Yüksel, U., . . . Romeo, T. (1997). The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli. *J Biol Chem*, *272*(28), 17502-17510. https://doi.org/10.1074/jbc.272.28.17502

Lott, S. C., Schäfer, R. A., Mann, M., Backofen, R., Hess, W. R., Voß, B., & Georg, J. (2018). GLASSgo - Automated and Reliable Detection of sRNA Homologs From a Single Input Sequence. *Front Genet*, *9*, 124. https://doi.org/10.3389/fgene.2018.00124

Lu, M., Jiao, S., Gao, E., Song, X., Li, Z., Hao, X., . . . Wei, G. (2017). Transcriptome Response to Heavy Metals in Sinorhizobium meliloti CCNWSX0020 Reveals New Metal Resistance Determinants That Also Promote Bioremediation by

Medicago lupulina in Metal-Contaminated Soil. *Appl Environ Microbiol*, *83*(20). https://doi.org/10.1128/aem.01244-17

Ma, Z., Jacobsen, F. E., & Giedroc, D. P. (2009). Coordination chemistry of bacterial metal transport and sensing. *Chemical reviews*, *109*(10), 4644-4681. https://doi.org/10.1021/cr900077w

Mackenzie, C., Eraso, J. M., Choudhary, M., Roh, J. H., Zeng, X., Bruscella, P., . . . Kaplan, S. (2007). Postgenomic Adventures with Rhodobacter sphaeroides. 283.

Maertens, L., Leys, N., Matroule, J.-Y., & Van Houdt, R. (2020). The Transcriptomic Landscape of Cupriavidus metallidurans CH34 Acutely Exposed to Copper. *Genes*, *11*(9). https://doi.org/10.3390/genes11091049

Malabirade, A., Habier, J., Heintz-Buschart, A., May, P., Godet, J., Halder, R., . . . Fritz, J. V. (2018). The RNA Complement of Outer Membrane Vesicles From Salmonella enterica Serovar Typhimurium Under Distinct Culture Conditions [10.3389/fmicb.2018.02015]. *Frontiers in Microbiology*, *9*, 2015.

Mank, N. N., Berghoff, B. A., Hermanns, Y. N., & Klug, G. (2012). Regulation of bacterial photosynthesis genes by the small noncoding RNA PcrZ. *Proceedings of the National Academy of Sciences*, *109*(40), 16306. https://doi.org/10.1073/pnas.1207067109

Mann, B., van Opijnen, T., Wang, J., Obert, C., Wang, Y.-D., Carter, R., . . . Rosch, J. W. (2012). Control of virulence by small RNAs in Streptococcus pneumoniae. *PLoS pathogens*, *8*(7), e1002788-e1002788. https://doi.org/10.1371/journal.ppat.1002788

Mann, M., Wright, P. R., & Backofen, R. (2017). IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. In (Vol. 45, pp. W435-W439).

Markowicz, A., Płociniczak, T., & Piotrowska-Seget, Z. (2010). Response of Bacteria to Heavy Metals Measured as Changes in FAME Profiles. *Polish Journal of Environmental Studies*, *19*(5), 957-965.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal; Vol 17, No 1: Next Generation Sequencing Data Analysis*. https://doi.org/10.14806/ej.17.1.200

Martínez-Salazar, J. M., Sandoval-Calderón, M., Guo, X., Castillo-Ramírez, S., Reyes, A., Loza, M. G., . . . Ramírez-Romero, M. A. (2009). The Rhizobium etli RpoH1 and RpoH2 sigma factors are involved in different stress responses. *Microbiology (Reading)*, *155*(Pt 2), 386-397. https://doi.org/10.1099/mic.0.021428-0

Mattheis, A., Eggenhofer, F., Wolff, J., Raden, M., Miladi, M., Mohamed, M. M., . . . Kundu, K. (2018). Freiburg RNA tools: a central online resource for RNA-focused research and teaching. *Nucleic Acids Research*, *46*(W1), W25-W29. https://doi.org/10.1093/nar/gky329

Melamed, S., Faigenbaum-Romm, R., Peer, A., Reiss, N., Shechter, O., Bar, A., . . . Margalit, H. (2018). Mapping the small RNA interactome in bacteria using RIL-seq. *Nat Protoc*, *13*(1), 1-33. https://doi.org/10.1038/nprot.2017.115

Mendel, J. (2019). *sRNA Target Prediction Tool Choice: A Biological Perspective*. [Unpublished manuscript]. Sam Houston State University.

Menyhart, O., Weltz, B., & Győrffy, B. (2021). MultipleTesting.com: A tool for life science researchers for multiple hypothesis testing correction. *PLOS ONE*, *16*(6), e0245824. https://doi.org/10.1371/journal.pone.0245824

Meyer, T. E., Kyndt, J. A., & Cusanovich, M. A. (2010). Occurrence and sequence of Sphaeroides Heme Protein and diheme cytochrome C in purple photosynthetic bacteria in the family Rhodobacteraceae. *BMC Biochem*, *11*, 24. https://doi.org/10.1186/1471-2091-11-24

Mingorance, J., Tamames, J., & Vicente, M. (2004). Genomic channeling in bacterial cell division. *J Mol Recognit*, *17*(5), 481-487. https://doi.org/10.1002/jmr.718

Mizuno, T., Chou, M.-Y., & Inouye, M. (1983). Regulation of Gene Expression by a Small RNA Transcript (micRNA) in Escherichia coli K-12. *Proceedings of the Japan Academy, Series B*, *59*(10), 335-338. https://doi.org/10.2183/pjab.59.335

Moore, M. D., & Kaplan, S. (1992). Identification of intrinsic high-level resistance to rare-earth oxides and oxyanions in members of the class Proteobacteria: characterization of tellurite, selenite, and rhodium sesquioxide reduction in Rhodobacter sphaeroides. *J Bacteriol*, *174*(5), 1505-1514. https://doi.org/10.1128/jb.174.5.1505-1514.1992

Morè, N., Martorana, A. M., Biboy, J., Otten, C., Winkle, M., Serrano, C. K. G., . . . Polissi, A. (2019). Peptidoglycan Remodeling Enables Escherichia coli To Survive Severe Outer Membrane Assembly Defect. *mBio*, *10*(1). https://doi.org/10.1128/mBio.02729-18

Müller, K. M. H., Berghoff, B. A., Eisenhardt, B. D., Remes, B., & Klug, G. (2016). Characteristics of Pos19 - A Small Coding RNA in the Oxidative Stress Response

of Rhodobacter sphaeroides. *PloS one*, *11*(9), e0163425-e0163425. https://doi.org/10.1371/journal.pone.0163425

Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C., & Jovine, L. (2003). Structure, function and evolution of the signal recognition particle. *The EMBO journal*, *22*(14), 3479-3485. https://doi.org/10.1093/emboj/cdg337

Naskulwar, K., & Peña-Castillo, L. (2021). sRNARFTarget: A fast machine-learning-based approach for transcriptome-wide sRNA Target Prediction. *bioRxiv*, 2021.2003.2005.433963. https://doi.org/10.1101/2021.03.05.433963

Nawrocki, E. P., Kolbe, D. L., & Eddy, S. R. (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics*, *25*(10), 1335-1337. https://doi.org/10.1093/bioinformatics/btp157

Nepple, B. B., Kessi, J., & Bachofen, R. (2000). Chromate reduction by Rhodobacter sphaeroides. *Journal of Industrial Microbiology and Biotechnology*, *25*(4), 198-203. https://doi.org/10.1038/sj.jim.7000049

Nguyen, C. C., Hugie, C. N., Kile, M. L., & Navab-Daneshmand, T. (2019). Association between heavy metals and antibiotic-resistant human pathogens in environmental reservoirs: A review. *Frontiers of Environmental Science & Engineering*, *13*(3), 46. https://doi.org/10.1007/s11783-019-1129-0

Nies, D. H. (1999). Microbial heavy-metal resistance. *Applied Microbiology and Biotechnology*, *51*(6), 730-750. https://doi.org/10.1007/s002530051457

Nuss, A. M., Glaeser, J., & Klug, G. (2009). RpoH(II) activates oxidative-stress defense systems and is controlled by RpoE in the singlet oxygen-dependent response in

Rhodobacter sphaeroides. *Journal of bacteriology*, *191*(1), 220-230. https://doi.org/10.1128/JB.00925-08

Ochsner, A. M., Hemmerle, L., Vonderach, T., Nüssli, R., Bortfeld-Miller, M., Hattendorf, B., & Vorholt, J. A. (2019). Use of rare-earth elements in the phyllosphere colonizer Methylobacterium extorquens PA1. *Mol Microbiol*, *111*(5), 1152-1166. https://doi.org/10.1111/mmi.14208

Oh, J.-I., Park, S.-J., Shin, S.-J., Ko, I.-J., Han Seung, J., Park Sae, W., . . . Kim Young, M. (2010). Identification of trans- and cis-Control Elements Involved in Regulation of the Carbon Monoxide Dehydrogenase Genes in Mycobacterium sp. Strain JC1 DSM 3803. *Journal of Bacteriology*, *192*(15), 3925-3933. https://doi.org/10.1128/JB.00286-10

Padalon-Brauch, G., Hershberg, R., Elgrably-Weiss, M., Baruch, K., Rosenshine, I., Margalit, H., & Altuvia, S. (2008). Small RNAs encoded within genetic islands of Salmonella typhimurium show host-induced expression and role in virulence. *Nucleic acids research*, *36*(6), 1913-1927. https://doi.org/10.1093/nar/gkn050

Pain, A., Ott, A., Amine, H., Rochat, T., Bouloc, P., & Gautheret, D. (2015). An assessment of bacterial small RNA target prediction programs. *RNA biology*, *12*(5), 509-513. https://doi.org/10.1080/15476286.2015.1020269

Papenfort, K., & Vogel, J. (2010). Regulatory RNA in Bacterial Pathogens. *Cell Host & Microbe*, *8*(1), 116-127. https://doi.org/https://doi.org/10.1016/j.chom.2010.06.008

Pavlova, N., Kaloudas, D., & Penchovsky, R. (2019). Riboswitch distribution, structure, and function in bacteria. *Gene*, *708*, 38-48. https://doi.org/https://doi.org/10.1016/j.gene.2019.05.036

Peluso, P., Herschlag, D., Nock, S., Freymann, D. M., Johnson, A. E., & Walter, P. (2000). Role of 4.5S RNA in Assembly of the Bacterial Signal Recognition Particle with Its Receptor. *Science*, *288*(5471), 1640-1643.

Peng, T., Berghoff, B. A., Oh, J.-I., Weber, L., Schirmer, J., Schwarz, J., . . . Klug, G. (2016). Regulation of a polyamine transporter by the conserved 3' UTR-derived sRNA SorX confers resistance to singlet oxygen and organic hydroperoxides in Rhodobacter sphaeroides. *RNA biology*, *13*(10), 988-999. https://doi.org/10.1080/15476286.2016.1212152

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*, *33*(3), 290-295. https://doi.org/10.1038/nbt.3122

Picard, F., Loubière, P., Girbal, L., & Cocaign-Bousquet, M. (2013). The significance of translation regulation in the stress response. *BMC Genomics*, *14*(1), 588. https://doi.org/10.1186/1471-2164-14-588

Pichon, C., & Felden, B. (2008). Small RNA gene identification and mRNA target predictions in bacteria. *Bioinformatics*, *24*(24), 2807-2813. https://doi.org/10.1093/bioinformatics/btn560

Pougach, K., Lopatina, A., & Severinov, K. (2012). CRISPR adaptive immunity systems of prokaryotes [Article]. *Molecular Biology*, *46*(2), 175-182. https://doi.org/10.1134/S0026893312020136

Prabhakaran, P., Ashraf, M., & Wan Mohd Noor, W. S. A. (2016). Microbial Stress Response to Heavy Metal in the Environment. *RSC Adv.*, *6*. https://doi.org/10.1039/C6RA10966G

Quendera, A. P., Seixas, A. F., dos Santos, R. F., Santos, I., Silva, J. P. N., Arraiano, C. M., & Andrade, J. M. (2020). RNA-Binding Proteins Driving the Regulatory Activity of Small Non-coding RNAs in Bacteria [10.3389/fmolb.2020.00078]. *Frontiers in Molecular Biosciences*, *7*, 78.

Ramasamy, K., Kamaludeen, & Banu, S. P. (2007). Bioremediation of Metals: Microbial Processes and Techniques. In S. N. Singh & R. D. Tripathi (Eds.), *Environmental Bioremediation Technologies* (pp. 173-187). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-34793-4_7

Rathoure, A. K., & Srivastava, M. (2016). *Cell Biology and Genetics* [Book]. Daya Publishing House.

Rau, M. H., Bojanovič, K., Nielsen, A. T., & Long, K. S. (2015). Differential expression of small RNAs under chemical stress and fed-batch fermentation in E. coli. *BMC Genomics*, *16*, 1051. https://doi.org/10.1186/s12864-015-2231-8

Reinbothe, C., Bakkouri, M. E., Buhr, F., Muraki, N., Nomata, J., Kurisu, G., . . . Reinbothe, S. (2010). Chlorophyll biosynthesis: spotlight on protochlorophyllide reduction. *Trends in Plant Science*, *15*(11), 614-624. https://doi.org/https://doi.org/10.1016/j.tplants.2010.07.002

Reinkensmeier, J., & Giegerich, R. (2015). Thermodynamic matchers for the construction of the cuckoo RNA family. *RNA biology*, *12*(2), 197-207. https://doi.org/10.1080/15476286.2015.1017206

Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., & Mondragón, A. (2010). Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*, *468*(7325), 784-789. https://doi.org/10.1038/nature09516

Richard, A. L., Withey, J. H., Beyhan, S., Yildiz, F., & DiRita, V. J. (2010). The Vibrio cholerae virulence regulatory cascade controls glucose uptake through activation of TarA, a small regulatory RNA. *Molecular microbiology*, *78*(5), 1171-1181. https://doi.org/10.1111/j.1365-2958.2010.07397.x

Richterich, P. (1998). Estimation of errors in "raw" DNA sequences: a validation study. *Genome research*, *8*(3), 251-259. https://doi.org/10.1101/gr.8.3.251

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139-140. https://doi.org/10.1093/bioinformatics/btp616

Romeo, T. (1998). Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Molecular Microbiology*, *29*(6), 1321-1330. https://doi.org/10.1046/j.1365-2958.1998.01021.x

Rossi, C. C., Bossé, J. T., Li, Y., Witney, A. A., Gould, K. A., Langford, P. R., & Bazzolli, D. M. S. (2016). A computational strategy for the search of regulatory small RNAs in Actinobacillus pleuropneumoniae. *RNA (New York, N.Y.)*, *22*(9), 1373-1385. https://doi.org/10.1261/rna.055129.115

Roth, A., & Breaker, R. R. (2009). The Structural and Functional Diversity of Metabolite-Binding Riboswitches. *Annual Review of Biochemistry*, *78*(1), 305-334. https://doi.org/10.1146/annurev.biochem.78.070507.135656

Rui, B., Shen, T., Zhou, H., Liu, J., Chen, J., Pan, X., . . . Shi, Y. (2010). A systematic investigation of Escherichia coli central carbon metabolism in response to superoxide stress. *BMC Syst Biol*, *4*, 122. https://doi.org/10.1186/1752-0509-4-122

Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeida, D., . . . Collado-Vides, J. (2019). RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. *Nucleic Acids Res*, *47*(D1), D212-d220. https://doi.org/10.1093/nar/gky1077

Schalk, I. J., Hannauer, M., & Braud, A. (2011). New roles for bacterial siderophores in metal transport and tolerance. *Environ Microbiol*, *13*(11), 2844-2854. https://doi.org/10.1111/j.1462-2920.2011.02556.x

Schauer, K., Rodionov, D. A., & de Reuse, H. (2008). New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? *Trends Biochem Sci*, *33*(7), 330-338. https://doi.org/10.1016/j.tibs.2008.04.012

Schübel, U., Kraut, M., Mörsdorf, G., & Meyer, O. (1995). Molecular characterization of the gene cluster coxMSL encoding the molybdenum-containing carbon monoxide dehydrogenase of Oligotropha carboxidovorans. *J Bacteriol*, *177*(8), 2197-2203. https://doi.org/10.1128/jb.177.8.2197-2203.1995

Sharma, C. M., & Vogel, J. (2009). Experimental approaches for the discovery and characterization of regulatory small RNA. *Current Opinion in Microbiology*, *12*(5), 536-546. https://doi.org/https://doi.org/10.1016/j.mib.2009.07.006

Silver, S., & Phung, L. T. (1996). Bacterial heavy metal resistance: New surprises [Article]. *Annual Review of Microbiology*, *50*(1), 753. https://doi.org/10.1146/annurev.micro.50.1.753

Sistrom, W. R. (1960). A requirement for sodium in the growth of Rhodopseudomonas spheroides. *Journal Of General Microbiology*, *22*, 778-785.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., . . . Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545-15550. https://doi.org/10.1073/pnas.0506580102

Sun, J., Zhou, J., Wang, Z., He, W., Zhang, D., Tong, Q., & Su, X. (2014). Multi-omics based changes in response to cadmium toxicity in Bacillus licheniformis A. *RSC Adv.*, *5*. https://doi.org/10.1039/C4RA15280H

Tambosi, R., Liotenberg, S., Bourbon, M.-L., Steunou, A.-S., Babot, M., Durand, A., . . . Shuman Howard, A. (2018). Silver and Copper Acute Effects on Membrane Proteins and Impact on Photosynthetic and Respiratory Complexes in Bacteria. *mBio*, *9*(6), e01535-01518. https://doi.org/10.1128/mBio.01535-18

Tang, G., Shi, J., Wu, W., Yue, X., & Zhang, W. (2018). Sequence-based bacterial small RNAs prediction using ensemble learning strategies. *BMC Bioinformatics*, *19*(20), 503. https://doi.org/10.1186/s12859-018-2535-1

Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic acids research*, *28*(1), 33-36. https://doi.org/10.1093/nar/28.1.33

Teitzel, G. M., Geddie, A., De Long, S. K., Kirisits, M. J., Whiteley, M., & Parsek, M. R. (2006). Survival and growth in the presence of elevated copper: transcriptional profiling of copper-stressed Pseudomonas aeruginosa. *Journal of bacteriology*, *188*(20), 7242-7256. https://doi.org/10.1128/JB.00837-06

The Rnacentral Consortium. (2019). RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Research*, *47*(D1), D221-D229. https://doi.org/10.1093/nar/gky1034

Thomas, C. M., & Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. (9), 711.

Trindade, I. B., Silva, J. M., Fonseca, B. M., Catarino, T., Fujita, M., Matias, P. M., . . . Louro, R. O. (2019). Structure and reactivity of a siderophore-interacting protein from the marine bacterium Shewanella reveals unanticipated functional versatility. *The Journal of biological chemistry*, *294*(1), 157-167. https://doi.org/10.1074/jbc.RA118.005041

Trotochaud, A. E., & Wassarman, K. M. (2005). A highly conserved 6S RNA structure is required for regulation of transcription. *Nature Structural & Molecular Biology*, *12*(4), 313-319. https://doi.org/10.1038/nsmb917

Tso, W. W., & Adler, J. (1974). Negative chemotaxis in Escherichia coli. *J Bacteriol*, *118*(2), 560-576. https://doi.org/10.1128/jb.118.2.560-576.1974

Ul Haq, I., Müller, P., & Brantl, S. (2020). Intermolecular Communication in Bacillus subtilis: RNA-RNA, RNA-Protein and Small Protein-Protein Interactions [10.3389/fmolb.2020.00178]. *Frontiers in Molecular Biosciences*, *7*, 178.

Vatansever, F., de Melo, W. C. M. A., Avci, P., Vecchio, D., Sadasivam, M., Gupta, A., . . . Hamblin, M. R. (2013). Antimicrobial strategies centered around reactive oxygen species--bactericidal antibiotics, photodynamic therapy, and beyond. *FEMS microbiology reviews*, *37*(6), 955-989. https://doi.org/10.1111/1574-6976.12026

Volpicella, M., Costanza, A., Palumbo, O., Italiano, F., Claudia, L., Placido, A., . . . Ceci, L. R. (2014). Rhodobacter sphaeroides adaptation to high concentrations of cobalt ions requires energetic metabolism changes. *FEMS Microbiology Ecology*, *88*(2), 345-357. https://doi.org/10.1111/1574-6941.12303

Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G., & Gottesman, S. (2001). Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev*, *15*(13), 1637-1651. https://doi.org/10.1101/gad.901001

Waters, L. S., & Storz, G. (2009). Regulatory RNAs in Bacteria. *Cell*, *136*(4), 615-628. https://doi.org/https://doi.org/10.1016/j.cell.2009.01.043

Weber, L., Thoelken, C., Volk, M., Remes, B., Lechner, M., & Klug, G. (2016). The Conserved Dcw Gene Cluster of R. sphaeroides Is Preceded by an Uncommonly Extended 5' Leader Featuring the sRNA UpsM. *PloS one*, *11*(11), e0165694-e0165694. https://doi.org/10.1371/journal.pone.0165694

Wright, P. R., Georg, J., Mann, M., Sorescu, D. A., Richter, A. S., Lott, S., . . . Backofen, R. (2014). CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. (W1), W119.

Wright, P. R., Richter, A. S., Papenfort, K., Mann, M., Vogel, J., Hess, W. R., . . . Georg, J. (2013). Comparative genomics boosts target prediction for bacterial small

RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(37), E3487-E3496. https://doi.org/10.1073/pnas.1303248110

Wu, Y., Xing, X., You, T., Liang, R., & Liu, J. (2017). RT-qPCR with chimeric dU stem-loop primer is efficient for the detection of bacterial small RNAs. *Applied microbiology and biotechnology*, *101*(11), 4561-4568. https://doi.org/10.1007/s00253-017-8181-0

Yu, H., Rao, X., & Zhang, K. (2017). Nucleoside diphosphate kinase (Ndk): A pleiotropic effector manipulating bacterial virulence and adaptive responses. *Microbiological Research*, *205*, 125-134. https://doi.org/https://doi.org/10.1016/j.micres.2017.09.001

Zannoni, D., Schoepp-Cothenet, B., & Hosler, J. (2009). Respiration and Respiratory Complexes. In C. N. Hunter, F. Daldal, M. C. Thurnauer, & J. T. Beatty (Eds.), *The Purple Phototrophic Bacteria* (pp. 537-561). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8815-5_27

Zavala, N., Baeza, L., Gonzalez, S., & Choudhary, M. (2019). The Effects of Different Carbon Sources on the Growth of Rhodobacter sphaeroides. *Advances in Microbiology*, *09*, 737-749. https://doi.org/10.4236/aim.2019.98045

Zeilstra-Ryalls, J., Gomelsky, M., Eraso, J. M., Yeliseev, A., O'Gara, J., & Kaplan, S. (1998). Control of photosystem formation in Rhodobacter sphaeroides. *Journal of bacteriology*, *180*(11), 2801-2809. https://doi.org/10.1128/JB.180.11.2801-2809.1998

Zeng, X., Choudhary, M., & Kaplan, S. (2003). A Second and Unusual pucBA Operon of Rhodobacter sphaeroides 2.4.1:

Genetics and Function of the Encoded Polypeptides. *Journal of Bacteriology*, *185*(20), 6171. https://doi.org/10.1128/JB.185.20.6171-6184.2003

Zhang, A., Wassarman, K. M., Rosenow, C., Tjaden, B. C., Storz, G., & Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq [https://doi.org/10.1046/j.1365-2958.2003.03734.x]. *Molecular Microbiology*, *50*(4), 1111-1124. https://doi.org/https://doi.org/10.1046/j.1365-2958.2003.03734.x

Zhao, C., Yang, Q., Chen, W., & Teng, B. (2012). Removal of hexavalent chromium in tannery wastewater by Bacillus cereus. *Can J Microbiol*, *58*(1), 23-28. https://doi.org/10.1139/w11-096

Experimentally validated and previously identified noncoding RNA in Escherichia coli K12 substrate MG1655

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
trans-encoded sRNA	tpke11	89	14080	14168	+	BSRD, Rfam	Hershberg 2003	Experimental; Northern blot (PMID: 11553332)
trans-encoded sRNA	sokC	55	16952	17006	+	BSRD, RegulonDB	Pederson 1999, Kawano 2005	Experimental; Northern blot (PMID: 10361310)
trans-encoded sRNA	nc9	112	28509	28620	+	BSRD, RegulonDB	Macvanin 2012	Experimental; co-IP with HU (PMID:22942248), tiling arrays (PMID:22942248)
ncRNA	ES003	52	29551	29603	+	Rfam	n/a	Similarity (87.4 bit score)
ncRNA	naRNA4	65	66675	66610	-	Rfam	n/a	Similarity (60.2 bit score)
ncRNA	naRNA4	65	66760	66695	-	Rfam	n/a	Similarity (60.2 bit score)
trans-encoded sRNA	SroA	93	75608	75516	-	BSRD, RegulonDB	Vogel 2003	Experimental; Northern blot (PMID: 14602901), RACE (PMID: 14602901)
trans-encoded sRNA	SgrS (RyaA)	227	77367	77593	+	BSRD, RegulonDB	Zhang 2003, Kawamoto 2005, Malecka 2015, Mihailovic 2018, Sun 2013, Vanderpool 2004	Experimental; Northern blot (PMID: 14622403)
gene;sRNA	PssrA	89	105224	105313	+	Rfam	n/a	Similarity (105.6 bit score)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
gene;sRNA	STnc40	69	110960	111029	+	Rfam	n/a	Similarity (84.3 bit score)
gene;sRNA	naRNA4	76	111433	111509	+	Rfam	n/a	Similarity (77.8 bit score)
gene;sRNA	naRNA4	77	111549	111626	+	Rfam	n/a	Similarity (78.4 bit score)
gene;sRNA	tp2	161	122697	122857	-	BSRD, Rfam	Rivas 2001	Experimental; Northern blot (PMID: 11553332) Similarity (102 bit score)
gene;sRNA	tff (T44)	136	189712	189847	+	BSRD, RegulonDB	Rivas 2001, Aseev 2008	Experimental; Northern blot (PMID: 11553332)
gene;sRNA	CssrA	109	190685	190794	+	Rfam	n/a	Similarity (136.6 bit score)
gene;sRNA	naRNA4	76	216058	216134	+	Rfam	n/a	Similarity (92.8 bit score)
gene;sRNA	C0067	125	238462	238586	+	BSRD	Tjaden 2002	Experimental; Microarray (PMID: 12202758)
ncRNA	naRNA4	77	244132	244209	+	Rfam	n/a	Similarity (74.1 bit score)
ncRNA	naRNA4	77	247461	247538	+	Rfam	n/a	Similarity (73.3 bit score)
ncRNA	naRNA4	77	248203	248280	+	Rfam	n/a	Similarity (77 bit score)
gene;sRNA	eyeA	75	272580	272654	+	BSRD, RegulonDB	Saetrom 2005, Raghavan 2011	Experimental; RNAseq (PMID: 21665928) Genbank ID:7751631
gene;sRNA	nc1	168	297304	297137	-	BSRD, RegulonDB	Macvanin 2012	Experimental; co-IP with HU (PMID:22942248), tiling arrays (PMID:22942248)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
ncRNA	naRNA4	69	339880	339811	-	Rfam	n/a	Similarity (78 bit score)
ncRNA	naRNA4	69	339973	339904	-	Rfam	n/a	Similarity (72.4 bit score)
ncRNA	naRNA4	69	340066	339997	-	Rfam	n/a	Similarity (78 bit score)
ncRNA	naRNA4	71	349752	349823	+	Rfam	n/a	Similarity (63.1 bit score)
ncRNA	naRNA4	71	349845	349916	+	Rfam	n/a	Similarity (63.1 bit score)
ncRNA	naRNA4	76	354690	354766	+	Rfam	n/a	Similarity (71.6 bit score)
ncRNA	naRNA4	77	375061	374984	-	Rfam	n/a	Similarity (79.3 bit score)
ncRNA	naRNA4	77	375085	375162	+	Rfam	n/a	Similarity (62.7 bit score)
ncRNA	naRNA4	77	375162	375085	-	Rfam	n/a	Similarity (85 bit score)
ncRNA	naRNA4	77	375186	375263	+	Rfam	n/a	Similarity (62 bit score)
ncRNA	naRNA4	77	375263	375186	-	Rfam	n/a	Similarity (87.2 bit score)
ncRNA	naRNA4	75	375362	375287	-	Rfam	n/a	Similarity (62.8 bit score)
ncRNA	naRNA4	77	377391	377468	+	Rfam	n/a	Similarity (79.6 bit score)
ncRNA	naRNA4	79	411134	411213	+	Rfam	n/a	Similarity (79.8 bit score)
gene;sRNA	sraA (PsrA3, T15)	57	458784	458728	-	BSRD, RegulonDB, Rfam	Argaman 2001, Raghavan 2011, Rivas 2001	Experimental; Northern blot (PMID: 11448770), RNAseq (PMID: 21665928)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
4.5S RNA, SRP	ffs	114	476448	476561	+	BSRD, RegulonDB	Koch 1999, Avdeeva 2002, Bailey 1979, Huang 1994, Jensen 1994, Malygin 1996, Phillips 1992, Zwieb 2005	Experimental; Northern blot (PMID: 10397756)
gene;sRNA	nc2	145	497801	497945	+	BSRD, RegulonDB	Macvanin 2012	Experimental; co-IP with HU (PMID:22942248), tiling arrays (PMID:22942248), Northern blot (PMID:22942248)
gene;sRNA	naRNA4	77	501447	501524	+	Rfam	n/a	Similarity (74.3 bit score)
gene; sRNA	sroB (ChiX)	84	507204	507287	+	BSRD, Rfam, RNAcentral, RegulonDB	Vogel 2003, Mandin 2009, Rasmussen 2009, Zhang 2003, Wassarman 2001	Experimental; Northern blot (PMID: 14602901), RACE (PMID: 14602901)
ncRNA	naRNA4	77	508585	508662	+	Rfam	n/a	Similarity (79.9 bit score)
ncRNA	naRNA4	77	508662	508585	-	Rfam	n/a	Similarity (64 bit score)
ncRNA	naRNA4	77	508686	508763	+	Rfam	n/a	Similarity (89.3 bit score)
ncRNA	naRNA4	77	508763	508686	-	Rfam	n/a	Similarity (63.6 bit score)
ncRNA	STnc480	66	543188	543254	+	Rfam	n/a	Similarity (86.9 bit score)
gene; sRNA	ipeX	167	573588	573754	-	BSRD, RegulonDB	Castillo-Keller 2006	Experimental; RT-PCR (PMID: 16385048)
gene; sRNA	sokE	59	607734	607792	+	BSRD, RegulonDB	Pederson 1999, Kawano 2005	Experimental; Northern blot (PMID: 15718303)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
ncRNA	STnc70	93	638707	638800	+	Rfam	n/a	Similarity (42.7 bit score)
gene;sRNA	sroC (HB_314)	163	686843	686681	-	Rfam, RegulonDB	Azam 2015, Bak 2015, Frohlich 2018, Vogel 2003, Lalaouna 2019	Experimental; Northern blot (PMID: 14602901), RACE (PMID: 14602901) Similarity (174.8 bit score)
ncRNA	ES036	42	740988	741030	+	Rfam	n/a	Similarity (75.1 bit score)
ncRNA	naRNA4	70	758482	758412	-	Rfam	n/a	Similarity (88.4 bit score)
ncRNA	naRNA4	71	762854	762925	+	Rfam	n/a	Similarity (59.8 bit score)
gene;sRNA	sdhX (RybD)	101	765050	765150	+	RegulonDB	De Mets 2019, Miyakoshi 2018, Sridhar 2007, Zhang 2003	Experimental; co-IP (PMID: 14622403)
ncRNA	naRNA4	70	805859	805929	+	Rfam	n/a	Similarity (60.7 bit score)
ncRNA	naRNA4	70	837658	837558	-	Rfam	n/a	Similarity (78.4 bit score)
gene;sRNA	RybA	340	853064	852725	-	RegulonDB	Gerstle 2012, Wassarman 2001	Experimental; Northern blot (PMID: 11445539)
gene;sRNA	RybB	78	888054	887976	-	Rfam, RegulonDB	Vogel 2003, Wassarman 2001, Mihailovic 2018, El-Mowafi 2014, etc.	Experimental; Northern blot (PMID: 14602901), RACE (PMID: 14602901) Similarity (105.1 bit score)
ncRNA	naRNA4	80	900752	900832	+	Rfam	n/a	Similarity (75.5 bit score)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
ncRNA	STnc130	134	940734	940868	+	Rfam	n/a	Similarity (139 bit score)
ncRNA	ES036	42	984395	984437	+	Rfam	n/a	Similarity (53 bit score)
ncRNA	naRNA4	78	1062488	1062410	-	Rfam	n/a	Similarity (69.1 bit score)
ncRNA	ES056	91	1103252	1103343	+	Rfam	n/a	Similarity (119.5 bit score)
ncRNA	ES036	42	1113480	1113438	-	Rfam	n/a	Similarity (47.9 bit score)
gene;sRNA	SraB (psrD, psrA4, pke20)	169	1146589	1146757	+	Rfam, RNAcentral, RegulonDB	Argaman 2001, Tjaden 2002	Experimental; Northern blot (PMID: 11448770), Microarray (PMID: 12202758)
gene;sRNA	C0293	73	1195937	1196009	+	sRNAMap, EcoCyc	Tjaden 2002	Experimental; Microarray (PMID: 12202758)
gene; sRNA	nc10	209	1203886	1203678	-	BSRD, RegulonDB	Macvanin 2012	Experimental; co-IP with HU, tiling arrays (PMID:22942248)
gene;sRNA	C0299	79	1230629	1230707	+	Rfam, sRNAMap, EcoCyc	Tjaden 2002	Experimental; Microarray (PMID: 12202758) Similarity (111.9 bit score)
gene;sRNA	rdlA	67	1269323	1269389	+	BSRD, RegulonDB	Kawano 2002, Kawano 2005	Experimental; cDNA cloning-based screen (PMID: 15718303)
gene;sRNA	nc3	80	1276858	1276937	+	BSRD, RegulonDB	Macvanin 2012	Experimental; co-IP with HU, tiling arrays, Northern blot (PMID:22942248)
gene;sRNA	rttR	171	1287236	1287066	-	RegulonDB	Bosi 1991	Experimental; Northern blot (PMID: 1840671?
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
ncRNA	STnc180	202	1335499	1335701	+	Rfam	n/a	Similarity (195.7 bit score)
gene;sRNA	nc7	191	1351101	1351291	+	BSRD, RegulonDB	Macvanin 2012	Experimental; co-IP with HU, tiling arrays (PMID:22942248)
gene;sRNA	McaS (IsrA, IS061)	95	1405751	1405656	-	Rfam, RegulonDB	Chen 2002, Boehm 2012, Jorgensen 2012, Kumar 2016, Malecka 2015, Mihailovic 2018	Experimental; Northern blot (PMID: 12069726), Similarity (118.4 bit score)
gene;sRNA	C0343	75	1407387	1407461	+	sRNAMap, EcoCyc	Tjaden 2002, Durand 2010	Experimental; Microarray (PMID: 12202758)
gene;sRNA	FnrS (RydA)	122	1409129	1409246	+	Rfam, RegulonDB	Boysen 2010	Experimental; Microarray and Northern blot (PMID: 20075074) Similarity (134 bit score)
gene;sRNA	RalA	179	1413556	1413734	+	Rfam, EcoCyc	Guo 2014	Experimental; qRT-PCR (PMID: 24748661) Similarity (225.1 bit score)
gene;sRNA	ralA	179	1413556	1413734	+	RegulonDB	Guo 2014	Experimental; qRT-PCR (PMID: 24748661)
gene;sRNA	MicC (IS063, tke8)	111	1437121	1437229	+	Rfam, RegulonDB	Chen 2002, Chen 2004, De La Cruz 2010, Urban 2007	Experimental; Northern blot (PMID: 12069726) Similarity (123.6 bit score)
gene;sRNA	RydC	63	1491506	1491443	-	Rfam, RNAcentral, RegulonDB	Antal 2005, Zhang 2003	Experimental; Northern blot (PMID: 15618228) Similarity (94.7 bit score)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
gene;sRNA	C0362	386	1550025	1550410	+	sRNAMap, EcoCyc	Tjaden 2002	Experimental; Microarray (PMID: 12202758)
ncRNA	STnc560	213	1622735	1622948	+	Rfam	n/a	Similarity (280.4 bit score)
gene;sRNA	mgrR	98	1622914	1622817	-	RegulonDB	Lee 2016, Moon 2009, Yin 2019	Experimental; Microarray (PMID: 14622403)
gene;sRNA	DicF	53	1649382	1649434	+	RegulonDB	Bouche 1989, Tetart 1992, Murashko 2017, Faubladier 1990, Balasubramanian 2016	Experimental; Northern blot (PMID: 2477663)
ncRNA	STnc550	390	1737843	1737453	-	Rfam	n/a	Similarity (396.9 bit score)
gene;sRNA	rydB	68	1764780	1764713	-	Rfam, RegulonDB	Chen 2002, Komasa 2011, Rivas 2001, Wassarman 2001, Hershberg 2003	Experimental; Northern blot (PMID: 11445539, PMID: 12069726)
gene;sRNA	RprA	106	1770372	1770477	+	Rfam, RNAcentral, RegulonDB	Argaman 2001, Majdalani 2001, Mihailovic 2018	Experimental; Northern blot (PMID: 11448770)
ncRNA	STnc540	158	1795311	1795153	-	Rfam	n/a	Similarity (67.4 bit score)
ncRNA	ES036	42	1816219	1816261	+	Rfam	n/a	Similarity (55.5 bit score)
3' UTR sRNA	spy3'	47	1823131	1823084	-	RegulonDB	Kawano 2005	Experimental; cDNA cloning-based screen, Northern blot (PMID: 15718303)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
gene;sRNA	sroD (p24)	86	1888102	1888017	-	Rfam, sRNAMap, EcoCyc	Vogel 2003	Experimental; Northern blot, RACE (PMID: 14602901) Similarity (115.1 bit score)
gene;sRNA	SdsR (RyeB, tkpe79)	99	1923207	1923104	-	Rfam, RegulonDB	Bak 2015, Kim 2015, Rivas 2001, Sridhar 2007, Vogel 2003, Wassarman 2001	Experimental; Northern blot (PMID: 14602901. PMID: 11445539), RNomics (PMID: 14602901), Microarray (PMID: 11445539) Similarity (114.3 bit score)
gene;sRNA	MicL-S (sirA, ryeF)	80	1958520	1958441	-	Rfam, RNAcentral, RegulonDB	Guo et. al. 2014, Klein 2014, Nicoloff 2017, Zhang 2003	Experimental; Northern blot (PMID: 25030700), Similarity (381.9 bit score)
gene;sRNA	C0465	78	1972739	1972816	+	Rfam, sRNAMap, EcoCyc	Tjaden 2002	Experimental; Microarray (PMID: 12202758) Similarity (111.6 bit score)
gene;sRNA	isrB (IS092)	160	1987998	1987839	-	sRNAMap, EcoCyc	Chen 2002, Hemm 2008	Experimental; Northern blot (PMID: 12069726)
ETS sRNA	3'ETSleuZ	67	1991814	1991748	-	RegulonDB	Lalaouna 2015, Sinha 2018	Experimental; MS2 affinity purification and RNA-seq (PMID: 25891076)
gene;sRNA	sdsN	137	1996921	1997057	+	RegulonDB	Hao 2016, Raghavan 2011	Experimental; RNA-seq (PMID: 21665928)
gene;sRNA	dsrA	86	2025227	2025313	-	Rfam, RegulonDB	Chen 2002, Malecka 2015, Mihailovic 2018, Peterman 2014,	Experimental, Cloning (PMID: 7534408) Similarity (102.8 bit score)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
gene;sRNA	rseX	89	2033649	2033738	+	Rfam, RegulonDB	Douchin 2006, Kim 2015, Mihailovic 2018	Experimental; 5' RACE and Northern blot (PMID: 16513633) Similarity (96.4 bit score)
ncRNA	STnc240	74	2087283	2087209	-	Rfam	n/a	Similarity (68.2 bit score)
ncRNA	naRNA4	70	2118524	2118594	+	Rfam	n/a	Similarity (72.9 bit score)
gene;sRNA	sibB (ryeD, QUAD1b, Tpe60)	136	2153646	2153781	+	BSRD, RegulonDB	Fozo 2008, Wassarman 2001	Experimental; Northern blot (PMID: 11445539), Microarray (PMID: 11445539)
gene;sRNA	CyaR (RyeE)	87	2167114	2167200	+	Rfam, RegulonDB	Wassarman 2001, Johansen 2008, De Lay 2009	Experimental; Northern blot (PMID: 11445539) Similarity (99.7 bit score)
ncRNA	naRNA4	70	2177481	2177411	-	Rfam	n/a	Similarity (64.5 bit score)
ncRNA	naRNA4	77	2236680	2236603	-	Rfam	n/a	Similarity (88.3 bit score)
5' UTR sRNA	yejG5'	240	2276520	2276280	-	RegulonDB	Raghavan 2011	Experimental; RNA-seq (PMID: 21665928)
ncRNA	naRNA4	83	2304420	2304503	+	Rfam	n/a	Similarity (63.7 bit score)
ncRNA	naRNA4	83	2304533	2304616	+	Rfam	n/a	Similarity (63.7 bit score)
ncRNA	naRNA4	83	2304646	2304729	+	Rfam	n/a	Similarity (76 bit score)
ncRNA	naRNA4	83	2304759	2304842	+	Rfam	n/a	Similarity (74.2 bit score)
ncRNA	naRNA4	83	2304872	2304955	+	Rfam	n/a	Similarity (76 bit score)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
ncRNA	naRNA4	82	2304985	2305067	+	Rfam	n/a	Similarity (61.9 bit score)
4.5S RNA, SRP	micF	93	2313084	2313176	+	BSRD, RegulonDB	Andersen 1987, Andersen 1989, Chen 2002, Delihas 1997, Komatsu 1990, Kumar 2016, Mihailovic 2018, Tkachenko 2006, Urban 2007	Experimental; Autoradiogram (PMID: 2478539), Northern blot (PMID: 12069726)
ncRNA	naRNA4	76	2316913	2316837	-	Rfam	n/a	Similarity (80.4 bit score)
ncRNA	naRNA4	76	2317027	2316951	-	Rfam	n/a	Similarity (77.5 bit score)
ncRNA	naRNA4	70	2347277	2347207	-	Rfam	n/a	Similarity (76.4 bit score)
gene;sRNA	RyeG	194	2470665	2470472	-	RegulonDB	Bak 2015, Mandin 2010, Zhang 2003	Experimental; Overexpression (PMID: 26469694), Bioinformatic prediction (PMID: 12069726)
gene;sRNA	tpke70	436	2496629	2496194	-	sRNAMap, EcoCyc	Rivas 2001, Mihailovic 2018	Experimental; Northern blot (PMID: 11553332)
5' UTR sRNA	ZipA5'	230	2529483	2529253	-	RegulonDB	Kawano 2005	Experimental; cDNA cloning-based screen (PMID: 15718303), Northern blot (PMID: 15718303)
ncRNA	naRNA4	81	2568269	2568188	-	Rfam	n/a	Similarity (71.8 bit score)
ncRNA	naRNA4	76	2592981	2593057	+	Rfam	n/a	Similarity (92.7 bit score)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
ncRNA	naRNA4	76	2593057	2592981	-	Rfam	n/a	Similarity (61.5 bit score)
gene;sRNA	sroE	92	2640686	2640595	-	Rfam, RegulonDB	Vogel 2003, Rivas 2001	Experimental; Northern blot & RACE (PMID: 14602901)
gene;sRNA	IS129	392	2652078	2651686	-	RegulonDB	Chen 2002	Experimental; Northern blot (PMID: 12069726)
gene;sRNA	IS128	209	2653515	2653723	+	Rfam, sRNAMap, EcoCyc	Chen 2002	Experimental; Northern blot (PMID: 12069726)
gene;sRNA	C0614	87	2653538	2653452	-	sRNAMap, EcoCyc	Tjaden 2002	Experimental; Microarray (PMID: 12202758)
gene;sRNA	ryfA	304	2653855	2654158	+	Rfam, RNAcentral, RegulonDB	Bak 2015, Rivas 2001, Rudd 1999	Experimental; Northern blot (PMID: 11445539) Similarity (313.9 bit score)
ncRNA	naRNA4	71	2662430	2662359	-	Rfam	n/a	Similarity (72.2 bit score)
ncRNA	naRNA4	71	2662521	2662450	-	Rfam	n/a	Similarity (72.2 bit score)
gene;sRNA	GlmY (sroF, tke1)	184	2691340	2691193	-	Rfam, RegulonDB	Rivas 2001, Vogel 2003, Andrade 2012, Gonzalez 2017	Experimental; Northern blot (PMID: 11553332)
gene;sRNA	ryfB (shoB)	319	2700377	2700059	-	sRNAMap, EcoCyc	Kawano 2005	Experimental; cloning based screen, Northern blot (PMID: 15718303)
ncRNA	ES036	42	2714346	2714304	-	Rfam	n/a	Similarity (49.9 bit score)
gene;sRNA	ryfD (Ysr155)	143	2734295	2734153	-	Rfam, RegulonDB	Bak 2015, Kawano 2005	Experimental; cDNA cloning-based screen (PMID: 15718303)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
5' UTR sRNA	rpsP5'	249	2744454	2744205	-	RegulonDB	Kawano 2005	Experimental; Northern blot (PMID: 15718303)
tmRNA	SsrA (10Sa, SipB)	363	2755593	2755955	+	RegulonDB	Kirby 1994, Muto 1996, Nakano 2001, Oh 1990, Ray 1979, Roche 1999, Williams 2003	Experimental; Northern blot (PMID: 2482406)
ncRNA	naRNA4	77	2808134	2808211	+	Rfam	n/a	Similarity (72.2 bit score)
gene;sRNA	MicA (sraD)	73	2814802	2814874	+	Rfam, RNAcentral, RegulonDB	Argaman 2001, Guo 2014, Hammann 2014, Mihailovic 2018, Moores 2014	Experimental; Northern blot (PMID: 11448770)
gene;sRNA	C0664	113	2835055	2835167	+	sRNAMap, EcoCyc	Tjaden 2002	Experimental; Microarray (PMID: 12202758)
gene;sRNA	sokX	56	2887353	2887408	+	BSRD, RegulonDB	Kawano 2005, Raghavan 2011, Pedersen 1999	Experimental; Northern blot (PMID: 15718303)
gene;sRNA	CsrB	369	2924524	2924156	-	Rfam, RNAcentral, RegulonDB	Liu 1997	Experimental; co-IP and cDNA sequencing (PMID: 9211896) Similarity (376.7 bit score)
gene; sRNA	GcvB (PsrA11)	206	2942696	2942900	+	RegulonDB	Argaman 2001, Urbanowski 2000, etc.	Experimental; Northern blot (PMID: 11448770)
ncRNA	ES036	42	2945959	2945917	-	Rfam	n/a	Similarity (46.5 bit score)
gene;sRNA	OmrA (RygA, SraE, PsrA12, t59)	88	2976189	2976102	-	Rfam, RNAcentral, BSRD, RegulonDB	Holmqvist 2010, Mihailovic 2018, Moon 2011, Rivas 2001	Experimental; Northern blot (PMID: 11448770) Similarity (95.5 bit score)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
gene;sRNA	OmrB (RygB, t59, sraE)	82	2976385	2976304	-	RegulonDB, Rfam, RNAcentral	Guillier 2006, Guillier 2008, Mihailovic 2018, Vogel 2003	Experimental; Northern blot and microarray (PMID: 16359331), Similarity (85.5 bit score)
ncRNA	naRNA4	77	3042395	3042472	+	Rfam	n/a	Similarity (75 bit score)
6S RNA	ssrS	183	3055983	3056165	+	RegulonDB	Brownlee 1971, Skylar 1975, Somasekhar 1983, Trown 1973, Wassarman 2000	Experimental; Northern blot (PMID: 10892648)
gene;sRNA	och5	158	3067187	3067344	+	RegulonDB	Bak 2015, Raghavan 2011	Experimental; RNA-seq (PMID: 21665928)
ncRNA	ES036	42	3070037	3070079	+	Rfam	n/a	Similarity (47.3 bit score)
ncRNA	naRNA4	70	3098534	3098464	-	Rfam	n/a	Similarity (60.9 bit score)
gene;sRNA	C0719	222	3121358	3121579	+	Rfam, sRNAMap, EcoCyc	Tjaden 2002	Experimental; Microarray (PMID: 12202758) Similarity (298.2 bit score)
ncRNA	naRNA4	76	3139624	3139700	+	Rfam	n/a	Similarity (92.6 bit score)
ncRNA	naRNA4	76	3139700	3139624	-	Rfam	n/a	Similarity (60.1 bit score)
ncRNA	naRNA4	70	3150740	3150670	-	Rfam	n/a	Similarity (61 bit score)
ncRNA	ES173	63	3156521	3156584	+	Rfam	n/a	Similarity (98.9 bit score)
gene;sRNA	sroG	149	3184718	3184570	-	sRNAMap, EcoCyc	Vogel 2003, Mihailovic 2018	Experimental; Northern blot, RACE (PMID: 14602901)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
gene;sRNA	sibD (tp8, C0730, IS156, QUAD1d, RygD)	145	3194865	3194721	-	RegulonDB, BSRD	Chen 2002, Fozo 2008, Rivas 2001, Rudd 1999, Wassarman 2001	Experimental; Northern blot (PMID: 11445539), Microarray (PMID: 11445539)
gene;sRNA	sibE (rygE, QUAD1e)	144	3195240	3195097	-	RegulonDB, BSRD	Fozo 2008, Rudd 1999, Wassarman 2001	Experimental; Northern blot (PMID: 18710431)
ncRNA	naRNA4	70	3203265	3203195	+	Rfam	n/a	Similarity (79.3 bit score)
ncRNA	naRNA4	81	3231382	3231301	-	Rfam	n/a	Similarity (72.8 bit score)
ncRNA	naRNA4	81	3231482	3231401	-	Rfam	n/a	Similarity (66.9 bit score)
gene;sRNA	psrN (sraF, psrA14, tpk1, IS160)	188	3238374	3238561	+	sRNAMap, EcoCyc	Argaman 2001, Rivas 2001, Nechooshtan 2009	Experimental; Northern blot (PMID: 11448770)
gene; sRNA	nc4	120	3265219	3265338	+	BSRD, RegulonDB	Macvanin 2012	Experimental; co-IP with HU (PMID:22942248), tiling arrays (PMID:22942248)
RNase P; ribozyme	rnpB (M1 RNA)	377	3270592	3270216	-	RegulonDB	Guerrier-Takada 1983, Altman 1990, Kole 1979, Kole 1981,	Experimental; Subcloning (PMID: 6183002)
gene;sRNA	sraG (P3, PsrA15, psrO)	216	3311183	3311398	+	RegulonDB, Rfam	Argaman 2001, Fontaine 2016	Experimental; Northern blot (PMID: 11448770), Similarity (164.2 bit score)
ncRNA	naRNA4	70	3313325	3313255	-	Rfam	n/a	Similarity (68.1 bit score)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
3' UTR sRNA	YrbL3'	39	3347168	3347207	+	RegulonDB	Kawano 2005	Experimental; cDNA cloning-based screen, Northern blot (PMID: 15718303)
3' UTR sRNA	YhcF3' (ES186)	157	3365792	3365635	-	RegulonDB	Raghavan 2011	Experimental; RNA-seq (PMID: 21665928)
3' UTR sRNA	rpsI3'	84	3375473	3375389	-	RegulonDB	Kawano 2005	Experimental; Northern blot (PMID: 15718303)
ncRNA	naRNA4	69	3392198	3392129	-	Rfam	n/a	Similarity (77.3 bit score)
ncRNA	naRNA4	69	3392289	3392220	-	Rfam	n/a	Similarity (76 bit score)
ncRNA	naRNA4	69	3392380	3392311	-	Rfam	n/a	Similarity (71.1 bit score)
gene;sRNA	crpT	300	3483973	3483673	-	RegulonDB	Hanamura 1991, Okamoto 1986, Tjaden 2002	Experimental (PMID: 3053643), Microarray (PMID: 12202758)
gene;sRNA	RyhB (sraI, IS176, PsrA18)	95	3581016	3580927	-	Rfam, RegulonDB	Argaman 2001, Arbel-Goren 2016, Baez 2017, Bos 2013, Chen 2002, Masse 2002, etc.	Experimental; Northern blot (PMID: 11917098, PMID: 11448770) (Similarity (83.3 bit score)
gene;sRNA	agrA	82	3648063	3648146	+	RegulonDB	Weel-Sneve 2013, Kristiansen 2016	Experimental; Northern blot (PMID: 23408903)
gene;sRNA	arrS (6H57)	70	3658054	3657986	-	RegulonDB	Aiso 2011, Aiso 2014	Experimental; Shotgun cloning
gene;sRNA	GadY (IS183)	106	3664864	3664969	+	Rfam, RegulonDB, sRNAMap, EcoCyc	Opdyke 2004, Opdyke 2011, Chen 2002, Kobayashi 2006	Experimental; Northern blot (PMID: 12069726, PMID: 15466020)
ncRNA	ES036	42	3705997	3706039	+	Rfam	n/a	Similarity (56.7 bit score)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
gene;sRNA	sokA	30	3722076	3722105	+	EcoCyc	Pederson 1999	Experimental; Northern blot (PMID: 10361310)
ncRNA	ES036	42	3741027	3740985	-	Rfam	n/a	Similarity (46.1 bit score)
5' UTR sRNA	rirA	73	3808238	3808166	-	RegulonDB	Klein 2016	Experimental; Subcloning (PMID: 27629414)
gene;sRNA	istR-1	75	3853192	3853118	-	sRNAMap, EcoCyc	Vogel 2004, Darfeuille 2007, Dorr 2010, Malecka 2015	Experimental; Northern blot (PMID: 15620655)
gene;sRNA	istR-2	140	3853257	3853118	-	sRNAMap, EcoCyc	Vogel 2004, Dorr 2010	Experimental; Northern blot (PMID: 15620655)
ncRNA	STnc410	157	3915284	3915441	+	Rfam	n/a	Similarity (188.4 bit score)
gene;sRNA	GlmZ (SraJ, PsrA20, RyiA)	207	3986432	3986638	+	Rfam, RNAcentral, RegulonDB	Argaman 2001, Wassarman 2001, Kalamorz 2007, Mihailovic 2018	Experimental; Northern blot (PMID: 11448770, PMID: 11445539)
5' UTR sRNA	YigE5'	143	4001191	4001334	+	RegulonDB	Raghavan 2011	Experimental; RNA-seq (PMID: 21665928)
gene;sRNA	EsrE	252	4019978	4020229	+	Rfam, RNAcentral, RegulonDB	Chen 2012	Experimental; RACE (PMID: 22575655) Similarity (319.5 bit score)
ncRNA	naRNA4	77	4027375	4027452	+	Rfam	n/a	Similarity (80.2 bit score)
ncRNA	naRNA4	77	4027473	4027550	+	Rfam	n/a	Similarity (92.1 bit score)
gene;sRNA	Spot_42 (spf, IS197)	111	4049899	4050009	+	Rfam, RegulonDB	Ikemura 1973, Sahagan 1979 Chen 2002, Mihailovic 2018, Moller 2002	Experimental; 2D-PAGE (PMID: 390161)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
gene;sRNA	CsrC (SraK)	253	4051036	4051289	+	Rfam, RegulonDB	Argaman 2001, Weilbacher 2003	Experimental; Northern blot (PMID: 11448770), Genetic screen (PMID: 12694612)
ncRNA	naRNA4	77	4053533	4053610	+	Rfam	n/a	Similarity (75.2 bit score)
gene;sRNA	glnA3	195	4054201	4054007	-	RegulonDB	Kawano 2005	Experimental; Northern blot (PMID: 15718303)
5' UTR sRNA	typA5'	71	4056194	4056265	+	RegulonDB	Raghavan 2011	Experimental; RNA-seq (PMID: 21665928)
ncRNA	ES222	107	4058218	4058325	+	Rfam	n/a	Similarity (135.3 bit score)
ncRNA	STnc370	68	4060215	4060283	+	Rfam	n/a	Similarity (57 bit score)
ncRNA	naRNA4	70	4094319	4094389	+	Rfam	n/a	Similarity (77 bit score)
ncRNA	naRNA4	70	4094411	4094481	+	Rfam	n/a	Similarity (69.3 bit score)
ncRNA	naRNA4	70	4094503	4094573	+	Rfam	n/a	Similarity (76.6 bit score)
ncRNA	naRNA4	70	4094595	4094665	+	Rfam	n/a	Similarity (76.6 bit score)
ncRNA	naRNA4	77	4103587	4103510	-	Rfam	n/a	Similarity (78.6 bit score)
ncRNA	ES036	41	4108633	4108592	-	Rfam	n/a	Similarity (48.2 bit score)
gene;sRNA	OxyS	110	4158394	4158285	-	Rfam, RNAcentral, RegulonDB	Akay 2015, Altuvia 1997, Storz 2016, Tjaden 2006, etc.	Experimental; Northern blot (PMID: 9230301) Similarity (139.3 bit score)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
gene;sRNA	sroH	161	4190487	4190327	-	Rfam, RNAcentral, RegulonDB	Vogel 2003, Hobbs 2010, Mihailovic 2018	Experimental; Northern blot (PMID: 14602901), RACE (PMID: 14602901) Similarity (202 bit score)
ncRNA	naRNA4	77	4218439	4218516	+	Rfam	n/a	Similarity (85.6 bit score)
5' UTR sRNA	LysC5'	525	4231337	4230812	-	RegulonDB	Kawano 2005	Experimental; Northern blot (PMID: 15718303)
ncRNA	naRNA4	70	4235424	4235494	+	Rfam	n/a	Similarity (79.5 bit score)
ncRNA	STnc430	149	4235709	4235560	-	Rfam	n/a	Similarity (77.3 bit score)
ncRNA	naRNA4	70	4245095	4245165	+	Rfam	n/a	Similarity (67.4 bit score)
ncRNA	naRNA4	70	4249324	4249394	+	Rfam	n/a	Similarity (84.4 bit score)
gene;sRNA	SraL (ryjA, PsrA24)	141	4278066	4277926	-	Rfam, RegulonDB	Argaman 2001, Wassarman 2001	Experimental; Northern blot (PMID: 11448770) Similarity (151.2 bit score)
ncRNA	naRNA4	70	4285305	4285375	+	Rfam	n/a	Similarity (84.1 bit score)
ncRNA	naRNA4	70	4295911	4295981	+	Rfam	n/a	Similarity (71.4 bit score)
ncRNA	naRNA4	70	4296024	4296094	+	Rfam	n/a	Similarity (77.1 bit score)
ncRNA	naRNA4	70	4296137	4296207	+	Rfam	n/a	Similarity (71.4 bit score)
ncRNA	naRNA4	70	4296250	4296320	+	Rfam	n/a	Similarity (71.4 bit score)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
ncRNA	naRNA4	68	4296363	4296431	+	Rfam	n/a	Similarity (63.7 bit score)
ncRNA	naRNA4	80	4323249	4323329	+	Rfam	n/a	Similarity (69.1 bit score)
gene;sRNA	nc5	90	4323897	4324000	+	BSRD, RegulonDB	Macvanin 2012	Experimental; co-IP with HU, tiling arrays, Northern blot (PMID:22942248)
gene;sRNA	naRNA6	76	4325881	4325805	-	Rfam, RegulonDB	Qian 2015	Experimental; RNA-seq (PMID: 26307168) Similarity (106.7 bit score)
gene;sRNA	naRNA5	76	4325981	4325905	-	Rfam, RegulonDB	Qian 2015	Experimental; RNA-seq (PMID: 26307168) Similarity (110.8 bit score)
gene;sRNA	naRNA4	77	4326081	4326005	-	Rfam, RegulonDB	Qian 2015	Experimental; RNA-seq (PMID: 26307168) Similarity (111 bit score)
gene;sRNA	naRNA3	77	4326181	4326105	-	Rfam, RegulonDB	Qian 2015	Experimental; RNA-seq (PMID: 26307168) Similarity (106.8 bit score)
gene;sRNA	naRNA2	77	4326281	4326205	-	Rfam, RegulonDB	Qian 2015	Experimental; RNA-seq (PMID: 26307168) (Similarity (111 bit score)
gene;sRNA	naRNA1	77	4326381	4326305	-	Rfam, RegulonDB	Qian 2015	Experimental; RNA-seq (PMID: 26307168), Similarity (95 bit score)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
ncRNA	STnc630	165	4332047	4332212	+	Rfam	n/a	Similarity (181.8 bit score)
ncRNA	ES239	122	4436566	4436688	+	Rfam	n/a	Similarity (157.4 bit score)
5' UTR sRNA	YtfL5'	105	4439353	4439248	-	RegulonDB	Raghavan 2011	Experimental; RNA-seq (PMID: 21665928)
gene;sRNA	G0-10706	106	4441330	4441225	-	RegulonDB	Raghavan 2011	Experimental; RNA-seq (PMID: 21665928)
ncRNA	STnc450	57	4441334	4441277	-	Rfam	n/a	Similarity (77.4 bit score)
ncRNA	naRNA4	78	4450947	4451025	+	Rfam	n/a	Similarity (70.2 bit score)
ncRNA	naRNA4	77	4457188	4457265	+	Rfam	n/a	Similarity (86.2 bit score)
5' UTR sRNA	MgtA5'	379	4464820	4465199	+	RegulonDB	Kawano 2005	Experimental; Northern blot (PMID: 15718303)
gene;sRNA	ryjB	90	4527977	4528066	+	RegulonDB	Raghavan 2011, Kawano 2005	Experimental; RNAseq (PMID: 21665928)
3' UTR sRNA	FimA3	48	4541230	4541277	+	RegulonDB	Kawano 2005	Experimental; Northern blot (PMID: 15718303)
gene;sRNA	symR (RyjC)	77	4579835	4579911	+	BSRD, RegulonDB	Kawano 2005, Kawano 2007	Experimental; cDNA cloning-based screen, Northern blot (PMID: 15718303)
ncRNA	naRNA4	78	4614260	4614338	+	Rfam	n/a	Similarity (75.4 bit score)
ncRNA	naRNA4	78	4614361	4614439	+	Rfam	n/a	Similarity (76.7 bit score)
ncRNA	naRNA4	78	4614462	4614540	+	Rfam	n/a	Similarity (77.3 bit score)
ncRNA	naRNA4	78	4614563	4614641	+	Rfam	n/a	Similarity (80.4 bit score)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
ncRNA	naRNA4	78	4628698	4628776	+	Rfam	n/a	Similarity (66.1 bit score)
gene;sRNA	rdlB	66	1269858	1269923	+	BSRD, RegulonDB	Bak 2015, Kawano 2002, Kawano 2005	Experimental; cDNA cloning (PMID: 15718303) Overexpression (PMID: 26469694)
gene;sRNA	rdlC	68	1270393	1270460	+	BSRD, RegulonDB	Bak 2015, Kawano 2002, Kawano 2005	Experimental; cDNA cloning (PMID: 15718303) Overexpression (PMID: 26469694)
5' UTR sRNA	oppA5' (RNA0-359)	254	1298697	1298951	+	RegulonDB	Raghavan 2011	Experimental; RNA-seq (PMID: 21665928)
gene;sRNA	sokB	56	1492119	1492174	+	RegulonDB	Faridani 2006, Kawano 2005, Pedersen 1999, Schneider 2000, Woods 2006	Experimental; Northern blot (PMID: 10361310)
gene;sRNA	SraC (RyeA)	272	1923066	1923337	+	Rfam, RegulonDB	Chen 2002, Choi 2018, Gottesman 2001, Hayes 2006, Peano 2015, Wassarman 2001, Wu 2017	Experimental; Northern blot (PMID: 11448770) Similarity (174.2 bit score)
gene;sRNA	isrC (IS102)	195	2071317	2071511	+	RegulonDB, Rfam	Chen 2002, Wallecha 2014	Experimental; Northern blot (PMID: 12069726), Similarity (253.5 bit score)
gene;sRNA	sibA (ryeC, Tp11, QUAD1a)	144	2153311	2153454	+	BSRD, RegulonDB	Fozo 2008, Wassarman 2001	Experimental; Northern blot (PMID: 11445539), Microarray (PMID: 11445539)
								(continued)

Classification	Name	Size	Coordinates		Strand	Database	Articles	Validation Method
			Start	Stop				
gene;sRNA	ryfC (ohsC)	79	2700520	2700598	+	RegulonDB	Fozo 2008, Kawano 2005	Experimental; Northern blot (PMID: 15718303)
gene;sRNA	sibC (RygC, QUAD1c, T27)	141	3056851	3056991	+	BSRD, RegulonDB	Fozo 2008, Wassarman 2001	Experimental; Northern blot (PMID: 11445539), Microarray (PMID: 11445539)
gene;sRNA	arcZ (sraH, ryhA)	121	3350577	3350697	+	RegulonDB	Papenfort 2009, Chen 2018, Argaman 2001	Experimental; Northern blot (PMID: 11448770), Similarity (120.3 bit score)
gene;sRNA	agrB	82	3648294	3648377	+	RegulonDB	Weel-Sneve 2013, Kristiansenv2016	Experimental; Northern blot (PMID: 23408903)
3' UTR sRNA	gadF (ES205)	91	3658992	3659082	+	RegulonDB, Rfam	Melamed 2015	Experimental; Northern blot & RIL-seq (PMID: 27588604),
gene;sRNA	nc8	148	3670918	3671173	+	BSRD, RegulonDB	Macvanin 2012	Experimental; co-IP with HU, tiling arrays (PMID:22942248)
gene;sRNA	rdlD	64	3700136	3700199	+	RegulonDB	Kawano 2002, Kawano 2005	Experimental; cDNA cloning (PMID: 15718303)
gene;sRNA	cpxQ	58	4106330	4106387	+	RegulonDB	Grabowicz 2016	Experimental; Northern blot, RNAseq (PMID: 26805574)
gene;sRNA	pspH	111	4263139	4263250	+	RegulonDB	Melamed 2016, Thomason 2015	Experimental; RIL-seq (PMID: 27588604)
gene;sRNA	nc6	112	4566418	4566529	+	BSRD, RegulonDB	Macvanin 2012	Experimental; co-IP with HU (PMID:22942248), tiling arrays (PMID:22942248)
								(continued)

APPENDIX B

Total RNAspace predictions for Rhodobacter sphaeroides 2.4.1

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	240	22716	22916	201	+	BLAST/CG-seq/RNAz	0.908889
C1	267	25550	25637	88	+	BLAST/CG-seq/RNAz	0.734562
C1	382	30466	30519	54	+	BLAST/CG-seq/RNAz	0.940261
C1	306	33562	33633	72	+	BLAST/CG-seq/RNAz	0.874518
C1	331	35209	35286	78	+	BLAST/CG-seq/RNAz	0.990409
C1	178	54070	54224	155	+	BLAST/CG-seq/RNAz	0.995462
C1	615	63096	63200	105	+	BLAST/CG-seq/RNAz	0.994116
C1	220	69342	69571	230	-	BLAST/CG-seq/RNAz	0.888444
C1	242	69932	70000	69	+	BLAST/CG-seq/RNAz	0.947818
C1	212	83493	83659	167	+	BLAST/CG-seq/RNAz	0.859236
C1	716	84263	84322	60	+	BLAST/CG-seq/RNAz	0.769433
C1	754	87120	87302	183	+	BLAST/CG-seq/RNAz	0.884772
C1	527	95706	95772	67	-	BLAST/CG-seq/RNAz	0.744894
C1	408	99287	99487	201	+	BLAST/CG-seq/RNAz	0.999116
C1	410	99394	99594	201	-	BLAST/CG-seq/RNAz	0.972387
C1	444	100884	100957	74	-	BLAST/CG-seq/RNAz	0.913827
C1	711	128132	128366	235	+	BLAST/CG-seq/RNAz	0.969701
C1	595	138616	138696	81	-	BLAST/CG-seq/RNAz	0.986766
C1	667	142058	142258	201	-	BLAST/CG-seq/RNAz	0.928321
C1	279	147384	147584	201	+	BLAST/CG-seq/RNAz	0.996045
C1	325	151870	152001	132	+	BLAST/CG-seq/RNAz	0.96896
C1	361	153739	153814	76	-	BLAST/CG-seq/RNAz	0.730332
C1	374	155013	155112	100	-	BLAST/CG-seq/RNAz	0.909915
C1	83	156480	156680	201	+	BLAST/CG-seq/RNAz	0.985631

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	834	166863	166951	89	+	BLAST/CG-seq/RNAz	0.93362
C1	855	168659	168849	191	+	BLAST/CG-seq/RNAz	0.90719
C1	574	170512	170633	122	+	BLAST/CG-seq/RNAz	0.867596
C1	289	184576	184776	201	+	BLAST/CG-seq/RNAz	0.961206
C1	776	193822	193887	66	+	BLAST/CG-seq/RNAz	0.948871
C1	770	194836	194982	147	+	BLAST/CG-seq/RNAz	0.954439
C1	385	201352	201420	69	+	BLAST/CG-seq/RNAz	0.986347
C1	66	207862	207927	66	+	BLAST/CG-seq/RNAz	0.968372
C1	552	211052	211247	196	+	BLAST/CG-seq/RNAz	0.905605
C1	548	212202	212304	103	+	BLAST/CG-seq/RNAz	0.909789
C1	546	213032	213232	201	+	BLAST/CG-seq/RNAz	0.988245
C1	177	223411	223464	54	-	BLAST/CG-seq/RNAz	0.987026
C1	649	224244	224454	211	+	BLAST/CG-seq/RNAz	0.999543
C1	343	234649	234718	70	-	BLAST/CG-seq/RNAz	0.985822
C1	831	237981	238033	53	+	BLAST/CG-seq/RNAz	0.878867
C1	828	238076	238133	58	-	BLAST/CG-seq/RNAz	0.9753
C1	421	243701	243900	200	-	BLAST/CG-seq/RNAz	0.86869
C1	860	243747	243850	104	-	INFERNAL	0.000000325
C1	118	246780	246878	99	+	BLAST/CG-seq/RNAz	0.891578
C1	590	248150	248363	214	+	BLAST/CG-seq/RNAz	0.70644
C1	383	260087	260170	84	+	BLAST/CG-seq/RNAz	0.922946
C1	778	268150	268190	41	+	BLAST/CG-seq/RNAz	0.996995
C1	70	280045	280091	47	-	BLAST/CG-seq/RNAz	0.864264
C1	256	288989	289167	179	+	BLAST/CG-seq/RNAz	0.921305
C1	733	290238	290399	162	+	BLAST/CG-seq/RNAz	0.95451
C1	646	292181	292442	262	+	BLAST/CG-seq/RNAz	0.999726
C1	347	293382	293625	244	+	BLAST/CG-seq/RNAz	0.968849

(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	342	297760	297840	81	+	BLAST/CG-seq/RNAz	0.923387
C1	339	302089	302153	65	+	BLAST/CG-seq/RNAz	0.915244
C1	821	306416	306489	74	+	BLAST/CG-seq/RNAz	0.999302
C1	493	307666	307737	72	+	BLAST/CG-seq/RNAz	0.988076
C1	424	309686	309807	122	+	BLAST/CG-seq/RNAz	0.98572
C1	422	312458	312696	239	+	BLAST/CG-seq/RNAz	0.997518
C1	124	315203	315254	52	+	BLAST/CG-seq/RNAz	0.853692
C1	116	317536	317656	121	+	BLAST/CG-seq/RNAz	0.997585
C1	589	319659	319832	174	+	BLAST/CG-seq/RNAz	0.995167
C1	71	319771	319864	94	+	INFERNAL	0.00412
C1	588	320605	320719	115	+	BLAST/CG-seq/RNAz	0.998697
C1	496	322257	322562	306	+	BLAST/CG-seq/RNAz	0.967807
C1	217	323127	323181	55	+	BLAST/CG-seq/RNAz	0.93218
C1	691	329313	329460	148	+	BLAST/CG-seq/RNAz	0.997921
C1	380	333485	333740	256	+	BLAST/CG-seq/RNAz	0.988789
C1	58	337598	337798	201	+	BLAST/CG-seq/RNAz	0.903576
C1	532	343222	343331	110	+	BLAST/CG-seq/RNAz	0.993107
C1	251	346847	346888	42	+	BLAST/CG-seq/RNAz	0.98395
C1	247	347972	348175	204	+	BLAST/CG-seq/RNAz	0.823268
C1	245	348752	348866	115	+	BLAST/CG-seq/RNAz	0.897797
C1	150	351336	351534	199	+	BLAST/CG-seq/RNAz	0.998377
C1	640	353360	353469	110	+	BLAST/CG-seq/RNAz	0.721976
C1	639	354472	354738	267	-	BLAST/CG-seq/RNAz	0.921043
C1	633	355696	355785	90	+	BLAST/CG-seq/RNAz	0.953721
C1	815	362452	362550	99	+	BLAST/CG-seq/RNAz	0.962137
C1	414	376047	376153	107	+	BLAST/CG-seq/RNAz	0.956445

(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	105	380685	380796	112	-	BLAST/CG-seq/RNAz	0.827109
C1	585	382090	382195	106	+	BLAST/CG-seq/RNAz	0.999946
C1	678	395470	395560	91	+	BLAST/CG-seq/RNAz	0.964932
C1	673	401769	401850	82	+	BLAST/CG-seq/RNAz	0.739576
C1	281	410229	410423	195	+	BLAST/CG-seq/RNAz	0.859255
C1	276	412725	412970	246	+	BLAST/CG-seq/RNAz	0.756037
C1	449	421683	421820	138	+	BLAST/CG-seq/RNAz	0.957951
C1	447	425730	425929	200	+	BLAST/CG-seq/RNAz	0.823943
C1	51	430755	430834	80	-	BLAST/CG-seq/RNAz	0.753709
C1	727	439408	439608	201	+	BLAST/CG-seq/RNAz	0.999596
C1	5	444939	445015	77	+	BLAST/Rfam_10.0_seed	2E-28
C1	298	444939	445015	77	+	INFERNAL	4.55E-13
C1	635	446246	446340	95	+	BLAST/CG-seq/RNAz	0.865384
C1	812	458545	458803	259	+	BLAST/CG-seq/RNAz	0.998088
C1	810	459485	459583	99	+	BLAST/CG-seq/RNAz	0.837129
C1	484	462910	463007	98	+	BLAST/CG-seq/RNAz	0.985679
C1	416	463227	463298	72	+	BLAST/CG-seq/RNAz	0.986006
C1	583	471550	471686	137	+	BLAST/CG-seq/RNAz	0.72903
C1	581	472470	472648	179	+	BLAST/CG-seq/RNAz	0.907337
C1	580	473814	474019	206	-	BLAST/CG-seq/RNAz	0.877842
C1	676	479730	479930	201	-	BLAST/CG-seq/RNAz	0.812476
C1	672	481331	481399	69	+	BLAST/CG-seq/RNAz	0.795657
C1	375	482477	482527	51	+	BLAST/CG-seq/RNAz	0.974332
C1	367	484790	484968	179	+	BLAST/CG-seq/RNAz	0.93901
C1	857	485137	485210	74	+	BLAST/CG-seq/RNAz	0.989898
C1	768	485697	485744	48	+	BLAST/CG-seq/RNAz	0.822556
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	446	490719	490919	201	-	BLAST/CG-seq/RNAz	0.870058
C1	629	505366	505566	201	+	BLAST/CG-seq/RNAz	0.970585
C1	628	507745	507997	253	+	BLAST/CG-seq/RNAz	0.80189
C1	407	522305	522387	83	+	BLAST/CG-seq/RNAz	0.726653
C1	404	524515	524594	80	+	BLAST/CG-seq/RNAz	0.833399
C1	100	527219	527486	268	-	BLAST/CG-seq/RNAz	0.938183
C1	575	532178	532430	253	+	BLAST/CG-seq/RNAz	0.978915
C1	567	534498	534612	115	+	BLAST/CG-seq/RNAz	0.999979
C1	216	536637	536887	251	-	INFERNAL	4.24E-20
C1	4	536753	536839	87	-	BLAST/Rfam_10.0_seed	0.000001
C1	474	538006	538118	113	+	BLAST/CG-seq/RNAz	0.9416
C1	200	542901	543042	142	-	BLAST/CG-seq/RNAz	0.879492
C1	366	547928	548196	269	+	BLAST/CG-seq/RNAz	0.780572
C1	363	548416	548580	165	+	BLAST/CG-seq/RNAz	0.75673
C1	271	550593	550793	201	+	BLAST/CG-seq/RNAz	0.988392
C1	753	554557	554666	110	-	BLAST/CG-seq/RNAz	0.88537
C1	445	557733	557927	195	-	BLAST/CG-seq/RNAz	0.878347
C1	521	565034	565093	60	-	BLAST/CG-seq/RNAz	0.824936
C1	518	570493	570597	105	+	BLAST/CG-seq/RNAz	0.966167
C1	232	571566	571673	108	+	BLAST/CG-seq/RNAz	0.896021
C1	226	573834	573932	99	+	BLAST/CG-seq/RNAz	0.745977
C1	717	578313	578444	132	+	BLAST/CG-seq/RNAz	0.998022
C1	327	581248	581288	41	-	BLAST/CG-seq/RNAz	0.93442
C1	319	583914	583979	66	-	BLAST/CG-seq/RNAz	0.968749
C1	472	592242	592347	106	+	BLAST/CG-seq/RNAz	0.983876
C1	95	600854	601031	178	+	BLAST/CG-seq/RNAz	0.95551
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	571	608197	608378	182	-	BLAST/CG-seq/RNAz	0.717416
C1	661	618690	618864	175	-	BLAST/CG-seq/RNAz	0.872106
C1	365	620088	620294	207	-	BLAST/CG-seq/RNAz	0.955156
C1	853	620854	621109	256	-	BLAST/CG-seq/RNAz	0.777545
C1	751	624398	624453	56	+	BLAST/CG-seq/RNAz	0.999445
C1	230	646166	646219	54	-	BLAST/CG-seq/RNAz	0.991722
C1	718	647000	647223	224	+	BLAST/CG-seq/RNAz	0.999127
C1	402	656814	656899	86	+	BLAST/CG-seq/RNAz	0.926095
C1	398	660160	660229	70	-	BLAST/CG-seq/RNAz	0.927698
C1	93	666760	666865	106	+	BLAST/CG-seq/RNAz	0.999999
C1	89	668098	668299	202	-	BLAST/CG-seq/RNAz	0.738042
C1	562	669627	669783	157	-	BLAST/CG-seq/RNAz	0.96842
C1	263	670042	670137	96	+	BLAST/CG-seq/RNAz	0.999559
C1	657	674356	674438	83	-	BLAST/CG-seq/RNAz	0.997872
C1	516	711577	711654	78	+	BLAST/CG-seq/RNAz	0.863489
C1	510	714785	714963	179	-	BLAST/CG-seq/RNAz	0.870564
C1	222	717771	717884	114	+	BLAST/CG-seq/RNAz	0.992488
C1	708	725078	725172	95	+	BLAST/CG-seq/RNAz	0.98274
C1	310	727902	728132	231	+	BLAST/CG-seq/RNAz	0.928348
C1	305	735141	735275	135	-	BLAST/CG-seq/RNAz	0.848948
C1	793	739435	739600	166	+	BLAST/CG-seq/RNAz	0.971455
C1	790	740060	740161	102	+	BLAST/CG-seq/RNAz	0.940299
C1	469	744364	744514	151	-	BLAST/CG-seq/RNAz	0.985774
C1	85	761504	761697	194	-	BLAST/CG-seq/RNAz	0.99444
C1	656	776197	776290	94	-	BLAST/CG-seq/RNAz	0.915088
C1	842	785026	785073	48	+	BLAST/CG-seq/RNAz	0.99047
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	840	786954	787057	104	-	BLAST/CG-seq/RNAz	0.978613
C1	434	794212	794272	61	+	BLAST/CG-seq/RNAz	0.991697
C1	134	798864	799064	201	+	BLAST/CG-seq/RNAz	0.996318
C1	136	798977	799177	201	-	BLAST/CG-seq/RNAz	0.949266
C1	606	799405	799586	182	+	BLAST/CG-seq/RNAz	0.999163
C1	181	804587	804674	88	+	INFERNAL	3.13E-09
C1	6	804597	804673	77	+	BLAST/Rfam_10.0_seed	6E-10
C1	182	804675	804751	77	+	INFERNAL	0.0000123
C1	511	809190	809378	189	+	BLAST/CG-seq/RNAz	0.962187
C1	798	832594	832694	101	-	BLAST/CG-seq/RNAz	0.954339
C1	427	864966	865103	138	+	BLAST/CG-seq/RNAz	0.942068
C1	125	872036	872140	105	+	BLAST/CG-seq/RNAz	0.993605
C1	122	874031	874120	90	+	BLAST/CG-seq/RNAz	0.816403
C1	704	885192	885480	289	-	BLAST/CG-seq/RNAz	0.711917
C1	598	890702	890884	183	+	BLAST/CG-seq/RNAz	0.993972
C1	41	890713	891110	398	+	BLAST/Rfam_10.0_seed	0
C1	268	890713	891110	398	+	INFERNAL	1.72E-86
C1	599	890984	891182	199	+	BLAST/CG-seq/RNAz	0.893004
C1	785	904526	904778	253	+	BLAST/CG-seq/RNAz	0.796469
C1	780	905400	905642	243	+	BLAST/CG-seq/RNAz	0.838171
C1	459	914509	914708	200	-	BLAST/CG-seq/RNAz	0.997845
C1	78	918808	919109	302	+	BLAST/CG-seq/RNAz	0.759894
C1	72	923955	924105	151	+	BLAST/CG-seq/RNAz	0.999483
C1	551	928449	928673	225	-	BLAST/CG-seq/RNAz	0.807833
C1	260	929568	929717	150	-	BLAST/CG-seq/RNAz	0.733212
C1	259	930636	930836	201	-	BLAST/CG-seq/RNAz	0.965343
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	648	937737	937910	174	-	BLAST/CG-seq/RNAz	0.827066
C1	350	939152	939197	46	-	BLAST/CG-seq/RNAz	0.717629
C1	830	953697	953956	260	+	BLAST/CG-seq/RNAz	0.763177
C1	738	956778	956978	201	+	BLAST/CG-seq/RNAz	0.871287
C1	121	964467	964540	74	+	BLAST/CG-seq/RNAz	0.835653
C1	596	964793	964910	118	+	BLAST/CG-seq/RNAz	0.998376
C1	501	967238	967340	103	+	BLAST/CG-seq/RNAz	0.793437
C1	703	989253	989482	230	-	BLAST/CG-seq/RNAz	0.916275
C1	697	990266	990458	193	-	BLAST/CG-seq/RNAz	0.970483
C1	389	992913	993095	183	+	BLAST/CG-seq/RNAz	0.963695
C1	782	1001915	1002000	86	+	BLAST/CG-seq/RNAz	0.989918
C1	460	1007161	1007330	170	+	BLAST/CG-seq/RNAz	0.998499
C1	180	1009754	1009874	121	+	BLAST/CG-seq/RNAz	0.946782
C1	74	1011310	1011392	83	+	BLAST/CG-seq/RNAz	0.999981
C1	68	1013704	1013905	202	+	BLAST/CG-seq/RNAz	0.828327
C1	62	1014846	1014919	74	+	INFERNAL	0.000725
C1	554	1014846	1015046	201	+	BLAST/CG-seq/RNAz	0.997047
C1	737	1021120	1021212	93	-	BLAST/CG-seq/RNAz	0.901678
C1	734	1023645	1023718	74	+	BLAST/CG-seq/RNAz	0.984083
C1	1	1030432	1030633	202	+	BLAST/Rfam_10.0_seed	2E-97
C1	106	1030432	1030632	201	+	INFERNAL	5.21E-34
C1	111	1030563	1030759	197	+	BLAST/CG-seq/RNAz	0.854286
C1	587	1035833	1035887	55	-	BLAST/CG-seq/RNAz	0.938399
C1	498	1039123	1039282	160	+	BLAST/CG-seq/RNAz	0.800408
C1	495	1041538	1041621	84	+	BLAST/CG-seq/RNAz	0.836594
C1	687	1055917	1056195	279	+	BLAST/CG-seq/RNAz	0.939774
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	680	1056706	1056892	187	+	BLAST/CG-seq/RNAz	0.999603
C1	379	1057643	1057723	81	-	BLAST/CG-seq/RNAz	0.953349
C1	774	1064572	1064616	45	+	BLAST/CG-seq/RNAz	0.877721
C1	543	1094477	1094534	58	+	BLAST/CG-seq/RNAz	0.970477
C1	172	1103368	1103565	198	+	BLAST/CG-seq/RNAz	0.994688
C1	645	1106306	1106500	195	+	BLAST/CG-seq/RNAz	0.997999
C1	642	1106819	1106864	46	-	BLAST/CG-seq/RNAz	0.967871
C1	334	1120551	1120751	201	-	BLAST/CG-seq/RNAz	0.978586
C1	823	1121535	1121626	92	+	BLAST/CG-seq/RNAz	0.998658
C1	814	1125361	1125582	222	+	BLAST/CG-seq/RNAz	0.87799
C1	420	1128389	1128536	148	-	BLAST/CG-seq/RNAz	0.931311
C1	418	1129630	1129870	241	+	BLAST/CG-seq/RNAz	0.987531
C1	114	1132773	1132994	222	+	BLAST/CG-seq/RNAz	0.997762
C1	109	1133172	1133241	70	+	BLAST/CG-seq/RNAz	0.990045
C1	210	1151865	1151954	90	-	BLAST/CG-seq/RNAz	0.993851
C1	682	1154309	1154403	95	+	BLAST/CG-seq/RNAz	0.918854
C1	376	1163795	1163856	62	+	BLAST/CG-seq/RNAz	0.772344
C1	772	1182824	1182901	78	+	BLAST/CG-seq/RNAz	0.795765
C1	765	1186566	1186659	94	+	BLAST/CG-seq/RNAz	0.891649
C1	451	1190925	1190993	69	+	BLAST/CG-seq/RNAz	0.943287
C1	64	1198274	1198474	201	+	BLAST/CG-seq/RNAz	0.990811
C1	56	1198730	1198809	80	+	BLAST/CG-seq/RNAz	0.991486
C1	538	1204581	1204781	201	+	BLAST/CG-seq/RNAz	0.99877
C1	253	1214637	1214782	146	-	BLAST/CG-seq/RNAz	0.893385
C1	341	1220082	1220313	232	-	BLAST/CG-seq/RNAz	0.922058
C1	827	1224021	1224210	190	+	BLAST/CG-seq/RNAz	0.992079
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	206	1238534	1238697	164	-	BLAST/CG-seq/RNAz	0.987434
C1	674	1245676	1245733	58	+	BLAST/CG-seq/RNAz	0.959368
C1	277	1251225	1251447	223	+	BLAST/CG-seq/RNAz	0.996118
C1	272	1254739	1254808	70	-	BLAST/CG-seq/RNAz	0.749954
C1	760	1254974	1255192	219	+	BLAST/CG-seq/RNAz	0.987565
C1	49	1259410	1259519	110	-	BLAST/CG-seq/RNAz	0.90255
C1	528	1264885	1265034	150	+	BLAST/CG-seq/RNAz	0.935787
C1	637	1276266	1276316	51	+	BLAST/CG-seq/RNAz	0.890929
C1	631	1276878	1276959	82	+	BLAST/CG-seq/RNAz	0.852304
C1	626	1278715	1278914	200	+	BLAST/CG-seq/RNAz	0.963488
C1	811	1282550	1282750	201	-	BLAST/CG-seq/RNAz	0.821487
C1	479	1287551	1287816	266	+	BLAST/CG-seq/RNAz	0.905389
C1	103	1297498	1297592	95	+	BLAST/CG-seq/RNAz	0.937053
C1	577	1305653	1305852	200	+	BLAST/CG-seq/RNAz	0.895508
C1	205	1311810	1311937	128	-	BLAST/CG-seq/RNAz	0.929024
C1	668	1317261	1317462	202	+	BLAST/CG-seq/RNAz	0.725785
C1	369	1318033	1318233	201	+	BLAST/CG-seq/RNAz	0.78738
C1	758	1328552	1328772	221	+	BLAST/CG-seq/RNAz	0.972288
C1	146	1336531	1336596	66	+	BLAST/CG-seq/RNAz	0.984397
C1	526	1343396	1343494	99	-	BLAST/CG-seq/RNAz	0.839761
C1	244	1347057	1347297	241	+	BLAST/CG-seq/RNAz	0.776979
C1	720	1354225	1354346	122	-	BLAST/CG-seq/RNAz	0.839831
C1	477	1369416	1369646	231	+	BLAST/CG-seq/RNAz	0.881417
C1	141	1415842	1415975	134	+	BLAST/CG-seq/RNAz	0.967752
C1	519	1420336	1420485	150	+	BLAST/CG-seq/RNAz	0.999279
C1	228	1425642	1425708	67	+	BLAST/CG-seq/RNAz	0.884419
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	715	1429957	1430144	188	-	BLAST/CG-seq/RNAz	0.972677
C1	618	1431976	1432047	72	+	BLAST/CG-seq/RNAz	0.983883
C1	307	1436799	1436947	149	-	BLAST/CG-seq/RNAz	0.774811
C1	801	1437659	1437739	81	+	BLAST/CG-seq/RNAz	0.989317
C1	799	1439113	1439311	199	+	BLAST/CG-seq/RNAz	0.962011
C1	795	1440208	1440288	81	+	BLAST/CG-seq/RNAz	0.997328
C1	401	1451299	1451496	198	-	BLAST/CG-seq/RNAz	0.883103
C1	97	1453358	1453616	259	+	BLAST/CG-seq/RNAz	0.920633
C1	566	1459540	1459736	197	-	BLAST/CG-seq/RNAz	0.854652
C1	2	1481651	1481845	195	+	BLAST/Rfam_10.0_seed	2E-92
C1	108	1481651	1481845	195	+	INFERNAL	8.86E-19
C1	359	1493277	1493477	201	-	BLAST/CG-seq/RNAz	0.988603
C1	360	1493525	1493725	201	+	BLAST/CG-seq/RNAz	0.87273
C1	851	1495380	1495578	199	+	BLAST/CG-seq/RNAz	0.991735
C1	439	1508773	1508823	51	+	BLAST/CG-seq/RNAz	0.991935
C1	139	1513820	1513984	165	+	BLAST/CG-seq/RNAz	0.992574
C1	852	1518246	1518390	145	-	BLAST/CG-seq/RNAz	0.905173
C1	517	1521996	1522036	41	-	BLAST/CG-seq/RNAz	0.988458
C1	713	1534990	1535145	156	+	BLAST/CG-seq/RNAz	0.998503
C1	617	1537763	1537937	175	-	BLAST/CG-seq/RNAz	0.814026
C1	315	1542969	1543201	233	+	BLAST/CG-seq/RNAz	0.973315
C1	400	1568330	1568612	283	+	BLAST/CG-seq/RNAz	0.920508
C1	91	1573856	1573970	115	+	BLAST/CG-seq/RNAz	0.791993
C1	569	1575441	1575699	259	-	BLAST/CG-seq/RNAz	0.904958
C1	563	1582321	1582395	75	-	BLAST/CG-seq/RNAz	0.75312
C1	848	1598989	1599081	93	+	BLAST/CG-seq/RNAz	0.923847
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	509	1602973	1603148	176	+	BLAST/CG-seq/RNAz	0.993469
C1	219	1616401	1616478	78	+	BLAST/CG-seq/RNAz	0.964628
C1	608	1622420	1622489	70	-	BLAST/CG-seq/RNAz	0.848696
C1	603	1624233	1624307	75	+	BLAST/CG-seq/RNAz	0.828853
C1	304	1627888	1628036	149	-	BLAST/CG-seq/RNAz	0.994028
C1	299	1631424	1631579	156	-	BLAST/CG-seq/RNAz	0.840815
C1	791	1632474	1632674	201	+	BLAST/CG-seq/RNAz	0.930592
C1	465	1635559	1635615	57	-	BLAST/CG-seq/RNAz	0.823662
C1	462	1635983	1636220	238	+	BLAST/CG-seq/RNAz	0.949791
C1	859	1639296	1639375	80	+	INFERNAL	0.0000242
C1	394	1639319	1639389	71	+	BLAST/CG-seq/RNAz	0.946187
C1	86	1639684	1639714	31	+	BLAST/CG-seq/RNAz	0.828366
C1	81	1642261	1642468	208	+	BLAST/CG-seq/RNAz	0.839229
C1	560	1644592	1644693	102	+	BLAST/CG-seq/RNAz	0.800878
C1	556	1652206	1652377	172	-	BLAST/CG-seq/RNAz	0.901236
C1	188	1654437	1654642	206	+	BLAST/CG-seq/RNAz	0.969051
C1	653	1660942	1661142	201	-	BLAST/CG-seq/RNAz	0.957099
C1	354	1663204	1663398	195	-	BLAST/CG-seq/RNAz	0.905319
C1	351	1664053	1664253	201	-	BLAST/CG-seq/RNAz	0.865607
C1	836	1668328	1668421	94	-	BLAST/CG-seq/RNAz	0.874912
C1	433	1674583	1674692	110	-	BLAST/CG-seq/RNAz	0.933107
C1	507	1701747	1701947	201	+	BLAST/CG-seq/RNAz	0.93514
C1	506	1703217	1703367	151	+	BLAST/CG-seq/RNAz	0.9496
C1	707	1710756	1710807	52	+	BLAST/CG-seq/RNAz	0.75597
C1	300	1721183	1721272	90	+	BLAST/CG-seq/RNAz	0.987651
C1	296	1722121	1722321	201	+	BLAST/CG-seq/RNAz	0.997234
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	788	1723272	1723473	202	+	BLAST/CG-seq/RNAz	0.927118
C1	787	1725377	1725567	191	+	BLAST/CG-seq/RNAz	0.959146
C1	464	1729093	1729319	227	+	BLAST/CG-seq/RNAz	0.977009
C1	79	1736044	1736178	135	+	BLAST/CG-seq/RNAz	0.947893
C1	261	1745168	1745368	201	+	BLAST/CG-seq/RNAz	0.853777
C1	655	1750911	1751006	96	-	BLAST/CG-seq/RNAz	0.769362
C1	832	1775273	1775465	193	+	BLAST/CG-seq/RNAz	0.93928
C1	747	1776948	1777185	238	+	BLAST/CG-seq/RNAz	0.778426
C1	82	1798478	1798670	193	-	BLAST/CG-seq/RNAz	0.815171
C1	372	1800977	1801122	146	+	BLAST/CG-seq/RNAz	0.999631
C1	499	1802395	1802495	101	+	BLAST/CG-seq/RNAz	0.99782
C1	148	1809066	1809215	150	+	BLAST/CG-seq/RNAz	0.910185
C1	430	1819300	1819534	235	-	BLAST/CG-seq/RNAz	0.838199
C1	677	1837107	1837282	176	-	BLAST/CG-seq/RNAz	0.897317
C1	829	1837916	1838029	114	+	BLAST/CG-seq/RNAz	0.747813
C1	613	1846484	1846718	235	+	BLAST/CG-seq/RNAz	0.957003
C1	399	1855369	1855524	156	+	BLAST/CG-seq/RNAz	0.841236
C1	524	1857135	1857333	199	-	BLAST/CG-seq/RNAz	0.878491
C1	191	1867308	1867509	202	+	BLAST/CG-seq/RNAz	0.758968
C1	273	1880812	1881033	222	-	BLAST/CG-seq/RNAz	0.999977
C1	349	1887762	1887906	145	+	BLAST/CG-seq/RNAz	0.986648
C1	659	1889194	1889329	136	+	BLAST/CG-seq/RNAz	0.989023
C1	808	1890926	1891117	192	+	BLAST/CG-seq/RNAz	0.962614
C1	174	1892060	1892095	36	+	BLAST/CG-seq/RNAz	0.954313
C1	731	1902145	1902281	137	+	BLAST/CG-seq/RNAz	0.976826
C1	223	1903716	1903755	40	+	BLAST/CG-seq/RNAz	0.94469
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	664	1909858	1910058	201	-	BLAST/CG-seq/RNAz	0.799645
C1	665	1909974	1910174	201	+	BLAST/CG-seq/RNAz	0.973944
C1	330	1911019	1911142	124	+	BLAST/CG-seq/RNAz	0.890941
C1	456	1912409	1912532	124	+	BLAST/CG-seq/RNAz	0.857544
C1	600	1913166	1913331	166	-	BLAST/CG-seq/RNAz	0.939178
C1	7	1920604	1920656	53	+	BLAST/Rfam_10.0_seed	0.000000002
C1	138	1930289	1930326	38	-	INFERNAL	0.00836
C1	426	1963553	1963690	138	-	BLAST/CG-seq/RNAz	0.780225
C1	721	1965447	1965669	223	+	BLAST/CG-seq/RNAz	0.978083
C1	50	1966699	1966899	201	-	BLAST/CG-seq/RNAz	0.809044
C1	473	1972061	1972205	145	-	BLAST/CG-seq/RNAz	0.956273
C1	739	1980219	1980391	173	+	BLAST/CG-seq/RNAz	0.989766
C1	87	1982421	1982543	123	+	BLAST/CG-seq/RNAz	0.716478
C1	654	1988258	1988409	152	+	BLAST/CG-seq/RNAz	0.751939
C1	152	1990440	1990654	215	+	BLAST/CG-seq/RNAz	0.932721
C1	317	1993006	1993165	160	+	BLAST/CG-seq/RNAz	0.998258
C1	593	2000996	2001243	248	+	BLAST/CG-seq/RNAz	0.938964
C1	746	2005123	2005228	106	-	BLAST/CG-seq/RNAz	0.990176
C1	101	2043018	2043211	194	+	BLAST/CG-seq/RNAz	0.996589
C1	255	2044592	2044792	201	+	BLAST/CG-seq/RNAz	0.930622
C1	391	2045733	2045780	48	+	BLAST/CG-seq/RNAz	0.976427
C1	540	2052779	2052927	149	+	BLAST/CG-seq/RNAz	0.92024
C1	847	2054554	2054764	211	+	BLAST/CG-seq/RNAz	0.923271
C1	54	2075864	2076021	158	+	BLAST/CG-seq/RNAz	0.855429
C1	357	2078465	2078654	190	+	BLAST/CG-seq/RNAz	0.951032
C1	476	2081364	2081416	53	+	BLAST/CG-seq/RNAz	0.896109
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	90	2086209	2086360	152	-	BLAST/CG-seq/RNAz	0.998695
C1	320	2098335	2098496	162	+	BLAST/CG-seq/RNAz	0.982788
C1	685	2106436	2106494	59	+	BLAST/CG-seq/RNAz	0.985581
C1	322	2111502	2111563	62	+	BLAST/CG-seq/RNAz	0.932174
C1	346	2112119	2112244	126	-	BLAST/CG-seq/RNAz	0.961923
C1	467	2112695	2112840	146	+	BLAST/CG-seq/RNAz	0.999989
C1	763	2116027	2116093	67	+	BLAST/CG-seq/RNAz	0.94506
C1	837	2121925	2122210	286	-	BLAST/CG-seq/RNAz	0.840299
C1	302	2151118	2151193	76	+	BLAST/CG-seq/RNAz	0.939042
C1	735	2153108	2153175	68	-	BLAST/CG-seq/RNAz	0.813984
C1	513	2161130	2161202	73	-	BLAST/CG-seq/RNAz	0.971335
C1	605	2165575	2165779	205	-	BLAST/CG-seq/RNAz	0.939862
C1	239	2174827	2174979	153	-	BLAST/CG-seq/RNAz	0.892436
C1	377	2176161	2176304	144	+	BLAST/CG-seq/RNAz	0.851612
C1	436	2185452	2185652	201	+	BLAST/CG-seq/RNAz	0.986903
C1	396	2193148	2193348	201	+	BLAST/CG-seq/RNAz	0.997869
C1	523	2197629	2197786	158	-	BLAST/CG-seq/RNAz	0.798038
C1	689	2198955	2199084	130	+	BLAST/CG-seq/RNAz	0.988509
C1	530	2215124	2215210	87	+	BLAST/CG-seq/RNAz	0.958972
C1	838	2220422	2220483	62	+	BLAST/CG-seq/RNAz	0.988633
C1	274	2235778	2235978	201	+	BLAST/CG-seq/RNAz	0.988465
C1	508	2250544	2250709	166	+	BLAST/CG-seq/RNAz	0.7821
C1	816	2259382	2259581	200	+	BLAST/CG-seq/RNAz	0.763376
C1	176	2262340	2262426	87	+	BLAST/CG-seq/RNAz	0.825456
C1	736	2274936	2275181	246	+	BLAST/CG-seq/RNAz	0.958713
C1	841	2283433	2283677	245	-	BLAST/CG-seq/RNAz	0.970528
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	196	2285425	2285703	279	+	BLAST/CG-seq/RNAz	0.95603
C1	258	2291779	2291842	64	-	BLAST/CG-seq/RNAz	0.844037
C1	755	2307507	2307700	194	+	BLAST/CG-seq/RNAz	0.95269
C1	60	2318496	2318561	66	+	BLAST/CG-seq/RNAz	0.987289
C1	362	2321373	2321569	197	-	BLAST/CG-seq/RNAz	0.996369
C1	481	2323208	2323279	72	-	BLAST/CG-seq/RNAz	0.760597
C1	783	2325367	2325540	174	+	BLAST/CG-seq/RNAz	0.748035
C1	502	2332426	2332515	90	+	BLAST/CG-seq/RNAz	0.93987
C1	658	2332807	2333056	250	-	BLAST/CG-seq/RNAz	0.838212
C1	803	2333264	2333386	123	-	BLAST/CG-seq/RNAz	0.76781
C1	189	2335923	2336006	84	+	BLAST/CG-seq/RNAz	0.994181
C1	234	2368363	2368499	137	+	BLAST/CG-seq/RNAz	0.82233
C1	698	2374152	2374237	86	-	BLAST/CG-seq/RNAz	0.943435
C1	845	2379953	2380191	239	+	BLAST/CG-seq/RNAz	0.981709
C1	535	2388506	2388603	98	-	INFERNAL	2.81E-12
C1	45	2388523	2388576	54	-	BLAST/Rfam_10.0_seed	0.0002
C1	333	2392392	2392458	67	-	BLAST/CG-seq/RNAz	0.954819
C1	705	2409984	2410067	84	-	BLAST/CG-seq/RNAz	0.890356
C1	757	2422974	2423150	177	+	BLAST/CG-seq/RNAz	0.939683
C1	431	2425229	2425327	99	+	BLAST/CG-seq/RNAz	0.996568
C1	482	2432169	2432274	106	+	BLAST/CG-seq/RNAz	0.965061
C1	643	2436544	2436777	234	+	BLAST/CG-seq/RNAz	0.987817
C1	592	2442083	2442381	299	+	BLAST/CG-seq/RNAz	0.966601
C1	744	2450515	2450682	168	+	BLAST/CG-seq/RNAz	0.999189
C1	42	2455876	2455977	102	+	BLAST/Rfam_10.0_seed	2E-42
C1	455	2455877	2455976	100	+	INFERNAL	7.16E-13
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	804	2460088	2460327	240	+	BLAST/CG-seq/RNAz	0.986901
C1	192	2461817	2462119	303	+	BLAST/CG-seq/RNAz	0.983759
C1	120	2464356	2464628	273	+	BLAST/CG-seq/RNAz	0.990155
C1	525	2468463	2468736	274	-	BLAST/CG-seq/RNAz	0.947078
C1	536	2490584	2490682	99	+	BLAST/CG-seq/RNAz	0.939146
C1	695	2491406	2491547	142	+	BLAST/CG-seq/RNAz	0.700885
C1	797	2495702	2495901	200	+	BLAST/CG-seq/RNAz	0.766515
C1	65	2495722	2495877	156	+	INFERNAL	0.000428
C1	143	2497132	2497338	207	+	BLAST/CG-seq/RNAz	0.940372
C1	254	2503846	2503934	89	-	BLAST/CG-seq/RNAz	0.92718
C1	819	2509055	2509291	237	+	BLAST/CG-seq/RNAz	0.964731
C1	183	2510972	2511116	145	+	BLAST/CG-seq/RNAz	0.981994
C1	3	2531207	2531414	208	+	BLAST/Rfam_10.0_seed	1E-101
C1	107	2531207	2531414	208	+	INFERNAL	6.55E-28
C1	504	2544521	2544579	59	+	BLAST/CG-seq/RNAz	0.929883
C1	295	2552367	2552580	214	-	BLAST/CG-seq/RNAz	0.803266
C1	195	2565728	2565928	201	-	BLAST/CG-seq/RNAz	0.767716
C1	699	2579042	2579242	201	+	BLAST/CG-seq/RNAz	0.824907
C1	492	2584697	2584855	159	+	BLAST/CG-seq/RNAz	0.941121
C1	651	2586705	2586809	105	-	BLAST/CG-seq/RNAz	0.884771
C1	283	2592751	2592904	154	+	BLAST/CG-seq/RNAz	0.975827
C1	564	2598156	2598356	201	+	BLAST/CG-seq/RNAz	0.765913
C1	723	2599947	2599997	51	+	BLAST/CG-seq/RNAz	0.984351
C1	52	2600127	2600291	165	+	BLAST/CG-seq/RNAz	0.853347
C1	393	2601267	2601467	201	+	BLAST/CG-seq/RNAz	0.931482
C1	184	2612472	2612709	238	+	BLAST/CG-seq/RNAz	0.97131
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	201	2628827	2628880	54	+	BLAST/CG-seq/RNAz	0.99909
C1	335	2629943	2630177	235	+	BLAST/CG-seq/RNAz	0.979833
C1	609	2633162	2633421	260	+	BLAST/CG-seq/RNAz	0.992926
C1	112	2636226	2636410	185	+	BLAST/CG-seq/RNAz	0.926238
C1	269	2640658	2640833	176	+	BLAST/CG-seq/RNAz	0.939593
C1	413	2642236	2642444	209	-	BLAST/CG-seq/RNAz	0.810538
C1	550	2645480	2645602	123	+	BLAST/CG-seq/RNAz	0.869008
C1	710	2646832	2646966	135	-	BLAST/CG-seq/RNAz	0.730124
C1	854	2649046	2649198	153	+	BLAST/CG-seq/RNAz	0.849133
C1	344	2652186	2652257	72	+	BLAST/CG-seq/RNAz	0.843463
C1	620	2659381	2659599	219	+	BLAST/CG-seq/RNAz	0.70746
C1	170	2660605	2660689	85	+	BLAST/CG-seq/RNAz	0.998695
C1	291	2662701	2662825	125	+	BLAST/CG-seq/RNAz	0.968223
C1	729	2665849	2666118	270	+	BLAST/CG-seq/RNAz	0.978214
C1	486	2672294	2672484	191	+	BLAST/CG-seq/RNAz	0.965474
C1	324	2680486	2680618	133	+	BLAST/CG-seq/RNAz	0.992425
C1	8	2687394	2687443	50	+	BLAST/Rfam_10.0_seed	0.000001
C1	99	2689609	2689807	199	-	BLAST/CG-seq/RNAz	0.826173
C1	693	2696136	2696214	79	+	BLAST/CG-seq/RNAz	0.984918
C1	458	2700274	2700325	52	-	BLAST/CG-seq/RNAz	0.948326
C1	514	2717228	2717443	216	+	BLAST/CG-seq/RNAz	0.812611
C1	825	2721067	2721267	201	+	BLAST/CG-seq/RNAz	0.944153
C1	186	2727718	2727771	54	+	BLAST/CG-seq/RNAz	0.947986
C1	308	2729855	2730083	229	+	BLAST/CG-seq/RNAz	0.997409
C1	741	2734726	2734935	210	+	BLAST/CG-seq/RNAz	0.961761
C1	237	2739050	2739201	152	+	BLAST/CG-seq/RNAz	0.998383
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	545	2739565	2739659	95	-	BLAST/CG-seq/RNAz	0.733967
C1	850	2741030	2741185	156	+	BLAST/CG-seq/RNAz	0.851242
C1	337	2742233	2742299	67	+	BLAST/CG-seq/RNAz	0.945204
C1	209	2762414	2762633	220	-	BLAST/CG-seq/RNAz	0.772473
C1	438	2770389	2770507	119	+	BLAST/CG-seq/RNAz	0.948357
C1	579	2771987	2772050	64	-	BLAST/CG-seq/RNAz	0.723802
C1	288	2776582	2776667	86	-	INFERNAL	0.00526
C1	221	2777682	2777734	53	-	BLAST/CG-seq/RNAz	0.888201
C1	249	2797570	2797809	240	+	BLAST/CG-seq/RNAz	0.849456
C1	387	2798634	2798788	155	+	BLAST/CG-seq/RNAz	0.895732
C1	622	2804313	2804508	196	+	BLAST/CG-seq/RNAz	0.947069
C1	47	2811633	2811912	280	+	BLAST/CG-seq/RNAz	0.978912
C1	215	2812414	2812626	213	+	BLAST/CG-seq/RNAz	0.842549
C1	411	2827392	2827487	96	+	BLAST/CG-seq/RNAz	0.97736
C1	231	2831856	2832036	181	+	BLAST/CG-seq/RNAz	0.983539
C1	652	2840435	2840651	217	+	BLAST/CG-seq/RNAz	0.757402
C1	285	2850707	2850761	55	+	BLAST/CG-seq/RNAz	0.999959
C1	311	2867241	2867439	199	+	BLAST/CG-seq/RNAz	0.945066
C1	611	2870990	2871190	201	+	BLAST/CG-seq/RNAz	0.97726
C1	572	2871028	2871089	62	+	INFERNAL	0.00177
C1	756	2872425	2872625	201	-	BLAST/CG-seq/RNAz	0.751032
C1	522	2887763	2887842	80	-	BLAST/CG-seq/RNAz	0.86996
C1	453	2894266	2894376	111	+	BLAST/CG-seq/RNAz	0.829752
C1	662	2903030	2903115	86	+	BLAST/CG-seq/RNAz	0.971096
C1	624	2906563	2906636	74	+	BLAST/CG-seq/RNAz	0.843428
C1	767	2912341	2912430	90	+	BLAST/CG-seq/RNAz	0.849137
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	126	2913858	2914021	164	-	BLAST/CG-seq/RNAz	0.799437
C1	137	2915460	2915500	41	-	INFERNAL	0.00094
C1	843	2921082	2921282	201	+	BLAST/CG-seq/RNAz	0.790266
C1	198	2922176	2922376	201	+	BLAST/CG-seq/RNAz	0.963436
C1	355	2924204	2924467	264	+	BLAST/CG-seq/RNAz	0.913558
C1	132	2929883	2930076	194	+	BLAST/CG-seq/RNAz	0.999996
C1	235	2938384	2938444	61	+	BLAST/CG-seq/RNAz	0.986502
C1	371	2939594	2939665	72	-	BLAST/CG-seq/RNAz	0.959503
C1	428	2950999	2951186	188	+	BLAST/CG-seq/RNAz	0.936369
C1	570	2956841	2957040	200	-	BLAST/CG-seq/RNAz	0.833364
C1	725	2957878	2957951	74	+	BLAST/CG-seq/RNAz	0.918897
C1	313	2969001	2969159	159	+	BLAST/CG-seq/RNAz	0.834552
C1	742	2970805	2970853	49	+	BLAST/CG-seq/RNAz	0.994172
C1	207	2984021	2984084	64	+	BLAST/CG-seq/RNAz	0.979392
C1	557	3004985	3005185	201	-	BLAST/CG-seq/RNAz	0.758528
C1	806	3019558	3019851	294	+	BLAST/CG-seq/RNAz	0.987074
C1	173	3020131	3020185	55	+	BLAST/CG-seq/RNAz	0.991264
C1	293	3021257	3021452	196	+	BLAST/CG-seq/RNAz	0.882442
C1	441	3021638	3021771	134	-	BLAST/CG-seq/RNAz	0.731583
C1	76	3028390	3028444	55	+	BLAST/CG-seq/RNAz	0.729054
C1	488	3038728	3038803	76	+	BLAST/CG-seq/RNAz	0.840287
C1	328	3041673	3041774	102	+	BLAST/CG-seq/RNAz	0.969016
C1	749	3044374	3044674	301	+	BLAST/CG-seq/RNAz	0.999922
C1	541	3073451	3073524	74	+	BLAST/CG-seq/RNAz	0.849059
C1	466	3109763	3109872	110	+	BLAST/CG-seq/RNAz	0.846712
C1	761	3113294	3113368	75	+	BLAST/CG-seq/RNAz	0.869379
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C1	43	3120985	3121036	52	+	BLAST/Rfam_10.0_seed	2E-16
C1	706	3120985	3121036	52	+	INFERNAL	3.55E-14
C1	558	3123867	3123964	98	+	BLAST/CG-seq/RNAz	0.95356
C1	470	3139082	3139138	57	+	BLAST/CG-seq/RNAz	0.730065
C1	442	3144003	3144281	279	+	BLAST/CG-seq/RNAz	0.985163
C1	225	3153043	3153096	54	+	BLAST/CG-seq/RNAz	0.928628
C1	490	3155900	3155983	84	+	BLAST/CG-seq/RNAz	0.989464
C1	601	3161941	3162078	138	+	BLAST/CG-seq/RNAz	0.997782
C1	534	3168810	3168909	100	+	INFERNAL	4.83E-15
C1	44	3168816	3168896	81	+	BLAST/Rfam_10.0_seed	1E-12
C1	670	3172100	3172148	49	+	BLAST/CG-seq/RNAz	0.978414
C1	701	3176348	3176477	130	+	BLAST/CG-seq/RNAz	0.94147
C1	352	3176604	3176770	167	+	BLAST/CG-seq/RNAz	0.997826
C1	130	3182882	3183183	302	+	BLAST/CG-seq/RNAz	0.756803
C1	265	3185797	3185856	60	+	BLAST/CG-seq/RNAz	0.990297
C2	78	17108	17269	162	+	BLAST/CG-seq/RNAz	0.978197
C2	167	17810	17903	94	+	BLAST/CG-seq/RNAz	0.991497
C2	63	47969	48025	57	+	BLAST/CG-seq/RNAz	0.997173
C2	130	68188	68308	121	-	BLAST/CG-seq/RNAz	0.782269
C2	101	85984	86113	130	+	BLAST/CG-seq/RNAz	0.975441
C2	190	93796	93909	114	+	BLAST/CG-seq/RNAz	0.732471
C2	202	99382	99468	87	+	BLAST/CG-seq/RNAz	0.968301
C2	136	106712	106849	138	+	BLAST/CG-seq/RNAz	0.963424
C2	120	113821	114021	201	+	BLAST/CG-seq/RNAz	0.984173
C2	125	114949	115146	198	+	BLAST/CG-seq/RNAz	0.735685
C2	73	121061	121166	106	+	BLAST/CG-seq/RNAz	0.794646
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C2	65	135527	135583	57	+	BLAST/CG-seq/RNAz	0.853896
C2	160	146016	146168	153	-	BLAST/CG-seq/RNAz	0.997547
C2	161	146676	146729	54	-	BLAST/CG-seq/RNAz	0.807382
C2	173	152923	153120	198	-	BLAST/CG-seq/RNAz	0.989724
C2	117	157386	157513	128	+	BLAST/CG-seq/RNAz	0.715498
C2	86	158825	159025	201	+	BLAST/CG-seq/RNAz	0.721445
C2	87	160332	160410	79	-	BLAST/CG-seq/RNAz	0.814753
C2	100	165537	165737	201	+	BLAST/CG-seq/RNAz	0.761175
C2	210	193669	193732	64	+	BLAST/CG-seq/RNAz	0.940432
C2	144	199856	200132	277	+	BLAST/CG-seq/RNAz	0.986408
C2	152	202133	202280	148	+	BLAST/CG-seq/RNAz	0.982515
C2	197	219628	219712	85	+	BLAST/CG-seq/RNAz	0.80334
C2	195	223142	223362	221	+	BLAST/CG-seq/RNAz	0.987603
C2	194	227901	227954	54	-	BLAST/CG-seq/RNAz	0.968061
C2	123	228022	228132	111	+	BLAST/CG-seq/RNAz	0.84169
C2	121	228559	228731	173	-	BLAST/CG-seq/RNAz	0.926779
C2	32	238542	238769	228	+	BLAST/CG-seq/RNAz	0.92596
C2	2	239508	239721	214	-	BLAST/Rfam_10.0_seed	1E-105
C2	44	239508	239721	214	-	INFERNAL	7.7E-29
C2	3	239516	239721	206	-	BLAST/Rfam_10.0_seed	1E-51
C2	4	239518	239706	189	-	BLAST/Rfam_10.0_seed	6E-42
C2	5	239522	239716	195	-	BLAST/Rfam_10.0_seed	3E-27
C2	6	239524	239703	180	-	BLAST/Rfam_10.0_seed	5E-18
C2	29	240411	240497	87	+	BLAST/CG-seq/RNAz	0.982652
C2	142	242082	242345	264	+	BLAST/CG-seq/RNAz	0.998025
C2	79	256004	256029	26	+	INFERNAL	0.00444
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C2	169	261589	261850	262	+	BLAST/CG-seq/RNAz	0.982998
C2	1	274276	274455	180	+	BLAST/Rfam_10.0_seed	8E-85
C2	43	274276	274455	180	+	INFERNAL	4.98E-24
C2	99	274300	274500	201	-	BLAST/CG-seq/RNAz	0.729882
C2	47	291161	291227	67	+	BLAST/CG-seq/RNAz	0.994331
C2	159	294874	294998	125	+	BLAST/CG-seq/RNAz	0.881369
C2	131	296870	297070	201	+	BLAST/CG-seq/RNAz	0.768564
C2	71	300933	301174	242	+	BLAST/CG-seq/RNAz	0.989905
C2	177	304935	304991	57	+	BLAST/CG-seq/RNAz	0.990107
C2	174	309933	310199	267	-	BLAST/CG-seq/RNAz	0.703551
C2	108	320179	320359	181	-	BLAST/CG-seq/RNAz	0.9263
C2	122	333690	333904	215	+	BLAST/CG-seq/RNAz	0.961461
C2	59	343749	343858	110	+	BLAST/CG-seq/RNAz	0.911413
C2	33	347722	347767	46	+	BLAST/CG-seq/RNAz	0.887293
C2	31	349064	349262	199	+	BLAST/CG-seq/RNAz	0.715033
C2	141	357121	357389	269	+	BLAST/CG-seq/RNAz	0.863136
C2	98	367824	368074	251	+	BLAST/CG-seq/RNAz	0.933148
C2	50	382994	383089	96	+	BLAST/CG-seq/RNAz	0.795331
C2	49	384926	385160	235	+	BLAST/CG-seq/RNAz	0.949558
C2	157	388316	388497	182	+	BLAST/CG-seq/RNAz	0.994746
C2	70	396267	396399	133	+	BLAST/CG-seq/RNAz	0.804657
C2	175	400453	400684	232	+	BLAST/CG-seq/RNAz	0.752901
C2	27	415065	415265	201	+	BLAST/CG-seq/RNAz	0.880353
C2	140	423549	423607	59	-	BLAST/CG-seq/RNAz	0.998521
C2	138	424572	424854	283	+	BLAST/CG-seq/RNAz	0.700803
C2	83	430136	430343	208	+	BLAST/CG-seq/RNAz	0.992161
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C2	165	436434	436484	51	+	BLAST/CG-seq/RNAz	0.995983
C2	164	437625	437801	177	+	BLAST/CG-seq/RNAz	0.789859
C2	208	444785	444867	83	+	BLAST/CG-seq/RNAz	0.979813
C2	114	454856	455027	172	+	BLAST/CG-seq/RNAz	0.9209
C2	7	458014	458214	201	-	BLAST/Rfam_10.0_seed	7E-98
C2	45	458014	458214	201	-	INFERNAL	2.02E-26
C2	8	458030	458200	171	-	BLAST/Rfam_10.0_seed	1E-19
C2	9	458032	458178	147	-	BLAST/Rfam_10.0_seed	6E-10
C2	46	458115	458315	201	-	BLAST/CG-seq/RNAz	0.979914
C2	42	462153	462225	73	-	BLAST/CG-seq/RNAz	0.930595
C2	156	463681	463781	101	+	BLAST/CG-seq/RNAz	0.924855
C2	68	476576	476635	60	+	BLAST/CG-seq/RNAz	0.982968
C2	56	506076	506341	266	+	BLAST/CG-seq/RNAz	0.837391
C2	26	506677	506762	86	+	BLAST/CG-seq/RNAz	0.999999
C2	134	512254	512347	94	+	BLAST/CG-seq/RNAz	0.985224
C2	81	521036	521094	59	+	BLAST/CG-seq/RNAz	0.990442
C2	80	522898	523036	139	-	BLAST/CG-seq/RNAz	0.78444
C2	115	572299	572418	120	+	BLAST/CG-seq/RNAz	0.862463
C2	113	579112	579174	63	-	BLAST/CG-seq/RNAz	0.885196
C2	40	586711	586911	201	+	BLAST/CG-seq/RNAz	0.998365
C2	154	588456	588663	208	+	BLAST/CG-seq/RNAz	0.959539
C2	67	599281	599572	292	+	BLAST/CG-seq/RNAz	0.706313
C2	107	613972	614037	66	-	BLAST/CG-seq/RNAz	0.91796
C2	192	620440	620642	203	+	BLAST/CG-seq/RNAz	0.911422
C2	119	622476	622547	72	-	BLAST/CG-seq/RNAz	0.979432
C2	96	663769	664011	243	+	BLAST/CG-seq/RNAz	0.782983

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C2	111	678418	678509	92	+	BLAST/CG-seq/RNAz	0.991694
C2	128	687277	687405	129	+	BLAST/CG-seq/RNAz	0.98874
C2	171	692809	692971	163	+	BLAST/CG-seq/RNAz	0.777791
C2	103	698323	698386	64	+	BLAST/CG-seq/RNAz	0.823306
C2	162	728729	728928	200	+	BLAST/CG-seq/RNAz	0.992489
C2	94	731269	731468	200	+	BLAST/CG-seq/RNAz	0.899022
C2	206	731840	732040	201	+	BLAST/CG-seq/RNAz	0.990691
C2	204	733515	733679	165	+	BLAST/CG-seq/RNAz	0.998623
C2	127	737970	738270	301	-	BLAST/CG-seq/RNAz	0.858871
C2	38	744217	744416	200	+	BLAST/CG-seq/RNAz	0.935313
C2	36	745198	745249	52	+	BLAST/CG-seq/RNAz	0.974821
C2	105	764676	764791	116	+	BLAST/CG-seq/RNAz	0.925898
C2	213	769700	769766	67	+	BLAST/CG-seq/RNAz	0.86766
C2	54	795895	796026	132	+	BLAST/CG-seq/RNAz	0.982569
C2	76	806720	807000	281	+	BLAST/CG-seq/RNAz	0.994262
C2	75	813215	813343	129	+	BLAST/CG-seq/RNAz	0.977474
C2	183	815098	815326	229	+	BLAST/CG-seq/RNAz	0.996399
C2	110	822091	822306	216	-	BLAST/CG-seq/RNAz	0.759432
C2	109	823357	823414	58	+	BLAST/CG-seq/RNAz	0.953336
C2	148	832216	832476	261	+	BLAST/CG-seq/RNAz	0.898718
C2	147	838660	838860	201	+	BLAST/CG-seq/RNAz	0.959161
C2	85	839916	839986	71	+	BLAST/CG-seq/RNAz	0.905565
C2	66	841879	842079	201	+	BLAST/CG-seq/RNAz	0.735025
C2	102	853846	854046	201	+	BLAST/CG-seq/RNAz	0.76762
C2	187	858743	858823	81	+	BLAST/CG-seq/RNAz	0.716251
C2	185	867727	867902	176	+	BLAST/CG-seq/RNAz	0.990106
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
C2	116	878428	878542	115	-	BLAST/CG-seq/RNAz	0.990433
C2	52	884603	884838	236	+	BLAST/CG-seq/RNAz	0.996229
C2	212	889171	889224	54	-	BLAST/CG-seq/RNAz	0.994819
C2	132	890005	890204	200	+	BLAST/CG-seq/RNAz	0.988904
C2	74	901771	901881	111	-	BLAST/CG-seq/RNAz	0.919178
C2	182	904132	904266	135	-	BLAST/CG-seq/RNAz	0.929126
C2	180	910056	910251	196	+	BLAST/CG-seq/RNAz	0.94196
C2	179	912467	912718	252	-	BLAST/CG-seq/RNAz	0.975299
C2	92	914675	914773	99	+	BLAST/CG-seq/RNAz	0.737093
C2	90	915290	915401	112	+	BLAST/CG-seq/RNAz	0.996676
C2	88	915610	915694	85	+	BLAST/CG-seq/RNAz	0.997439
C2	200	917369	917456	88	+	BLAST/CG-seq/RNAz	0.958964
C2	199	918844	919078	235	-	BLAST/CG-seq/RNAz	0.763218
C2	126	921539	921843	305	+	BLAST/CG-seq/RNAz	0.989363
C2	34	934125	934331	207	+	BLAST/CG-seq/RNAz	0.770762
C2	150	937946	938145	200	+	BLAST/CG-seq/RNAz	0.899686
pA	20	4092	4215	124	+	BLAST/CG-seq/RNAz	0.991687
pA	71	20711	20962	252	+	BLAST/CG-seq/RNAz	0.781946
pA	73	25649	25694	46	+	BLAST/CG-seq/RNAz	0.816305
pA	1	28823	29023	201	+	BLAST/CG-seq/RNAz	0.992996
pA	3	28934	29134	201	+	BLAST/CG-seq/RNAz	0.950091
pA	9	33811	33891	81	+	BLAST/CG-seq/RNAz	0.993738
pA	11	33959	34237	279	+	BLAST/CG-seq/RNAz	0.86861
pA	52	37955	38078	124	-	BLAST/CG-seq/RNAz	0.907302
pA	34	52715	52763	49	-	BLAST/CG-seq/RNAz	0.724016
pA	59	58894	59093	200	+	BLAST/CG-seq/RNAz	0.996393

Location	Sequence ID	Start	End	Size	Strand	Program	Score
pA	63	61664	61743	80	+	BLAST/CG-seq/RNAz	0.878299
pA	69	70576	70783	208	+	BLAST/CG-seq/RNAz	0.872803
pA	40	79877	80176	300	+	BLAST/CG-seq/RNAz	0.999836
pA	49	91512	91571	60	+	BLAST/CG-seq/RNAz	0.768981
pA	50	100366	100539	174	-	BLAST/CG-seq/RNAz	0.831594
pA	23	106010	106190	181	+	BLAST/CG-seq/RNAz	0.976707
pB	51	1443	1502	60	+	BLAST/CG-seq/RNAz	0.768981
pB	15	3469	3641	173	+	BLAST/CG-seq/RNAz	0.915799
pB	55	4932	5075	144	+	BLAST/CG-seq/RNAz	0.939529
pB	32	13267	13467	201	+	BLAST/CG-seq/RNAz	0.99816
pB	67	20353	20486	134	+	BLAST/CG-seq/RNAz	0.905258
pB	70	22613	22768	156	-	BLAST/CG-seq/RNAz	0.976306
pB	5	24253	24521	269	-	BLAST/CG-seq/RNAz	0.721987
pB	61	36867	37145	279	+	BLAST/CG-seq/RNAz	0.986877
pB	65	39184	39275	92	+	BLAST/CG-seq/RNAz	0.983339
pB	53	57740	57878	139	+	BLAST/CG-seq/RNAz	0.997445
pB	29	58899	59129	231	+	BLAST/CG-seq/RNAz	0.996646
pB	18	68033	68323	291	+	BLAST/CG-seq/RNAz	0.986548
pB	57	91729	91770	42	+	BLAST/CG-seq/RNAz	0.991819
pB	42	96161	96319	159	+	BLAST/CG-seq/RNAz	0.982437
pB	38	113810	113928	119	+	BLAST/CG-seq/RNAz	0.994864
pC	8	62679	62764	86	-	BLAST/CG-seq/RNAz	0.974515
pC	13	73770	74045	276	+	BLAST/CG-seq/RNAz	0.915286
pD	25	23	215	193	+	BLAST/CG-seq/RNAz	0.999956
pD	14	25878	26182	305	-	BLAST/CG-seq/RNAz	0.895995
pD	35	34094	34208	115	+	BLAST/CG-seq/RNAz	0.778745
							(continued)

Location	Sequence ID	Start	End	Size	Strand	Program	Score
pD	19	61553	61609	57	+	BLAST/CG-seq/RNAz	0.973823
pD	27	65240	65399	160	+	BLAST/CG-seq/RNAz	0.999975
pE	17	1	162	162	+	BLAST/CG-seq/RNAz	0.705201
pE	10	566	792	227	+	BLAST/CG-seq/RNAz	0.952285
pE	22	4415	4531	117	+	BLAST/CG-seq/RNAz	0.758372
pE	16	7245	7328	84	-	BLAST/CG-seq/RNAz	0.933619
pE	30	16853	16894	42	-	BLAST/CG-seq/RNAz	0.720275
pE	4	19232	19400	169	+	BLAST/CG-seq/RNAz	0.900764
pE	7	21004	21204	201	-	BLAST/CG-seq/RNAz	0.940929
pE	36	22144	22221	78	-	INFERNAL	0.00494
							(continued)

VITA

Jovinna Laryssa Mendel

Education

Sam Houston State University, Huntsville, Texas
M.S. (Biology)
Summa Cum Laude (GPA: 4.0)
Graduation Date: December 2021
Relevant Coursework:

>Biochemical Analysis of Proteins
>Genomics and Bioinformatics
>Advanced Genetics: Molecular Genetics
>Genetic Analysis of Human Disease
>Virology
>Cell Structure & Physiology
>Microscopy and Imaging Technologies

Sam Houston State University, Huntsville, Texas
B.S. (Major: Biology; Minor: Forensic Science)
Graduated: May 2018
Summa Cum Laude (GPA: 4.0)
Relevant Coursework:

>Organic Chemistry I & II
>Microbiology
>Genetics
>Cell Biology
>Advanced Molecular Genetics
>Biochemistry
>Molecular Biology

Related Experience

Department of Biological Sciences, Sam Houston State University
REU-GSCB Program Graduate Student Assistant
June 2021 – August 2021

- Mentored students through office hours with questions about projects.
- Communicated effectively with faculty and staff involved in the program.
- Helped with administrative support, coordinating files, and ordering material.
- Greeted incoming office visitors, answered questions, and directed individuals to office locations.

Department of Biological Sciences, Sam Houston State University
Introductory Genetics Head Teaching Assistant
January 2020 – December 2020
- Prepared lab curriculum for labs, both in-person and online.
- Prepared laboratory materials/equipment for all lab sections.
- Conducted weekly lab meetings with teaching assistants to discuss curriculum.
- Answered questions from students in lab, facilitated lab discussion/lab work, and graded weekly lab exercises and quizzes.

Department of Biological Sciences, Sam Houston State University
Graduate Student Teaching Assistant
August 2018 – January 2020
- Lab courses taught: Introductory Genetics, Contemporary Biology, Cell Biology.
- Prepared laboratory materials/equipment.
- Answered questions from students in lab, facilitated lab discussion/lab work, and graded weekly lab exercises, quizzes, and presentations.
- Conducted midterm/final exam for lab.

Department of Biological Sciences, Sam Houston State University
EURECA Grant Student Research Assistant
May 2018 – August 2018
- Prepared soil, fish tissue, and plant samples for analysis.
- Performed mercury digestion and isotope analysis for soil samples.
- Utilized equipment at TRIES facility in Huntsville, Texas for soil analysis.

Department of Biological Sciences, Sam Houston State University
Coastal Chemical Company, LLC – Canyon, Texas
Student Research Assistant
May 2018 – August 2018
- Prepared nutrient-rich culture media with various sugar substitutes for experiment.
- Prepared bacterial culture samples for analysis and storage.
- Attended weekly meetings for research progress and updates.

Academic Success Center, Sam Houston State University
Supplemental Instruction Leader
August 2015 – May 2018
- Facilitated student-led study sessions for science courses with a high history of D/F/W rates, particularly General Botany, General Chemistry I & II, and Elementary Functions
- Attended weekly meetings with other SI leaders and supervisors to discuss updates.
- Generated monthly reports of student's success to report to supervisor.

Department of Chemistry, Sam Houston State University
Stockroom Assistant
May 2016 – August 2016
- Maintained inventory of chemicals and glassware.
- Cleaned used glassware from laboratories and organized stockroom supplies.
- Interacted with students and faculty for chemical/glassware needs.

Presentations

Three Minute Thesis (3MT) Competition – March 2020
Sam Houston State University, Huntsville, Texas
Identification of sRNA in Rhodobacter sphaeroides under heavy metal tolerance

Texas Branch American Society of Microbiology Conference – November 2018
Texas A&M University of Corpus Christi, Corpus Christi, Texas
Computational identification of regulatory small RNAs in Rhodobacter sphaeroides

Texas Branch American Society of Microbiology Conference – October 2017
Texas A&M University, College Station, Texas
Identification of regulatory small RNAs and their corresponding target genes in Rhodobacter sphaeroides

McNair Scholars Program Research Conference – September 2017
Sam Houston State University, Huntsville, Texas
Role of Small RNA (sRNA) in Protein Expression of Bacteria

Publications

Mendel J., Jha T., Cho H., Choudhary M., (2022). Bioinformatic prediction of sRNA in Rhodobacter sphaeroides and identification of sRNAs in gold chloride stress. *BMC Microbiology*. **In Preparation**.

Mendel J., Choudhary M., (2022). Differential Gene Expression of Rhodobacter sphaeroides in gold chloride stress. *RNA Biology*. **In Preparation**.

Jha T., **Mendel J**., Cho H., Choudhary M., (2021). Prediction of Bacterial sRNAs using Sequence-Derived Features and Machine Learning. *Bioinformatics and Biology Insights*. **In Press.**

Prabhu Balaraman R., **Mendel J**., Flores L., Choudhary M. (2021) Nanoparticle Biosynthesis and Interaction with the Microbial Cell, Antimicrobial and Antibiofilm Effects, and Environmental Impact. In: Sharma N., Sahi S. (eds) Nanomaterial Biointeractions at the Cellular, Organismal and System Levels. Nanotechnology in the Life Sciences. Springer, Cham. https://doi.org/10.1007/978-3-030-65792-5_15

Honors and Awards

- Sam Houston State University O.R.S.P. Grant Recipient - Spring 2020
- Finalist; Three Minute Thesis Competition - Spring 2020
- A.S.P.I.R.E. Scholar - Fall 2019
- Sam Houston State University Graduate Studies General Scholarship ($1,000) - Spring 2018
- Sam Houston State University College of Science and Engineering Technology Advisement Fee Scholarship ($1,000) - Fall 2018
- Tri-Beta National Honor Society Delta Tau Chapter, Regular Member
- Sam Houston State University McNair Scholars Program Completion Award