Transcriptomic studies of the bacterium *Rhodobacter capsulatus*

by

Marc Grüll, BSc

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biology

Faculty of Science

Memorial University of Newfoundland

St. John's, Newfoundland and Labrador, Canada

February 2019

Abstract

Rhodobacter capsulatus is a purple, non-sulfur alpha-proteobacterium that is studied for different aspects of physiology and its ability to produce the bacteriophage like particle called gene transfer agent (RcGTA). These particles are capable of transferring ~4 kb of random host DNA to other *R. capsulatus* through a process analogous to transduction. The genes that encode the RcGTA particles are located in a region of ~15 kb called the RcGTA structural gene cluster. The exact regulatory mechanisms involved in the production of RcGTA remain elusive. I investigated the regulation of GTA regulatory genes on a global scale using next generation sequencing methods. I performed investigations to analyse small RNA (sRNA) sequences genome-wide using RNA sequencing. I compared an *R. capsulatus* wild-type strain to a *ctrA* null mutant due to previous studies that suggested the global response regulator CtrA might controll a sRNA that regulates RcGTA gene expression. Using the latest bioinformatic approach, 422 putative sRNAs were detected in *R. capsulatus* during early stationary phase. To get a more detailed insight into the expression of the RcGTA structural gene cluster I performed total RNA sequencing on a RcGTA overproducer mutant strain of *R. capsulatus*. The previously developed differential RNA sequencing (dRNA-seq) protocol, which can distinguish between primary and processed transcripts, has been modified to identify transcription start sites (TSS) and to predict transcription termination sites (TTS). The combination of total RNA and 5’ -3’ specific sequencing data allowed for the prediction of transcriptional units and to analyze operon complexity in *R. capsulatus*. The analysis revealed a complex operon architecture, similar to other bacterial species, with

operons having multiple TSS and TTS, genomic regions of high transcriptional activity and novel transcripts. Finally, to gain further insight into how RcGTA-associated genes affect each other, I performed qPCR experiments on eight RcGTA-associated genes within and outside of the RcGTA structural gene cluster and within five different *R. capsulatus* mutant strains. The results indicate that the absence of any one of the known and putative GTA regulators investigated can have effects on the transcription of GTA loci.

Acknowledgements

Many people have helped and supported my thesis research. First and foremost, I would like to acknowledge my supervisors Dr. Andrew S. Lang and Dr. Martin E. Mulligan for their continuous guidance, patience and mentorship. They always gave me a new perspective to look at experiments and challenged me to think outside of my comfort zone. Their interest and accessibility as well as their involvement and support made my research possible. I would also like to thank my co-supervisor Dr. Lourdes Peña-Castillo for her continuous involvement in my various research projects. Her efforts in the bioinformatic aspects of my research made a lot of the analysis that was performed possible and she was always thriving to improve the quality of the analysis in every aspect. I would like to thank all members, current and alumni, of the Lang lab for their help in times of crisis. The positive attitude of every lab member always made my stay in the Lang lab more than enjoyable. I would like to make a special mention of Maggie McGuire, who was my anchor and soulmate outside of the lab. Finally, I must thank my family who has supported me throughout my career. They made it possible for me to start university and move on to greater things in Canada. Although it has not always been easy to be separated by such a great distance, their continued interest in my work and mental support have made me the person and researcher that I am today, and I am forever grateful to them.

My doctoral program and the research I have performed has been supported by a graduate student fellowship from the Memorial University School of Graduate Studies, and by a grant from the Dean of Science to Dr. M. Mulligan. The research in Dr. Lang's

laboratory was supported by grants from the Natural Sciences and Engineering Research Council (NSERC).

Table of Contents

List of Tables

List of Figures

List of Abbreviations, Symbols and Nomenclature

Abbreviation	Meaning
%	per cent
5'-P	5'-monophosphate
5'-PPP	5'-triphosphate
bp	basepair
bps	basepairs
CRISPR	clustered interspaced short palindromic repeats
DNA	deoxyribonucleic acid
dRNA-seq	differential RNA sequencing
ETS	external transcribed spacer
Fur	ferric uptake regulator
GTA	gene transfer agent
HGT	horizontal gene transfer
ITS	internal transcribed spacer
kb	kilo bases
kbp	kilo base pairs
Mb	Mega bases
mRNA	messenger RNA
nm	nanometers
nts	nucleotides
OMP	outer membrane protein
ORF	open reading frame
ORI	origin of replication
RBS	ribosome binding site
RcGTA	Rhodobacter capsulatus gene transfer agent
RNA	ribonucleic acid
rpm	revolutions per minute
rRNA	ribosomal RNA
s	seconds
sRNA	small RNA
tRNA	transfer RNA
TSS	transcription start site
TTS	transcription termination site
UTR	untranslated region
YPS	yeast extract/peptone/salts medium

List of Appendices

Co-authorship statement

Chapter 2 is a version of a manuscript published in RNA Biology (Marc P. Grüll, Lourdes Peña-Castillo, Martin E. Mulligan, and Andrew S. Lang (2017) Genome-wide identification and characterization of small RNAs in *Rhodobacter capsulatus* and identification of small RNAs affected by loss of the response regulator CtrA. RNA Biology 14(7): 914-925; DOI: 10.1080/15476286.2017.1306175). A.S. Lang, M.E. Mulligan and I designed the research, I carried out the RNA sequencing experiments, as well as parts of the data analysis with the assistance of L. Peña-Castillo. The manuscript was drafted by me, L. Peña-Castillo and A.S. Lang with subsequent editorial input from M.E. Mulligan.

Chapter 3 is a version of a manuscript that is currently being prepared for submission. Research in this chapter was proposed and designed by me and A.S. Lang. For this research I developed a new protocol for 5' and 3' targeted sequencing using the Ion Torrent PGM sequencing platform and I carried out the RNA sequencing experiments. I performed parts of the data analysis with the assistance of L. Peña-Castillo. The manuscript was drafted by me, L. Peña-Castillo and A.S. Lang with subsequent editorial input from M.E. Mulligan.

Chapter 4 is a draft version of a future manuscript. Experiments in this chapter were proposed and designed by me and A.S. Lang. I carried out the RNA isolations, primer design and qPCR experiments, as well as all data analysis. The manuscript was drafted and prepared by me and A.S. Lang with subsequent editorial input from M.E. Mulligan.

Chapter 1 – Introduction and Overview

1.1 Horizontal gene transfer

Horizontal gene transfer (HGT) is the transfer of genetic material between different genomes and is considered as one of the most powerful forces in the evolution of bacteria and archaea (de la Cruz & Davies, 2000, Ochman *et al.*, 2000, Lawrence, 2002). HGT is widely recognized as the mechanism responsible for the widespread distribution of antibiotic resistance genes, gene clusters encoding biodegradative pathways, and pathogenicity determinants (de la Cruz & Davies, 2000). HGT involves different mechanisms and mobile genetic elements, and it is believed to enable bacteria to quickly adapt to various environments (Hacker & Carniel, 2001, Burrus & Waldor, 2004, Beiko *et al.*, 2005, Kloesges *et al.*, 2011, Lang *et al.*, 2012). New techniques in the field of genome sequencing have resulted in the characterization of many bacterial genomes, the results of which support the concept of bacterial genome plasticity due to HGT (Sorensen *et al.*, 2005). Three main conditions need to be fulfilled in order for HGT to be successful: there must be a way for the donor DNA to be delivered into the recipient cell and a transferred gene has to be stably integrated into a cell's genome; regulatory or genetic structures cannot be disturbed; and, a functional protein has to be expressed and produced (Ochman *et al.*, 2000, Susanna *et al.*, 2006). It has been estimated that ~14% of a bacterium's open reading frames (ORFs) could represent recently horizontally transferred genes (Nakamura *et al.*, 2004, Zhaxybayeva *et al.*, 2006).

There are three canonical mechanisms by which HGT can occur in prokaryotes: conjugation (direct cell to cell transfer of elements such as plasmids), natural transformation (uptake of free DNA from the environment), and transduction (transfer by phage particles). Conjugation is the

transfer of genetic material through direct cell to cell contact. In Gram-negative bacteria this is often achieved using a type IV secretion apparatus that produces an extracellular structure called a pilus (Lederberg & Tatum, 1946, Thomas & Nielsen, 2005). The conjugative transfer systems observed in Gram-positive bacteria involve the production of multiple small, hydrophobic peptides that act as signals for interbacterial communication (Chandler & Dunny, 2004). As a result of the activation of these transfer genes, the production of a membrane protein that promotes aggregation of the donors and recipients occurs (Waters & Dunny, 2001). Natural transformation is the stable uptake and integration of free DNA from the environment by a bacterial cell (Chen & Dubnau, 2004, Thomas & Nielsen, 2005). For natural transformation to be operative, cells expressing proteins that are required for DNA uptake enter a physiological condition known as competence (Chen *et al.*, 2005). Transduction is a bacteriophage (phage)-facilitated transfer of genetic material from one bacterial host to a recipient cell (Zinder & Lederberg, 1952, Ochman *et al.*, 2000). Two classes of transduction are recognized, specialized and generalized, depending on how the phage propagates within the host cell. In specialized transduction, host DNA that is located adjacent to a temperate phage's integration site is packaged into the phage particle, as is the case with phage λ that can transduce either *bio* or *gal* operon sequences (Morse *et al.*, 1955, Del Campillo-Campbell *et al.*, 1967). During generalized transduction, which is more typically carried out by lytic phages, random segments of the bacterial genome that are approximately the same size as the phage genome can be incorporated into the phage particle, resulting in phages that carry host genetic material (Fogg *et al.*, 2011, Lang *et al.*, 2012).

1.2 Bacteriophages and prophages

Bacteriophages (phages) are viruses that exclusively infect bacteria. Phages are highly diverse, with variations in genome composition, structure and complexity (Krupovic *et al.*, 2010). They represent the most abundant biological entity on our planet, and they occur at densities that are on average 10-fold more abundant than their bacterial hosts (Bergh *et al.*, 1989, Ashelford *et al.*, 2003). Studies have estimated ~10^{31} phage virions to be present in the biosphere (Suttle, 2007). Direct counts made with electron and epifluorescence microscopy have shown phages to be abundant in water from marine (Bergh *et al.*, 1989, Bratbak *et al.*, 1990, Hara *et al.*, 1991) and freshwater habitats (Hennes & Simon, 1995).

The high abundance of phages can influence microbial communities, as the overall phage population can determine host cell numbers (Hennes *et al.*, 1995, Winter *et al.*, 2004). However, despite the high abundance and biological importance of phages, their roles in controlling microbial community structure and in driving microbial population dynamics are unclear. Most phages have a narrow host range and are restricted to specific species (Santamaria *et al.*, 2014). Some phages have a broader host range and can therefore infect multiple species. One example of a phage with a broad host spectrum is the phage Mu, which can infect bacteria from the *Enterobacteriaceae* family, including *Citrobacter, Escherichia, Erwinia, Salmonella* and *Shigella* (Paolozzi & Ghelardini, 2006).

In general, tempered phages have two modes of development for phages. One mode is a lytic development in which a phage infects a cell, replicates within the cell, and then causes cell lysis to escape into the environment. The released phages have the potential to infect neighboring cells and repeat the process of penetration, replication, maturation and lysis. One example of a lytic phage is enterobacteriophage T4 (Streisinger *et al.*, 1961). Lytic phages specifically

regulate their gene expression. Phage T4, for example, relies heavily on multiple phage-encoded factors (e.g. MotA, AsiA) for the temporal regulation of early, middle and late phage genes (Hinton, 2010). The lytic development cycle of a phage appears to be the most efficient means of genome replication and seems to be favored when the host cells are healthy and numerous (Dodd *et al.*, 2005).

Another mode of development is lysogeny, which is carried out by temperate phages. In lysogeny, the temperate phage's DNA is replicated as a prophage that is integrated into the host cell genome or maintained as a plasmid, with the expression of most viral genes essentially shut off. One example of a temperate phage is the enterobacteriophage λ (Dodd *et al.*, 2005). Typically, phage replication is regulated by phage-encoded transcription factors which determine appropriate timing of phage gene expression. One such transcription factor is the λ CI repressor, which binds the operator sequences and prevents the expression of genes that are required for lytic development (Ptashne *et al.*, 1980). The activation of host DNA repair systems can lead to proteolytic cleavage of CI, and expression of lytic genes (Little, 1984). Temperate phages, such as λ, can switch between lysogenic and lytic modes of action.

It is believed that phages play a pivotal role in the evolution of prokaryotic genomes and microbial diversity (Filée *et al.*, 2003, Pal *et al.*, 2007, Rodriguez-Valera *et al.*, 2009). This is supported by the abundance of phage-derived genetic information in bacterial genomes (Canchaya *et al.*, 2003, Casjens, 2003), as well as the detection of high numbers of bacterial genes in viral metagenome studies (Rosario & Breitbart, 2011). Today we know that most bacterial genomes contain a number of prophages that can be functional or non-functional phage-like elements (Stanton, 2007), and that these can account for up to 20% of a bacterial genome (Casjens, 2003). Some bacteria, like *Escherichia coli*, *Xylella fastidiosa* and *Shigella flexneri*,

contain between 7 and 20 prophages within their genomes (Blattner *et al.*, 1997, Simpson *et al.*, 2000, Hayashi *et al.*, 2001, Qi *et al.*, 2002). Functional prophages can be stimulated to enter the lytic cycle by a variety of stressors, such as UV light and chemicals (e.g. Mitomycin C). A non-functional prophage is incapable of self-replication but may still provide some biological function and serve as an important contributor to genome evolution (Stanton, 2007). The lysogenic phage cycle is believed to have an important influence on the evolution of both the phage and the host. The influence on the phage can be through recombination with other phages that enter the cell, while the influence for the cell can be through the provision of novel genes (Krupovic *et al.*, 2011).

1.3 Phage-like gene transfer agents (GTAs)

Gene transfer agents (GTAs) were first discovered in 1974 in the bacterium *Rhodobacter capsulatus* as virus-sized, DNase-resistant, and protease-sensitive mediators of cell contact-independent genetic exchange (Marrs, 1974, Solioz *et al.*, 1975). Since the initial discovery of the *R. capsulatus* GTA (RcGTA), GTAs have been identified in diverse prokaryotic species such as *Brachyspria hyodysenteriae* (Humphrey et al., 1997), *Bartonella henselae* (Anderson *et al.*, 1994), *Methanococcus voltae* (Eiserling *et al.*, 1999) and *Desulfovibrio desulfuricans* (Rapp & Wall, 1987). GTAs are tailed phage-like particles that contain DNA from the producing cell's genome and which are released into the extracellular environment via lysis of the producing cell (Figure 1.1) (Stanton *et al.*, 2008, Lang *et al.*, 2012, Lang *et al.*, 2017).

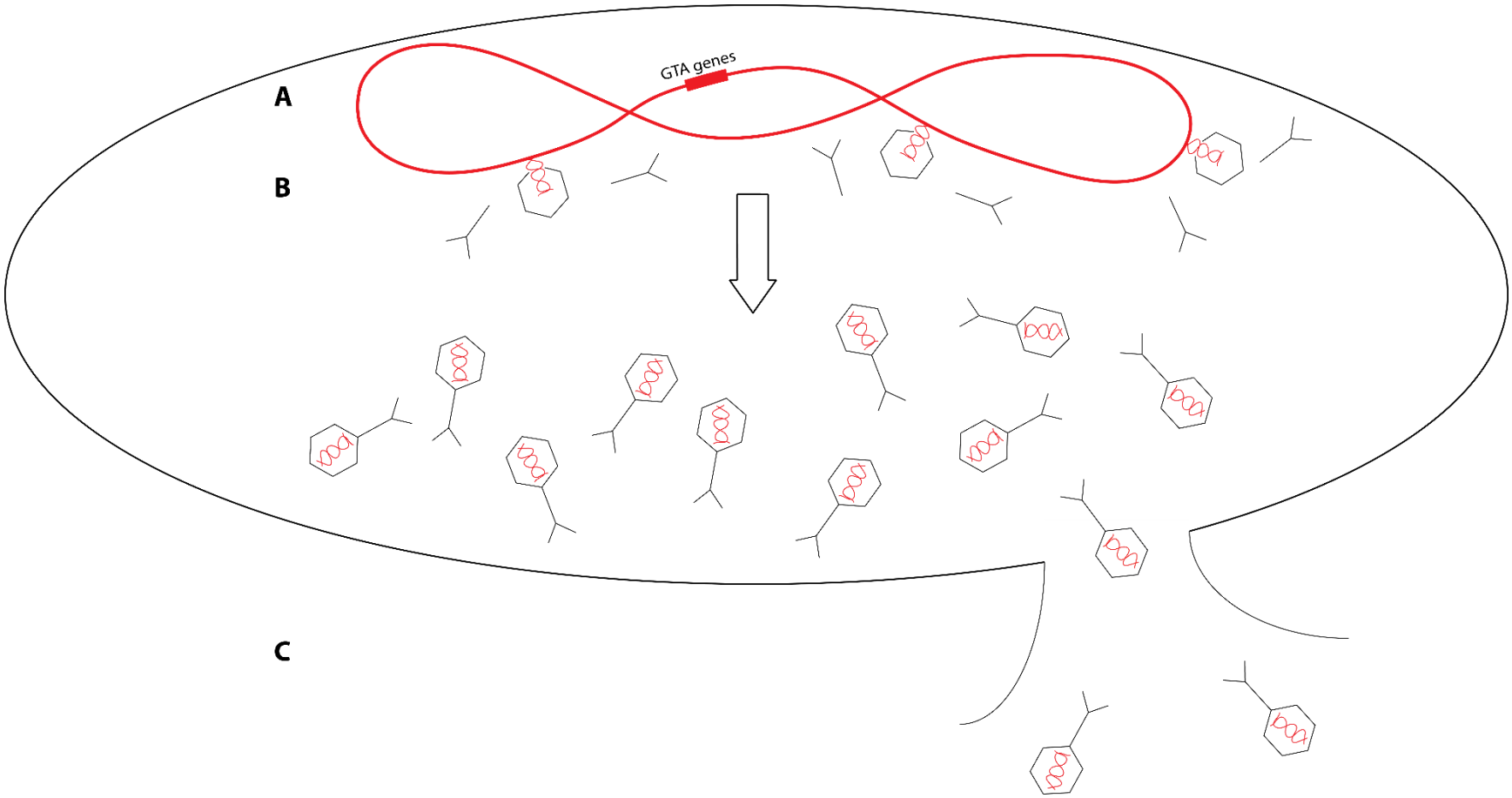

Figure 1.1: Production of gene transfer agents. (A) An array of genes that are typically organized in clusters encode the GTA particles. (B) Genomic DNA is randomly incorporated into the phage-like particles. (C) The GTAs are then released out of the cell by cell lysis.

GTAs can vary in capsid size and cargo size between different bacterial species. Capsid sizes vary from 30 to 80 nm and they contain from 4 to 14 kb of DNA packaged in the protein capsid shell (Yen *et al.*, 1979, Rapp & Wall, 1987, Humphrey *et al.*, 1995, Eiserling *et al.*, 1999, Barbian & Minnick, 2000, Nagao *et al.*, 2015, Tomasch *et al.*, 2018). Phages usually package their entire genomes and can replicate in suitable infected host cells. By contrast, GTA particles do not package sufficient DNA to encode their own production and no GTA-directed replication occurs in infected cells (Lang & Beatty, 2000, Matson *et al.*, 2005). GTAs contain fragments of DNA from across the entire genome of the host cell that vary in size and most GTA particles do not contain any GTA-encoding genes (Lang & Beatty, 2007, Stanton, 2007, Lang *et al.*, 2012). This means that GTAs are non-replicative. The production of GTA particles is dependent on the host's physiology, and the factors controlling their production appear to differ between different organisms. For example, in *R. capsulatus* the highest production of GTA particles is observed in the stationary phase of lab cultures (Solioz *et al.*, 1975) whereas in *B. hyodysenteriae* production of the particles is increased in response to certain DNA-damaging antibiotics (Humphrey *et al.*, 1995, Stanton *et al.*, 2008).

1.4 *Rhodobacter capsulatus* and RcGTA

R. capsulatus is classified within the *Rhodobacterales* order in the class of α-proteobacteria. *R. capsulatus* is a model organism for the study of GTAs and has also been used previously for studying photosynthesis, cellular energetics, and nitrogen fixation due to its metabolic diversity (Imhoff, 1995, Madigan, 1995, Pemberton *et al.*, 1998, Haselkorn *et al.*, 2001). This species can grow aerobically as well as anaerobically using various sources of organic carbon and terminal electron acceptors (Imhoff, 1995). It is also capable of growing photoheterotrophically using

light as an energy source (Pemberton *et al.*, 1998). The *R. capsulatus* genome is comprised of a 3.7-Mb chromosome as well as a 134-kb circular plasmid. Both the chromosome and the plasmid have relatively high GC content of 66.6% (Haselkorn *et al.*, 2001, Strnad *et al.*, 2010). Within its genome there are large regions of functionally clustered genes, including 8 clustered interspaced short palindromic repeat (CRISPR) regions, a photosynthesis gene cluster of ~45 kb, and 237 phage-like genes, including the RcGTA gene cluster (Strnad *et al.*, 2010). Most of the genes encoding RcGTA are organized in a ~15 kb cluster (RCAP_rcc01682 to RCAP_rcc01698) (Lang & Beatty, 2002) comprising 17 ORFs in an operon that encodes most of the GTA particle proteins (Lang & Beatty, 2000), however at least five additional RcGTA related genes are encoded at two distant loci (Hynes *et al.*, 2012, Hynes *et al.*, 2016, Westbye *et al.*, 2016) (Figure 1.2). A search for the 15 GTA genes from *R. capsulatus* in complete and draft bacterial genomes showed that they are widely present in most α-proteobacteria, with an especially high rate of complete gene clusters present in the *Rhodobacterales* order (Lang & Beatty, 2007, Biers *et al.*, 2008)

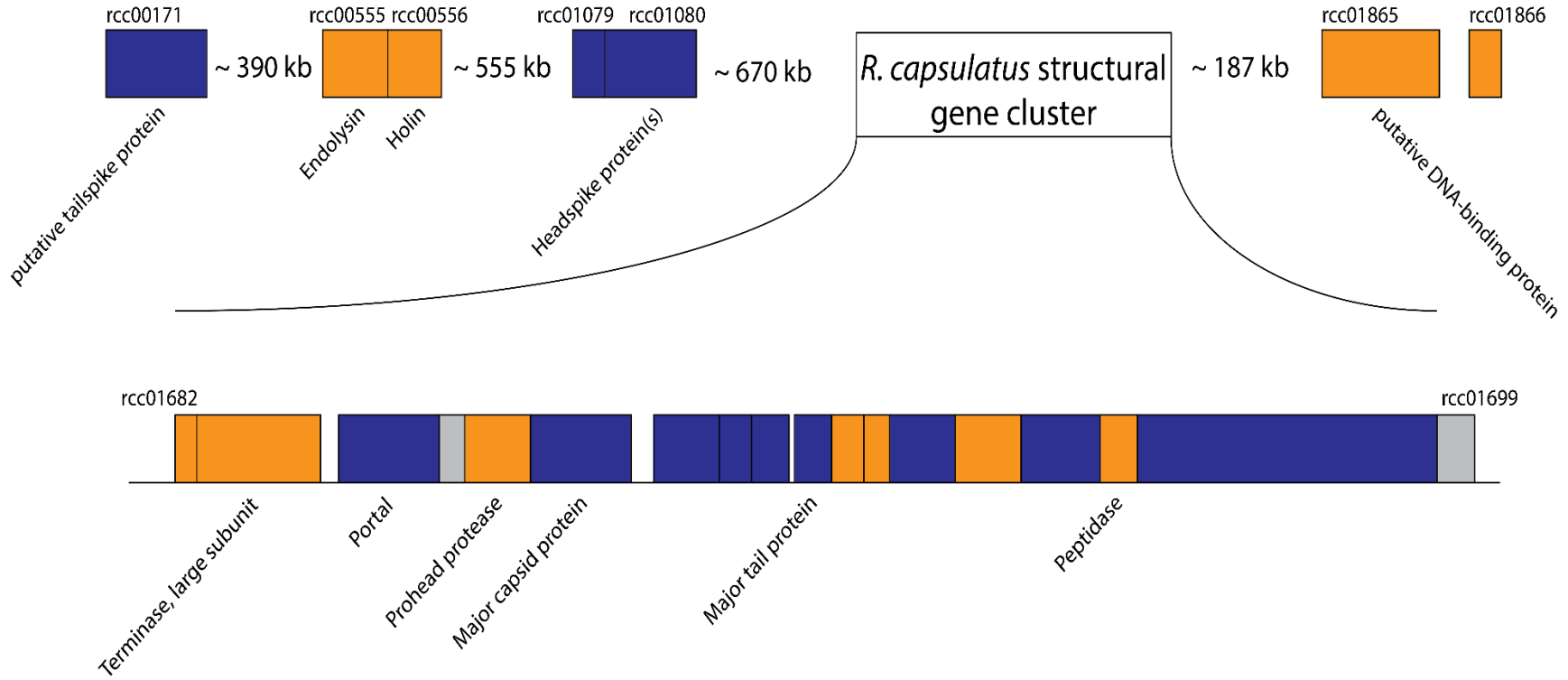

Figure 1.2: The RcGTA structural gene cluster (bottom) and additional genes involved in RcGTA particle production (top). Individual ORFs are drawn to scale and the distances between loci are shown. Putative roles of encoded gene products are shown below each gene. The genes represented in orange were not experimentally found associated with purified RcGTA particles in a previous proteomic study whereas genes represented in blue encode proteins associated with the purified RcGTA particles (Chen *et al.*, 2008).

The GTA particles of *R. capsulatus* resemble small tailed phages and have a head diameter of ~30 nm and a tail length of ~50 nm (Yen *et al.,* 1979). Only approximately 4 kb of genomic DNA is incorporated into its GTA particles (Lang & Beatty, 2000). This amount of DNA is not sufficient to contain all the GTA structural genes necessary to encode the production of new particles inside a recipient cell (Lang & Beatty, 2000, Matson *et al.,* 2005). Transcription levels of the GTA genes and the amount of GTA particles released into the extracellular milieu are dependent upon several cellular regulatory systems (Mercer *et al.*, 2012, Lang *et al.*, 2017, Westbye *et al.*, 2017).

One of the cellular regulatory systems on which the production of RcGTAs is dependent is a two-component signal transduction pathway that involves the global response regulator protein CtrA. Two-component signaling systems are bacterial signal transduction pathways that can mediate rapid alterations in gene expression in response to environmental and intracellular changes (Galperin & Gomelsky, 2005). The CtrA protein was first characterized in *Caulobacter crescentus* (Quon *et al.,* 1996), where it acts as a master regulator of the cell cycle and is essential for cell viability (Skerker & Laub, 2004). CtrA has also been found to control the expression of small RNAs (sRNAs) in *C. crescentus* (Landt *et al.*, 2008). In contrast to *C. crescentus*, the CtrA protein is not essential in *R. capsulatus* and does not seem to be involved in cell cycle processes. However, it has been found that CtrA is required for the transcription of the RcGTA gene cluster (Mercer *et al.*, 2010). In the absence of *ctrA*, a complete loss of RcGTA expression can be observed (Lang & Beatty, 2002). Although no predicted CtrA binding sites have been identified upstream of the RcGTA gene cluster, it seems likely that CtrA somehow directly controls the RcGTA genes or it regulates a binding factor that affects the RcGTA gene cluster. Since CtrA in *C. crescentus* was found to be involved in sRNA expression, it is

conceivable that CtrA in *R. capsulatus* has a similar mechanism, regulating a sRNA that itself regulates RcGTA production in the presence of CtrA. To date only a few sRNAs have been identified that are involved in the regulation of GTA regulating genes, which makes more research necessary to illuminate the role of sRNAs, and RNAs in general, in the production and release of RcGTAs.

1.5 Distribution of known gene transfer agents

Over the last decade we have seen an increase in sequenced bacterial genomes, which has allowed for a more comprehensive search for the distribution of GTAs. The evolutionary relationships between species that have been identified to possess complete or incomplete GTA gene clusters have been described in phylogenetic studies using RcGTA as a model (Lang & Beatty, 2007). It was found that the most highly scoring nucleotide BLAST (Basic Local Alignment Search Tool) hits corresponding to RcGTA sequences were from α-proteobacteria, in which most species contained at least some of the genes that are required to encode GTA particles (Lang & Beatty, 2007).

It is believed that RcGTA-like genes have been transmitted vertically from a GTA-containing ancestor strain (Lang & Beatty, 2007). The fact that GTAs are evolutionary conserved helps support the hypothesis that GTAs provide a selective advantage that benefits the host species (Lang *et al.*, 2012). However, to date, the sole known function for GTAs is to mediate gene transfer.

More than 25% of marine prokaryotic communities may be represented by the *Roseobacter* lineage in the Rhodobacterales, and numerous isolates of these bacteria have been found to

express GTA capsid proteins (Fu *et al.*, 2010). Studies showed that up to 47% of laboratory culturable recipient strains can receive genetic material through GTA-mediated transfer, which suggests that cross-species gene transfer via GTAs may occur in the natural environment (McDaniel *et al.*, 2010). The fact that many bacterial species have been found to possess the necessary genes that are required to encode GTAs and the potential that functional particles are actually produced by many of these species, suggests that GTAs may play a pivotal role in the overall evolution of bacterial genomes and the shaping of microbial communities in general.

The GTAs produced by different organisms can vary greatly in some aspects of their molecular biology. The GTA of *B. hyodysenteriae* for example is encoded in two co-regulated gene clusters (Matson *et al.*, 2005, Stanton *et al.*, 2009). Unlike *R. capsulatus*, the *B. hyodysenteriae* GTA structural genes show no sequence homology with known phage genes. The GTAs of *B. henselae* can encapsulate up to 14 kb of linear DNA (Anderson *et al.*, 1994). In general, DNA segments from throughout the producing cells' genomes are incorporated into the GTA particles in most organisms (Yen *et al.*, 1979, Rapp & Wall, 1987, Humphrey *et al.*, 1997, Bertani, 1999, Hynes *et al.*, 2012). The first documented clear deviation from random packaging comes from the bacterium *Dinoroseobacter shibae* where chromosomal position and GC content are associated with differences in packaging, and DNA from different replicons in the cells are packaged at very different rates (Tomasch *et al.*, 2018). DNA packaging into tailed phage particles is performed by the terminase enzyme (Casjens, 2011), and the *Bartonella, R. capsulatus* and *D. shibae* GTA clusters contain predicted terminase-encoding genes (Lang & Beatty, 2000, Berglund *et al.*, 2009, Tomasch *et al.*, 2018). The sequence specificity of this enzyme, or lack thereof, possibly combined with other features of the genomic DNA, are expected to determine exactly what DNA is packaged into the GTA particles. The organization

of the GTA structural genes are not yet known for *Methanococcus voltae* and *Desulfovibrio desulfuricans*.

It was recently pointed out that GTAs carrying DNA would mimic viral particles under epifluorescence microscopy if they contain enough DNA to be detected (Forterre *et al.*, 2013, Soler *et al.*, 2015). It has been estimated that there are up to a billion virus particles in every milliliter of seawater (Suttle, 2005). GTAs could potentially represent a portion of these particles (Lang & Beatty, 2007, Rohwer & Thurber, 2009, Kristensen *et al.*, 2010). Therefore, further work is needed to specifically quantify GTAs in environmental samples to understand how much of a contribution they make to what we currently view as the viral fraction within environmental samples (Forterre 2013). This will then also allow us to understand better the contributions of GTAs to gene exchange in natural environments. One of the fundamental unanswered questions regarding GTAs remains if and how they contribute to adaptations and increased fitness of the organisms in nature. Based on what is known about GTAs today, it appears that they function solely to mediate gene exchange and therefore act as a driving force in the evolution of bacterial genomes. Regulation of one GTA by cellular systems, such as quorum-sensing, supports the possibility that GTA production is purposeful and beneficial.

1.6 RNAs and their involvement in bacterial gene regulation

Bacteria have developed diverse mechanisms of regulation that can act at several different levels of gene expression. The first level of gene regulation is via controlling the initiation of transcription, and most transcription-regulating pathways involve proteins that bind to DNA, which cause genes to be expressed or repressed, depending on a protein's unique

interaction with DNA. However, it is now known that RNAs also act as regulatory molecules for gene expression regulation.

The involvement of sRNAs in gene regulation was first discovered in the early 1980s in regulation of expression of to the outer membrane proteins OmpF and OmpC in *E. coli* (Mizuno *et al.*, 1984). That study identified the interaction of a sRNA encoded by the *ompC* gene with an *ompF* mRNA. It was proposed that the sRNA OmpC inhibits the translation of the *ompF* mRNA by means of hybridization, causing premature termination of translation and/or destabilization of the *ompF* mRNA. Another study investigated the involvement of a plasmid-derived antisense sRNA (RNAα) in the regulation of the expression of iron transport genes in *Vibrio anguillarum* (Chen & Crosas, 1996). The authors found that synthesis of RNAα was dependent on the ferric uptake regulator (Fur) at the level of transcription initiation, and independent of the iron status of the cell. The study was able to show that RNAα plays a role in the negative regulation of the expression of the iron transport genes under physiological conditions. In the early 2000s, new computational approaches allowed for systematic genome-wide screenings for regulatory RNA molecules (Rivas & Eddy, 2001, Rivas *et al.*, 2001). These studies successfully identified many novel structural RNA genes and sRNAs based on comparative sequence analysis algorithms. In addition to bioinformatics, traditional cloning-based techniques and microarray experiments have been used to screen for sRNAs and have played a key role in confirming predictions from computational approaches.

Bacterial regulatory sRNAs can modulate transcription or translation (Storz, 2002). The mechanisms by which sRNAs do this include: changes within RNA conformation, protein binding, base pairing with other RNAs, and interactions with DNA (Waters & Storz, 2009).

Regulatory sRNAs can be divided into two major classes with respect to their mechanism of action. The first class of regulatory sRNAs act post-transcriptionally by binding to proteins and modifying the protein's activity. An example of a protein-binding sRNA is CsrB (Romeo, 1998). CsrB is a sRNA involved in the global carbon storage regulator (Csr) system and is important for growth efficiency in bacteria. The sRNA CsrB is essential to the Csr system as it works to inactivate the RNA-binding protein, CsrA, which ultimately leads to the translation of CsrA-targeted mRNAs. The sRNA CsrB is also an example for a group of sRNAs that can have protein-specific binding sites that allow them to capture several proteins at once. In the case of CsrB, the sRNA can form a large globular ribonucleoprotein complex with approximately 18 CsrA proteins, which antagonizes the effects of CsrA *in vivo* (Romeo, 1998).

The second class of regulatory sRNAs act by base-pairing with a target mRNA to modify the mRNA's translation or stability. This can be achieved by either *cis*- or *trans*-encoded base-pairing (Figure 1.3).

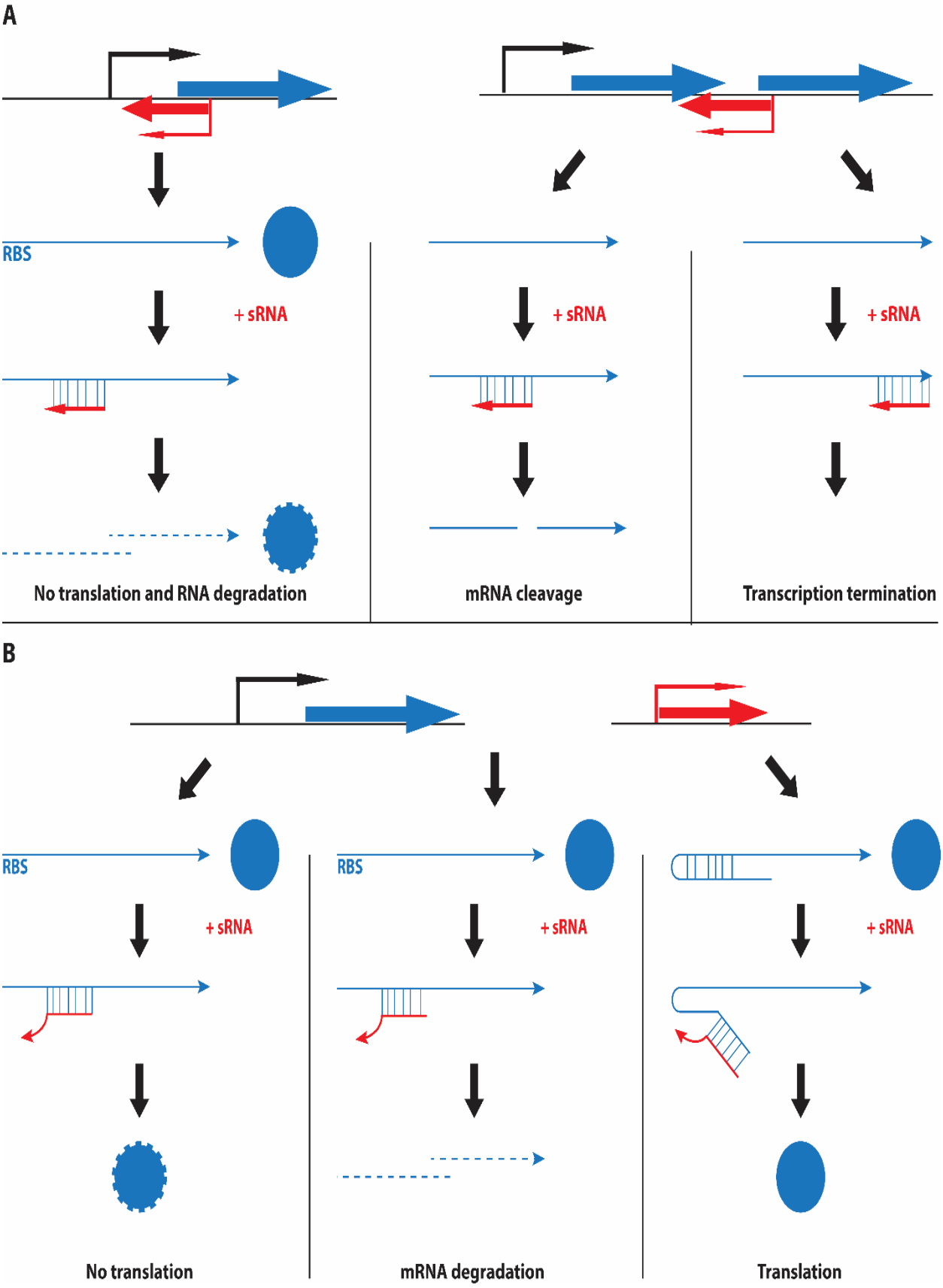
A
RBS
+ sRNA
No translation and RNA degradation
+ sRNA
mRNA cleavage
+ sRNA
Transcription termination
B
RBS
+ sRNA
No translation
RBS
+ sRNA
mRNA degradation
+ sRNA
Translation

Figure 1.3: Gene arrangement and regulatory functions of base-pairing small regulatory RNAs. (A) Two possible configurations of *cis*-encoded antisense sRNAs (red) and their target RNAs (blue), which share extensive complementarity. Left panel: a sRNA encoded opposite to the 5'-UTR of its target mRNA. Base pairing inhibits ribosome binding and often leads to target mRNA degradation. Right panel: a sRNA encoded opposite to the sequence separating two genes in an operon. Base pairing of the sRNA can target RNases to the region and cause mRNA cleavage, with various regulatory effects, or the sRNA can cause transcriptional termination by binding to the ribosome binding site (RBS), leading to reduced levels of expression of the downstream genes. (B) Genes encoding *trans*-encoded sRNAs (red) are located separately from the genes encoding their target RNAs (blue) and only have limited complementarity. *Trans*-encoded sRNAs can act negatively with the 5'-UTR and block ribosome-binding (left panel) and/or targeting the sRNA-mRNA duplex for degradation by RNases (middle panel). *Trans*-encoded sRNAs can act positively by preventing the formation of an inhibitory structure, which sequesters the ribosome binding site (right panel) (based on a figure from (Waters & Storz, 2009)).

Cis-encoded anti-sense sRNAs are encoded on the DNA strand opposite the target RNA and share extended regions of complete complementarity with their target, often 75 nucleotides or more (Wagner *et al.*, 2002). *Cis*-encoded anti-sense sRNAs can cause the activation or inhibition of translation, or the degradation of target mRNAs. *Trans*-encoded sRNAs are encoded in a different locus than their targets and may share only limited complementarity with their target mRNA. In some cases, only two or so short (7-9 nucleotides) regions of complementarity are involved in the interaction between the RNAs. This limited complementarity has been found to enable *trans*-encoded sRNA to regulate multiple target mRNAs (Masse *et al.*, 2007). The majority of known *cis*- or *trans*-encoded anti-sense sRNAs regulate their target mRNA by inhibiting translation (Gottesman, 2005, Aiba, 2007). In the case of positive regulation, the sRNA binds to its target mRNA and causes stabilization of the target

and initiation of translation. Like *cis*-encoded sRNAs, *trans*-encoded sRNAs can regulate the translation or stability of their target mRNAs. In most cases in which a sRNA interacts with a target mRNA, the chaperone protein Hfq facilitates the association with the mRNA target and, in some cases, protects the sRNA from ribonuclease cleavage by regulating access of the sRNA to RNase E (Moller *et al.*, 2002, Zhang *et al.*, 2002, Geissmann & Touati, 2004, Vogel & Luisi, 2011, De Lay *et al.*, 2013). Regulatory sRNAs have been predicted to play a major role in improving bacterial adaptation to environmental changes (Guillier *et al.*, 2006, Vogel & Papenfort, 2006, Valentin-Hansen *et al.*, 2007).

Another group of sRNAs that have regulatory effects on their targets, called sponge RNAs, have been identified and described in recent years. The first sponge RNAs in prokaryotes were discovered in 2009 in *Salmonella enterica* (Figueroa-Bossi *et al.*, 2009) and in *E. coli* (Overgaard *et al.*, 2009). Bacterial sRNA promoters can be so powerful that it becomes challenging for the cell to repress them, even under full repression conditions (Lalaouna *et al.*, 2015). This can cause transcriptional noise from the sRNA promoters. Therefore, bacterial cells have developed sponge RNAs that are capable of sequestering sRNAs when they are present in excess. One example of such a sponge RNA is the tRNA-derived sponge RNA 3'ETS^{leuZ} of *E. coli* (Lalaouna *et al.*, 2015). Bacterial transfer RNAs (tRNAs) are processed from long primary transcripts containing 5'- and 3'external transcribed spacers (ETS), as well as internal transcribed spacers (ITS). These spacers separate tRNA genes into polycistronic transcripts. The maturation of a tRNA involves several endonucleolytic cleavage events that are followed by exonucleolytic trimming, resulting in the mature tRNA. The ETS and ITS remnants are thought to undergo rapid degradation (Li & Deutscher, 2002). 3'ETS^{leuZ} is derived from the *glyW-cysT-*

leuZ operon, which encodes three tRNA genes that are co-transcribed as a tricistronic precursor, which is subsequently cleaved into the mature forms of the three tRNAs (Li & Deutscher, 2002, Ow & Kushner, 2002). It was found that 3'ETS^{leuZ} contains a conserved sequence that is capable of binding both RybB and RyhB by base pairing of short stretches of RNA. The sRNA RybB is involved in the regulation of a cell envelope stress response under the control of the alternative sigma factor σ^E. RybB regulates multiple mRNA targets in the cell that encode some of the major outer membrane proteins (OMPs), such as porins and transporters, by base pairing with their 5'-UTR, thus causing repression of their synthesis (Johansen *et al.*, 2006, Gogol *et al.*, 2011). The sRNA RyhB coordinates bacterial responses to iron starvation (Masse *et al.*, 2005, Salvail *et al.*, 2010) and is in turn controlled by Fur. In iron-rich conditions, *ryhB* is repressed by Fur. Under iron starvation conditions, however, Fur is inactive and *ryhB* becomes highly expressed, inducing the rapid degradation of several mRNAs that produce non-essential Fe-using proteins (Masse *et al.*, 2005, Jacques *et al.*, 2006). 3'ETS^{leuZ} only sequesters *rybB* and *ryhB* when the sRNAs are present in large stochiometric excess. This way 3'ETS^{leuZ} sets a concentration threshold that either sRNA must reach to become effective in regulation. Since both *rybB* and *ryhB* sRNAs are competing for the same sponge, the threshold that is being dictated by 3'ETS^{leuZ} has been predicted to vary depending on the level of expression of RybB to RyhB or vice versa. At the same time, while 3'ETS^{leuZ} sequesters its target sRNAs, the levels of the mature tRNA from the same precursor remain unaffected, which suggests that 3'ETS^{leuZ} itself is a functional RNA.

1.7 Next-generation sequencing and transcriptomics

Sequencing technologies have applications ranging from genome sequencing, epigenetics and metagenomics to the discovery of sRNAs and protein binding sites (Hall, 2007, Holt & Jones, 2008, Marguerat *et al.*, 2008, Snyder *et al.*, 2009). The introduction of next-generation sequencing technology, which became first available in 2005 (Margulies *et al.*, 2005), initially aimed to (re)sequence genomes in a shorter time and at a lower cost than traditional Sanger sequencing (Sanger *et al.*, 1977). The ability of this new technology in the early stages was limited to short reads of around 36 to 100 nucleotides, which was adequate for resequencing applications, but too short for *de novo* assembly (MacLean *et al.*, 2009). In the last decade, new high-throughput next-generation sequencing technologies, all using different underlying biochemistries, have been developed that are capable of read lengths of more than 300 bp and of producing millions of DNA sequence reads in a single run (Mardis, 2008).

Before the commercial application of next-generation sequencing, microarrays were used to investigate the bacterial transcriptome. Microarrays use probes to simultaneously analyze the expression of thousands of genes. Each probe thereby targets a unique sequence within a transcript of interest. The results obtained with this method are a snapshot of actively expressed genes and transcripts at a given point in time. One of the drawbacks of most microarrays is that the snapshot of the transcriptome they provide is incomplete. They cannot detect previously unidentified genes or transcripts. Therefore, microarrays will miss changes in the expression in those genes, which may be associated with a phenotype of interest. Next-generation sequencing on the other hand offers a more comprehensive sequence coverage through single-read or paired-end sequencing, enabling the rapid and deep profiling of the transcriptome.

The benefits of high-throughput RNA sequencing (RNA-seq) over microarrays include the ability to detect novel transcripts because of its independence from species- or transcript-specific probes (Wang *et al.*, 2009, Wilhelm & Landry, 2009). RNA-seq has a wider dynamic range ($>10^5$ for RNA-seq vs. 10^3 for microarrays) because of its ability to measure gene expression as discrete, digital sequencing read counts (Zhao *et al.*, 2014). RNA-seq has a higher specificity and sensitivity for differentially expressed genes, especially those with low expression, compared to microarrays (Wang *et al.*, 2014, Liu *et al.*, 2015, Li *et al.*, 2016). In the last decade, RNA-seq methods have been employed to identify RNAs on a genome-wide scale in numerous bacterial species, e.g. *E. coli* (Shinhara *et al.*, 2011), *Pseudomonas aeruginosa* (Wurtzel *et al.*, 2012), *Clostridium difficile* (Soutourina *et al.*, 2013), and *R. sphaeriodes* (Berghoff *et al.*, 2009, Berghoff *et al.*, 2011).

1.8 Research goals

The initial hypothesis for this thesis research is that CtrA might regulate transcription of a sRNA that affects expression of the RcGTA genes, which could have previously been missed using microarray experiments. The possibility of sRNA involvement will be approached by utilizing a next-generation sRNA-seq approach to perform a comparative analysis of sRNA expression in wild-type and *ctrA* mutant strains of *R. capsulatus* (Chapter 2). I will investigate the whole transcriptome of *R. capsulatus* using total RNA-seq to gain insight into the transcription of all *R. capsulatus* genes during the early stationary phase of growth. To gain greater insight into the expression of RcGTA-related genes, I will perform a comparative analysis of wild-type and RcGTA-overproducer mutant strains for my sequencing experiments. I will combine these data with a newly developed 5' and 3' end-targeted sequencing approach to

define the operon structure in *R. capsulatus* (Chapter 3), since most of the previous knowledge about operon structure in this bacterium is based on computational predictions and a limited amount of experimental data. Lastly, I will perform a study on the transcription of different RcGTA loci to investigate the effect that loss of function of RcGTA-regulating genes has on the different RcGTA loci (Chapter 4). I will utilize the data from Chapter 3 to predict transcripts from the genes of interest to optimize this analysis. The results of the work presented in this thesis further elucidate the regulatory complexity involved in controlling the *R. capsulatus* transcriptome and provide starting points for future work on the regulation of RcGTA gene transcription.

1.9 References

Aiba H (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. *Current Opinion in Microbiology* **10**: 134-139.

Anderson B, Goldsmith C, Johnson AP, Padmalayam I & Baumstark B (1994) Bacteriophage-Iike particle of *Rochalimaea henselae*. *Molecular Microbiology* **31**: 67-73.

Ashelford KE, Day MJ & Fry JC (2003) Elevated abundance of bacteriophage infecting bacteria in soil. *Applied and Environmental Microbiology* **69**: 285-289.

Barbian KD & Minnick MF (2000) A bacteriophage-like particle from *Bartonella bacilliformis*. *Microbiology* **146**: 599-609.

Beiko RG, Harlow TJ & Ragan MA (2005) Highways of gene sharing in prokaryotes. *Proceedings of the National Academy of Sciences USA* **102**: 14332-14337.

Bergh O, Borsheim KY, Bratbak G & Heldal M (1989) High abundance of viruses found in aquatic environments. *Nature* **340**: 467-468.

Berghoff BA, Glaeser J, Sharma CM, Vogel J & Klug G (2009) Photooxidative stress-induced and abundant small RNAs in *Rhodobacter sphaeroides*. *Molecular Microbiology* **74**: 1497-1512.

Berghoff BA, Glaeser J, Sharma CM, Zobawa M, Lottspeich F, Vogel J & Klug G (2011) Contribution of Hfq to photooxidative stress resistance and global regulation in *Rhodobacter sphaeroides*. *Molecular Microbiology* **80**: 1479-1495.

Berglund EC, Frank AC, Calteau A, Vinnere Pettersson O, Granberg F, Eriksson AS, Naslund K, Holmberg M, Lindroos H & Andersson SG (2009) Run-off replication of host-adaptability genes is associated with gene transfer agents in the genome of mouse-infecting *Bartonella grahamii*. *PLoS Genetics* **5**: e1000546.

Bertani G (1999) Transduction-like gene transfer in the methanogen *Methanococcus voltae*. *Journal of Bacteriology* **181**: 2992-3002.

Biers EJ, Wang K, Pennington C, Belas R, Chen F & Moran MA (2008) Occurrence and expression of gene transfer agent genes in marine bacterioplankton. *Applied and Environmental Microbiology* **74**: 2933-2939.

Blattner FR, Plunkett III G, Bloch CA, *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453-1462.

Bratbak G, Heldal M, Norland S & Thingstad TF (1990) Viruses as partners in spring bloom microbial trophodynamics. *Applied and Environmental Microbiology* **56**: 1400-1405.

Burrus V & Waldor MK (2004) Shaping bacterial genomes with integrative and conjugative elements. *Research in Microbiology* **155**: 376-386.

Canchaya C, Proux C, Fournous G, Bruttin A & Brussow H (2003) Prophage genomics. *Microbiology and Molecular Biology Reviews* **67**: 238-276.

Casjens S (2003) Prophages and bacterial genomics: what have we learned so far? *Molecular Microbiology* **49**: 277-300.

Casjens SR (2011) The DNA-packaging nanomotor of tailed bacteriophages. *Nature Reviews Microbiology* **9**: 647-657.

Chandler JR & Dunny GM (2004) *Enterococcal* peptide sex pheromones: synthesis and control of biological activity. *Peptides* **25**: 1377-1388.

Chen F, Spano A, Goodman BE, Blaster KR, Sabat A, Jeffery E, Norris A, Shabanowitz J, Hunt

DE & Lebedev N (2008) Proteomic analysis and identification of the structural and regulatory proteins of the *Rhodobacter capsulatus* gene transfer agent. *Journal of Proteome Research* **8**: 967-973.
Chen I & Dubnau D (2004) DNA uptake during bacterial transformation. *Nature Reviews Microbiology* **2**: 241-249.
Chen I, Christie PJ & Dubnau D (2005) The ins and outs of DNA transfer in bacteria. *Science* **310**: 1456-1460.
Chen Q & Crosas JH (1996) Antisense RNA, Fur, iron, and the Regulation of Iron Transport Genes in *Vibrio anguillarum*. *Journal of Biological Chemistry* **271**: 18885-18891.
de la Cruz F & Davies J (2000) Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends in Microbiology* **8**.
De Lay N, Schu DJ & Gottesman S (2013) Bacterial small RNA-based negative regulation: Hfq and its accomplices. *Journal of Biological Chemistry* **288**: 7996-8003.
Del Campillo-Campbell A, Kayajanian G, Campbell A & Adhya S (1967) Biotin-requiring mutants of *Escherichia coli* K-12. *Journal of Bacteriology* **94**: 2065-2066.
Dodd IB, Shearwin KE & Egan JB (2005) Revisited gene regulation in bacteriophage lambda. *Current Opinion in Genetics and Development* **15**: 145-152.
Eiserling F, Pushkin A & Bertani G (1999) Bacteriophage-like particles associated with the gene transfer agent of *Methanococcus voltae* PS. *Journal of General Virology* **80**: 3305-3308.
Figueroa-Bossi N, Valentini M, Malleret L, Fiorini F & Bossi L (2009) Caught at its own game: regulatory small RNA inactivated by an inducible transcript mimicking its target. *Genes and Development* **23**: 2004-2015.
Filée J, Forterre P & Laurent J (2003) The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. *Research in Microbiology* **154**: 237-243.
Fogg PC, Hynes AP, Digby E, Lang AS & Beatty JT (2011) Characterization of a newly discovered Mu-like bacteriophage, RcapMu, in *Rhodobacter capsulatus* strain SB1003. *Virology* **421**: 211-221.
Forterre P, Soler N, Krupovic M, Marguet E & Ackermann HW (2013) Fake virus particles generated by fluorescence microscopy. *Trends in Microbiology* **21**: 1-5.
Fu Y, MacLeod DM, Rivkin RB, Chen F, Buchan A & Lang AS (2010) High diversity of *Rhodobacterales* in the subarctic North Atlantic Ocean and gene transfer agent protein expression in isolated strains. *Aquatic Microbial Ecology* **59**: 283-293.
Galperin MY & Gomelsky M (2005) Bacterial signal transduction modules: from genomics to biology. *ASM News* **71**: 326-333.
Geissmann TA & Touati D (2004) Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO Journal* **23**: 396-405.
Gogol EB, Rhodius VA, Papenfort K, Vogel J & Gross CA (2011) Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proceedings of the National Academy of Sciences USA* **108**: 12875-12880.
Gottesman S (2005) Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends in Genetics* **21**: 399-404.
Guillier M, Gottesman S & Storz G (2006) Modulating the outer membrane with small RNAs. *Genes and Development* **20**: 2338-2348.
Hacker J & Carniel E (2001) Ecological fitness, genomic islands and bacterial pathogenicity.

EMBO Reports **2**: 376-381.
Hall N (2007) Advanced sequencing technologies and their wider impact in microbiology. *Journal of Experimental Biology* **210**: 1518-1525.
Hara S, Terauchi K & Koike I (1991) Abundance of viruses in marine waters: assessment by epifluorescence and transmission electron microscopy. *Applied and Environmental Microbiology* **57**: 2731-2734.
Haselkorn R, Lapidus A, Kogan Y, *et al.* (2001) The *Rhodobacter capsulatus* genome. *Photosynthesis Research* **70**: 43-52.
Hayashi T, Makino K, Ohnishi M, *et al.* (2001) Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Research* **8**: 11-22.
Hennes KP & Simon M (1995) Significance of bacteriophages for controlling bacterioplankton growth in a mesotrophic lake. *Applied and Environmental Microbiology* **61**: 333-340.
Hennes KP, Suttle CA & Chan AM (1995) Fluorescently labeled virus probes show that natural virus populations can control the structure of marine microbial communities. *Applied and Environmental Microbiology* **61**: 3623-3627.
Hinton DM (2010) Transcriptional control in the prereplicative phase of T4 development. *Virology Journal* **7**.
Holt RA & Jones SJ (2008) The new paradigm of flow cell sequencing. *Genome Research* **18**: 839-846.
Humphrey SB, Stanton TB & Jensen NS (1995) Mitomycin C induction of bacteriophages from *Serpulina hyodysenteriae* and *Serpulina innocens*. *FEMS Microbiology Letters* **134**: 97-101.
Humphrey SB, Stanton TB, Jensen NS & Zuerner RL (1997) Purification and characterization of VSH-1, a generalized transducing bacteriophage of Serpulina hyodysenteriae. *Journal of Bacteriology* **179**: 323-329.
Hynes A, Shakya M, Mercer R, *et al.* (2016) Functional and evolutionary characterization of a gene transfer agent's multi-locus "genome". *Molecular Biology and Evolution* **33**: 2530-2543.
Hynes AP, Mercer RG, Watton DE, Buckley CB & Lang AS (2012) DNA packaging bias and differential expression of gene transfer agent genes within a population during production and release of the *Rhodobacter capsulatus* gene transfer agent, RcGTA. *Molecular Microbiology* **85**: 314-325.
Illumina (2011) Transitioning from microarrays to mRNA-Seq. *Illumina Community Newsletter*.
Imhoff JF (1995) Taxonomy and physiology of phototrophic purple bacteria and green sulfur bacteria. *Anoxygenic Photosynthetic Bacteria Advances in Photosynthesis and Respiration* **2**: 1-15.
Jacques JF, Jang S, Prevost K, Desnoyers G, Desmarais M, Imlay J & Masse E (2006) RyhB small RNA modulates the free intracellular iron pool and is essential for normal growth during iron limitation in *Escherichia coli*. *Molecular Microbiology* **62**: 1181-1190.
Johansen J, Rasmussen AA, Overgaard M & Valentin-Hansen P (2006) Conserved small non-coding RNAs that belong to the sigmaE regulon: role in down-regulation of outer membrane proteins. *Journal of Molecular Biology* **364**: 1-8.
Kloesges T, Popa O, Martin W & Dagan T (2011) Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. *Molecular Biology and Evolution* **28**: 1057-1074.

Kristensen DM, Mushegian AR, Dolja VV & Koonin EV (2010) New dimensions of the virus world discovered through metagenomics. *Trends in Microbiology* **18**: 11-19.
Krupovic M, Forterre P & Bamford DH (2010) Comparative analysis of the mosaic genomes of tailed archaeal viruses and proviruses suggests common themes for virion architecture and assembly with tailed viruses of bacteria. *Journal of Molecular Biology* **397**: 144-160.
Krupovic M, Prangishvili D, Hendrix RW & Bamford DH (2011) Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiology and Molecular Biology Reviews* **75**: 610-635.
Lalaouna D, Carrier MC, Semsey S, Brouard JS, Wang J, Wade JT & Masse E (2015) A 3' external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. *Molecular Cell* **58**: 393-405.
Landt SG, Abeliuk E, McGrath PT, Lesley JA, McAdams HH & Shapiro L (2008) Small non-coding RNAs in *Caulobacter crescentus*. *Molecular Microbiology* **68**: 600-614.
Lang AS & Beatty JT (2000) Genetic analysis of a bacterial genetic exchange element: The gene transfer agent of *Rhodobacter capsulatus*. *Proceedings of the National Academy of Sciences USA* **97**: 859-864.
Lang AS & Beatty JT (2002) A bacterial signal transduction system controls genetic exchange and motility. *Journal of Bacteriology* **184**: 913-918.
Lang AS & Beatty JT (2007) Importance of widespread gene transfer agent genes in alpha-proteobacteria. *Trends in Microbiology* **15**: 54-62.
Lang AS, Zhaxybayeva O & Beatty JT (2012) Gene transfer agents: phage-like elements of genetic exchange. *Nature Reviews Microbiology* **10**: 472-482.
Lang AS, Westbye AB & Beatty JT (2017) The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. *Annual Review of Virology* **4**: 87-104.
Lawrence JG (2002) Gene transfer in bacteria: speciation without species? *Theoretical Population Biology* **61**: 449-460.
Lederberg J & Tatum EL (1946) Gene recombination in the bacterium *Escherichia coli*. *Nature* **158**.
Li J, Hou R, Niu X, Liu R, Wang Q, Wang C, Li X, Hao Z, Yin G & Zhang K (2016) Comparison of microarray and RNA-Seq analysis of mRNA expression in dermal mesenchymal stem cells. *Biotechnology Letters* **38**: 33-41.
Li Z & Deutscher MP (2002) RNase E plays an essential role in the maturation of *Escherichia coli* tRNA precursors. *RNA* **8**: 97-109.
Little JW (1984) Autodigestion of lexA and phage λ repressors. *Proceedings of the National Academy of Sciences USA* **81**: 1375-1379.
Liu Y, Morley M, Brandimarto J, *et al.* (2015) RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics* **105**: 83-89.
MacLean D, Jones JDG & Studholme DJ (2009) Application of 'next-generation' sequencing technologies to microbial genetics. *Nature Reviews Microbiology* **7**: 96-97.
Madigan MT (1995) Microbiology of nitrogen fixation by anoxygenic photosynthetic bacteria. *Anoxygenic Photosynthetic Bacteria Advances in Photosynthesis and Respiration* **2**.
Maggi RG & Breitschwerdt EB (2005) Isolation of bacteriophages from *Bartonella vinsonii* subsp. berkhoffii and the characterization of Pap31 gene sequences from bacterial and phage DNA. *Journal of Molecular Microbiology and Biotechnology* **9**: 44-51.
Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**: 133-141.

Marguerat S, Wilhelm BT & Bahler J (2008) Next-generation sequencing: applications beyond genomes. *Biochemistry Society Transactions* **36**: 1091-1096.
Margulies M, Egholm M, Altman WE*, et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.
Marrs B (1974) Genetic recombination in *Rhodopseudomonas capsulata*. *Proceedings of the National Academy of Sciences USA* **71**: 971-973.
Masse E, Vanderpool CK & Gottesman S (2005) Effect of RyhB small RNA on global iron use in *Escherichia coli*. *Journal of Bacteriology* **187**: 6962-6971.
Masse E, Salvail H, Desnoyers G & Arguin M (2007) Small RNAs controlling iron metabolism. *Current Opinion in Microbiology* **10**: 140-145.
Matson EG, Thompson MG, Humphrey SB, Zuerner RL & Stanton TB (2005) Identification of genes of VSH-1, a prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. *Journal of Bacteriology* **187**: 5885-5892.
McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB & Paul JH (2010) High frequency of horizontal gene transfer in the oceans. *Science* **330**: 50-50.
Mercer R & Lang AS (2014) Identification of a predicted partner-switching system that affects production of the gene transfer agent RcGTA and stationary phase viability in *Rhodobacter capsulatus*. BMC Microbiology 14.
Mercer RG, Quinlan M, Rose AR, Noll S, Beatty JT & Lang AS (2012) Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. *FEMS Microbiology Letters* **331**: 53-62.
Mercer RG, Callister SJ, Lipton MS, Pasa-Tolic L, Strnad H, Paces V, Beatty JT & Lang AS (2010) Loss of the response regulator CtrA causes pleiotropic effects on gene expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *Journal of Bacteriology* **192**: 2701-2710.
Mizuno T, Chou M-Y & Inouye M (1984) A unique mechanism regulating gene expression: Translational inhibition by complementary RNA transcript (micRNA). *Biochemistry* **81**: 1966-1970.
Moller T, Franch T, Hojrup P, Keene DR, Bächinger HP, Brennan RG & Valentin-Hansen P (2002) Hfq: A bacterial Sm-like protein that mediates RNA-RNA interaction. *Molecular Cell* **9**: 23-30.
Morse ML, Ledeberg EM & Lederberg J (1955) Transduction in *Escherichia coli* K-12. *Genetics* **41**: 142-156.
Nagao N, Yamamoto J, Komatsu H, Suzuki H, Hirose Y, Umekage S, Ohyama T & Kikuchi Y (2015) The gene transfer agent-like particle of the marine phototrophic bacterium *Rhodovulum sulfidophilum*. *Biochemistry and Biophysics Reports* **4**: 369-374.
Nakamura Y, Itoh T, Matsuda H & Gojobori T (2004) Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nature Genetics* **36**: 760-766.
Ochman H, Lawrence JG & Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**.
Overgaard M, Johansen J, Moller-Jensen J & Valentin-Hansen P (2009) Switching off small RNA regulation with trap-mRNA. *Molecular Microbiology* **73**: 790-800.
Ow MC & Kushner SR (2002) Initiation of tRNA maturation by RNase E is essential for cell viability in E. coli. *Genes and Development* **16**: 1102-1115.
Pal C, Macia MD, Oliver A, Schachar I & Buckling A (2007) Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **450**: 1079-1081.

Paolozzi L & Ghelardini P (2006) The bacteriophage Mu. *The Bacteriophages* 469-496.
Pemberton JM, Horne IM & McEwan AG (1998) Regulation of photosynthetic gene expression in purple bacteria. *Microbiology* **144**: 267-278.
Ptashne M, Jeffrey A, Johnson AD, Maurer R, Meyer BJ, Pabo CO, Roberts TM & Sauer RT (1980) How the λ repressor and Cro work. *Cell* **19**: 1-11.
Qi J, Zhenghong Y, Jianguo X, *et al.* (2002) Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157. *Nucleic Acids Research* **30**: 4432-4441.
Quon KC, Marczynski GT & Shapiro L (1996) Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell* **84**: 83-93.
Rapp BJ & Wall JD (1987) Genetic transfer in *Desulfovibrio desulfuricans*. *Proceedings of the National Academy of Sciences USA* **84**: 9128-9130.
Rivas E & Eddy SR (2001) Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* **2**.
Rivas E, Klein RJ, Jones TA & Eddy SR (2001) Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Brief Communication* **11**: 1369-1373.
Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F & Mira A (2009) Explaining microbial population genomics through phage predation. *Nature Reviews in Microbiology* **7**: 828-836.
Rohwer F & Thurber RV (2009) Viruses manipulate the marine environment. *Nature* **459**: 207-212.
Romeo T (1998) Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Molecular Microbiology* **29**: 1321-1330.
Rosario K & Breitbart M (2011) Exploring the viral world through metagenomics. *Current Opinion in Virology* **1**: 289-297.
Salvail H, Lanthier-Bourbonnais P, Sobota JM, Caza M, Benjamin JA, Mendieta ME, Lepine F, Dozois CM, Imlay J & Masse E (2010) A small RNA promotes siderophore production through transcriptional and metabolic remodeling. *Proceedings of the National Academy of Sciences USA* **107**: 15223-15228.
Sanger F, Nicklen S & Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA* **74**: 5463-5467.
Santamaria RI, Bustos P, Sepulveda-Robies O, *et al.* (2014) Narrow-host-range bacteriophages that infect *Rhizobium etli* associate with distinct genomic types. *Applied and Environmental Microbiology* **80**: 446-454.
Shinhara A, Matsui M, Hiraoka K, Nomura W, Hirano R, Nakahigashi K, Tomita M, Mori H & Kanai A (2011) Deep sequencing reveals as-yet-undiscovered small RNAs in *Escherichia coli*. *BMC Genomics* **12**: 428-2164-2112-2428.
Simpson AJG & Reinach FC & Arruda P, *et al.* (2000) The genome sequence of the plant pathogen Xylella fastidiosa. *Nature* **406**: 151.
Skerker JM & Laub MT (2004) Cell-cycle progression and the generation of asymmetry in *Caulobacter crescentus*. *Nature Reviews in Microbiology* **2**: 325-337.
Snyder LA, Loman N, Pallen MJ & Penn CW (2009) Next-generation sequencing--the promise and perils of charting the great microbial unknown. *Microbial Ecology* **57**: 1-3.
Soler N, Krupovic M, Marguet E & Forterre P (2015) Membrane vesicles in natural environments: a major challenge in viral ecology. *ISME Journal* **9**: 793-796.
Solioz M, Yen HC & Marrs B (1975) Release and uptake of gene transfer agent by

Rhodopseudomonas capsulata. *Journal of Bacteriology* **123**: 651-657.
Sorensen SJ, Bailey M, Hansen LH, Kroer N & Wuertz S (2005) Studying plasmid horizontal transfer in situ: a critical review. *Nature Reviews in Microbiology* **3**: 700-710.
Soutourina OA, Monot M, Boudry P, *et al.* (2013) Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. *PLoS Genetics* **9**: e1003493.
Stanton TB (2007) Prophage-like gene transfer agents-novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. *Anaerobe* **13**: 43-49.
Stanton TB, Humphrey SB, Sharma VK & Zuerner RL (2008) Collateral effects of antibiotics: carbadox and metronidazole induce VSH-1 and facilitate gene transfer among *Brachyspira hyodysenteriae* strains. *Applied and Environmental Microbiology* **74**: 2950-2956.
Stanton TB, Humphrey SB, Bayles DO & Zuerner RL (2009) Identification of a divided genome for VSH-1, the prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. *Journal of Bacteriology* **191**: 1719-1721.
Storz G (2002) An expanding universe of noncoding RNAs. *Science* **296**: 1260-1263.
Streisinger G, Mukai F, Dreyer WJ, Miller B & Horiuchi S (1961) Mutations affecting the lysozyme of phage T4. Vol. 26 p. 25-30. Cold Spring Harbor Laboratory Press.
Strnad H, Lapidus A, Paces J, Ulbrich P, Vlcek C, Paces V & Haselkorn R (2010) Complete genome sequence of the photosynthetic purple nonsulfur bacterium *Rhodobacter capsulatus* SB 1003. *Journal of Bacteriology* **192**: 3545-3546.
Susanna KA, den Hengst CD, Hamoen LW & Kuipers OP (2006) Expression of transcription activator ComK of *Bacillus subtilis* in the heterologous host *Lactococcus lactis* leads to a genome-wide repression pattern: a case study of horizontal gene transfer. *Applied and Environmental Microbiology* **72**: 404-411.
Suttle CA (2005) Viruses in the sea. *Nature* **437**: 356-361.
Suttle CA (2007) Marine viruses--major players in the global ecosystem. *Nature Reviews Microbiology* **5**: 801-812.
Thomas CM & Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews in Microbiology* **3**: 711-721.
Tomasch J, Wang H, Hall ATK, *et al.* (2018) Packaging of *Dinoroseobacter shibae* DNA into gene transfer agent particles is not random. *Genome Biology and Evolution*.
Valentin-Hansen P, Johansen J & Rasmussen AA (2007) Small RNAs controlling outer membrane porins. *Current Opinion in Microbiology* **10**: 152-155.
Vogel J & Papenfort K (2006) Small non-coding RNAs and the bacterial outer membrane. *Current Opinion in Microbiology* **9**: 605-611.
Vogel J & Luisi BF (2011) Hfq and its constellation of RNA. *Nature Reviews Microbiology* **9**: 578-589.
Wagner EGH, Altuvia S & Romby P (2002) Antisense RNAs in bacteria and their genetic elements. *Advances in Genetics* **46**: 361-398.
Wang C, Gong B, Bushel PR, *et al.* (2014) The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nature Biotechnology* **32**: 926-932.
Wang Z, Gerstein M & Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**: 57-63.
Waters CM & Dunny GM (2001) Analysis of functional domains of the *Enterococcus faecalis*

pheromone-induced surface protein aggregation substance. *Journal of Bacteriology* **183**: 5659-5667.
Waters LS & Storz G (2009) Regulatory RNAs in bacteria. *Cell* **136**: 615-628.
Westbye AB, Beatty JT & Lang AS (2017) Guaranteeing a captive audience: coordinated regulation of gene transfer agent (GTA) production and recipient capability by cellular regulators. *Current Opinion in Microbiology* **38**: 122-129.
Westbye AB, Kuchinski K, Yip CK & Beatty JT (2016) The gene transfer agent RcGTA contains head spikes needed for binding to the *Rhodobacter capsulatus* polysaccharide cell capsule. *Journal of Molecular Biology* **428**: 477-491.
Wilhelm BT & Landry JR (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* **48**: 249-257.
Winter C, Herndl GJ & Weinbauer MG (2004) Diel cycles in viral infection of bacterioplankton in the North Sea. *Aquatic Microbial Ecology* **35**: 207-216.
Wurtzel O, Yoder-Himes DR, Han K, Dandekar AA, Edelheit S, Greenberg EP, Sorek R & Lory S (2012) The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. *PLoS Pathogens* **8**: e1002945.
Yen HC, Hu NT & Marrs BL (1979) Characterization of the gene transfer agent made by an overproducer mutant of *Rhodopseudomonas capsulata*. *Journal of Molecular Biology* **131**: 157-168.
Zhang A, Wassarman KM, Ortega J, Steven AC & Storz G (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Molecular Cell* **9**: 11-22.
Zhao S, Fung-Leung WP, Bittner A, Ngo K & Liu X (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PloS One* **9**: e78644.
Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF & Papke RT (2006) Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Research* **16**: 1099-1108.
Zinder ND & Lederberg J (1952) Genetic exchange in *Salmonella*. *Journal of Bacteriology* **64**: 679-699.

Chapter 2 - Genome-wide identification and characterization of small RNAs in *Rhodobacter capsulatus* and identification of small RNAs affected by loss of the response regulator CtrA

Abstract

Small non-coding RNAs (sRNAs) are involved in the control of numerous cellular processes through various regulatory mechanisms, and in the past decade many studies have identified sRNAs in a multitude of bacterial species using RNA sequencing (RNA-seq). Here, we present the first genome-wide analysis of sRNA sequencing data in *Rhodobacter capsulatus*, a purple non-sulfur photosynthetic alpha proteobacterium. Using a recently developed bioinformatics approach, sRNA-Detect, we detected 422 putative sRNAs from *R. capsulatus* RNA-seq data. Based on their sequence similarity to sRNAs in a sRNAs collection, consisting of published putative sRNAs from 23 additional bacterial species, and RNA databases, the sequences of 124 putative sRNAs were conserved in at least one other bacterial species; and, 19 putative sRNAs were assigned a predicted function. We bioinformatically characterized all putative sRNAs and applied machine learning approaches to calculate the probability of a nucleotide sequence to be a bona fide sRNA. The resulting quantitative model was able to correctly classify 95.2% of sequences in a validation set. We found that putative cis-targets for antisense and partially overlapping sRNAs were enriched with protein-coding genes involved in primary metabolic processes, photosynthesis, compound binding, and with genes forming part of macromolecular complexes. We performed differential expression analysis to compare the wild-type strain to a mutant lacking the response regulator CtrA, an important regulator of gene

expression in *R. capsulatus*, and identified 18 putative sRNAs with differing levels in the two strains. Finally, we validated the existence and expression patterns of four novel sRNAs by Northern blot analysis.

2.1 Introduction

Bacterial small non-coding RNAs (sRNAs) are regulatory RNAs that are heterogeneous in size (generally approximately 50 to 250 nucleotides) and structure. sRNAs are known to function in a number of regulatory processes such as inhibition and activation of translation, degradation and stabilization of mRNA, transcriptional interference, and control of protein activity. sRNAs are usually classified into five categories based on their regulatory mechanisms. *Cis*-encoded base-pairing RNAs are those that bind to their mRNA target with the highest degree of complementarity. An example of this type of sRNA is GadY, which is involved in the regulation of the acid response system of *Escherichia coli (*Castanie-Cornet *et al., 1999,* De Biase *et al., 1999).* Riboswitches are *cis*-regulatory elements that directly bind a metabolite when abundance of this metabolite exceeds a threshold level. This binding induces a conformational change in the RNA to form a structure that affects transcription termination or translation initiation (Serganov & Nudler, 2013).Some riboswitches also function as sRNAs and are able to act *in trans*, such as the S-adenosylmethionine (SAM) riboswitches SreA and SreB of *Listeria monocytogenes (*Loh *et al., 2009).*These two riboswitches regulate the expression of the virulence regulator PrfA by pairing with the 5' untranslated region (UTR) of its mRNA (Loh *et al.,* 2009). *Trans*-encoded base-pairing small RNAs have limited complementarity to their target mRNA(s) and can, in some cases, regulate more than one target. A well-characterized example of a *trans*-encoded regulatory sRNA is RyhB, which is involved in the regulation of intracellular

iron usage in bacteria such as *E. coli (*Hantke, *2001)*. Protein modulator sRNAs are ones that counter the activities of mRNA-binding proteins. An example is CsrB, which is part of the carbon storage regulator (Csr) system in *E. coli (*Romeo, *1998)*. The final category consists of the clustered regularly interspaced short palindromic repeat (CRISPR) RNAs (crRNAs), which are palindromes interspaced with short unique spacer sequences that act as a defence mechanism against homologous foreign DNA, such as that from viruses (Storz *et al.*, 2011).

Numerous cellular processes, such as metabolic reactions, quorum sensing, biofilm formation, stress responses, and pathogenesis, are controlled by sRNAs in various species of bacteria (Michaux *et al.*, 2014). In the last decade, high-throughput RNA sequencing (RNA-seq) methods have been employed to identify sRNAs on a genome-wide scale in numerous bacterial species (see Table 2.1 for some examples). In this work, we used sRNA-Detect (Pena-Castillo *et al.*, 2016) to perform the first genome-wide detection of sRNAs from RNA-seq data in the purple non-sulfur alpha proteobacterium *Rhodobacter capsulatus.* This is an organism of interest for its metabolic versatility (Imhoff, 2005) and production of a gene transfer agent (Lang *et al.*, 2012).We performed comparative RNA-seq targeting sRNAs in the *R. capsulatus* wild-type strain, SB1003, and a mutant strain, SBRM1, lacking the histidyl-aspartyl phosphorelay response regulator CtrA, and identified 422 putative sRNAs expressed in *R. capsulatus* in the early stationary growth phase when growing in photoheterotrophic conditions. Among these 422 putative sRNAs, we identified 18 sRNAs with differing levels in the two strains. Based on significant matches to sequences in the Rfam database (Burge *et al.*, 2013), in the RNAcentral database (Consortium, 2015), and in the bacterial small regulatory RNA database (BSRD) (Li *et al.*, 2013), 19 of the 422 putative sRNAs were assigned a predicted function. The transcript levels for selected sRNA candidates were validated by Northern blot analysis.

Table 2.1: List of bacterial species and sRNAs used for comparative analysis.

Species	**Phylum or class**	**Genome assembly accession number**	**Number of sRNAs**	**Reference**
Chlamydia trachomatis L2b/UCH-1/proctitis	*Chlamydiae*	NC_010280.2	46	(Albrecht *et al.*, 2010)
Clostridium difficile 630	*Firmicutes*	NC_009089.1	253	(Soutourina *et al.*, 2013)
Streptococcus pneumoniae TIGR4	*Firmicutes*	NC_003028.3	88	
Bacillus subtilis subsp. subtilis str. 168	*Firmicutes*	NC_000964.3	84	(Rasmussen *et al.*, 2009)
Corynebacterium glutamicum ATCC 13032	*Actinobacteria*	NC_003450.3	805	(Mentz *et al.*, 2013)
Mycobacterium tuberculosis H37Rv	*Actinobacteria*	NC_000962.3	258	(Miotto *et al.*, 2012)

Propionibacterium acnes KPA171202	*Actinobacteria*	AE017283.1	79	(Lin *et al.*, 2013)
Streptomyces venezuelae ATCC 10712	*Actinobacteria*	NC_018750.1	175	(Moody *et al.*, 2013)
Streptomyces avermitilis MA-4680	*Actinobacteria*	NC_003155.4	199	(Moody *et al.*, 2013)
Streptomyces coelicolor A3	*Actinobacteria*	NC_003888.3	92	(Moody *et al.*, 2013)
Campylobacter jejuni RM1221, 81-176, 81116, and NCTC11168	*Epsilonproteobacteria*	NC_003912.7, NC_008787.1, NC_009839.1, NC_002163.1	102	(Dugar *et al.*, 2013)
Helicobacter pylori 26695	*Epsilonproteobacteria*	NC_000915.1	276	(Sharma *et al.*, 2010)
Neisseria gonorrhoeae FA 1090	*Betaproteobacteria*	NC_002946.2	231	(McClure *et al.*, 2014)
Caulobacter crescentus sp. K31	*Alphaproteobacteria*	NC_010338.1	29	(Landt *et al.*, 2008)

Rhodobacter capsulatus SB1003	*Alphaproteobacteria*	NC_014034.1 NC_014035.1	422	This work
Agrobacterium tumefaciens	*Alphaproteobacteria*	NC_003062.2, NC_003063.2	187	(Wilms *et al.*, 2012)
Rhodobacter sphaeroides 2.4.1	*Alphaproteobacteria*	NC_007493.2, NC_007494.2	28	(Berghoff *et al.*, 2009, Berghoff *et al.*, 2011)
Sinorhizobium meliloti 1021	*Alphaproteobacteria*	NC_003047.1	150	(Schluter *et al.*, 2010)
Vibrio cholerae O1 biovar El Tor str. N16961	*Gammaproteobacteria*	NC_002505.1, NC_002506.1	480	(Bradley *et al.*, 2011)
Pseudomonas aeruginosa UCBPP-PA14	*Gammaproteobacteria*	NC_008463.1	165	(Wurtzel *et al.*, 2012)
Escherichia coli str. K-12 substr. MG1655	*Gammaproteobacteria*	NC_000913.2	309	(Shinhara *et al.*, 2011)
Erwinia amylovora ATCC 49946	*Gammaproteobacteria*	NC_013971.1	40	(Zeng & Sundin, 2014)

Yersinia pestis KIM10+	*Gammaproteobacteria*	NC_004088.1	31	(Beauregard *et al.*, 2013)
Salmonella enterica subsp. enterica serovar Typhimurium str. SL1344	*Gammaproteobacteria*	NC_016810.1	113	(Kroger *et al.*, 2012)

We collected genome sequences and published putative sRNAs from 23 additional bacterial species, which included representatives from the phyla *Chlamydiae*, *Firmicutes*, and *Actinobacteria*, and the *Alpha-*, *Beta-*, *Gamma-, and Epsilonproteobacteria* classes of the phylum *Proteobacteria*. This yielded a collection of 4,725 predicted sRNAs. Based on sequence comparisons, 124 of the 422 putative *R. capsulatus* sRNAs were conserved in at least one other bacterial species.

Finally, we characterized all putative sRNAs for four bioinformatics characteristics and then applied machine learning approaches to develop a quantitative model to calculate the probability of a given RNA sequence to be a bona fide sRNA. The model was able to correctly classify 95.2% of sequences in a validation set.

2.2 Materials and Methods

2.2.1 *R. capsulatus* growth and RNA isolation

R. capsulatus cultures were grown under anaerobic phototrophic conditions at 35°C in complex YPS medium (Weaver *et al.*, 1975) until four hours after reaching stationary phase. The

culture was mixed 5:1 with 95% ethanol and 5% saturated phenol (Jahn *et al.*, 2008), the cells were pelleted by centrifugation, the supernatant was removed, and the cell pellets were frozen in dry ice/ethanol and stored at -80ºC until RNA isolations were performed. sRNA purification was performed with the NucleoSpin® miRNA kit (MACHEREY-NAGEL) following the manufacturer's protocol for purification of the small RNA fraction (<200 nts).

2.2.2 Library preparation and sequencing

The isolated small RNA fraction was used for RNA library preparation for sequencing using an Ion Torrent Personal Genome Machine (PGM; Thermo-Fisher). The RNA quality was checked prior to library preparation using an Agilent Bioanalyzer (Agilent Technologies). Library preparation followed the manufacturer's recommendations for small RNA libraries with the RNA-seq Kit v2 (Thermo-Fisher). The library was amplified using an Ion Torrent One Touch 2 system. The samples were loaded individually on 316 v2 chips and sequenced with the number of flows set to 550.

2.2.3 Processing of RNA-seq data

The RNA-seq data quality was verified using the FastQC tool (version 0.10.0) and reads were filtered and trimmed using the fastq_quality_trimmer available in the FASTX Toolkit (version 0.0.13.2) with a quality threshold of 22 and minimum read length of 28 nucleotides. Filtered and trimmed reads were mapped to the *R. capsulatus* genome using the Torrent mapper tmap (version 3.0.1), executed with the parameters: -B 18 -a 2 -v stage1 map1 map2 map3. Mapping statistics were obtained using samtools (Li *et al.*, 2009).

2.2.4 Detection of sRNAs from RNA-seq data

sRNAs were predicted from mapped RNA-seq data using sRNA-Detect (Pena-Castillo *et al.*, 2016). sRNA-Detect constructs a coverage vector using the function GenomicArray available in HTSeq (Anders *et al.*, 2015) (version 0.5.4p5) and then goes through the genomic intervals in the coverage vector and finds continuous segments between 20 and 250 nucleotides long with similar numbers of reads, with a maximum percentage change of 3% allowed in the average number of reads. A minimum of 10 reads across all samples was required to consider a transcript as expressed. sRNA-Detect is available at www.cs.mun.ca/~lourdes. Predicted transcripts overlapping to tRNAs and rRNAs were removed from the putative sRNA set using the tool intersectBed available in BEDtools (Quinlan & Hall, 2010) (version 2.25).

2.2.5 Collection and analysis of sRNAs from other bacterial species

Published studies performing genome-wide identification of sRNAs using RNA-seq data were identified (Table 2.1), genomic coordinates of the putative sRNAs were collected, and the corresponding sRNA sequences were obtained using the tool fastaFromBed available in BEDtools.

2.2.6 Bioinformatic analysis of sRNAs

Sequence conservation of putative sRNAs was determined by identifying reciprocal best BLAST matches between pairs of species (Table S1). The program blastn (version 2.2.30+) was executed with an E-value cut-off of 1E-4, a best_hit_overhang of 0.1 and task mode of “blastn”.

Rfam matches were obtained using the batch search functionality in the Rfam database (version 12.1). If a sRNA had multiple Rfam matches only the most significant match was considered. All 30581 sRNA sequences in BSRD were downloaded (May 2015) and a BLAST database was created with these sequences. BSRD matches per sRNA were obtained using blastn with the same settings as for the homology search. If a sRNA had multiple BSRD matches only the match with the lowest E-value was considered. The RNAcentral database (release 5) was downloaded (May 2016) and nhmmer (Wheeler & Eddy, 2013) (version 3.1b2) with an E-value cutoff of 1E-3 was used to identify RNAcentral matches for each putative sRNA. If a sRNA had multiple RNAcentral matches only the most significant match was considered. CentroidFold with parameters -e "CONTRAfold" and -g 4 was used to predict the secondary structure of putative sRNAs and random genomic sequences. Sequences of the sRNAs including 150 nts upstream of the predicted 5' end were obtained using slopBed and fastaFromBed and promoter sites were predicted using BPROM with default values. Rho-independent terminators in *R. capsulatus* genome were predicted using TransTermHP with default values and providing an annotation file with the coordinates of the protein-coding genes. The numbers of reads mapped to the putative sRNAs per strain were calculated using htseq-count available in HTSEQ. Normalized log2 fold changes between the two *R. capsulatus* strains were obtained using edgeR (Robinson *et al.*, 2010) (version 3.12.1). All results were compiled, processed and visualized using R (version 3.2.4).

To apply machine learning approaches, we represented a putative sRNA or a random genomic sequence as a numerical vector X consisting of seven numerical predictors (input variables); namely, free energy of the secondary structure, distance ranging from [-150, 20] nts to the -10 predicted promoter site (if no promoter site was predicted in that range a value of -

1000 was used), distance to terminator ranging from [0,500] nts (if no terminator was predicted within this distance range a value of 1000 was used), distance (-∞, 0] nts to closest left ORF, a binary number indicating whether the RNA is transcribed on same strand as its left ORF (1 if transcribed on same strand), distance [0, +∞) to closest right ORF, and a binary number indicating whether the RNA is transcribed on same strand as its right ORF. For training the classifiers, 33 of the 41 putative sRNAs deemed as bona fide sRNAs were randomly selected as positive instances, and 98 of the 4420 random genomic sequences were randomly selected as negative instances. The remaining sequences were used for testing. Logistic regression was applied using the R function glm (with family = binomial), and cross-validation was performed using the function cv.glm from the R package boot (version 1.3-18). LDA and QDA were applied using the lda and qda functions from the R package MASS (version 7.3-45). Performance measurements were calculated using the R package ROCR (Sing *et al.*, 2005) (version 1.0-7). For the classifiers' performance comparison, we used recall and false positive rates. Recall indicates the proportion of testing positive instances that are predicted to be bona fide sRNAs by a given approach at a certain probability threshold (i.e., true positives (TP) divided by the total number of positive instances (P)). The false positive rate is the proportion of negative instances that are predicted to be bona fide sRNAs by a given approach at a certain probability threshold (i.e., false positives (FP) divided by the total number of negative instances (N)). The logistic regression estimates the parameter θ to model $p(X) = e^{(\theta 0 + \theta 1 X 1 + ... + \theta p X p)} / (1 + e^{(\theta 0 + \theta 1 X 1 + ... + \theta p X p)})$ where X is the vector of attributes representing an instance, e is the base of the natural logarithm, p is the number of attributes in X, and Xi is the value of attribute i. The parameter θ was chosen to maximize the likelihood function. The value of the estimated parameters was $\theta = [-2.02, -0.037, -5.8e-4, -2.1e-3, -0.011, 0.25, 5e-3, 0.38]$.

Intuitively, the model learnt makes sense; for instance, decreasing the free energy increases the probability of being a bona fide sRNA, and decreasing the distance to a terminator increases the probability of being a bona fide sRNA. We used these parameters' values to calculate the probability of being a bona fide sRNA for all putative sRNAs. A probability cut-off of 0.6 was chosen as the optimal cut-off to have high recall while maintaining the false positive rate low.

2.2.7 Detection of sRNAs by Northern blotting

Purified sRNA was eluted in 30 µl of nuclease free water. The water was subsequently evaporated using a vacuum centrifuge (Thermo Scientific, Savant DNA120 SpeedVac Concentrator) for 30 minutes at high vacuum setting. The RNA was then dissolved in 20 µl of nuclease free water to increase the initial concentration.

A denaturing 15% polyacrylamide gel containing 7 M urea was used to separate the sRNAs. The gel was pre-run for 30 minutes at 18 mA (100 V) in 1X TBE buffer (89mM Tris, 89 mM boric acid, 20 mM EDTA; pH 8.0). A total of 10 µg of RNA was prepared in a 10 µl sample for electrophoresis and mixed with 5 µl of 3X loading buffer (95% (v/v) formamide, 20 mM EDTA, some Bromphenole blue and Cyanol xylene) such that paired wild-type and mutant samples contained the same amount of RNA. The low-range single stranded RNA ladder (NEB; N0364S) was included for size reference. The gels were run for 80 minutes at 18 mA (100 V) in 1X TBE buffer. After electrophoresis, the lanes containing one set of samples with a corresponding ladder were cut from the gel and stained in ethidium bromide for 15 minutes before taking an image. The remaining gel was cut into pieces containing paired wild-type and mutant sRNA samples and each pair transferred to a Hybond-N^+ nylon membrane (Amersham)

by electro-blotting in 1X SSC buffer (3M NaCl, 30 mM Sodium Citrate) for 2 hours at 250 mA (4 V). The RNA was cross-linked to the membranes by exposing them to 120000 μJ/ cm^{-2} (UVC500 UV Crosslinker; Hoefer) and the membranes were then dried at 50 °C for 30 minutes.

The membranes were rolled with hybridization mesh and pre-hybridized for 3 hours in 10 ml pre-hybridization solution containing 10 μg ml^{-1} of salmon sperm DNA at 40 °C in a hybridization oven (Model 5420; VWR). After the pre-hybridization step, 50 μg ml^{-1} of biotin-labelled probe (Huang *et al.*, 2014) was added directly to the pre-hybridization solution and the membranes were hybridized with the probe for 16 hours at 40 °C. Probe sequences are given in Table 2.2. After hybridization, the membranes were washed twice in 2X SSC/0.1% SDS for 15 minutes at 40 °C, with a final wash in 0.1X SSC for 15 minutes at room temperature (Rio, 2014). The Chemiluminescent Nucleic Acid Detection Module (catalog # 89880; Thermo-Fisher) was used for probe detection following the manufacturer's recommendations. Images were captured using an ImageQuant LAS4000 (General Electric Canada). The resulting images were adjusted for brightness and contrast using Adobe Photoshop CC 2017. The images of the ethidium bromide-stained portions of the corresponding gels were used to construct standard curves to allow estimation of the sizes of bands detected on the blots.

Table 2.2: Biotin-labeled probes for detection of selected sRNAs on Northern blots.

sRNA	**Oligo sequence (5' to 3')**
sRNA00385	BIO-GCGCAGTTGACGCGCCGTCT
sRNA01029	BIO-GGAAACCGGGCGCGGGAACC
sRNA00848	BIO-TCAAGCCTCTGAGGAAGGTC
sRNA01129	BIO-GGGGCTGTTGACCGCCCGCC

2.3 Results and discussion

2.3.1 Sequencing and detection of *R. capsulatus* sRNAs

We grew cultures under photoheterotrophic conditions to early in the stationary phase of growth so that the data collected would match with our most comprehensive collection of transcriptomic data from previous microarray studies (Mercer *et al.*, 2010, Pena-Castillo *et al.*, 2014). Sequencing of size-selected RNA, ≤ 200 nucleotides, from the genome-sequenced strain, SB1003, and its derived *ctrA* mutant strain, SBRM1 (Lang & Beatty, 2002), generated a total of 4.45 million reads. From these reads, 93.5% were uniquely mapped to the *R. capsulatus* genome. These sequence data have been submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE82056.

Recently, we showed that sRNA-Detect, a new computational program for the detection of bacterial small transcripts from RNA-seq data, exhibits higher recall rates at comparable specificity levels than other standalone computational approaches (Pena-Castillo *et al.*, 2016). We used sRNA-Detect on our sequence data, and after removal of detected small transcripts located within annotated transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), we detected 422 potential sRNAs in *R. capsulatus.*

2.3.2 sRNAs with predicted functions or homologs

To annotate *R. capsulatus* putative sRNAs with predicted functions, we retrieved significant matches to *R. capsulatus* sRNAs from the Rfam, RNAcentral and BSRD databases. Based on these matches, we annotated 19 sRNAs with predicted functions (Table 2.3 and Table S2). There were six riboswitches (including those binding thiamine pyrophosphate and

cobalamin), three segments of transfer-messenger RNA (tmRNA), three segments of the catalytic RNA of ribonuclease P (RNase P RNA), the signal recognition particle (SRP) RNA (ffs), 6S RNA, an α-operon ribosome binding site, the *cspA* thermoregulator, the upstream sRNA of *mraZ* (UpsM) (Weber *et al.*, 2016) and an sRNA homologous to a validated *Rhodobacter sphaeroides* sRNA (RSs1386) (Berghoff *et al.*, 2009). We detected several sRNAs corresponding to fragments of the tmRNA and the RNase P RNA due to differences in read depth coverage across the full length of these transcripts.

Table 2.3: List of functionally annotated sRNAs.

Identifier(s)	**Inferred Annotation**
sRNA00627	TPP riboswitch
sRNA00822	Signal recognition particle (SRP) RNA (ffs)
sRNA00687, sRNA00526, sRNA00688,	Cobalamin riboswitch
sRNA00123	α-operon ribosome binding site
sRNA00598	Bacterial small signal recognition particle RNA
sRNA01208	*cspA* thermoregulator
sRNA01158, sRNA01157, sRNA01156	Transfer-messenger mRNA (fragment of)
sRNA01077	Upstream sRNA of mraZ, UpsM
sRNA00648	6S RNA
sRNA00470	Homologous sRNA to the *Rhodobacter sphaeroides*
sRNA01141, sRNA01140, sRNA01139	Catalytic RNA ribonuclease P (P RNA) (fragment of)

To investigate the extent of sequence conservation of putative *R. capsulatus* sRNAs in different bacterial species, we obtained sRNA sequences identified in recent studies of 23 other bacterial species (Table 2.1) and used BLAST (version 2.2.30+) (Shiryev *et al.*, 2007) to search for pairwise reciprocal best matches between the sRNAs of each of the other 23 bacterial species and the *R. capsulatus* sRNAs from this study. As differences in the characteristics of each study, including but not limited to differences in sequencing platforms, growth conditions, RNA extraction methods, and sRNA identification methods, lead to limitations in this analysis, we also searched for sequence conservation of *R. capsulatus* sRNAs in the genomes of these 23 other bacterial species. In total 124 (or 29%) of the 422 putative sRNAs had homologous sRNAs or were found to be conserved in the genome of at least one other bacterial species (Figure 2.1). We organized these 124 sRNAs based on our level of confidence in their conservation. We referred to sRNAs with matches in at least one of the three RNA databases (Rfam, RNAcentral and BSRD) as hypothetical equivalogs, which represented 24 sRNAs that likely belong to a set of sRNAs conserved with respect to function. This category includes the 19 sRNAs for which we inferred a function. We classified sRNAs with homologs found in bacterial species belonging to other genera as inter-taxa homologs, which represented 40 sRNAs that are likely to be true functional sRNAs. The sRNAs whose sequences were only present in the genome of the related bacterial species *R. sphaeroides* were classified as intra-genus homologs, which represented 60 sRNAs. The remaining 298 putative *R. capsulatus* sRNAs appear to be species-specific. Not surprisingly, there are more intra-genus than inter-taxa homologs and, as already pointed out by Gomez-Lozano *et al*. (Gomez-Lozano *et al.*, 2015), there is limited sRNA sequence conservation across different species.

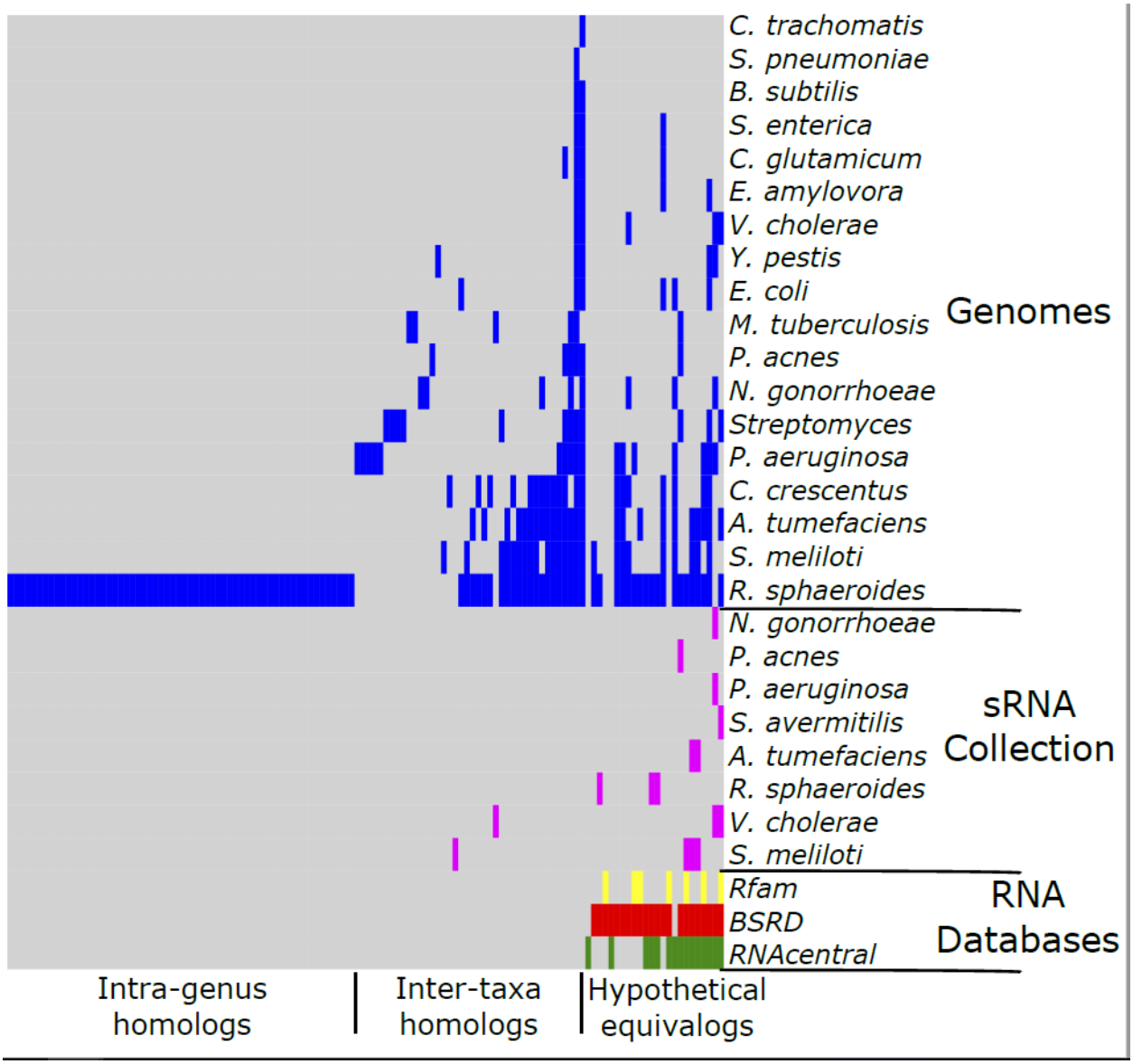

Figure 2.1**:** Map of 124 sRNAs in *R. capsulatus* with sequence conservation in other bacterial species. Sequence similarity searches were performed for all putative *R. capsulatus* sRNAs against three RNA databases and a panel of 23 bacterial species including representatives from the *Chlamydiae*, *Firmicutes*, and *Actinobacteria* phyla and the *Alpha-*, *Beta-*, *Gamma-*, and *Epsilonproteobacteria* classes of the phylum *Proteobacteria*. From right to left, three classes (hypothetical equivalogs, 24 sRNAs; inter-taxa homologs, 40 sRNAs; and intra-genus homologs, 60 sRNAs) proceed from nearly complete certainty about a putative sRNA's function to no functional information. Gray indicates no homologs (matches) were found for the sRNA in that organism or database.

2.3.3 Bioinformatic characterization of putative sRNAs in *R. capsulatus*

We characterized all 422 putative sRNAs in terms of their predicted secondary structures, their proximities to predicted promoter sites, their proximities to predicted Rho-independent terminators and their genomic contexts. To be able to compare the features of the putative sRNAs with a null distribution, we randomly extracted sequences matching the length and strand of putative sRNAs from the *R. capsulatus* genome. There were at least 10 random sequences for each putative sRNA sequence. We used CentroidFold (Hamada *et al.*, 2009) to predict the secondary structures of both the sRNA sequences and the random sequences, and to calculate the free energies of the folded structures. We found that the distribution of free energies of the sRNAs' secondary structures was shifted towards lower values than the distribution of free energies of the random sequences' secondary structures (Figure 2.2A). The difference between the free energies of the sRNAs' secondary structures and the free energies of the random sequences' secondary structures was statistically significant ($p=5.9E-12$, Mann-Whitney test). This indicates that our putative sRNAs tend to adopt more stable conformations than random genomic sequences.

Using the BPROM program (Solovjev & Salamov, 2011), we searched for putative promoters in the region spanning 150 nucleotides (nts) upstream to 20 nts downstream from the predicted 5' ends of both the putative sRNAs and the random genomic sequences. Of the 422 putative sRNAs, 183 (43%) had predicted promoter sites, in contrast to 18.6% of the random sequences. Furthermore, there was a distinct peak at position -21.5 in the probability density function for the -10 promoter positions of putative sRNAs, whereas the random sequences had a uniform probability distribution for the -10 promoter positions (Figure 2.2B). As sRNA-Detect tends to predict transcripts that lie within the boundaries of the actual sRNA (i.e., it misses some

nucleotides at the 5’ and 3’ ends of the sRNAs) (Pena-Castillo *et al.*, 2016), the average distance to the -10 and -35 promoter sites from the actual 5’ end of the sRNAs would be less than as estimated above. Our data indicates that many of the putative sRNAs have proximal promoter sites and supports the notion that they are independently transcribed.

Next, we used TransTermHP (Kingsford *et al.*, 2007), a computational method to detect Rho-independent transcription terminators, to predict the locations of terminators in the *R. capsulatus* genome. We associated a terminator to a putative sRNA if the terminator was within 500 nts downstream from the predicted 3’ end of the sRNA as described by Kingsford *et al.* (Kingsford *et al.*, 2007). Of the 422 putative sRNAs, 130 (31%) had an associated predicted Rho-independent transcription terminator, whereas only 8.15% of the random genomic sequences did. Moreover, as depicted in Figure 2.2C, there was a distinct peak in the probability that the 3’ ends of the sRNAs were located 7 nts from the closest downstream terminator, whereas the random sequences’ density function had a uniform distribution.

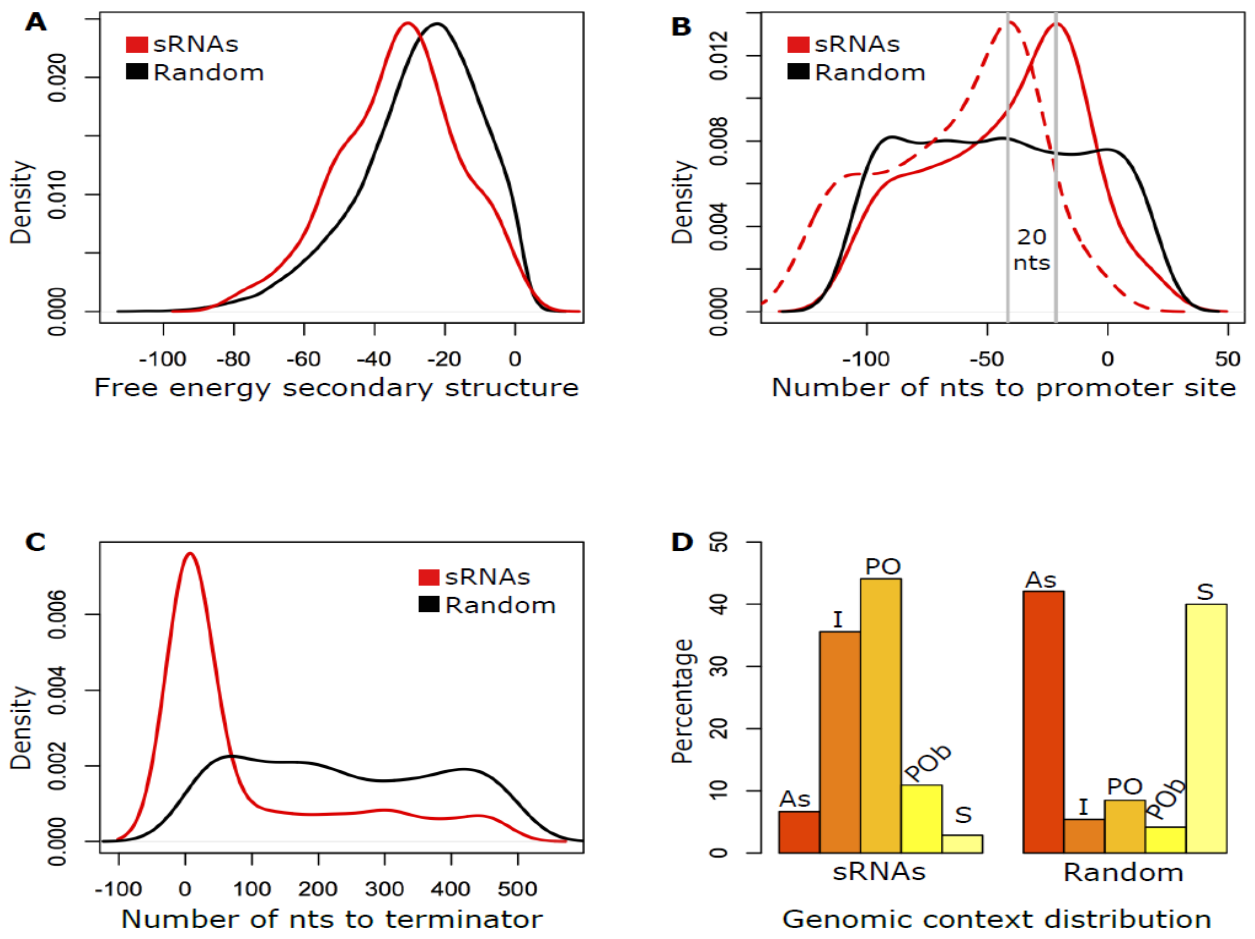

Figure 2.2: Characteristics of putative sRNAs in comparison with the null distribution. A) Probability distribution of the free energy of the predicted secondary structures for the putative sRNAs (red line) and 4,400 random genomic sequences of matching length and orientation (black line). The average free energy of the sRNAs' predicted secondary structures is statistically significantly lower than the average free energy of the random sequences' secondary structures (p=5.902E-12, Mann-Whitney test). B) Density function of the number of nucleotides (nts) upstream from the predicted 5' end of the putative sRNAs to -10 (solid red line) and -35 (dashed red line) predicted promoter sites in comparison with the number of nts from 5' end of random genomic sequences to -10 predicted promoter sites (solid black line). C) As B, but number of nts downstream from the predicted 3' end of the candidate sRNAs (red line) and of random genomic sequences (black line) to predicted Rho-independent terminators. D) Proportion of sRNAs (left) and random genomic sequences (right) in a specific class of genomic context (antisense (AS), 28 sRNAs; intergenic (I), 150 sRNAs; partial overlapping (PO), 186 sRNAs; partial overlapping on both ends (POb), 46 sRNAs; and sense (S), 12 sRNAs).

Based on the putative sRNAs' genomic contexts, we classified the sRNAs as either "intergenic" if located in intergenic regions (IGRs), "antisense" if located within an annotated gene and transcribed on the strand opposite to this gene, "partially overlapping" if the 5' or 3' end of the sRNA overlaps the 5' or 3' end of an annotated gene, "partially overlapping on both ends" if the 5' end of the sRNA overlaps an annotated gene and the 3' end of the sRNA overlaps another annotated gene, or "sense" if located within an annotated gene and transcribed on the same strand as this gene (Figure 2.3). 150 sRNAs were intergenic, 186 were partially overlapping and 46 were partially overlapping on both ends. These amounts were 6.6, 5.2 and 2.6 times more than expected, respectively, if the locations were randomly distributed over the genome (Figure 2.2D). In contrast, 28 antisense sRNAs (asRNAs) and 12 sense sRNAs were detected, which is 6.3 and 14.1 times less than expected, respectively (Figure 2.2D).

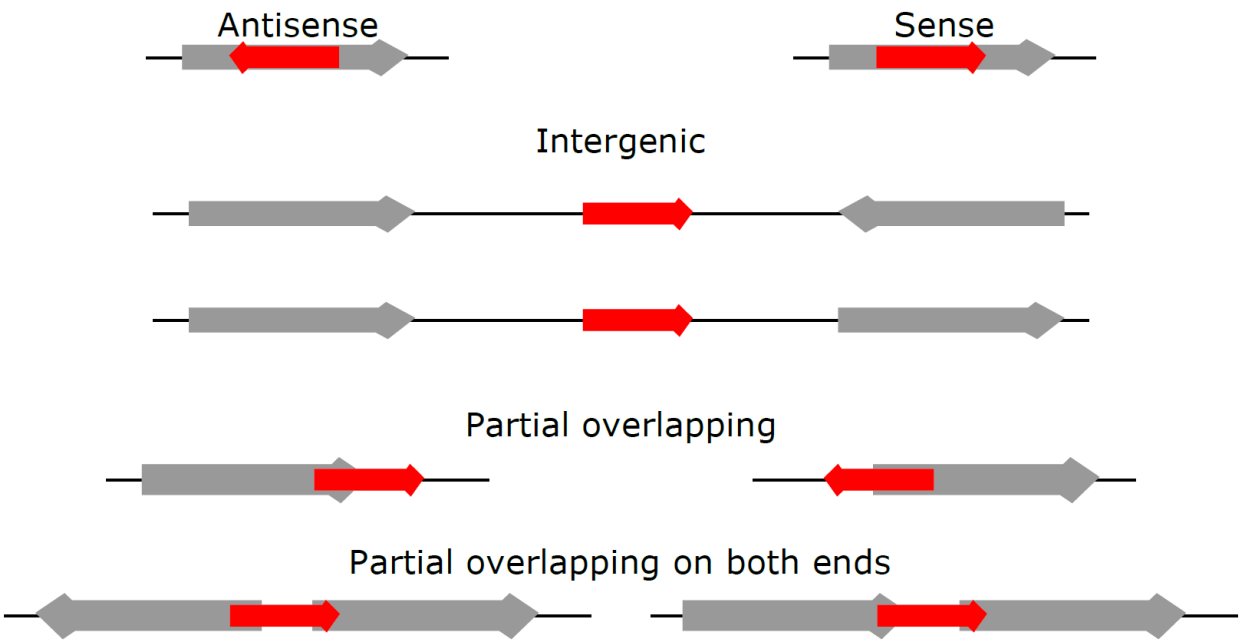

Figure 2.3: Schematic illustration of the different classes of genomic contexts of sRNAs. Genes are depicted as thick arrows with open reading frames (ORFs) shown in grey and sRNAs shown in red. Only a subset of all possible direction of transcription combinations are shown. Antisense RNAs (asRNAs) are within an ORF that is transcribed on the strand opposite to the asRNAs. Sense sRNAs are within an ORF that is transcribed on the same strand as the sRNA. Intergenic sRNAs are found in intergenic regions (IGRs) between ORFs. Partial overlapping sRNAs occur when the 5' or 3' end of the sRNA overlaps with the 5' or 3' end of an ORF. Partial overlapping on both ends sRNAs occur when the 5' of the sRNA overlaps the 5' or 3' end of an ORF and the 3' of the sRNA with the 5' or 3' end of another ORF.

As putative sRNAs had clearly distinct characteristics from random sequences, we decided to apply machine learning approaches (classifiers) to obtain a model to quantify the probability of a sequence being a bona fide sRNA. To derive the model, we selected as predictors (attributes) the free energy of the predicted secondary structure of the sRNA, the distance to a predicted promoter site, the distance to a Rho-independent terminator, and the sRNA genomic context. The genomic context included distance to the closest "left"

neighbouring ORF, distance to the closest "right" neighbouring ORF, and whether the sRNA was on the same strand as the closest neighbouring annotated ORFs. We refer to an annotated ORF located at the 5' end of a sRNA on the forward strand or an annotated ORF located at the 3' end of a sRNA on the reverse strand as "left", and an annotated ORF located at the 3' end of a sRNA on the forward strand or an annotated ORF located at the 5' end of a sRNA on the reverse strand as "right" (illustrated in Figure S1). To create the model, we considered those sRNAs with inter-taxa homologs in the sRNAs collection or conserved in the genome of at least two other bacterial species, and sRNAs with hypothetical equivalogs (Figure 2.1) as "bona fide sRNAs". We randomly chose 33 of these 41 bona fide sRNAs as positive instances and 98 random sequences as negative instances to train the classifiers. We then evaluated the classifiers' performances on the remaining 8 bona fide sRNAs and 4322 random sequences. We applied three machine learning approaches, namely, logistic regression, linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) (James *et al.*, 2014). Among these three methods, logistic regression had the highest recall rates at the lowest false positive rate (Figure S2A-C). Details about the logistic regression model obtained are given in the Materials and Methods section. At a probability cut-off of 0.6, the logistic regression model retrieved 66.25% of the positive test instances and only 4.7% of the random sequences. We then calculated the probability of being a bona fide sRNA using the logistic regression model for all 422 putative sRNAs. Of the 422 putative sRNAs, 109 (26%) scored a probability >0.6 (Figure S2D). At the estimated false positive rate, only five of these 109 sRNAs would be expected to be false positives. We expect that assigning a confidence estimate for being a bona fide sRNA to a given putative sRNA will help prioritize sRNAs for experimental validation. A limitation of this analysis is that, as the majority of positive instances used to learn the logistic regression model

were intergenic or partially overlapping sRNAs, the logistic regression model underestimates the probability of asRNAs being bona fide sRNAs. These analyses need to be replicated in other bacterial species with a larger number of confirmed sRNAs to corroborate these findings and obtain better performance estimates. Table S2 contains the full description of all putative sRNAs, including their estimated probabilities of being bona fide sRNAs.

2.3.4 Identification of a putative tRNA-derived sRNA locus

We observed a putative intergenic sRNA (sRNA00295) found to be conserved in the genomes of 16 other bacterial species without a homologous sRNA in the sRNAs collection or the RNA databases. As this sRNA lacked a homologous sRNA, its function could not be inferred. Nevertheless, we decided to inspect it. The sequence of sRNA00295 was identical to the 3' region of the four tRNA-Met genes found in the *R. capsulatus* chromosome. The homology with the tRNAs makes interpreting the RNA-seq read data somewhat challenging, as reads originating from the tRNAs could be mapping onto this putative sRNA locus, and *vice versa*. However, a promoter site and Rho-independent terminator were predicted to flank this putative sRNA. We also checked this region in an additional unpublished dataset based on differential RNA-seq (dRNA-seq), which identifies 5' ends of RNAs that originate from transcription initiation as opposed to RNA processing (Sharma & Vogel, 2014), and a 5' end was identified at this location (Grüll et al., unpublished). There have been recent discoveries of tRNA-derived sRNAs, which have been implicated in different regulatory processes (Lee *et al.*, 2009, Raina & Ibba, 2014). If genuine, this sRNA would instead represent an independent tRNA-derived fragment locus, and this warrants future investigation. Figure 2.4 depicts sRNA00295's genomic context and predicted secondary structure.

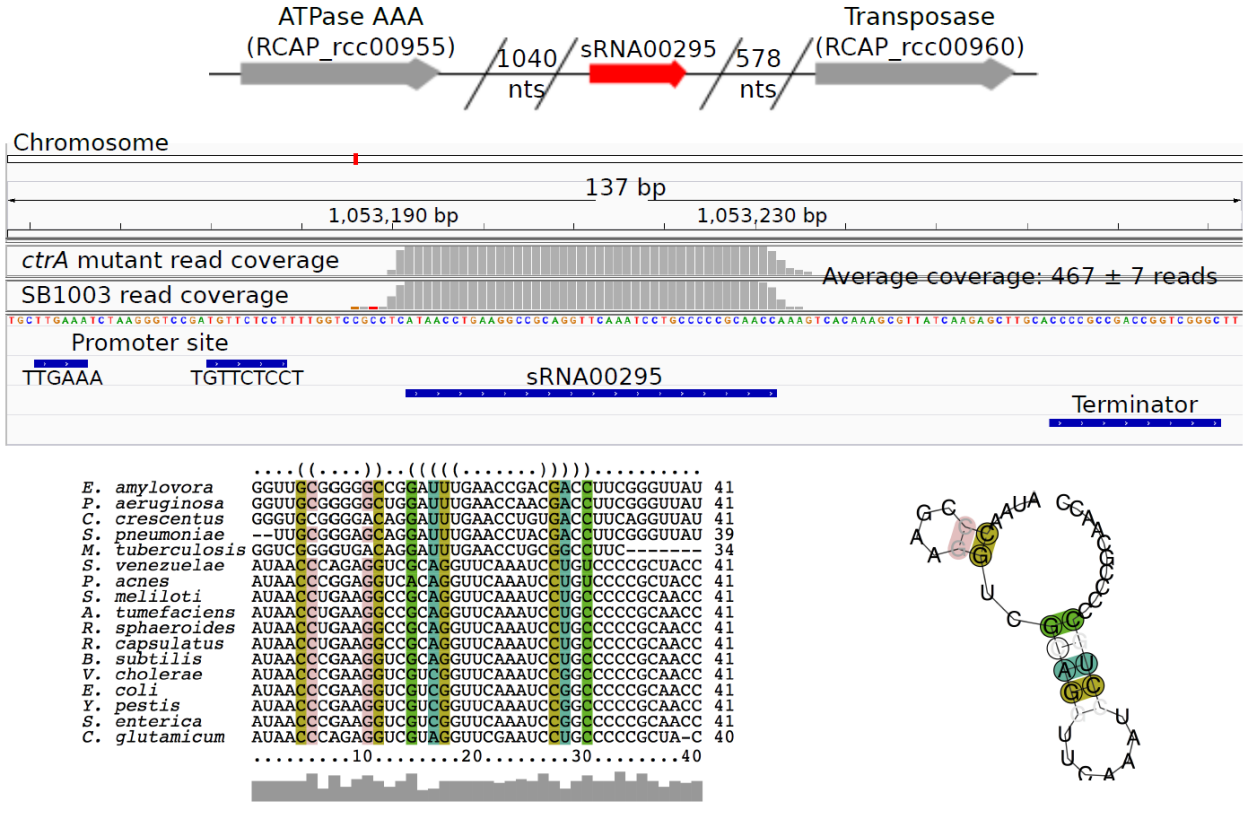

Figure 2.4: Genomic context and predicted secondary structure of a putative tRNA-derived sRNA locus (sRNA00295). The top panel shows sRNA00295's genomic context indicating distance to the closest protein-coding genes. The middle panel illustrates read depth coverage, location of predicted promoter site, and location of Rho-independent terminator (panel generated using Integrative Genomics Viewer version 2.3.72). The numbers of reads mapped for the SB1003 and the ctrA mutant strain are 472 and 462, respectively, as calculated by htseq-count. The bottom panel shows a multiple sequence alignment and predicted consensus secondary structure obtained using LocARNA. Coloured nucleotides indicate correspondence between positions in the alignment and the RNA structure.

To gain insight into the likely functional role of this putative sRNA, we used the CopraRNA web server (Wright *et al.*, 2013) to predict sRNA00295's targets. Despite recent advances, most sRNA target prediction programs have a high false positive rate; CopraRNA,

which requires at least three homologous sequences to predict targets, has twice the prediction accuracy of other sRNA target prediction programs (Pain *et al.*, 2015). Table 2.4 shows the top 10 sRNA00295 targets predicted by CopraRNA (the complete CopraRNA results, which includes all 76 predicted targets are listed in Table S3). To quantify protein interactions among sRNA00295's 76 predicted targets, we used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, version 10.0) database of physical and functional interactions (Szklarczyk *et al.*, 2015). Compared with similarly sized randomly selected protein sets, sRNA00295's 76 predicted targets have significantly more interactions among themselves (PPI enrichment p=1.13E-08), with 28 interactions as compared with eight for random protein sets. We also tested for functional enrichment among sRNA00295's 76 predicted targets using STRING, but no functional enrichment was found.

Table 2.4: Top 10 targets predicted by CopraRNA for sRNA00295 a putative tRNA-derived sRNA locus.

Gene ID	**Description**
RCAP_rcc01474	amino acid permease
RCAP_rcp00009	LacI family transcriptional regulator
RCAP_rcc00101	ABC transporter permease
RCAP_rcc02606	mammalian cell entry domain-containing protein
RCAP_rcc00024	glutaryl-CoA dehydrogenase
RCAP_rcc01400	signal transduction histidine kinase
RCAP_rcc00616	acriflavine resistance protein B
RCAP_rcc00505	type II secretion system protein E

RCAP_rcc01291	kinetochore Spc7 domain-containing protein
RCAP_rcc02771	TetR family transcriptional regulator

2.3.5 Functional and protein-interaction enrichment of potential *cis*-targets of putative antisense and partially overlapping sRNAs

To obtain insight into the biological processes potentially regulated by the antisense and partially overlapping putative sRNAs, we assumed that they were *cis*-acting and examined the 265 overlapping protein-coding mRNAs for functional and protein interaction enrichment using STRING. As antisense and partially overlapping sRNAs have been shown to also regulate gene expression *in trans (*Caldelari *et al., 2013)*, this approach likely missed additional regulatory targets of these putative sRNAs. Nevertheless, the set of *cis*-targets showed a significant enrichment of genes involved in primary metabolic process (28 genes, FDR-corrected p=1.97E-5), photosynthesis (16 genes, FDR-corrected p=3.98E-5), compound binding (24 genes, FDR-corrected p=0.004), and of genes encoding parts of macromolecular complexes (17 genes, FDR-corrected p=3.2E-7). The complete functional enrichment results are provided in Table S4. We also investigated whether putative *cis*-targets were co-expressed based on previously determined *R. capsulatus* gene co-expression modules (Pena-Castillo *et al.*, 2014), and found that *cis*-targets showed a significant accumulation in two gene co-expression modules (13 genes in the midnightblue module, FDR-corrected p=0.002; and 7 genes in the salmon4 module, FDR-corrected p=0.003). Additionally, there were significantly higher interactions among the network of *cis*-targets (PPI enrichment p=0), with 528 interactions as compared to 204 for random protein sets. This indicates that several of the likely *cis*-targets interact and are co-expressed and supports the notion that sRNAs play a regulatory role in these processes.

2.3.6 Effects of loss of *ctrA* on sRNA expression

We investigated whether putative sRNAs were differentially expressed between two *R. capsulatus* strains: the genome-sequenced strain, SB1003, and its *ctrA* null mutant derivative, SBRM1. CtrA is a two-component/histidyl-aspartyl phosphorelay response regulator that affects many processes in *R. capsulatus* such as motility and gene transfer agent production (Mercer *et al.*, 2012). In *Caulobacter crescentus*, where it is an essential protein and controls many cell cycle-related processes, CtrA was shown to regulate expression of sRNAs as part of its regulon (Landt *et al.*, 2008). Figure S3 illustrates the distribution of the normalized log2 fold change of the sRNAs' read counts between the two strains. Although more samples are required to have enough statistical power to identify statistically differentially expressed sRNAs, the vast majority of sRNAs do not appear to be differentially expressed. However, 18 sRNAs had an absolute log2 fold change >3, suggesting possible differential expression between the strains. Among these 18 sRNAs, there are 2 asRNAs, 7 intergenic, 8 partially overlapping, and 2 partially overlapping on both ends sRNAs. Nine of the 14 ORFs overlapped by the antisense and partially overlapping sRNAs were previously identified as affected by the loss of CtrA (Mercer *et al.*, 2010) (p=4.14E-10, Hypergeometric test), including genes encoding the flagellar protein MotB (*rcc00006*), the flagellar hook-associated protein FlgK (*rcc00008*), the Hpt domain-containing protein (*rcc00180*), and the DNA protecting protein DprA (*rcc03098*). We investigated whether these 14 ORFs overlapped by potential differentially expressed sRNAs were co-expressed based on previously determined *R. capsulatus* gene co-expression modules and, indeed, they were significantly over-represented in two modules: pink (6 genes, FDR-corrected p=1.05E-5) and orange (3 genes, FDR-corrected p=2.9E-4). The orange module was identified as associated with

the production of RcGTA (Pena-Castillo *et al.*, 2014) and the DprA protein is required for uptake of DNA from RcGTA particles by recipient cells (Brimacombe *et al.*, 2014), thereby adding sRNAs as another regulatory mechanism involved in controlling RcGTA-mediated gene exchange in *R. capsulatus (*Lang *et al., 2012).* As all of the potentially differentially expressed sRNAs are *R. capsulatus*-specific, we were unable to use CopraRNA to predict potential targets of the intergenic sRNAs.

2.3.7 Experimental validation of putative sRNAs using Northern blot analysis

We chose four putative sRNAs to evaluate by Northern blotting. These were sRNA00385, sRNA01029, sRNA00848, and sRNA01129, representing four intergenic *R. capsulatus*-specific sRNAs, three of which showed differential expression between the wild-type and *ctrA* mutant strains, as evaluated by read counts in the RNA-seq data. We purposefully chose three of the targets due to their predicted differential expression to help with interpretation of the Northern blots as previous studies have detected multiple bands on Northern blots probed for sRNA detection (Thomason *et al.*, 2015). These differentially expressed sRNAs are candidates for future investigation for potential roles in the regulation of CtrA-affected processes, such as the production of RcGTA. As expected due to the program's limitation with respect to correctly identifying the 5' and 3' boundaries of sRNAs (Pena-Castillo *et al.*, 2016), the bands detected for each of the sRNAs were larger than predicted by sRNA-Detect. Manual inspection of the sequence read data allowed us to estimate the boundaries and sizes of these sRNAs more accurately (Figure 2.5) to match the sizes estimated on the Northern blots, and we identified putative promoter sequences for these sRNAs (Figure 2.6) that agree with a previously identified *R. capsulatus* consensus promoter sequence (Leung *et al.*, 2013).

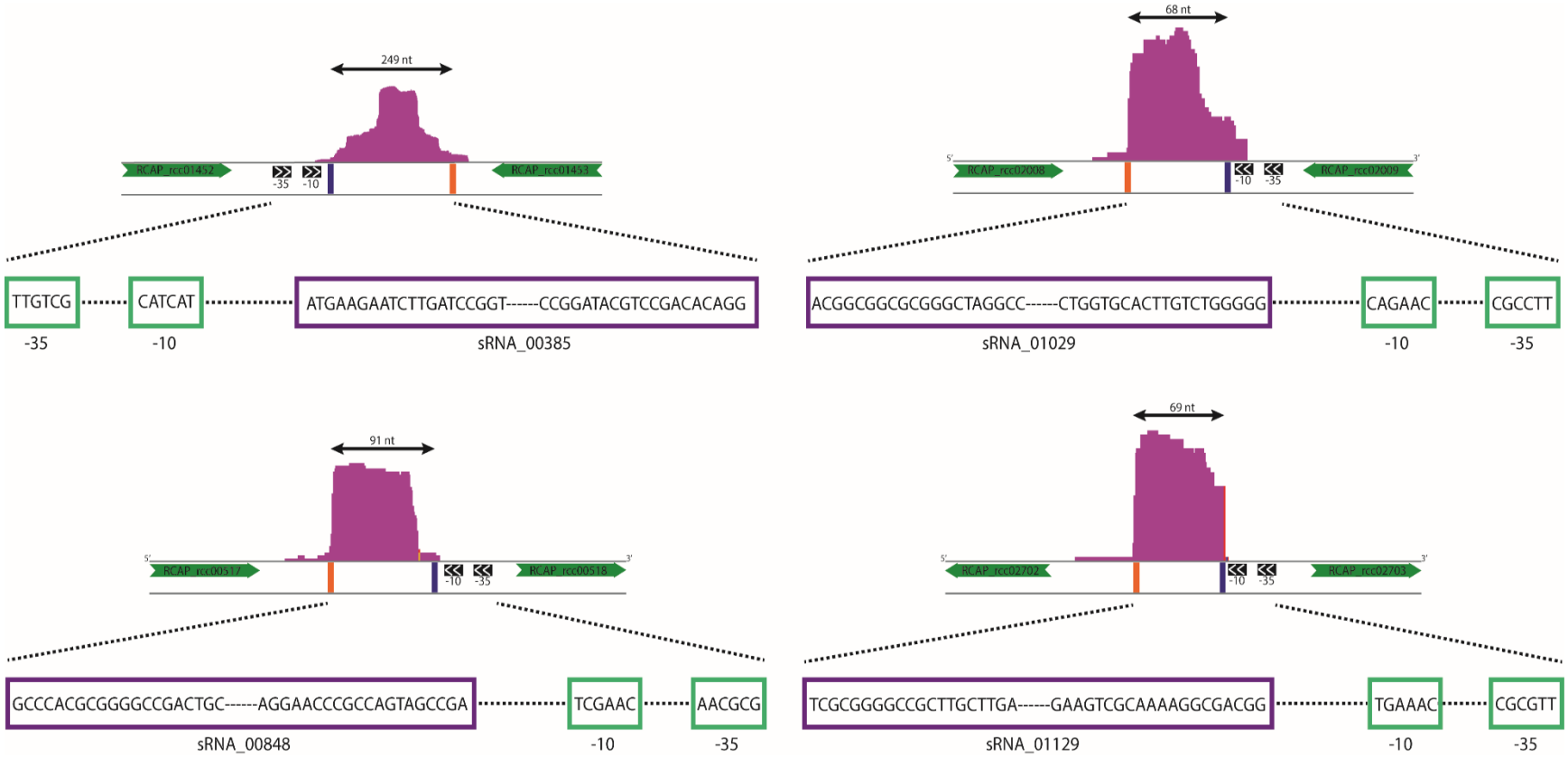

Figure 2.5: Read depth coverage plots and genomic locations for the experimentally confirmed sRNAs. Parts of the neighboring genes are shown with green arrows indicating their direction in the genome. Their relative distance to the coverage plots is not to scale. Predicted promoter -10 and -35 elements are depicted as black boxes with white arrows inside, and the sequences are given in Figure 2.6 and below. The distance of the promoter relative to the 5' end of the corresponding sRNA is also not to scale. The sRNA sequencing reads are presented as purple plots with an indication of the sRNA's predicted size on top. Blue bars mark the predicted 50 ends and orange bars the predicted 3' ends of the sRNAs. The sRNAs' 50 and 30 end sequences and the -10/-35 elements are shown underneath the plots in their respective 5'–3' orientations.

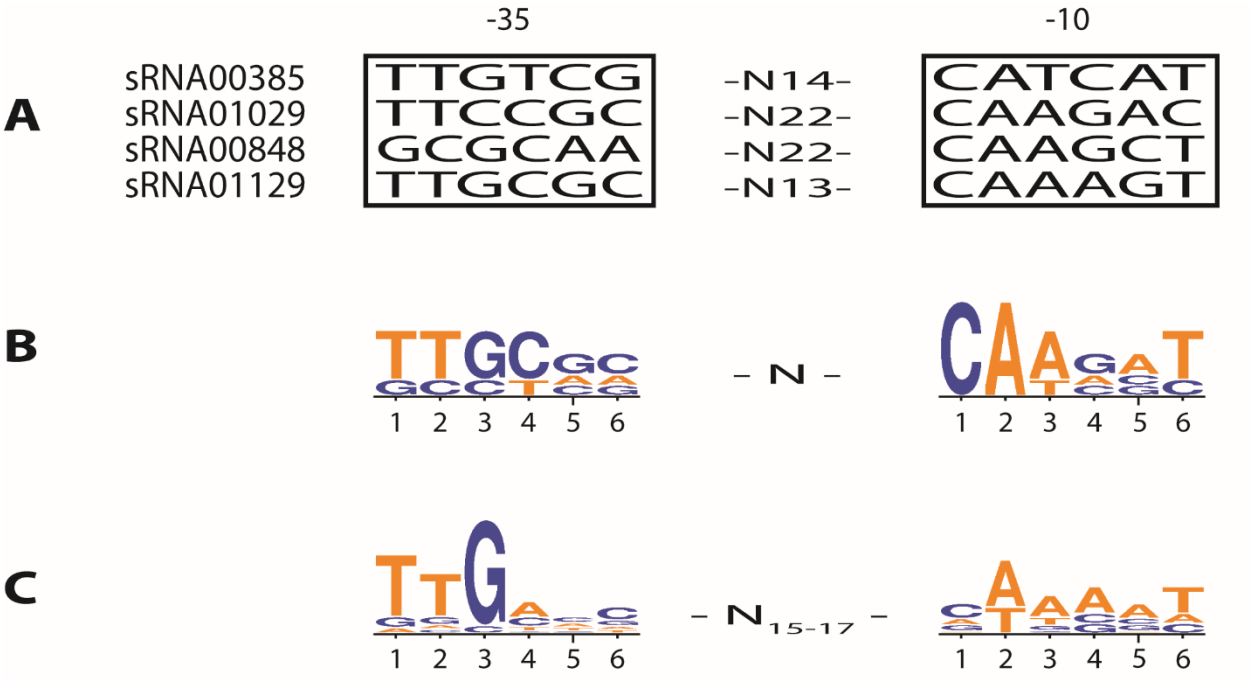

Figure 2.6: Identification of putative promoter -10 and -35 sequences for four experimentally confirmed sRNAs. A) Predicted promoter sites upstream of each of the sRNAs. The nucleotide spacing between the motifs is indicated. B) The frequency of bases found at each position is indicated by the size of the colored letters, created with Weblogo 3.0 (Crooks *et al.*, 2004). C) Consensus promoter sequence based on promoters identified in a previous study (Leung *et al.*, 2013).

sRNA00385 was predicted to have a size of 189 nts based on sRNA-Detect. Examination of the sequence reads for this region suggested an actual size of 249 nts (Figure 2.5). A putative promoter site was found upstream of the predicted 5' end (Figure 2.5 andFigure 2.6) although in this case the -10 site was centered 18 nts upstream of the predicted 5' end, possibly indicating either poor read coverage at the 5' end as frequently found in RNA-seq (Wang *et al.*, 2011), or variable length spacing in the promoter elements (Hook-Barnard & Hinton, 2007, Guzina & Djordjevic, 2016). This putative sRNA showed similar, high levels of expression in the RNA-seq data from both strains. The Northern blot showed a major band at approximately 230 nts (Figure

2.7). There were several additional bands detected on this blot, most of which were present in both strains. These presumably result from non-specific hybridization of the probe to additional RNAs, as has been observed in previous studies detecting sRNAs by this method (Thomason *et al.*, 2015).

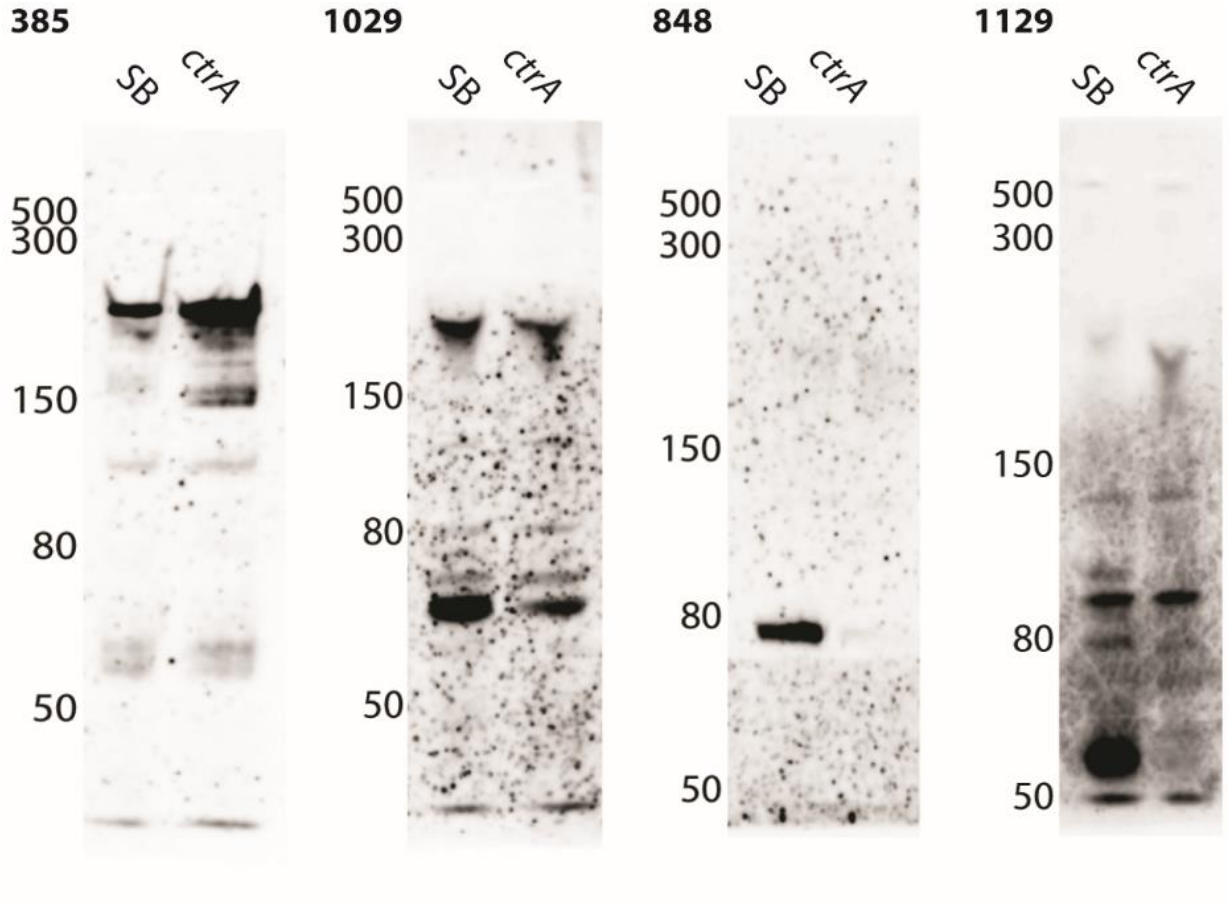

Figure 2.7: Northern blot images of the experimentally confirmed sRNAs. RNA from the genome-sequenced strain, SB1003, and the *ctrA* mutant strain were run in the left and right lane, respectively, of each gel. The sequences of the biotin-labelled probes are given in Table 2.2. The sizes for the corresponding ladder bands are indicated on the left of each blot image, and the number on top of each image identifies the corresponding sRNA probe.

sRNA01029 was predicted to have a size of 52 nts by sRNA-Detect. Inspection of the sequencing data for this sRNA suggested a size of 68 nts and a -10 element was identified centered 12 nts upstream of the predicted sRNA's 5' end (Figure 2.5 andFigure 2.6). The sRNA was predicted to be more highly expressed in the wild-type strain based on read count data (approximately 3:1, Table S2). The Northern blot showed several bands in both strains, with one band at approximately 65 nts that was present at higher levels in the wild-type strain in comparison to the *ctrA* mutant (Figure 2.7).

sRNA00848 was predicted to have a size of 71 nts by sRNA-Detect, with inspection of the sequencing data suggesting a size of 91 nts. A putative promoter sequence was identified upstream of the predicted 5' end (Figure 2.5 andFigure 2.6) but, as with sRNA00385, the -10 sequence was centered more than 10 nts upstream (25 nts). This sRNA was only detected in the RNA-seq data from the wild-type strain and the Northern blot showed a band at approximately 78 nts only in RNA from the wild-type strain (Figure 2.7).

sRNA01129 was predicted to have a size of 69 nts based on sRNA-Detect, and this matched the predicted size from manual inspection of the sequencing data. We found a putative promoter with a -10 element centered 7 nts upstream of the predicted 5' end (Figure 2.5 andFigure 2.6). This sRNA was detected at a much higher level in the wild-type strain RNA-seq data compared to the *ctrA* mutant (28:1, Table S2). The Northern blot showed a band at approximately 60 nts that was present in the wild-type strain but not detected in the *ctrA* mutant (Figure 2.7).

2.4 Conclusion

Using RNA-seq data we have identified 422 putative sRNAs in *R. capsulatus:* 24 sRNAs with hypothetical equivalogs, 40 sRNAs with putative inter-taxa homologs, 60 sRNAs with putative intra-genus homologs and 298 potential *R. capsulatus*-specific sRNAs. To help prioritize further investigations into these sRNAs, we have bioinformatically characterized these sRNAs and used logistic regression to quantify the probability of a putative sRNA being a bona fide sRNA. Using the logistic regression model, 109 (or 26%) of the 422 putative sRNAs were assigned a probability greater than 0.6 of being a bona fide sRNA; at the estimated false positive rate of 4.8%, only five out of these 109 sRNAs are expected to be false positives. Analysis of a strain lacking the important response regulator CtrA identified 18 putative sRNAs that were differentially expressed relative to the wild-type strain. This indicates that effects on the levels of sRNAs is another means by which the CtrA phosphorelay regulates processes in *R. capsulatus*. We experimentally confirmed the existence of four of the putative sRNAs by Northern blot analysis and validated the differential expression that was predicted from the RNA-seq data analysis for three of these. The abundance of sRNAs detected in *R. capsulatus* indicates that a potential extra layer of regulatory complexity exists in this species. Revealing the functional roles of these sRNAs will improve our understanding of the mechanisms *R. capsulatus* employs to regulate its physiology.

2.5 References

Albrecht M, Sharma CM, Reinhardt R, Vogel J & Rudel T (2010) Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. *Nucleic acids research* **38**: 868-877.

Anders S, Pyl PT & Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166-169.

Beauregard A, Smith EA, Petrone BL, Singh N, Karch C, McDonough KA & Wade JT (2013) Identification and characterization of small RNAs in *Yersinia pestis*. *RNA Biology* **10**: 397-405.

Berghoff BA, Glaeser J, Sharma CM, Vogel J & Klug G (2009) Photooxidative stress-induced and abundant small RNAs in *Rhodobacter sphaeroides*. *Molecular Microbiology* **74**: 1497-1512.

Berghoff BA, Glaeser J, Sharma CM, Zobawa M, Lottspeich F, Vogel J & Klug G (2011) Contribution of Hfq to photooxidative stress resistance and global regulation in *Rhodobacter sphaeroides*. *Molecular Microbiology* **80**: 1479-1495.

Bradley ES, Bodi K, Ismail AM & Camilli A (2011) A genome-wide approach to discovery of small RNAs involved in regulation of virulence in *Vibrio cholerae*. *PLoS Pathogens* **7**: e1002126.

Brimacombe CA, Ding H & Beatty JT (2014) *Rhodobacter capsulatus* DprA is essential for RecA-mediated gene transfer agent (RcGTA) recipient capability regulated by quorum-sensing and the CtrA response regulator. *Molecular Microbiology* **92**: 1260-1278.

Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP & Bateman A (2013) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Research* **41**: D226-232.

Caldelari I, Chao Y, Romby P & Vogel J (2013) RNA-mediated regulation in pathogenic bacteria. *Cold Spring Harbor Perspectives in Medicine* **3**: a010298.

Castanie-Cornet MP, Penfound TA, Smith D, Elliott JF & Foster JW (1999) Control of acid resistance in *Escherichia coli*. *Journal of Bacteriology* **181**: 3525-3535.

Consortium RN (2015) RNAcentral: an international database of ncRNA sequences. *Nucleic Acids Research* **43**: D123-129.

Crooks GE, Hon G, Chandonia JM & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Research* **14**: 1188-1190.

De Biase D, Tramonti A, Bossa F & Visca P (1999) The response to stationary-phase stress conditions in *Escherichia coli*: role and regulation of the glutamic acid decarboxylase system. *Molecular Microbiology* **32**: 1198-1211.

Dugar G, Herbig A, Forstner KU, Heidrich N, Reinhardt R, Nieselt K & Sharma CM (2013) High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. *PLoS Genetics* **9**: e1003495.

Gomez-Lozano M, Marvig RL, Molina-Santiago C, Tribelli PM, Ramos JL & Molin S (2015) Diversity of small RNAs expressed in *Pseudomonas* species. *Environmental Microbiology Reports* **7**: 227-236.

Guzina J & Djordjevic M (2016) Promoter recognition by extracytoplasmic function sigma factors: analyzing DNA and protein interaction motifs. *Journal of Baceriology* **198**: 1927-1938.

Hamada M, Kiryu H, Sato K, Mituyama T & Asai K (2009) Prediction of RNA secondary

structure using generalized centroid estimators. *Bioinformatics* **25**: 465-473.
Hantke K (2001) Iron and metal regulation in bacteria. *Current Opinion in Microbiology* **4**: 172-177.
Hook-Barnard IG & Hinton DM (2007) Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regulation and Systems Biology* **1**: 275-293.
Huang Q, Mao Z, Li S, Hu J & Zhu Y (2014) A non-radioactive method for small RNA detection by northern blotting. *Rice* **7**.
Imhoff JF (2005) Genus I. *Rhodobacter*. *Bergey's Manual of Systematic Bacteriology,* (Brenner DJ, Krieg NR, Staley JT & Garrity GM, eds.), p.161. Springer, New York.
Jahn CE, Charkowski AO & Willis DK (2008) Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *Journal of Microbiological Methods* **75**: 318-324.
James G, Witten D, Hastie T & Tibshirani R (2014) An Introduction to Statistical Learning: with Applications in R. Springer, New York.
Kingsford CL, Ayanbule K & Salzberg SL (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biology* **8**: R22.
Kroger C, Dillon SC, Cameron AD, *et al.* (2012) The transcriptional landscape and small RNAs of *Salmonella enterica* serovar *Typhimurium*. *Proceedings of the National Academy of Sciences USA* **109**: E1277-1286.
Landt SG, Abeliuk E, McGrath PT, Lesley JA, McAdams HH & Shapiro L (2008) Small non-coding RNAs in *Caulobacter crescentus*. *Molecular Microbiology* **68**: 600-614.
Lang AS & Beatty JT (2002) A Bacterial Signal Transduction System Controls Genetic Exchange and Motility. *Journal of Bacteriology* **184**: 913-918.
Lang AS, Zhaxybayeva O & Beatty JT (2012) Gene transfer agents: phage-like elements of genetic exchange. *Nature Reviews Microbiology* **10**: 472-482.
Lee EY, Choi DY, Kim DK, *et al.* (2009) Gram-positive bacteria produce membrane vesicles: proteomics-based characterization of *Staphylococcus aureus*-derived membrane vesicles. *Proteomics* **9**: 5425-5436.
Leung MM, Brimacombe CA & Beatty JT (2013) Transcriptional regulation of the *Rhodobacter capsulatus* response regulator CtrA. *Microbiology* **159**: 96-106.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R & Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
Li L, Huang D, Cheung MK, Nong W, Huang Q & Kwan HS (2013) BSRD: a repository for bacterial small regulatory RNA. *Nucleic Acids Research* **41**: D233-238.
Lin Y-f, A DR, Guan S, Mamanova L & McDowall KJ (2013) A combination of improved differential and global RNA-seq reveals pervasive transcription initiation and events in all stages of the life-cycle of functional RNAs in *Propionibacterium acnes*, a major contributor to wide-spread human disease. *BMC Genomics* **14**.
Loh E, Dussurget O, Gripenland J, Vaitkevicius K, Tiensuu T, Mandin P, Repoila F, Buchrieser C, Cossart P & Johansson J (2009) A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes. *Cell* **139**: 770-779.
McClure R, Tjaden B & Genco C (2014) Identification of sRNAs expressed by the human

pathogen *Neisseria gonorrhoeae* under disparate growth conditions. *Frontiers in Microbiology* **5**: 456.
Mentz A, Neshat A, Pfeifer-Sancar K, Puhler A, Ruckert C & Kalinowski J (2013) Comprehensive discovery and characterization of small RNAs in *Corynebacterium glutamicum* ATCC 13032. *BMC Genomics* **14**: 714-2164-2114-2714.
Mercer RG, Quinlan M, Rose AR, Noll S, Beatty JT & Lang AS (2012) Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. *FEMS Microbiology Letters* **331**: 53-62.
Mercer RG, Callister SJ, Lipton MS, Pasa-Tolic L, Strnad H, Paces V, Beatty JT & Lang AS (2010) Loss of the response regulator CtrA causes pleiotropic effects on gene expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *Journal of Baceriology* **192**: 2701-2710.
Michaux C, Verneuil N, Hartke A & Giard JC (2014) Physiological roles of small RNA molecules. *Microbiology* **160**: 1007-1019.
Miotto P, Forti F, Ambrosi A, Pellin D, Veiga DF, Balazsi G, Gennaro ML, Di Serio C, Ghisotti D & Cirillo DM (2012) Genome-wide discovery of small RNAs in *Mycobacterium tuberculosis*. *PloS One* **7**: e51950.
Moody MJ, Young RA, Jones SE & Elliot MA (2013) Comparative analysis of non-coding RNAs in the antibiotic-producing *Streptomyces* bacteria. *BMC Genomics* **14**: 558.
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**: 509-513.
Pena-Castillo L, Gruell M, Mulligan ME & Lang AS (2016) Detection of bacterial small transcripts from Rna-Seq data: a comparative assessment. *Pacific Symposium on Biocomputing* **21**: 456-467.
Pena-Castillo L, Mercer RG, Gurinovich A, Callister SJ, Wright AT, Westbye AB, Beatty JT & Lang AS (2014) Gene co-expression network analysis in *Rhodobacter capsulatus* and application to comparative expression analysis of *Rhodobacter sphaeroides*. *BMC Genomics* **15**: 730-2164-2115-2730.
Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.
Raina M & Ibba M (2014) tRNAs as regulators of biological processes. *Frontiers in Genetics* **5**: 171.
Rasmussen S, Nielsen HB & Jarmer H (2009) The transcriptionally active regions in the genome of *Bacillus subtilis*. *Molecular Microbiology* **73**: 1043-1057.
Rio DC (2014) Northern blots for small RNAs and microRNAs. *Cold Spring Harbor Protocols* **2014**: 793-797.
Robinson MD, McCarthy DJ & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.
Romeo T (1998) Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Molecular Microbiology* **29**: 1321-1330.
Schluter JP, Reinkensmeier J, Daschkey S, Evguenieva-Hackenberg E, Janssen S, Janicke S, Becker JD, Giegerich R & Becker A (2010) A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. *BMC Genomics* **11**: 245-2164-2111-2245.
Serganov A & Nudler E (2013) A decade of riboswitches. *Cell* **152**: 17-24.

Sharma CM & Vogel J (2014) Differential RNA-seq: the approach behind and the biological insight gained. *Current Opinion in Microbiology* **19**: 97-105.
Sharma CM, Hoffmann S, Darfeuille F, *et al.* (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**: 250-255.
Shinhara A, Matsui M, Hiraoka K, Nomura W, Hirano R, Nakahigashi K, Tomita M, Mori H & Kanai A (2011) Deep sequencing reveals as-yet-undiscovered small RNAs in *Escherichia coli*. *BMC Genomics* **12**: 428-2164-2112-2428.
Shiryev SA, Papadopoulos JS, Schaffer AA & Agarwala R (2007) Improved BLAST searches using longer words for protein seeding. *Bioinformatics* **23**: 2949-2951.
Sing T, Sander O, Beerenwinkel N & Lengauer T (2005) ROCR: visualizing classifier performance in R. *Bioinformatics* **21**: 3940-3941.
Solovjev V & Salamov A (2011) Automatic Annotation of Microbial Genomes and Metagenomic Sequences. *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies,* (Li RW, ed.) p. 61. Nova Science Publishers.
Soutourina OA, Monot M, Boudry P, *et al.* (2013) Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. *PLoS Genetics* **9**: e1003493.
Storz G, Vogel J & Wassarman KM (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Molecular Cell* **43**: 880-891.
Szklarczyk D, Franceschini A, Wyder S, *et al.* (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **43**: D447-452.
Thomason MK, Bischler T, Eisenbart SK, Forstner KU, Zhang A, Herbig A, Nieselt K, Sharma CM & Storz G (2015) Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *Journal of Bacteriology* **197**: 18-28.
Wang W, Wei Z, Lam TW & Wang J (2011) Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. *Scientific Reports* **1**: 55.
Weaver PF, Wall JD & Gest H (1975) Characterization of *Rhodopseudomonas capsulata*. *Archives of Microbiology* **105**: 207-216.
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**: e0165694.
Wheeler TJ & Eddy SR (2013) nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**: 2487-2489.
Wilms I, Overloper A, Nowrousian M, Sharma CM & Narberhaus F (2012) Deep sequencing uncovers numerous small RNAs on all four replicons of the plant pathogen *Agrobacterium tumefaciens*. *RNA Biology* **9**: 446-457.
Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R & Georg J (2013) Comparative genomics boosts target prediction for bacterial small RNAs. *Proceedings of the National Academy of Sciences USA* **110**: E3487-E3496.
Wurtzel O, Yoder-Himes DR, Han K, Dandekar AA, Edelheit S, Greenberg EP, Sorek R & Lory S (2012) The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. *PLoS Pathogens* **8**: e1002945.
Zeng Q & Sundin GW (2014) Genome-wide identification of Hfq-regulated small RNAs in the fire blight pathogen *Erwinia amylovora* discovered small RNAs with virulence regulatory function. *BMC Genomics* **15**: 414-2164-2115-2414.

Chapter 3 - Operon complexity in *Rhodobacter capsulatus* revealed by integration of total and end-targeted RNA-seq data

Abstract

RNA sequencing (RNA-seq) has been used as a powerful tool to examine the genome-wide expression of genes in bacteria for many years. In recent years new protocols have been developed that enable us to describe the full transcriptome of prokaryotic organisms. Here, we present an analysis of the transcriptome of two strains of the alpha-proteobacterium *Rhodobacter capsulatus* by strand-specific, single-nucleotide resolution, RNA-seq. We modified the protocol of the previously developed differential RNA-seq (dRNA-seq) method, which can distinguish between primary and processed transcripts, to use it on an Ion Torrent PGM sequencing machine. This allowed the identification of transcription start sites (TSS) and, by making alterations to the protocol, we were able to use dRNA-seq for 3' end-targeted sequencing to predict transcription termination sites (TTS). We integrated these dRNA-seq data with total RNA-seq data to predict transcriptional units and to analyze operon complexity in *R. capsulatus*. Our analyses revealed a complex operon architecture, where some operons have multiple TTS and TTS, genomic regions of high transcriptional activity, and novel transcripts.

3.1 Introduction

As a purple non-sulfur alpha-proteobacterium, *Rhodobacter capsulatus* is metabolically versatile and many aspects of its physiology, such as phototrophy and hydrogen production, are actively studied. *R. capsulatus* is of interest because it is one of the organisms known to produce

bacteriophage-like particles called gene transfer agents (GTAs), which package small pieces of the cell's genomic DNA that can be transferred to other cells in the population. Like the photosynthesis genes, the GTA genes are under complex regulation in this bacterium.

Advances in high-throughput sequencing technologies enabled the development of RNA sequencing (RNA-seq) and its application to transcriptomic studies (Hor *et al.*, 2018). RNA-seq has been used successfully in recent years to define the transcriptomes of many bacteria such as *Escherichia coli* (Ettwiller *et al.*, 2016), *Helicobacter pylori* (Sharma *et al*., 2010, Bischler *et al*., 2014), *Synechocystis* sp. PCC6803 (Mitschke *et al.*, 2011), and *Campylobacter jejuni* (Dugar *et al.*, 2013). Advantages of RNA-seq over microarrays, the previous predominant transcriptomic tool, are the possibilities for obtaining higher-resolution data with respect to transcript boundaries and the discovery of unexpected transcripts. In terms of resolving transcript boundaries and processing sites, different RNA-seq methods have been developed to identify 5' ends. These include strand-specific cDNA sequencing (ssRNA-seq) (Perkins *et al.*, 2009), Cappable-seq (Ettwiller *et al.*, 2016), and differential RNA-seq (dRNA-seq) (Borries *et al.*, 2012). dRNA-seq takes advantage of the presence of a triphosphate at the 5' end of unprocessed RNAs and uses enzymatic treatment to degrade processed RNAs, followed by sequencing of the remaining RNAs for the detection of transcription start sites. Based on the high throughput and increased read length of the latest generation of sequencing machines it is possible to identify the 3' end of a given RNA, but no distinction can be made between original and processed transcripts. A previous transcriptomic study in *R. capsulatus* based on microarray data resulted in the identification of 40 gene co-expression modules that lead to new information for functional annotation in *R. capsulatus* (Mercer *et al.*, 2010, Pena-Castillo *et al.*, 2014). A recent study using an RNA-seq approach investigated the presence of small RNAs (sRNAs) genome-wide in *R.*

capsulatus, resulting in the identification of a total of 422 putative sRNAs (Grüll *et al.*, 2017). Despite these previous studies and what was learned therein, more whole-transcriptome investigations and analyses are necessary to get a better understanding of the regulation of gene expression and cellular processes in *R. capsulatus*.

We combined the results from 5’ and 3’ end-targeted sequencing with data from total RNA-seq to identify putative transcriptional units, including their start sites (TSS, 5’ ends) and termination sites (TTS, 3’ ends). This involved development of customized dRNA-seq protocols for 5’ and 3’ end-targeted sequencing using the Ion Torrent Personal Genome Machine. The cellular RNA pool in bacteria contains both primary transcripts that carry a 5’-triphosphate (5’-PPP) as well as processed RNAs that can either have a 5’-monophosphate (5’-P) or a 5’-hydroxyl (5’-OH) group. For 5’ targeted sequencing, only transcripts that have an intact 5’-triphosphate end are of interest. Therefore, RNA 5’ ends that originate from processing of a longer transcript need to be depleted. For 3’ targeted sequencing, only original transcripts are of interest as well, and we developed a new workflow for characterizing these ends by RNA-seq. Our analyses were performed with the standard model strain, SB1003, along with a GTA-overproducing strain, DE442.

3.2 Materials and Methods

3.2.1 *R. capsulatus* growth and RNA isolation

R. capsulatus strains SB1003 and DE442 were grown under anaerobic phototrophic conditions at 35°C in complex YPS medium (Weaver *et al.*, 1975) until four hours after reaching stationary phase. Cultures were mixed 5:1 with a mixture of 95% ethanol and 5% saturated

phenol (Jahn *et al.*, 2008), cells were pelleted by centrifugation and pellets were frozen on dry ice/ethanol and stored at -80ºC. Total RNA isolation was performed with the DNA, RNA, and protein purification kit (MACHEREY-NAGEL) following the manufacturer's protocol for purification of total RNA.

3.2.2 Library preparation for differential RNA sequencing (dRNA-seq)

To detect transcription initiation and termination sites in the *R. capsulatus* genome, dRNA-seq was performed to sequence 5' and 3' RNA ends using an Ion Torrent Personal Genome Machine (PGM; Thermo-Fisher). dRNA-seq was originally developed for 5' end-targeted sequencing using the Illumina sequencing platform (Borries *et al.*, 2012, Sharma & Vogel, 2014) and therefore had to be modified for the Ion Torrent platform, as described below. To perform 5' targeted dRNA-seq, the total cellular RNA pool was treated with 7 U of Terminator 5´-Phosphate-Dependent Exonuclease (TEX) (1 U/µl, Epicentre). A TEX- sample that did not receive the TEX treatment was used as a negative control. After the TEX treatment, the TEX+ samples were treated with 2.5 U of RNA 5' Pyrophosphohydrolase (RppH) (5000 U/ml, NEB) to convert the original 5'-triphosphate (5'-PPP) to 5'-monophosphate (5'-P) for Ion Torrent adaptor ligation. Both TEX+ and TEX- samples then underwent a 5' adaptor ligation step, followed by a fragmentation step to reduce the size of RNAs to Ion Torrent sequencing specifications. 3' adaptors were subsequently ligated. Reverse transcription and the remaining library preparation steps were performed according to the manufacturer's recommendations for total RNA sequencing using the RNA-seq Kit v2 (Thermo Fisher). Libraries were amplified using an Ion Torrent One Touch 2 system (OT2, Life Technologies), loaded individually on 316

v2 chips (Life Technologies) and sequenced with the number of flows set to 550 on an Ion Torrent Personal Genome Machine (PGM; Thermo Fisher) (Figure 3.1).

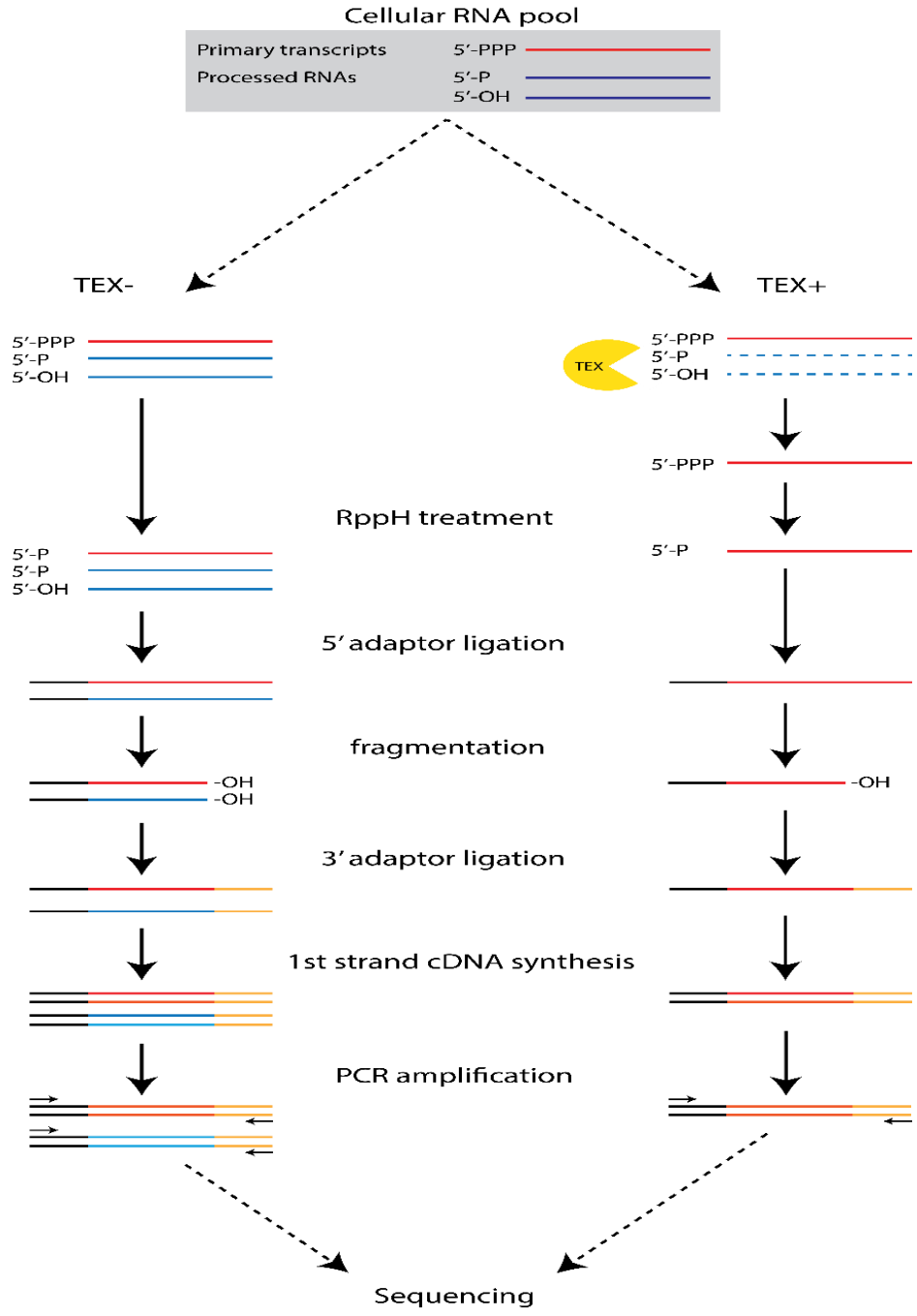

Figure 3.1: Workflow for a 5' targeted sequencing experiment. For dRNA-seq library construction the cellular RNA pool is split into two parts. The pool that is not treated with TEX functions as a negative control. The positive pool gets treated with TEX to deplete processed RNAs. Both samples are then treated with RppH to convert 5'-PPP to 5'-P. Subsequently, 5' and 3' sequencing adaptors are ligated to the RNAs to construct the sequencing libraries.

To perform 3’ targeted dRNA-seq, the initial TEX treatment was still performed to deplete processed transcripts. After the TEX treatment the 3’ adaptors were ligated to the TEX+ and TEX- RNA samples. Adaptor ligation was followed by a fragmentation step. Because the 5’ ends of fragmented RNAs have a 5’-P, no RppH treatment was necessary, and 5’ adaptors were directly ligated. Reverse transcription and the remaining library preparation steps were performed according to the manufacturers’ recommendations for total RNA sequencing using the RNA-seq Kit v2 (Thermo Fisher) (Figure 3.2).

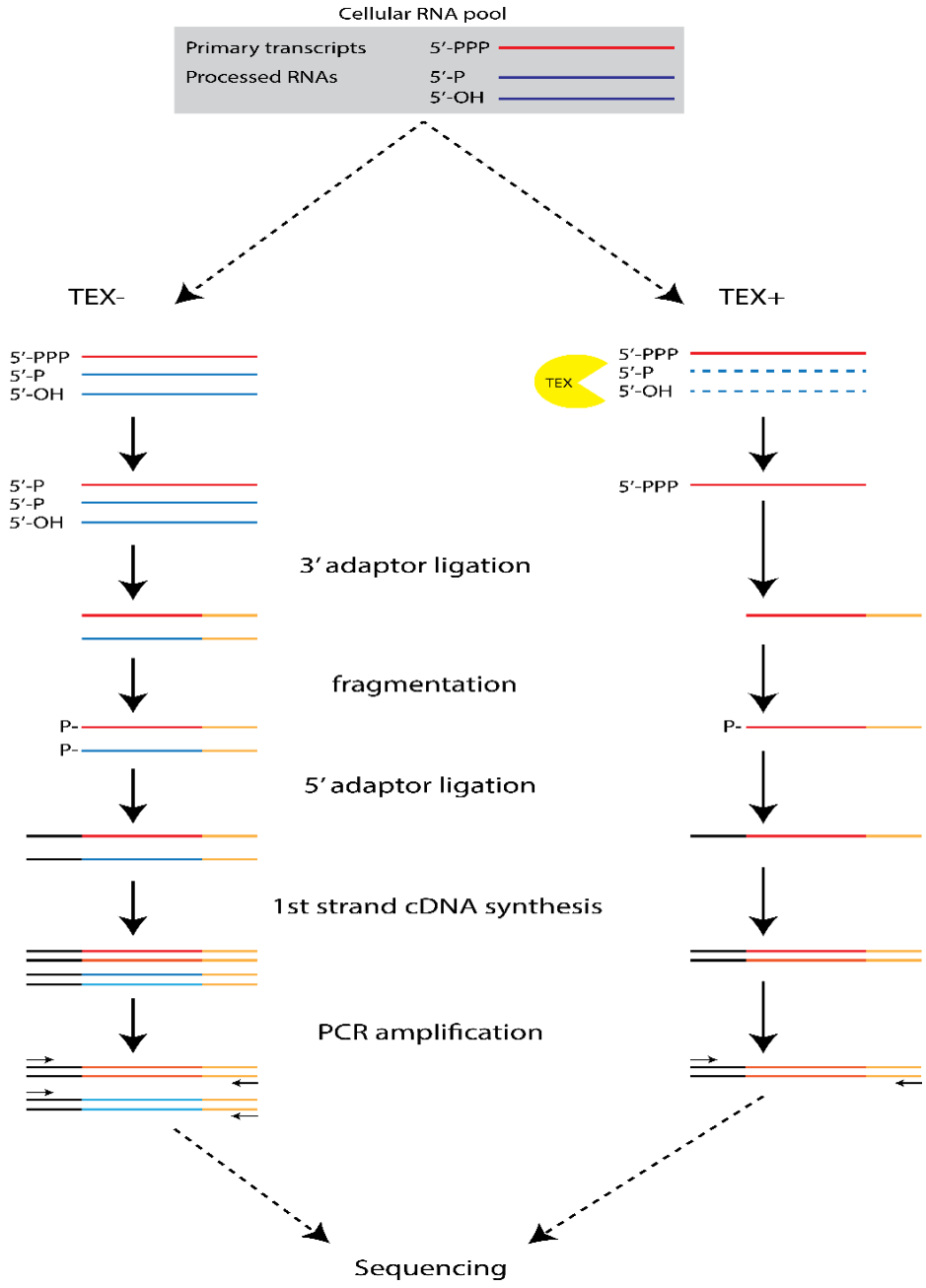

Figure 3.2: Workflow for a 3' targeted sequencing experiment. The cellular RNA pool is split into two parts. The pool that is not treated with TEX functions as a negative control. The positive pool gets treated with TEX to deplete processed RNAs. 3’ adaptors are ligated to the RNAs, followed by fragmentation. Fragmented RNA 5’ ends have 5’-P, therefore no RppH treatment is necessary. Subsequently, 5’ adaptors are ligated to the RNAs and the sequencing library is constructed.

3.2.3 Processing of RNA-seq data

The RNA-seq data quality was verified using the FastQC tool (version 0.10.0) and reads were filtered and trimmed using the fastq_quality_trimmer available in FASTX Toolkit (version 0.0.13.2) with a quality threshold of 22 and minimum read length of 28 nucleotides. Filtered and trimmed reads were mapped to the *R. capsulatus* genome using the Torrent mapper tmap (version 3.0.1), executed with the parameters: -B 18 -a 2 -v stage1 map1 map2 map3. Mapping statistics were obtained using samtools (Li *et al.*, 2009).

3.2.4 Operon detection pipeline

Our Python (version 3.5) operon detection pipeline to determine bacterial operon structure requires as input the candidate positions for transcriptional start sites (TSS), candidate positions for transcriptional termination sites (TTS), and a single-nucleotide resolution whole-genome read-depth coverage vector. The candidate positions for the TSS are used as seeds to search for putative transcriptional units. To obtain candidate TSS and TTS, we used the program TSSAR (Amman *et al.*, 2014) with the dRNA-seq data. We set TSSAR parameters as follows: p-value to 0.1, noise threshold to 2, and maximum range to merge to 1. As TSSAR is designed to detect the 5’ end location of sequencing reads, we reversed the strand of the reads of the 3’ end data prior to giving it as input to TSSAR. Once we received the analyzed data from TSSAR, we then reversed back the strand of the locations detected to be able to use the data to map the 3’ ends. From the total RNA-seq data, we obtained a whole-genome read-depth coverage vector with a single nucleotide resolution using the GenomicArray collection available in HTSeq (Anders *et al.*, 2015).

As proposed in a previous publication (Conway *et al.*, 2014), we used three features to define a transcriptional unit: the TSS (5’ end), the TTS (3’ end), and the read-depth coverage between both ends. We defined a set of criteria to deem a detected transcriptional unit as *bona fide*. The set of criteria includes the level of promoter efficiency, the terminator efficiency, the mean read-depth coverage and the percentage of sequence between the TSS and the TTS with read coverage. Promoter efficiency is calculated as the log2 of the ratio of the average read count of the 10 bases upstream of the 5’ end and the average read count of the 10 bases downstream of the 5’ end (Creecy & Conway, 2015). We required at least a 20% increase in read counts to deem a promoter as active. Likewise, terminator efficiency is calculated as the log2 of the ratio of the average read count of the 25 bases upstream of the 3’ end and the average read count of the 25 bases downstream of the 3’ end (Creecy & Conway, 2015). We required at least a 15% decrease in read counts to deem a terminator as active. 70% of the length of a detected transcriptional unit must be covered by reads to be considered as likely real. The value of these parameters was set based on inspection of the sequencing data for annotated transcripts. A nucleotide is considered to be covered if this nucleotide is contained in at least three reads. Detected transcriptional units with a length of less than 20 nts were discarded. If a 3’ end was not detected near a transcriptional unit, the TTS of this transcriptional unit was considered to be at the start of an at least 10 nts long segment with a read coverage of less than 3. Overlapping transcriptional units transcribed on the same strand were considered to form an operon. As output, our pipeline generates a table with the transcriptional units detected, a file with the genomic location of the predicted TSS, and a file with the genomic location of the predicted TTS.

To find the closest 5' end of an annotated gene or putative sRNA to each predicted TSS, and the closest 3' end of an annotated gene or putative sRNA to each predicted TTS, first we

obtained the location of the 5' and 3' ends of annotated genes and putative sRNAs using the BEDTools (version 2.27) flank command (Quinlan & Hall, 2010) and then used the BEDTools closestBed command with parameters -s -D "b" -k 1 -t "first" given as file -a the BED file with the TSS (TTS) and as file -b the BED file with the 5' (3') ends. Further analyses were done in R (version 3.3.2).

3.2.5 Prediction of conserved promoter motifs

To identify conserved promoter sequence motifs, we extracted the genomic sequence 40 nts upstream of a predicted TSS. Sequences on the same strand within 10 nts of each other were merged using the merge command available in BEDtools. We selected 40 nts upstream of a predicted TSS (as done by (Wittchen *et al.*, 2018)). The merged sequences were used as input for the GLAM2 software (version 4.11.1) (Frith *et al.*, 2008). We ran GLAM2 with the parameters -z 2 -a 2 -b 50 -w 6 -r 5 -n 10000 -D 0.1 -E 2.0 -I 0.02 -J 1.0.

3.2.6 5’ Rapid Amplification of cDNA Ends (RACE)

5’ RACE experiments were performed with the 5’ RACE System (Invitrogen). Total RNA was isolated from the *R. capsulatus* RcGTA overproducer strain DE442 with the NucleoSpin® miRNA kit (Macherey-Nagel). All total RNA samples for use in 5’ RACE experiments were prepared in the same way as samples that were used in dRNA-seq experiments. Total RNA was then reverse transcribed using a gene-specific primer. RACE experiments followed the manufacturer’s protocol for 5’ RACE using custom gene-specific primers (Table 3.1).

Table 3.1: Primers used in 5’ RACE experiments

Primer name	Primer sequence (5’ - 3’)
250 GSP1	TAATTGCGGAACAGA
250 GSP2	GGACGACACATGGCCGAAGCTCAG
250 GSP3	CTTCAGATGCGAGGTGACGGTGATC
493 GSP1	GTTGTGGTAATGCAG
493 GSP2	GCGAGACGACGGATGAGATCGATCT
493 GSP3	GGCGACAAGGTTCTGGATCACCATT
1682 GSP1	CGAAGTTTTTCAACAC
1682 GSP2	CACCCTTTCTTCCATCACCATCTGGAA
1682 GSP3	GAACCCCTTTCATCGCCAGGGCCAGTT
1874 GSP1	GGAATAGATCTGCCC
1874 GSP2	ATCTGCCCGACCGAGCGCGATTTC
1874 GSP3	TAGCCCGACAGCAGCATATAGGGTTC

Primers for 5’ RACE experiments were designed to confirm predicted TSS located at the 5’ end of transcripts.

3.3 Results and discussion

3.3.1 Integrative analysis reveals complex operon structure in *R. capsulatus*

Total RNA was extracted and used for different versions of RNA-seq on the Ion Torrent platform: total RNA-seq, dRNA-seq targeting 5’ ends, and dRNA-seq targeting 3’ ends (described in Methods). A recent study investigated the potential of the Ion Torrent instrument

for transcriptomic analysis in comparison to the Illumina platform and concluded that the platforms are equally suited for transcriptional analyses (Lahens *et al.*, 2017). Duplicate sequencing runs were performed for each strain and RNA-seq method, generating the numbers of reads indicated in Table 3.2.

Table 3.2: Number of reads generated for each strain. 85.65 % of all 5' and 3' reads were uniquely mapped to the *R. capsulatus* chromosome (NCBI accession number NC_014034.1). 95.7% of all total RNA reads mapped uniquely to the *R. capsulatus* chromosome.

	SB1003	**DE442**
5' TEX+	5.84†	4.73†
5' TEX–	4.68†	4.71†
3' TEX+	4.08†	3.84†
3' TEX–	4.66†	4.08†
Total RNA	2.0†	2.8†

†Data values represent millions of reads.

All sequence data have been submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE117537.

We developed a Python-based pipeline (described in Methods) to integrate the 5' and 3' dRNA-seq data and total RNA-seq data for each strain. This Operon Detection Pipeline (ODP) identified putative TSS, TTS, and transcriptional units. Putative transcriptional units have (i) a TSS and a TTS with an efficiency above the minimum efficiency score, and (ii) a minimum read depth coverage between the TSS and the TTS (see Methods).

Overlapping transcriptional units were grouped in operons. Our ODP identified 4045 TSS, 2906 TTS and 2141 operons for SB1003, and 5374 TSS, 3279 TTS and 2286 operons for

DE442. Putative transcriptional units were described in terms of their length, percentage of length covered by reads, promoter efficiency, terminator efficiency, and average read count.

We classified the operons detected in four complexity classes: those with a single TSS and a single TTS, those with multiple TSS and multiple TTS, those with a single TSS and multiple TTS, and those with multiple TSS and a single TTS. Table 3.3 shows the number of operons per class for each strain.

Table 3.3: Number of operons per complexity class for each strain.

Operon class	**SB1003**	**DE442**
Single TSS, single TTS	1059 (49.46%)	1162 (50.83%)
Multiple TSS, multiple TTS	562 (26.25%)	671 (29.35%)
Single TSS, multiple TTS	518 (24.19%)	452 (19.77%)
Multiple TSS, single TTS	2 (0.09%)	1 (0.04%)
Total	2141 (100%)	2286 (100%)

Simple operons with a single TSS and a single TTS constitute approximately half of the operons for each strain. We hypothesized that more complex operons were likely to be longer and more highly expressed than those with fewer TSS and/or TTS, and indeed both the expression level (as reflected in read coverage) and the length of the operon correlated positively with the number of TSS and TTS detected (Figures 3.3 and 3.4).

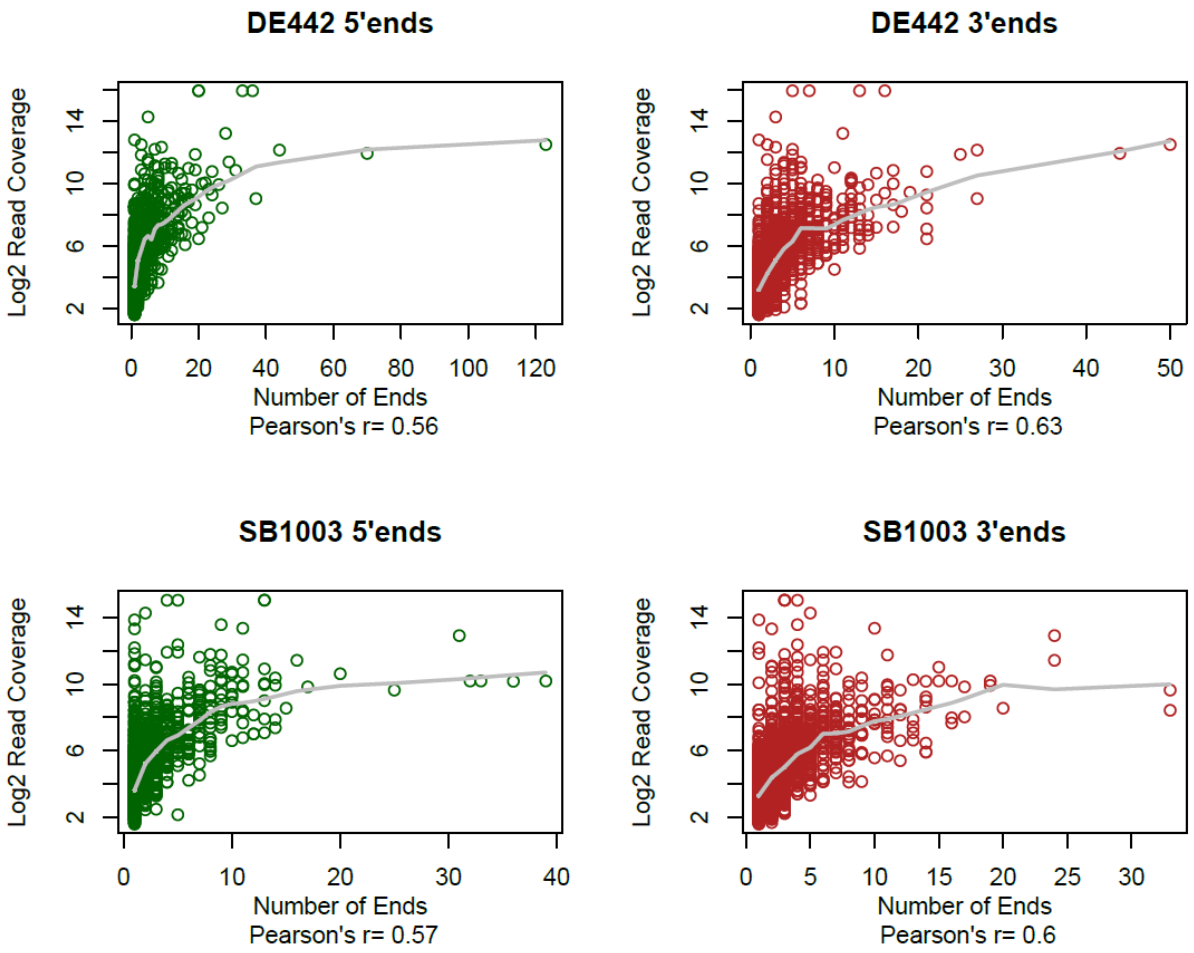

Figure 3.3: Operon read-depth coverage as a function of the number of TSS (TTS) predicted. Each scatterplot shows the log2 read-depth coverage of each operon as a function of the number of TSS predicted (green) and as a function of the number of TTS predicted (red) for that operon. Data is shown for each strain. Below each plot the Pearson correlation coefficient between the log2 read-depth coverage and the number of TSS (TTS) is provided. The grey line is the Locally Weighted Scatterplot Smoothing (LOWESS) line with a smoother span of 0.05.

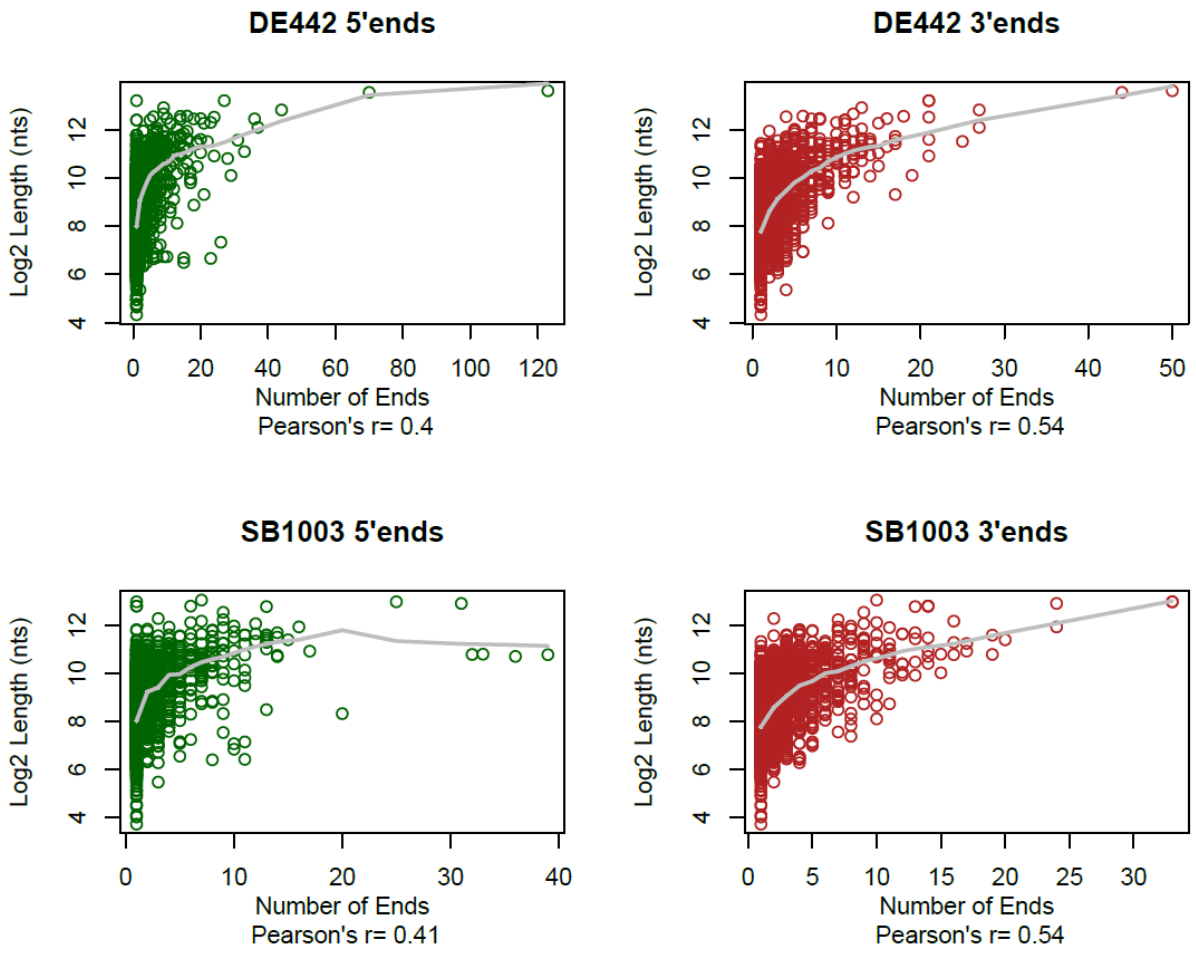

Figure 3.4: Operon length as a function of the number of ends predicted. Each scatterplot shows the log2 length in nucleotides (nts) of each operon as a function of the number of TSS predicted (green) and as a function of the number of TTS predicted (red) for that operon. Data is shown for each strain. Below each plot the Pearson correlation coefficient between the log2 operon length and the number of TSS (TTS) is provided. The grey line is the Locally Weighted Scatterplot Smoothing (LOWESS) line with a smoother span of 0.05.

One example of a complex operon is operon r5, which contains genes encoding flagellar proteins (RCAP_rcc03514-RCAP_rcc03520). We detected high expression levels of this genomic region in both strains. High expression of motility genes during anaerobic growth in

rich medium has previously been reported by Schindel *et al*. (Schindel & Bauer, 2016). We detected one predicted TSS at the start of this operon and a total of 3 additional TSS in SB1003 and 8 additional TSS in DE442. We detected a total of 6 TTS in SB1003 and 4 TTS in DE442. Visual inspection of the organization of the predicted TSS and TTS reveals that their locations are spaced out in a way that would allow part of the operon to be transcribed from an internally located TSS after termination at an upstream TTS.

Another example of a complex operon includes genes RCAP_rcc01822-RCAP_rcc01834. This genomic region is highly expressed in both sequenced strains. We detected a total of 15 TSS and 18 TTS in SB1003 and 29 TSS and 23 TTS in DE442. The presence of this unexpectedly high number of predicted TSS and TTS sites could be attributed to pervasive transcription. Pervasive transcription refers to the idea that transcription is not restricted by the position of annotated protein-coding genes but can initiate in almost any genomic context (Wade & Grainger, 2014). Such pervasive transcripts rarely have an assigned function and often occur within genes, resulting in non-coding RNAs (Kapranov *et al.*, 2007). However, it has been proposed that pervasive transcripts could be more than transcriptional noise, with important functions in gene regulation and genome evolution (Wade & Grainger, 2014).

In addition to highly complex operons we looked at operons with >30 predicted TSS or TTS and found that they correspond to genomic regions of high transcriptional activity (Table 3.4).

Table 3.4. Identified genomic regions with high transcriptional activity in *R. capsulatus*.

Start	End	Strand	Strain	Genes in region
159,978	161,744	+	SB1003	RCAP_rcc00001
219,691	222,710	-	DE442	RCAP_rcc00184 - RCAP_rcc00185
722,065	729,261	-	DE442	RCAP_rcc00656 - RCAP_rcc00663
755,175	767,657	+	DE442, SB1003	RCAP_rcc00684 - RCAP_rcc00697
1,003,719	1,005,510	+	SB1003	RCAP_rcr00004
1,003,894	1,009,459	+	DE442	RCAP_rcr00004 - RCAP_rct00014
1,110,924	1,115,279	-	DE442	RCAP_rcc01034 - 01036
1,513,758	1,515,535	+	SB1003	RCAP_rcr00007
1,969,972	1,981,866	-	DE442	RCAP_rcc01822 - RCAP_rcc01834
3,580,302	3,582,473	-	DE442, SB1003	RCAP_rct00050 - RCAP_rcr00012
3,627,685	3,635,871	+	SB1003	RCAP_rcc03426 - RCAP_rcc03433

3.3.2 Features of identified TSS and TTS

To compare predicted TSS and TTS with the current genome annotation of *R. capsulatus* (NCBI genome build accession GCF_000021865.1) and with a set of putative sRNAs (Grüll *et al.*, 2017), we calculated the distance from each predicted TSS/TTS to the closest 5'/3' end, respectively, of an annotated gene or putative sRNA transcribed on the same strand. Figure 3.5 shows the density distribution of the distances within ±500 nts of TSS and TTS from the 5' and 3’ ends, respectively, of genes and sRNAs.

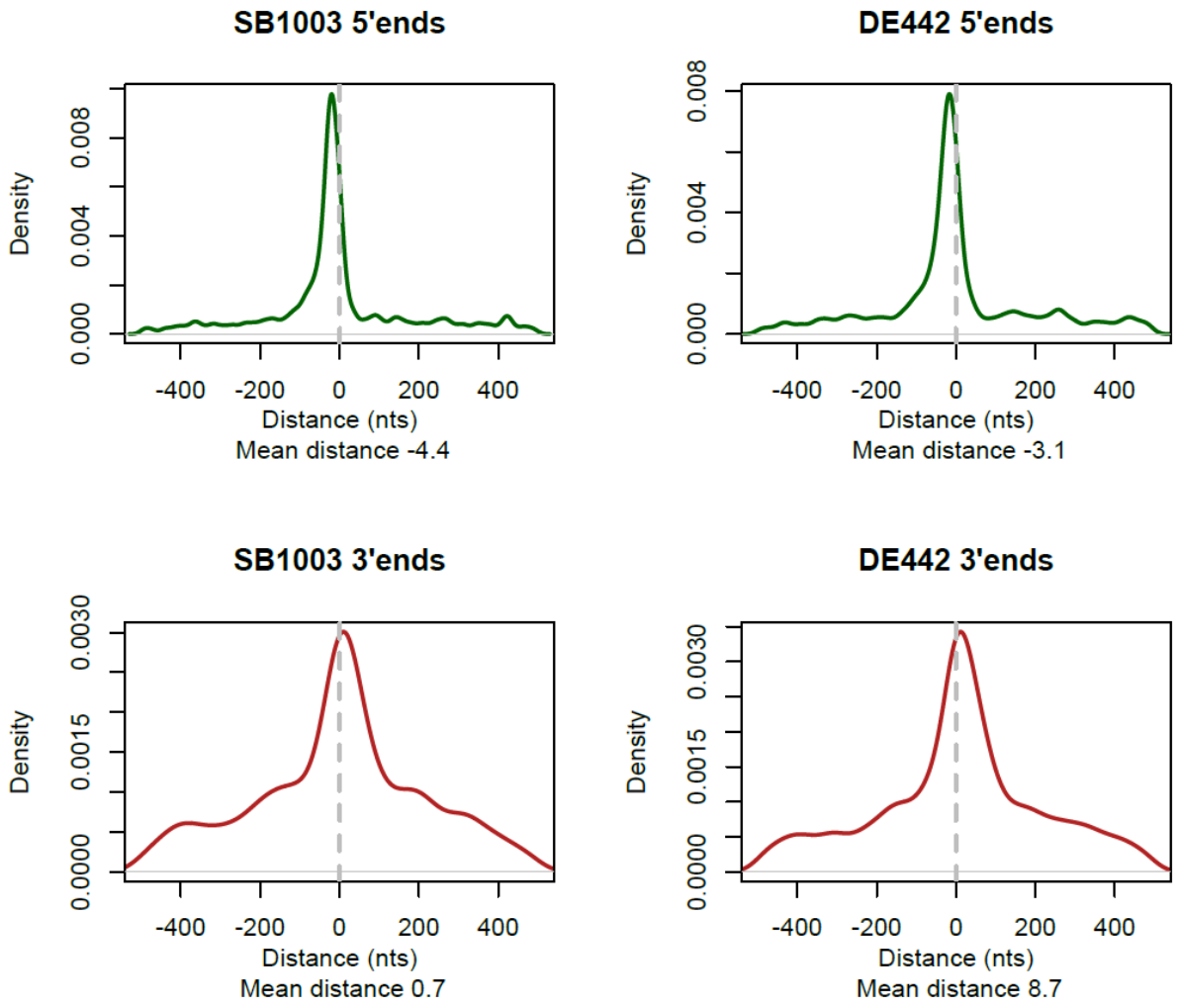

Figure 3.5: Density distribution of the distance in nucleotides of predicted TSS (green) and predicted TTS (red) to the closest 5' and 3' end, respectively, of an annotated gene or putative sRNA. The vertical grey line indicates the location of the 5' (or 3') ends. The mean distance is provided below each plot. A negative distance indicates that the TSS (or TTS) is upstream of the closest 5' (or 3') end of an annotated gene or putative sRNA.

There is a clear peak in the TSS and TTS distance distributions close to the annotated 5' and 3’ ends, respectively, of the genes or sRNAs. TSS with an absolute distance of more than 500 nts from the closest 5' end of an annotated gene or putative sRNA were statistically significantly less efficient than those within 500 nts of the closest 5' end (Mann-Whitney test; p = 7.5e-34 and 8.9e-45 for SB1003 and DE442, respectively). Similarly, TTS with an absolute distance of more than 500 nts from the closest 3' end of an annotated gene or putative sRNA were significantly less efficient than those within 500 nts of the closest 3' end (Mann-Whitney test; p = 4.3e-6 and 9.7e-10 for SB1003 and DE442, respectively). The vast majority of TSS and TTS were within 500 nts of the 5' and 3' end, respectively, of an annotated gene or putative sRNA (Table 3.5).

Table 3.5: Number of TSS and TTS in various genomic contexts relative to annotated genes or detected sRNAs for each strain.

Genomic context	**SB1003**		**DE442**	
	TSS	**TTS**	**TSS**	**TTS**
[-500, 500]	3310 (81.83%)	2221 (76.43%)	4379 (81.48%)	2570 (78.38%)
absolute (distance) < 500 and intragenic (internal)	672 (16.61%)	649 (22.33%)	917 (17.06%)	658 (20.07%)
absolute (distance) > 500 and intergenic	63 (1.56%)	36 (1.24%)	78 (1.45%)	51 (1.56%)
Total	4045 (100%)	2906 (100%)	5374 (100%)	3279 (100%)

We investigated the relationship between the TSS/TTS efficiency and their absolute distance to the closest 5' or 3' end, respectively, of an annotated gene or putative sRNA. There

are negative correlations between TSS and TTS efficiency (Pearson's correlation coefficients of -0.3 and -0.2, respectively) and absolute distance to the closest 5' or 3’ end of an annotated gene or putative sRNA, respectively (Figure 3.6).

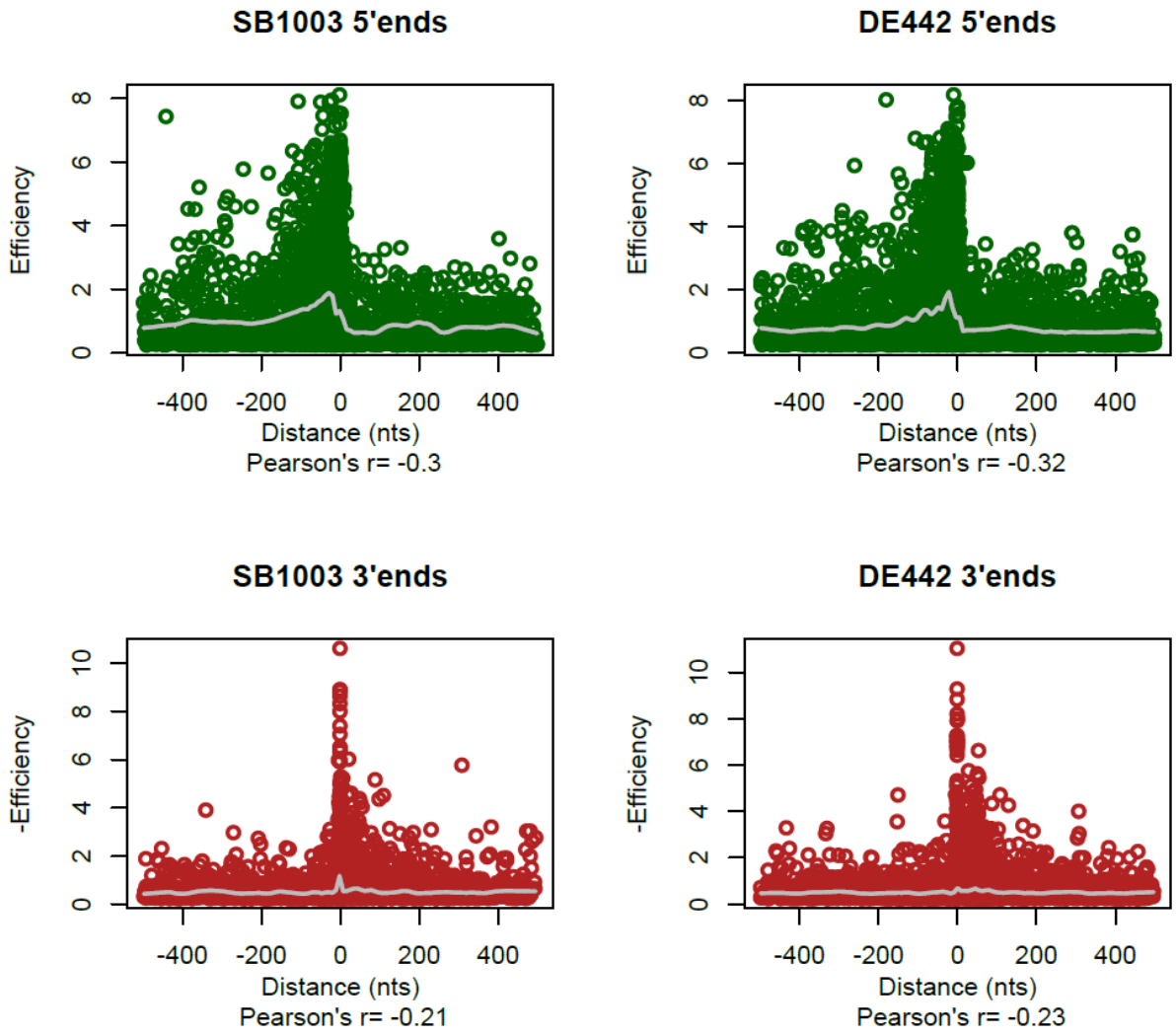

Figure 3.6: Efficiency of predictions as a function of distance to the nearest TSS or TTS. Top: Efficiency of predicted TSS as a function of distance to the closest 5' end of an annotated gene or putative sRNA for each strain. Bottom: Minus efficiency of predicted TTS as a function of their distance to the closest 3' end of an annotated gene or putative sRNA for each strain. Efficiency is the log2 of the ratio of the number of reads upstream of the TSS (or TTS) and the number of reads downstream of the TSS (or TTS). The grey lines are the Locally Weighted Scatterplot Smoothing (LOWESS) lines with a smoother span of 0.05. The Pearson correlation coefficients between the efficiency (or – efficiency for TTS) and the absolute distance are provided below each plot.

TSS upstream of the closest 5' end of an annotated gene or putative sRNA are significantly more efficient than those located downstream of the closest 5' end of an annotated gene or putative sRNA (Mann-Whitney test; p = 1.90e-71 and 2.33e-81 for SB1003 and DE442, respectively). Fitting a local linear polynomial to the distance versus efficiency data points (Figure 3.6) allowed visualization of an accumulation in the TSS efficiency upstream of the closest 5' end of an annotated gene or putative sRNA with a sudden drop in efficiency at distance zero. We did not observe this phenomenon for the TTS (Figure 3.6), but TTS downstream of the closest 3' end of annotated gene or putative sRNA were significantly more efficient than those located upstream of the closest 3' end (Mann-Whitney test; p = 0.0009 and 0.0003 for SB1003 and DE442, respectively).

Out of 415 putative sRNAs encoded on the *R. capsulatus* chromosome (Grüll *et al.*, 2017), 307 (74%) have either a TSS or a TTS within 50 nts of their 5' or 3' end, respectively, in at least one of the strains sequenced, 124 (30%) have both a TSS and a TTS within 50 nts of their 5' and 3' end, respectively, in at least one of the strains sequenced, and 50 (12%) have both a TSS and a TTS within 50 nts of their 5' and 3' end, respectively, in both strains. We explored whether

there were differences between TSS and TTS associated with the putative sRNAs and those associated with annotated genes, which showed that sRNAs have fewer internal TSS and TTS than annotated genes (Supplementary Figures S4, S5).

3.3.3 Novel transcripts

As there were intergenic TSS and TTS not associated with any annotated gene or putative sRNA, we wondered whether we could confidently identify novel transcripts in *R. capsulatus*. To do that, we selected transcriptional units flanked by intergenic TSS and TTS with a minimum read depth coverage of 10 reads in both SB1003 and DE442 and with <10% sequence overlap with annotated genes or putative sRNAs. We found four transcriptional units meeting these criteria (Figure 3.7).

Figure 3.7: Example of an identified novel transcript. The first four rows on the left show the total RNA coverage, predicted 5’ and 3’ ends and the total RNA read coverage for the wild-type strain SB1003. Rows five to eight show the same data for the DE442 mutant strain. The last three rows show the predicted novel transcript for SB1003 and DE442, respectively, with the neighboring annotated genes.

All four transcripts were investigated using the basic local alignment search tool (BLAST) (Altschul *et al.*, 1990) to find conservation in other bacterial species. Based on these results, transcript T933787 is conserved in the chromosomes of at least four other bacterial species, transcript T1025900 showed conservation in the chromosomes of at least nine other bacterial species, transcript T1805883 was found to be conserved in the chromosomes of at least nine other bacterial species and transcript T2265248 was found to be conserved in the chromosomes of more than 20 bacterial species. None of the search results revealed possible functions for the predicted transcripts. All predicted transcripts were then used as input for a search using Rfam (Kalvari *et al.*, 2018, Kalvari *et al.*, 2018), which returned no matches to any known transcripts. Using the HMMER software (Finn *et al.*, 2011) to search for protein sequence similarity resulted in no known matches for any of the predicted transcripts. The reason for the lack of known RNAs or protein sequences could be that no study has previously identified or characterized these intergenic RNAs. Further experimentation would be required to begin to identify possible functions for these transcripts.

3.3.4 Genome-wide prediction of conserved promoter motifs

For each identified TSS, we extracted the sequence corresponding to 40 nts upstream of the TSS. We merged overlapping sequences to obtain a total of 3172 sequences for SB1003 and 3877 sequences for DE442. Using the GMAL2 software (Frith *et al.*, 2008) on the upstream sequences of each strain, we identified a single strong motif (Figure 3.8).

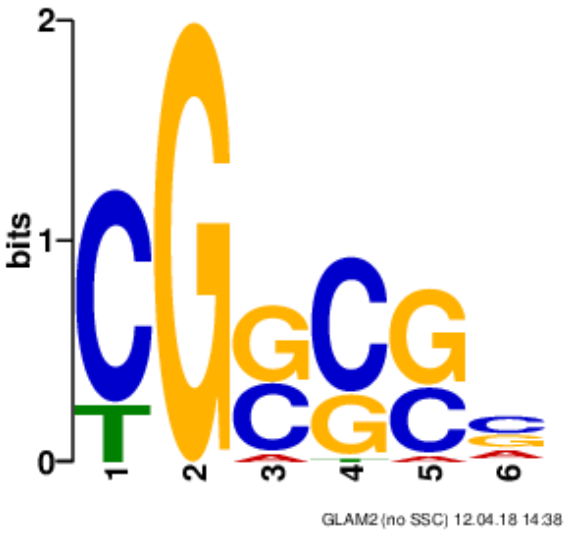

Figure 3.8: Promoter DNA sequence motif identified. The frequency of bases found at each position is indicated by the size of the colored letters, created with GLAM2 (Frith *et al.*, 2008).

3.3.5 Experimental validation of predicted TSS using 5' RACE

We chose four predicted locations for validation of TSS using 5' RACE. These were RCAP_rcc00250, RCAP_rcc00493, RCAP_rcc01682 and RCAP_rcc01874. We purposefully chose these targets because of their high expression in the DE442 GTA-overproducer strain. Promoter predictions were made based on the motifs identified in this study as well as motifs that have previously been identified (Leung *et al.*, 2013).

RCAP_rcc00250 was predicted to have several TSS. We identified a TSS using RACE at position 286,242 with a potential promoter starting at position 286,283. The closest predicted RNA-seq-based TSS was at position 286,215, 27 bp upstream of the site identified with RACE, with a predicted promoter at position 286,289 (Figure 3.9 A). Based on the sequencing coverage data the 5' end of this gene was expected to be located closer to the start of the gene.

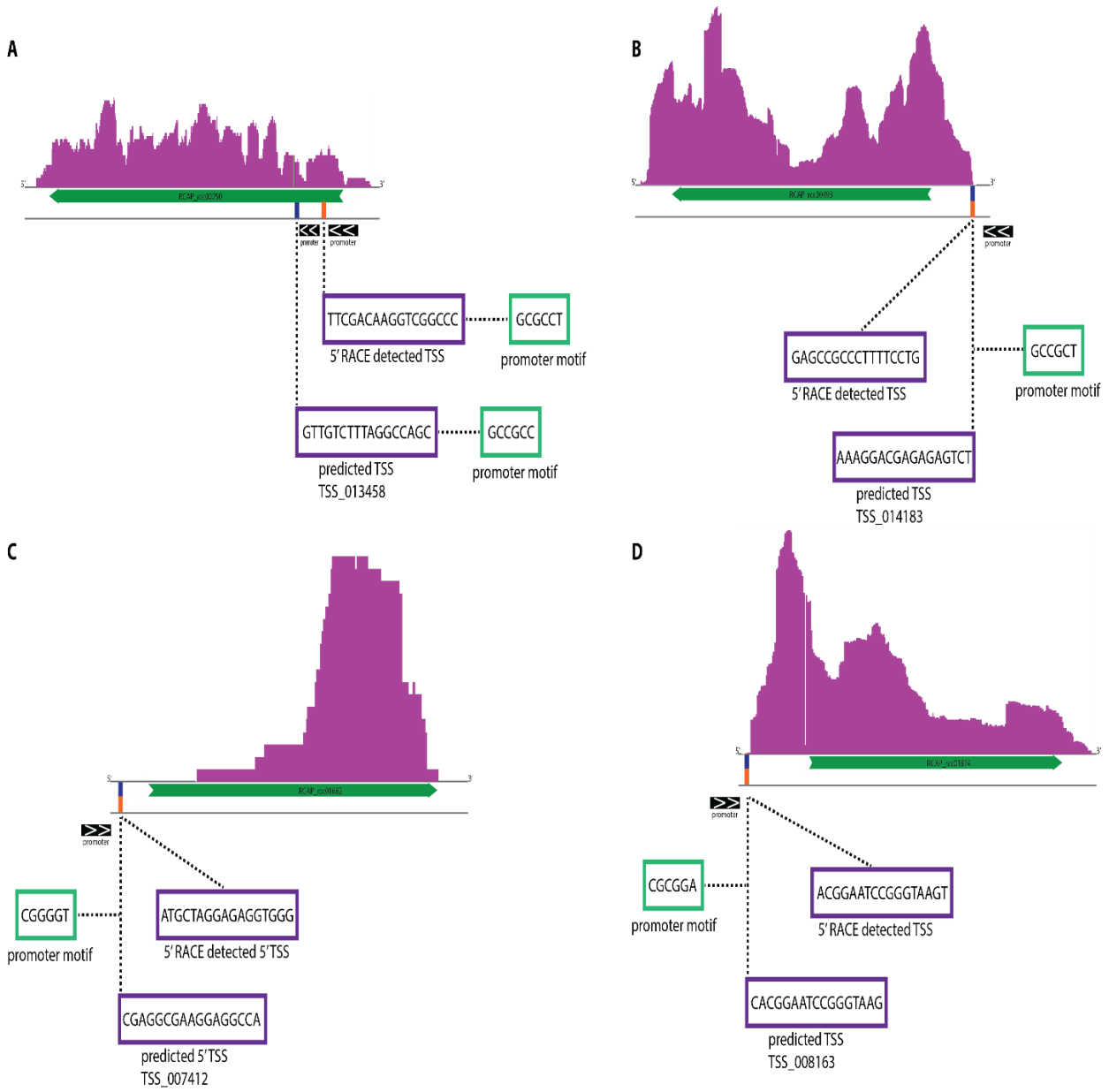
A
TTCGACAAGGTCGGCCC
GCGCCT
5′ RACE detected TSS
promoter motif
GTTGTCTTTAGGCCAGC
GCCGCC
predicted TSS
TSS_013458
promoter motif
B
GAGCCGCCCTTTTCCTG
GCCGCT
5′ RACE detected TSS
promoter motif
AAAGGACGAGAGAGTCT
predicted TSS
TSS_014183
C
CGGGGT
promoter motif
ATGCTAGGAGAGGTGGG
5′ RACE detected 5′TSS
CGAGGCGAAGGAGGCCA
predicted 5′TSS
TSS_007412
D
CGCGGA
promoter motif
ACGGAATCCGGGTAAGT
5′ RACE detected TSS
CACGGAATCCGGGTAAG
predicted TSS
TSS_008163

Figure 3.9: Read depth coverage plots and genomic locations for the experimentally confirmed TSS. Parts of the neighboring genes are shown with green arrows indicating their direction in the genome. Their relative distance to the coverage plots is not to scale. Predicted promoters are depicted as black boxes with white arrows inside, and the sequences are given below. The distance of the promoter relative to the TSS is also not to scale. The total RNA-seq reads are presented as purple plots. Blue bars mark the predicted TSS and orange bars the experimentally validated TSS. The start sequences of TSS locations and the promoters are shown underneath the plots in their respective 5'–3' orientations. Panel (A) shows the predicted and experimental results for gene RCAP_rcc00250, panel (B) shows results for gene RCAP_rcc0493, panel (C) shows results for gene RCAP_rcc01682 and panel (D) represents gene RCAP_rcc01874.

RCAP_rcc00493 was predicted to have a primary TSS at position 526,319 with a predicted promoter starting at position 526,371. We were able to confirm the position of a TSS at position 526,309 with RACE, 10 bp upstream of the RNA-seq-based site (Figure 3.9 B).

A previous study predicted a promoter site for RCAP_rcc01682 at position 1,818,559 (also referred to as *orfg1* for the RcGTA gene cluster), 129 bp upstream of its predicted start codon at position 1,818,688. Initial RACE experiments identified a 5' end at position 1,818,643, five nts downstream of a predicted stem-loop structure at position 1,818,615-1,818,637 (Mercer & Lang, 2014). It is likely that this stem-loop structure interfered with the reverse transcription at the beginning of the 5' RACE procedure. A subsequent 5' RACE experiment with added DMSO resulted in reverse transcription to pass the predicted stem-loop structure, resulting in a 5' end at position 1,818,595, 35 bp downstream of the previously predicted promoter sequence. This position agrees with a previous prediction (Mercer & Lang, 2014) The closest predicted RNA-seq-based TSS was at position 1,818,593, 2 bp upstream of the site identified with RACE (Figure 3.9 C). Although the location of the predicted 5' end agrees with previous predictions, the identified site using 5' RACE does not match with the sequencing coverage data.

RCAP_rcc01874 was predicted to have a primary TSS at position 2,027,818 with a predicted promoter at position 2,027,767. We were able to confirm the position of the TSS with RACE, identifying a TSS at position 2,027,819, 1 bp downstream of the predicted site (Figure 3.9 D).

With the exception of RCAP_rcc00250, the remaining three RACE-detected 5' TSS were in accordance with the sequencing coverage data. The TSS detected for RCAP_rcc00250 could have been caused by RNA processing prior to sequencing, despite the TEX treatment, or the RACE reaction could have been prematurely terminated.

3.4 Conclusion

Integrating the sequencing data obtained by an altered protocol of dRNA-seq for 5' and 3' end-targeted sequencing and total RNA-seq data, we were able to get an insight into the whole transcriptome of two strains of *R. capsulatus* during early stationary phase of growth. Similar to recent studies in other bacterial species (Sharma *et al.*, 2010, Thomason *et al.*, 2015, Ettwiller *et al.*, 2016), we found a high level of complexity in the *R. capsulatus* transcriptome.

Approximately half of the *R. capsulatus* operons were simple operons with one identified TSS and one identified TTS, but most of the remaining operons have multiple TSS and multiple TTS, or a single TSS and multiple TTS. Approximately 80% of the TSS and TTS identified are within 500 nts of the 5' and 3' end, respectively, of an annotated gene or putative sRNA. The majority of the remaining TSS and TTS are internal to an annotated gene. We identified a single strong promoter motif. Additionally, we were able to identify four novel transcripts and a number of genomic regions of high transcriptional activity. Experimental validation of predicted

TSS using 5’ RACE showed that we were able to confirm TSS in close proximity to the RNA-seq-predicted sites.

3.5 References

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**: 403-410.

Amman F, Wolfinger MT, Lorenz R, Hofacker IL, Stadler PF & Findeiss S (2014) TSSAR: TSS annotation regime for dRNA-seq data. *BMC Bioinformatics* **15**.

Anders S, Pyl PT & Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166-169.

Bischler T, Kopf M & Voß B (2014) Transcript mapping based on RNA-seq data. *BMC Bioinformatics* **15**.

Borries A, Vogel J & Sharma CM (2012) Differential RNA sequencing (dRNA-Seq): deep-sequencing-based analysis of primary transcriptomes. *Tag-Based Next Generation Sequencing*,p. 109-121. Wiley-VCH Verlag GmbH & Co. KGaA.

Conway T, Creecy JP, Maddox SM, *et al.* (2014) Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *mBio* **5**: e01442-01414.

Creecy JP & Conway T (2015) Quantitative bacterial transcriptomics with RNA-seq. *Current Opinion in Microbiology* **23**: 133-140.

Dugar G, Herbig A, Forstner KU, Heidrich N, Reinhardt R, Nieselt K & Sharma CM (2013) High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. *PLoS Genetics* **9**: e1003495.

Ettwiller L, Buswell J, Yigit E & Schildkraut I (2016) A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics* **17**: 199.

Finn RD, Clements J & Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**: W29-W37.

Frith MC, Saunders NF, Kobe B & Bailey TL (2008) Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Computational Biology* **4**: e1000071.

Grüll MP, Pena-Castillo L, Mulligan ME & Lang AS (2017) Genome-wide identification and characterization of small RNAs in *Rhodobacter capsulatus* and identification of small RNAs affected by loss of the response regulator CtrA. *RNA Biology* 1-12.

Hor J, Gorski SA & Vogel J (2018) Bacterial RNA biology on a genome scale. *Molecular cell* **70**: 785-799.

Jahn CE, Charkowski AO & Willis DK (2008) Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *Journal of Microbiological Methods* **75**: 318-324.

Kalvari I, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, Bateman A & Petrov AI (2018) Non-Coding RNA Analysis Using the Rfam Database. *Current Protocols in Bioinformatics* **62**: e51.

Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD & Petrov AI (2018) Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research* **46**: D335-D342.

Kapranov P, Cheng J, Dike S, *et al.* (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**: 1484-1488.

Lahens NF, Ricciotti E, Smirnova O, Toorens E, Kim EJ, Baruzzo G, Hayer KE, Ganguly T, Schug J & Grant GR (2017) A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. *BMC Genomics* **18**: 602.

Leung MM, Brimacombe CA & Beatty JT (2013) Transcriptional regulation of the *Rhodobacter capsulatus* response regulator CtrA. *Microbiology* **159**: 96-106.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R & Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Mercer R & Lang AS (2014) Identification of a predicted partner-switching system that affects production of the gene transfer agent RcGTA and stationary phase viability in *Rhodobacter capsulatus*. *BMC Microbiology* **14**.

Mercer RG, Callister SJ, Lipton MS, Pasa-Tolic L, Strnad H, Paces V, Beatty JT & Lang AS (2010) Loss of the response regulator CtrA causes pleiotropic effects on gene expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *Journal of Bacteriology* **192**: 2701-2710.

Mitschke J, Georg J, Scholz I, *et al.* (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *PNAS* **108**: 2124-2129.

Pena-Castillo L, Mercer RG, Gurinovich A, Callister SJ, Wright AT, Westbye AB, Beatty JT & Lang AS (2014) Gene co-expression network analysis in *Rhodobacter capsulatus* and application to comparative expression analysis of *Rhodobacter sphaeroides*. *BMC Genomics* **15**: 730-2164-2115-2730.

Perkins TT, Kingsley RA, Fookes MC, *et al.* (2009) A Strand-specific RNA–Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *PLoS Genetics* **5**: 1-13.

Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Schindel HS & Bauer CE (2016) The RegA regulon exhibits variability in response to altered growth conditions and differs markedly between *Rhodobacter* species. *Microbial Genomics*.

Sharma CM & Vogel J (2014) Differential RNA-seq: the approach behind and the biological insight gained. *Current Opinion in Microbiology* **19**: 97-105.

Sharma CM, Hoffmann S, Darfeuille F, *et al.* (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**: 250-255.

Thomason MK, Bischler T, Eisenbart SK, Forstner KU, Zhang A, Herbig A, Nieselt K, Sharma CM & Storz G (2015) Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *Journal of Bacteriology* **197**: 18-28.

Wade JT & Grainger DC (2014) Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nature Reviews in Microbiology* **12**: 647-653.

Weaver PF, Wall JD & Gest H (1975) Characterization of *Rhodopseudomonas capsulata*. *Archives of Microbiology* **105**: 207-216.

Wittchen M, Busche T, Gaspar AH, Lee JH, Ton-That H, Kalinowski J & Tauch A (2018) Transcriptome sequencing of the human pathogen *Corynebacterium diphtheriae* NCTC 13129 provides detailed insights into its transcriptional landscape and into DtxR-mediated transcriptional regulation. *BMC Genomics* **19**: 82.

Chapter 4 - Relative expression analysis of *Rhodobacter capsulatus* gene transfer agent-associated genes and investigation of their potential genetic interactions

Abstract

The purple non-sulfur alpha-proteobacterium *Rhodobacter capsulatus* produces phage-like particles called gene transfer agents (RcGTA). These particles package random ~4-kb pieces of the cell's genome, which can then be transferred to other cells and lead to gene exchange. Most of the genes encoding proteins found in RcGTA particles are encoded by a structural gene cluster comprising 17 annotated open reading frames. Transcriptomic and proteomic analyses led to the discovery of additional genes, located outside of the structural gene cluster, that were found to be involved in regulation and/or functionality and/or release of RcGTA particles. Here I use real-time quantitative polymerase chain reaction analysis to examine the relative gene expression of eight RcGTA-associated genes in *R. capsulatus*. The goal of this study was to determine the effects of loss of various RcGTA regulators on transcript levels from the different RcGTA loci. Our results show that loss of *rcc01865* or *rcc01749* (*cckA*) had the strongest effect on the expression of genes encoding RcGTA head-spikes and tail-fibers. Loss of gene *rcc01663* (*ctrA*) affected the head-spike and tail-fiber genes as well as the GTA structural gene cluster. Loss of any GTA-associated gene always affected the GTA-associated lysis gene *rcc00555*.

4.1 Introduction

The purple non-sulfur alpha-proteobacterium *Rhodobacter capsulatus* is studied for different aspects of physiology, such as anoxygenic photosynthesis, and for its ability to produce

a bacteriophage-like particle called a gene transfer agent, RcGTA (Marrs, 1974). These RcGTA particles contain ~4 kb of random host genomic DNA sequences (Solioz & Marrs, 1977, Yen *et al.*, 1979). It is known that RcGTA structural proteins are mostly encoded by a ~15-kb cluster of genes in the *R. capsulatus* chromosome (RCAP_rcc01682-RCAP_rcc01698) (Lang & Beatty, 2000, Chen *et al.*, 2008). This cluster comprises 17 open reading frames (ORFs) arranged in an apparent operon in a phage head-to-tail organization (Casjens *et al.*, 1992, Lang *et al.*, 2002).

Results from sRNA sequencing comparing two *R. capsulatus* strains (Chapter 2) and investigation of operon complexity (Chapter 3), including the gene expression patterns of a RcGTA overproducer mutant, showed that mutations in the *R. capsulatus* genome can affect the expression of RcGTA-associated genes. Additional genes that are located outside of the GTA structural gene cluster are known to be involved in the regulation, production and release of RcGTA. CtrA, a response regulator, ChpT, a histidine phosphotransferase, and CckA, a histidine kinase, are part of a phosphorelay that is involved in the regulation of GTA particle production and release (Lang & Beatty, 2000, Mercer *et al.*, 2012, Westbye *et al.*, 2013). CtrA has been thoroughly studied for its role as a master regulator of the cell cycle in *Caulobacter crescentus* (Quon *et al.* 1996, Skerker and Laub 2004) and, although CtrA has a very different role in *R. capsulatus* (Mercer *et al.*, 2010), it is essential for the production of RcGTA as transcription of RcGTA structural cluster genes is absent in a *ctrA* knock-out mutant (Lang and Beatty 2000). It was also found that phosphorylated as well as unphosphorylated CtrA can activate RcGTA gene expression, while phosphorylated CtrA is required for the release of the GTA particles from the cell (Mercer *et al.*, 2012). CckA appears to be required for maximal expression of another GTA-associated gene located outside of the GTA structural gene cluster that encodes an endolysin (*rcc00555*) for the release of GTA particles (Westbye *et al.*, 2013 Hynes *et al.*, 2012).

Since GTA gene expression and production in *R. capsulatus* differ dependent on growth phases, the transcriptomes during different growth phases and comparisons of the wild-type strain to different mutant strains affected for RcGTA production were used to generate a list of ORFs with differential transcript levels that matched the pattern of the RcGTA structural gene cluster (Hynes et al 2016). This identified 4 additional loci that are involved in RcGTA function: *rcc00171*, *rcc00555-rcc00556*, *rcc01079-rcc01080*, and *rcc01865-rcc01866*. *rcc00555* was previously shown to be involved in GTA particle release by encoding an endolysin (Hynes *et al.*, 2012, Westbye *et al.*, 2013). The remaining ORFs were investigated further for a possible connection with respect to RcGTA production (Hynes et al 2016), with mutant strains constructed to evaluate the effect of loss of each candidate gene on RcGTA gene transfer activity. Loss of *rcc00171* abolished RcGTA gene transfer, but not production, and this gene was predicted to encode a tail-spike protein based on the presence of a conserved domain (DUF2793) (Schulz & Ficner, 2011, Hynes *et al.*, 2016). Another locus, *rcc01079-rcc01080*, was previously predicted to encode tail fibre proteins (Lang et al 2012) and loss of these genes affected RcGTA binding to cells (Hynes *et al*, 2016), but it was found that the proteins made up head spikes on the particles (Westbye *et al*., 2016). *rcc01865*, a putative DNA-binding protein, and *rcc01866*, with no predicted function, were hypothesized to be non-structural contributors to RcGTA function based on the phenotypes of the knockout strains (Hynes *et al.*, 2016). *rcc01865* was found to be required for the expression of the RcGTA structural gene cluster whereas *rcc01866* appears to be important for the maturation of functional RcGTA and its release from cells.

It has been proposed that Rcc01866 and the sensor kinase CckA might act in a common regulatory pathway, based on almost identical RcGTA phenotypes obtained by the disruption of either gene (Mercer *et al.*, 2012, Westbye *et al.*, 2013, Hynes *et al.*, 2016). While strains with

disruptions of *rcc01079,* and *rcc01866* exhibited impaired RcGTA transfer activity, disruption of *rcc00171* or *rcc01865* completely abolished gene transfer activity (Hynes *et al.,* 2016). GTA binding activity assays showed that GTAs produced by the *rcc00171* knock-out strain had about the same binding activity as GTAs released from wild-type cells but, according to activity data, were not able to transmit their genetic content. GTAs produced by the *rcc01079* knock-out strain were found to have poor GTA binding activity and reduced GTA transfer ability.

It is still unclear how these different RcGTA loci are interconnected in terms of their regulation. In this study I evaluate the effects of the loss of individual RcGTA regulatory genes on expression of the different RcGTA loci. For this purpose, a quantitative polymerase chain reaction (qPCR) assay was developed to measure the expression of eight RcGTA-associated genes. The relative transcript levels of these genes in five strains of *R. capsulatus* were compared with data on RcGTA gene transfer activity and capsid protein levels within and outside of cells. For a list of all genes investigated in this study, see Table 4.1.

Table 4.1: Genes used in this study and their predicted functions.

Gene	**Predicted function**	**Reference**
rcc00171	Putative tail-spike protein	Schulz & Ficner, 2011, Hynes *et al.*, 2016
rcc00555	endolysin	Fogg *et al.*, 2012, Hynes *et al.*, 2012, Westbye *et al.*, 2013
rcc01079	Head-spike protein	Lang *et al*, 2012
rcc01683	Terminase (large subunit)	Lang & Beatty, 2000

rcc01684	Portal	Lang & Beatty, 2000
rcc01686	Prohead protease	Lang & Beatty, 2000
rcc01865	Putative DNA-binding protein	Hynes *et al.*, 2016
rcc01866	-	Hynes *et al.*, 2016

4.2 Materials and Methods

4.2.1 Strains and growth conditions

All cultures were grown under anaerobic, phototrophic conditions at 35°C in complex YPS medium (Weaver *et al.*, 1975) until four hours after reaching stationary phase, as measured by cell culture turbidity. Cultures were mixed at a ratio of 5:1 with a mixture of 95% ethanol and 5% saturated phenol (Jahn *et al*., 2008). Cells were then pelleted by centrifugation and pellets were frozen on dry ice/ethanol and stored at -80ºC as described in Grüll *et al*. (2017). For a list of *R. capsulatus* strains used in qPCR experiments, see Table 4.2.

Table 4.2: Strains used in qPCR experiments

Strain	**Mutated gene**	**Reference**
SB1003	None, parental strain	Strnad *et al.*, 2010
SBRM1	*rcc01663* (*ctrA*)	Mercer *et al.*, 2010
SBcckA	*rcc01749* (*cckA*)	Lang *et al.*, 2002
SB1865	*rcc01865*	Hynes *et al.*, 2016
SB1866	*rcc01866*	Hynes *et al.*, 2016

4.2.2 Design of primers for qPCR

Primers were designed to target the genes of interest (*rcc00171*, *rcc00555*, *rcc01079*, *rcc01683*, *rcc01684*, *rcc01686*, *rcc01865*, and *rcc1866*) using the Primer3 software (Koressaar & Remm, 2007, Untergasser *et al.*, 2012). Primer sets were also designed specific to the DNA-directed RNA polymerase omega subunit gene, *rpoZ*, and the DNA-directed RNA polymerase alpha subunit gene, *rpoA*, which were used as reference genes to normalize expression values for all other genes. *rpoZ* was chosen as a reference gene because its expression has been shown to be invariable across growth conditions in microarray experiments performed on *Rhodobacter sphaeroides* (Gomelsky *et al.*, 2003), a close relative of *R. capsulatus*. *rpoA* was chosen as a second reference gene because it has been used as a reference housekeeping gene in several other qPCR studies in bacteria, such as *Campylobacter jejuni* (Ritz *et al.*, 2009), *Clostridium ljungdahlii* (Tremblay *et al.*, 2012) and *Escherichia coli* (Wagner *et al.*, 2009). Primers were designed to amplify approximately 120 bp of the target gene with primer lengths varying between 20 and 24 nt. Primer sequences are provided in Table 4.3.

Table 4.3: Primers used in qPCR experiment

ORF	**Primer pair sequences (5'-3')**
rcc03318 (*rpoZ*)	TCTGGGTCTGGTTGCTTTCGATCA ACCACCGGGTTCTTGTCATTGTCA
rcc00326 (*rpoA*)	TGAACACCGGCAAGGGCTATGT ATTCGCCGGTGAAGAAGGTCGC
rcc00171	CAACACCACGCATGACACCACG CCTGCCCACCGATACCGCAAG
rcc00555	GCCAGTATTTCAAGGCGCCGAA CCGCATTCTGCAGGAGCTGTT

rcc01079	ATTACGAGATCACGCCCAGCGA GACCGGCATCACCGTGACCTAT
rcc01683	GGGCCTATCTGGCGGAAAGCTT TCTGCTGGAGGATGTCGAGGGG
rcc01684	TCAAGAGCTTCCATCCGACCGA AAGGCGCTTTTGGACAATGCCG
rcc01686	CGCGGGCTTTGGGTCAAGGG ATCGGATATCGCACCATCGCCG
rcc01865	TCAAGGCCTCCGCTTCCAACAG TGCACCAAGATCCAGCCGTTCA
rcc01866	ATCTGCGCCGACTTTTGGCGA CGGAGAGGCCGGCTACAGTTT

4.2.3 RNA Sample preparation and qPCR analysis

Total RNA was extracted using the NucleoSpin® miRNA kit (MACHEREY-NAGEL) following the manufacturer's protocol. Residual DNA in all samples was depleted using DNase I (NEB). Total RNA integrity before and after DNase I treatment was confirmed by gel electrophoresis, and the quality of the RNA after DNase I treatment was checked using a NanoDrop ND-1000 spectrophotometer. DNase I-treated total RNA was then converted to cDNA using the ProtoScript® II Reverse Transcriptase (NEB) following the manufacturer's standard protocol for first strand cDNA synthesis using the random primer mix. All cDNA samples were then used as template for amplification using all primer pair sets designed for qPCR experiments, and the products were examined by agarose gel electrophoresis to confirm that the correctly sized product was amplified.

PowerUp™ SYBR® Green Master Mix (Thermo Fisher) was used to monitor amplification and to quantify the amount of PCR products using the StepOnePlus™ Real-Time PCR System (Thermo Fisher) with software version 2.1. The threshold cycle (C_Q) was determined manually by selecting the beginning of exponential amplification of each sample. qPCR reactions (10 µL) were made according to the StepOnePlus protocol and contained PowerUp Master mix, 1 mM of each primer, and 5 ng cDNA. The thermal cycler protocol was as follows: initial UDG activation at 50°C for 2 min, activation of Dual-Lock™ DNA polymerase at 95°C for 2 min, followed by 40 cycles of 95°C for 3 s and 60°C for 30 s. The fluorescence signal was measured at the end of each extension step. After the amplification, a melting curve analysis was performed with the following settings: 95°C for 15 s (ramp speed 1.6°C/s), 60°C for 1 min (ramp speed 1.6°C/s) and 95°C for 15 s (ramp speed 0.15°C/s) to confirm that only one product was amplified. Each qPCR experiment was performed in triplicate.

4.2.4 Relative gene expression analysis

Relative quantification was performed using the $\Delta\Delta C_q$ method as described (Livak & Schmittgen, 2001). In this method, $\Delta C_q = C_q$ (target gene) – C_q (reference gene), and $\Delta\Delta C_q = \Delta C_q$ (target strain) – ΔC_q (reference strain). The target genes were *rcc00171*, *rcc00555*, *rcc01079*, *rcc01683*, *rcc01684*, *rcc01686*, *rcc01865*, and *rcc01866*. The reference genes were *rpoZ* and *rpoA*. The experimental strains were *R. capsulatus* SBRM1, SBcckA, SB1865, and SB1866, with the parental strain for these knock-out strains, SB1003, as the reference.

4.2.5 Statistical analysis

All statistical analysis was performed using GraphPad Prism software 7.04 (GraphPad software, 2018). The data were $\log_{10}$-transformed to test for normal distribution and to confirm that the assumptions for a parametric test were met. Analysis of variance (ANOVA) and Dunnett's multiple comparisons test were performed to test for significance of differences in mean expression of target genes in the different *R. capsulatus* strains (for full list of results see Supplementary Table S5).

4.3 Results

4.3.1 Quality control

Extracted and DNase I-treated RNA, which was used for qPCR experiments, were run on agarose gels to verify RNA integrity. The gel images showed good quality RNA for all samples based on the strong signal for the 16S and 14S rRNA bands. In *R. capsulatus*, the 23S rRNA is cleaved into 16S and 14S rRNA molecules at a predicted hairpin-loop structure that is removed from the 23S rRNA during a processing step (Kordes *et al.*, 1994). Another indicator for good RNA quality was intense smearing representing the variably sized other RNAs present inside cells (Figure 4.1).

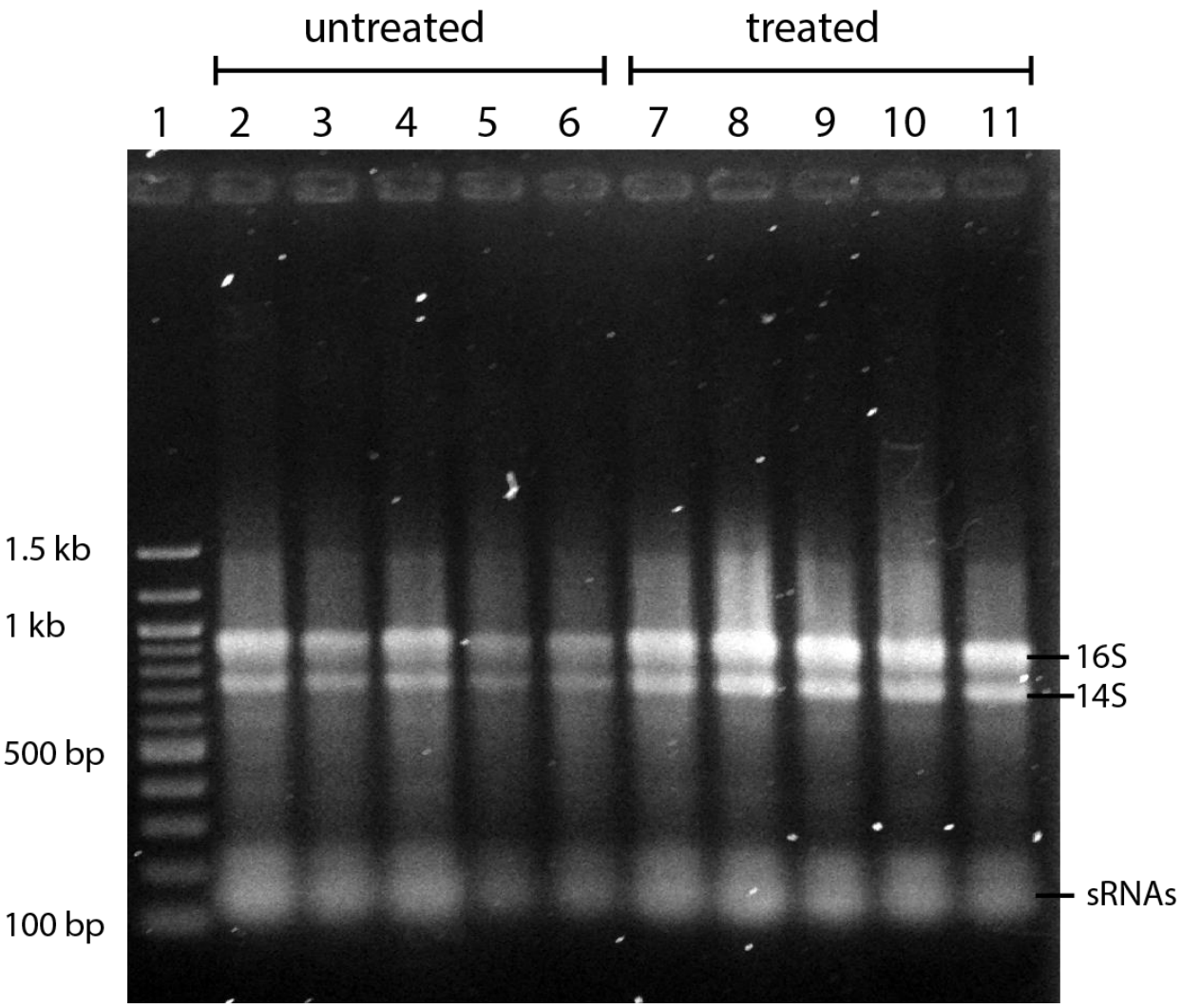

Figure 4.1: Agarose gel image showing extracted RNA samples before (untreated) and after (treated) DNase I treatment. Total RNA samples from the same strain before (2-6; loaded in order: SB1003, SB1865, SB1866, SBcckA, SBRM1) and after DNase I treatment (7-11; loaded in order SB1003, SB1865, SB1866, SBcckA, SBRM1) showed good quality based on a strong signal for 16S and 14S rRNAs, a visible smear and the presence of sRNAs. RNA sizes are indicated on the left.

NanoDrop readings were taken to verify the RNA quality after DNase I treatment. All readings presented positive for good quality RNA with 260/280 and 260/230 readings of approximately 2.0.

4.3.2 Standard and melting curves

The cDNA from all 5 strains was pooled and amplified with each set of primer pairs separately to generate standard and melt curves. This was done to verify that each set of primer pairs amplified their target specifically and efficiently.

Sample runs for the creation of standard curves were performed in triplicate. Replicates for each primer set showed similar amplification curves (see Figure 4.2) and confirmed the reaction efficiency for all primer pairs.

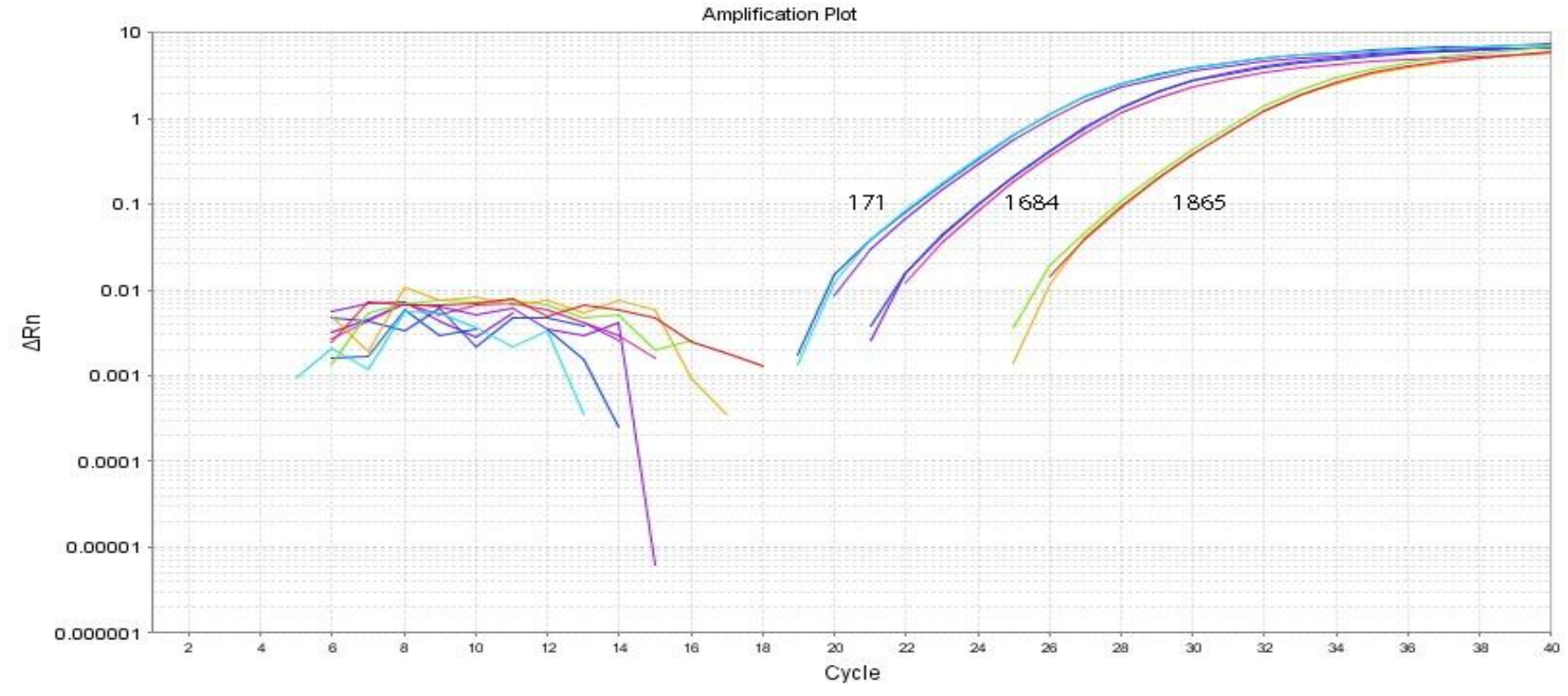

Figure 4.2: Representative amplification plots for three different primer sets. Data are plotted as normalized intensity of the reporter dye (ΔR_n) over number of PCR cycles. Pooled cDNAs, generated from DNase I-treated total RNA from all *R. capsulatus* strains used in these experiments, were tested for reaction efficiency. Three replicate amplification curves are shown for genes *rcc00171* (left), 1684 (middle) and 1865 (right). Similar results were obtained for the other 7 primer pair sets used (not shown).

To achieve exact quantification of specific PCR products, the generation of non-specific products must be avoided (Wilhelm *et al.*, 2003). To rule out non-specific binding, a melting curve analysis is frequently performed (Ririe *et al.*, 1997). The process of melting double-stranded DNA causes a sharp drop in the fluorescence signal from the SYBR Green dye around the melting temperature (T_m) of the PCR product, resulting in a clear peak in the negative derivative of the melting curve ($-dF/dT$). Fragments with different melting temperatures, including primer-dimers, will therefore appear as separate peaks (Ririe *et al.*, 1997). No amplification of non-specific products was observed (Figure 4.3).

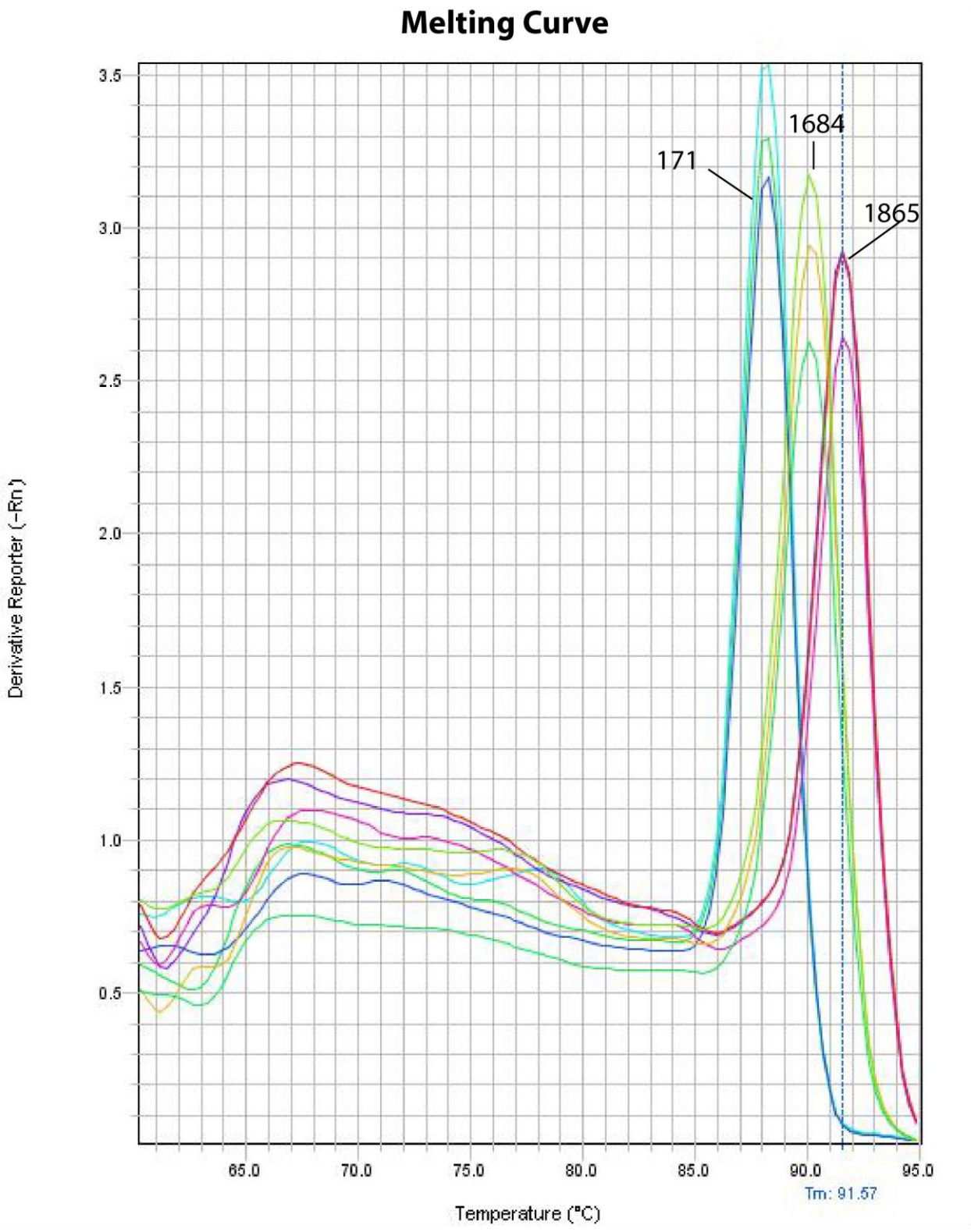
Melting Curve
Derivative Reporter (-Rn')
Temperature (°C)
171
1684
1865
Tm: 91.57
3.5
3.0
2.5
2.0
1.5
1.0
0.5
65.0
70.0
75.0
80.0
85.0
90.0
95.0

Figure 4.3: Representative melting curves for three different primer sets/ -genes. Pooled cDNA, generated from DNase I-treated total RNA from all *R. capsulatus* strains used in these experiments, was tested for non-specific amplification of targets. Three replicate melt curves are shown for the amplification products for genes *rcc00171* (left), *rcc01684* (middle) and *rcc01865* (right). Similar results were obtained for the other 7 primer pairs used (not shown).

4.3.5 Effect of loss of *rcc01865*

Analysis of transcript levels for the different genes by qPCR in the SB1865 strain revealed that most of the GTA-associated genes are down-regulated in the absence of *rcc01865* (Figure 4.4). The greatest effect was observed for *rcc01686,* which encodes the RcGTA major capsid protein, whose transcript levels were reduced approximately 700-fold. The transcript levels of genes *rcc00171, rcc00555, rcc01079, rcc01683* and *rcc01684* were reduced 5-80-fold. There was minimal or no effect on expression of *rcc01866.* All changes in gene expression, with the exception of *rcc01866,* were statistically significant.

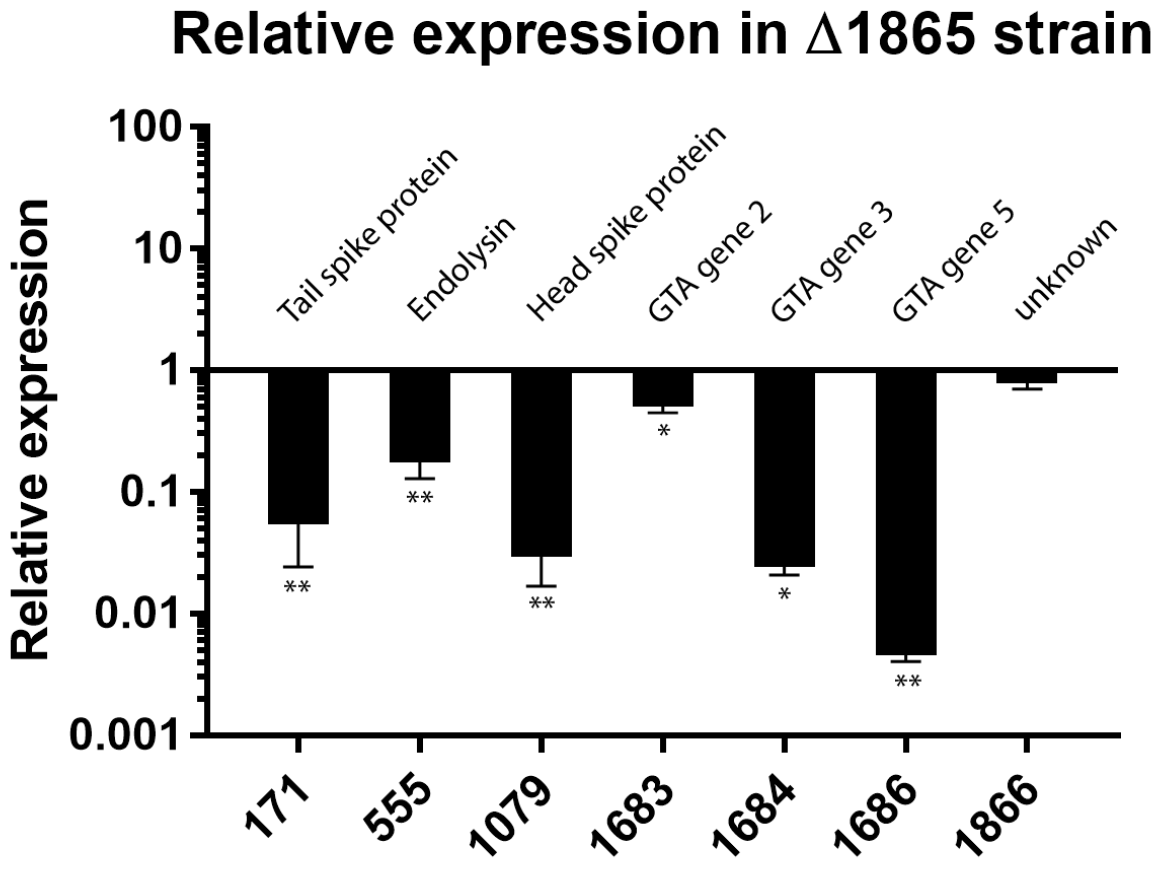

Figure 4.4: Effect of loss of gene *rcc01865* on expression of other GTA-associated genes within and outside the RcGTA structural gene cluster. The fold change of gene expression is shown relative to the wild-type (SB1003). Error bars represent standard deviations. Asterisks (*) indicate statistically significant differences from the wild-type (** $p < 0.0001$, * $p<0.005$; Dunnett's multiple comparisons test).

4.3.6 Effect of loss of *rcc01866*

Analysis of the SB1866 strain revealed up-regulation of most investigated genes in the absence of *rcc01866* (Figure 4.5). The highest up-regulation was observed on *rcc01865*, which was increased ~14 fold. Expression of gene *rcc00171* increased ~2.5-fold. Expression of all

other genes did not increase by more than ~1.7-fold. All changes in gene expression, with the exception of *rcc01079, rcc01683* and *rcc01686,* were statistically significant.

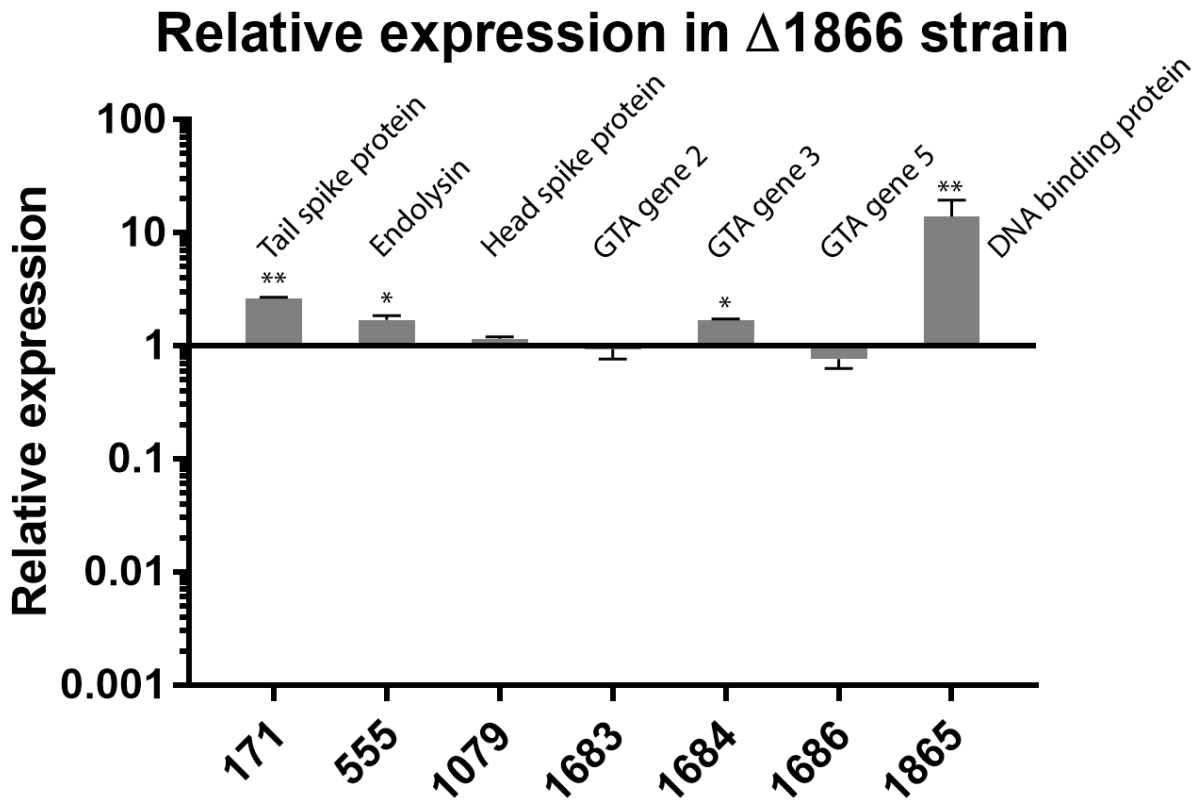

Figure 4.5: Effect of loss of gene *rcc01866* on expression of other GTA-associated genes within and outside the RcGTA structural gene cluster. The fold change of gene expression is shown relative to the wild-type. Error bars represent standard deviation. Asterisks (*) indicate statistically significant differences from the wild-type (** $p < 0.0001$, * $p<0.005$; Dunnett's multiple comparisons test).

4.3.7 Effect of loss of CckA

qPCR analysis of the SBcckA strain revealed that all of the GTA-associated genes are down-regulated in the absence of *cckA* (Figure 4.6). The greatest effect was seen on gene *rcc01079*, which was decreased ~100-fold. Expression of genes *rcc00171*, *rcc00555*, *rcc01683*, *rcc1684*, *rcc01686*, *rcc01865* and *rcc01866* was reduced 5-80-fold. All changes in gene expression, with the exception of *rcc01865*, were statistically significant.

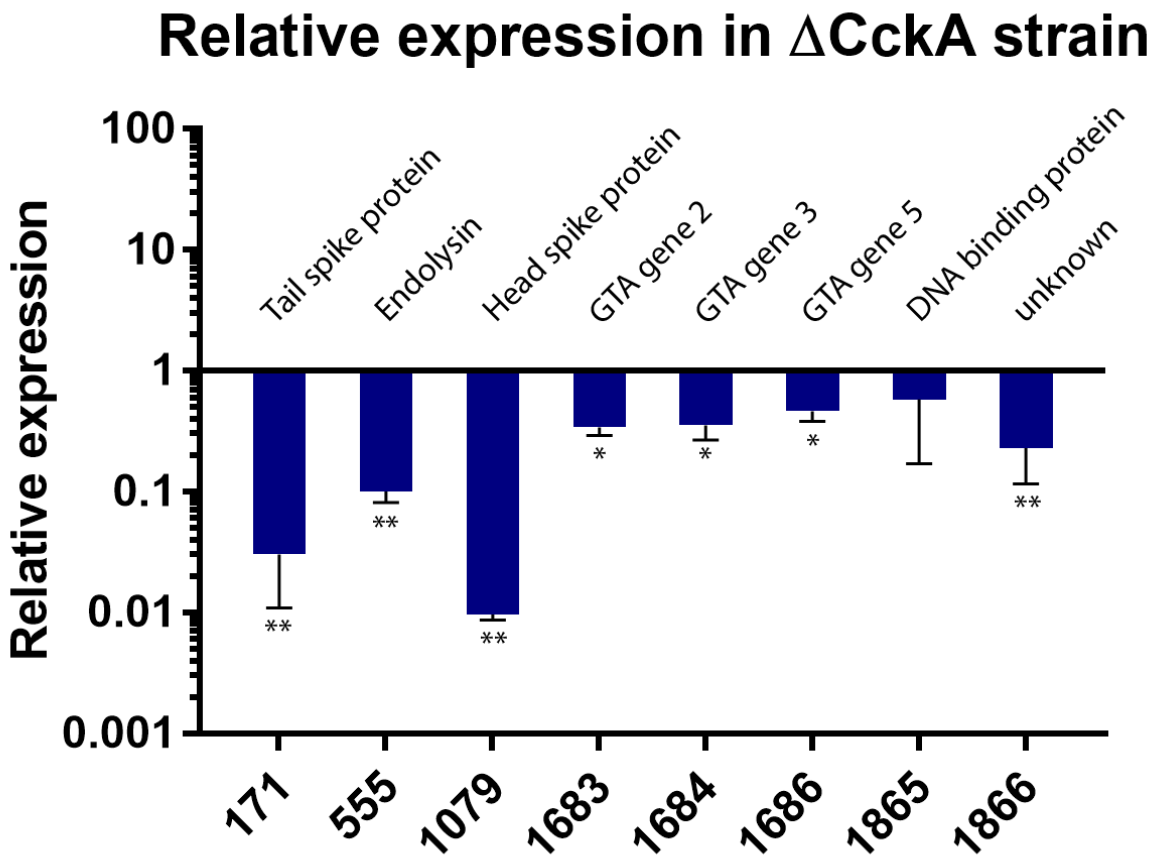

Figure 4.6: Effect of loss of gene *rcc01749* on expression of other GTA-associated genes within and outside the RcGTA structural gene cluster. The fold change of gene expression is shown relative to the wild-type. Error bars represent standard deviation. Asterisks (*) indicate statistically significant differences from the wild-type (** $p < 0.0001$, * $p<0.005$; Dunnett's multiple comparisons test).

4.3.8 Effect of loss of CtrA

The expression pattern of the SBRN1 strain revealed that all of the GTA-associated genes are down-regulated in the absence of *ctrA* (Figure 4.7). The greatest effect was seen on gene *rcc01686*, which encodes the major capsid protein and whose expression was reduced ~600-fold. Expression of genes *rcc00171*, *rcc00555*, *rcc01079*, *rcc01683*, *rcc1684*, *rcc01865* and *rcc01686* was reduced ~9-90-fold. All changes in gene expression, with the exception of *rcc01865*, were statistically significant.

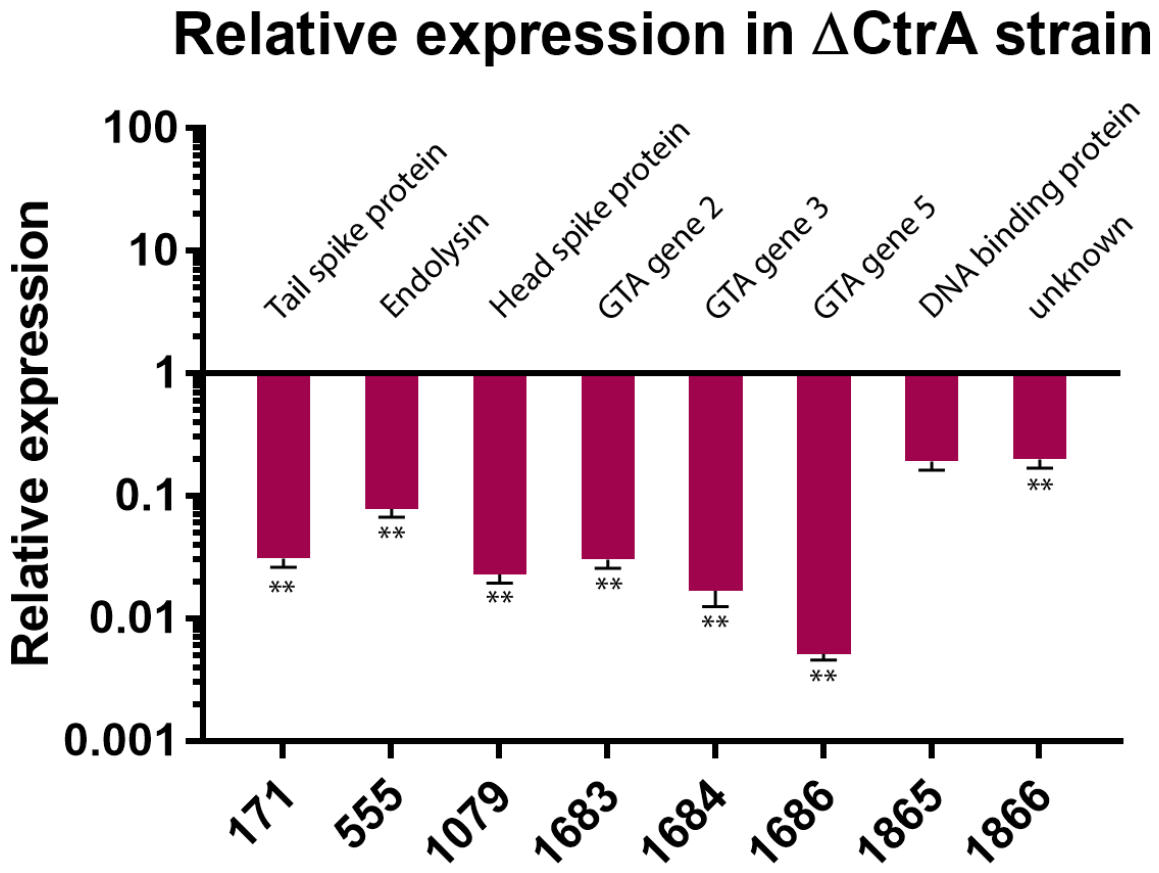

Figure 4.7: Effect of loss of gene *rcc01490* on expression of other GTA-associated genes within and outside the RcGTA structural gene cluster. The fold change of gene expression is shown relative to the wild-type. Error bars represent standard deviation. Asterisks (*) indicate statistically significant differences from the wild-type (** $p < 0.0001$, * $p<0.005$; Dunnett's multiple comparisons test).

4.4 Discussion

Our results show that losses of the putative and known GTA regulator genes studied here lead to decreased transcript levels of multiple GTA loci that I investigated in this study (Figure 4.8). Loss of *rcc01866* lead to increased transcript levels of some of the genes.

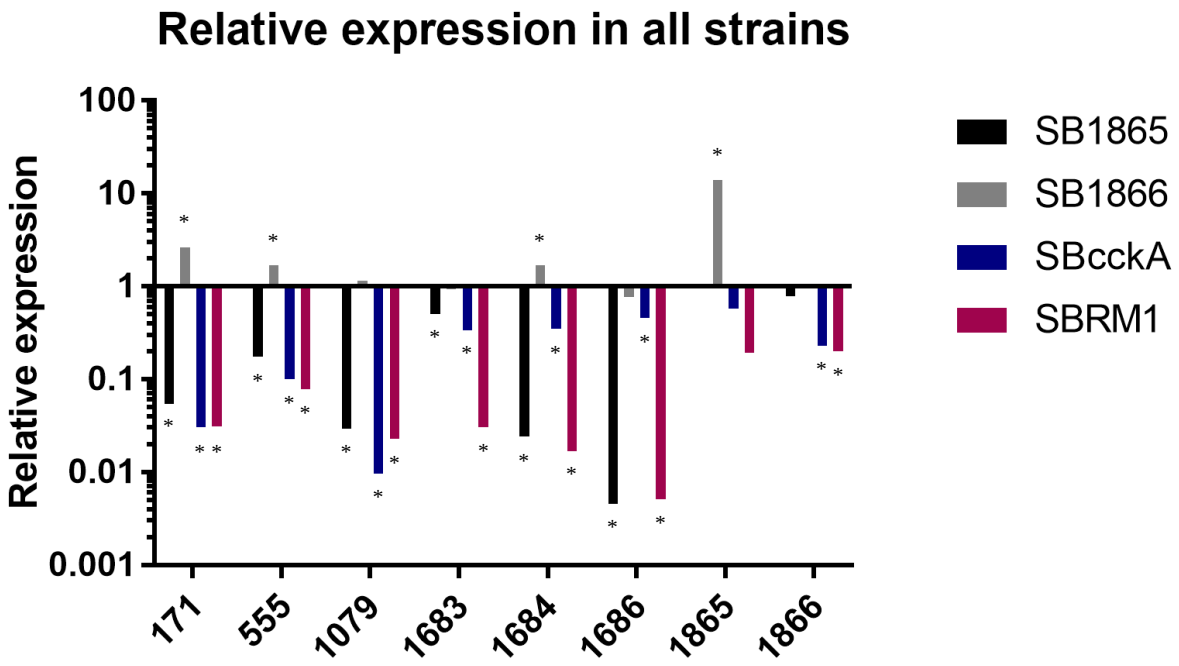

Figure 4.8: Comparison of relative expression of all genes in all strains used in this study. The fold change of gene expression is shown relative to the parental strain, SB1003. Error bars represent standard deviation. Asterisks (*) indicate statistically significant differences from the wild-type ($p < 0.001$; Dunnett's multiple comparisons test).

Loss of *rcc01865* and *ctrA* had the greatest effect on the RcGTA structural gene cluster, whereas loss of *cckA* had the largest effect on genes involved in RcGTA particle release and maturation.

4.4.1 Loss of *rcc01865*

Our results show that the GTA structural gene cluster genes are differentially expressed compared to each other in the absence of expression of *rcc01865*. I detected only slight down-regulation of *rcc01683*, the second gene of the GTA structural cluster, but there was a much greater (90-700 fold) down-regulation of gene *rcc01684* and *rcc01686*, the third and fifth genes of the GTA cluster. Considering the qPCR results and the predicted role of Rcc01865 as a DNA binding protein, it is possible that it acts as an upstream activator for the GTA structural cluster genes. However, it is not clear why the downstream genes are more affected by its absence. Inspection of the nucleic acid sequence between *rcc01683* and *rcc01684* reveals that a potential Rho-independent transcription terminator stem loop structure is present in this region. One possible explanation for the qPCR data is that Rcc01865 could be required for transcription anti-termination at this structure, thereby allowing transcription to continue into the downstream genes. Another explanation is that Rcc01865 also acts as a transcriptional activator between *rcc01683* and *rcc01684*, but none of my RNA-seq data (Chapter 3) support the initiation of transcription in this location. All the other GTA-associated loci were negatively affected by the loss of *rcc01865*, indicating that Rcc01865 plays an important role in GTA gene expression.

A previous study on the *rcc01865* knockout strain detected no GTA capsid protein within cells or in culture supernatants (Hynes *et al.*, 2016), which is consistent with the large down-regulation of the capsid gene I observed by qPCR.

4.4.2 Loss of *rcc01866*

Genes *rcc01865* and *rcc01866* are neighboring but divergent genes in the *R. capsulatus* genome. Although I observed a minimal change in the expression of gene *rcc01866* in the absence of *rcc01865*, the absence of *rcc01866* has a much greater effect on gene *rcc01865* with an up-regulation of *rcc01865* by ~14-fold. The *rcc01866* gene product has no predicted function but based on our results it seems likely that *rcc01866* functions as a negative regulator of *rcc01865*. Rcc01866 also has an effect on genes *rcc00171* and *rcc00555*, indicating that it could be involved in the regulation of maturation and release of GTA particles. The slight up-regulation of gene *rcc01684* could have been caused by the up-regulation of gene *rcc01865*. As I observed, in the absence of *rcc01865*, gene *rcc01684* was down-regulated ~90-fold. An up-regulation of *rcc01865* in the absence of *rcc01866* is therefore likely to cause an up-regulation of gene *rcc01864*. Overall, the loss of Rcc01866 had mostly modest effects on the transcript levels of the different genes.

Previous experiments with this strain showed the presence of GTA capsid proteins within cells, but not in culture supernatants, and the capsid protein observed inside the cells did not represent the presence of functional GTAs (Hynes *et al.*, 2016). Based on this, I expected to observe a decrease in lysis gene expression in this strain, and possibly for the head-spike and/or tail-spike genes. However, this was not observed in the qPCR results. Fitting with expectation, the capsid gene was not affected in this strain.

4.4.3 Loss of CckA

The decrease of genes *rcc00171* (~80-fold) and *rcc01079* (~100-fold), encoding the predicted tail- (Hynes *et al.*, 2016) and documented head-spike proteins (Westbye *et al.*, 2016), respectively, fits with previous data showing the GTA particles made by this strain are largely defective (Mercer *et al*., 2012; Westbye *et al*., 2013). Expression of the endolysin-encoding gene *rcc00555* was ~10-fold lower than in the *R. capsulatus* wild-type strain, which is consistent with previous findings that GTA particles are not released out of the *cckA* mutant (Mercer *et al*., 2012; Westbye *et al*., 2013).

4.4.4 Loss of CtrA

CtrA is a global response regulator that affects expression of more than 200 genes in *R. capsulatus* (Mercer *et al*., 2010). Loss of CtrA seems to affect the expression of the three GTA structural genes somewhat differently. I observed a higher down-regulation of gene *rcc01686* compared to the other two GTA structural genes. This is similar to the effect seen in the *rcc01865* mutant strain, and it is possible that CtrA is also somehow involved in anti-termination between *rcc01683* and *rcc01684*. I observed a decrease in transcript levels for *rcc01865* in the *ctrA* mutant, which further indicates a connection between *rcc01865* and CtrA.

The *rcc00171, rcc00555* and *rcc01079* genes are greatly affected by the loss of CtrA, with a reduction of between ~20 and 90-fold in expression.

4.5 Conclusion

Production and release of fully functional RcGTA particles requires structural proteins from three loci and lysis genes at an additional locus. From previous studies it appeared that these loci were differentially regulated by *rcc01865*, *rcc01866*, CckA and CtrA. Using qPCR analysis, I was able to examine the effects of loss of these different regulators of RcGTA production on transcript levels for the different RcGTA-associated loci.

In the absence of *rcc01865* I saw the greatest effect on the expression of genes *rcc00171*, *rcc01079*, *rcc01684* and *rcc01686*. These results show that the absence of Rcc01865 affects genes regulating the production of RcGTA particles and the formation of tail- and head-spikes, which are needed for successful attachment of GTAs to bacterial cells. The absence of gene rcc01866 had the least effect compared to the loss of other investigated GTA-associated genes in this study. The greatest effects were observed on genes *rcc00171*, *rcc00555*, *rcc01864* and *rcc01865*. The absence of CckA had the greatest effect on genes *rcc00171*, *rcc00555* and *rcc01079*, which are not essential for the production of RcGTA. This suggests that CckA may be more strongly involved in the regulation of RcGTA particle maturation and release. In the absence of the global response regulator CtrA, I saw the greatest effect on the RcGTA structural genes *rcc01683*, *rcc01684* and *rcc01686*, as well as genes *rcc00171*, *rcc00555* and *rcc01079*. Based on our results it is not clear if CtrA is directly responsible for the down-regulation of the genes outside of the RcGTA structural gene cluster, or if the down-regulation of these genes is a secondary effect caused by the stop of RcGTA particle production.

I was able to demonstrate that the loss of one GTA-associated gene can have a great effect on the expression of other GTA-associated genes. The expression patterns of the

investigated genes are in accordance with predictions based on the RcGTA production phenotypes of the mutant strains.

Future studies might repeat these qPCR assays with cultures grown under different conditions to gain additional insight into RcGTA production in *R. capsulatus*. Since the overall expression of GTA-associated genes is low in wild-type *R. capsulatus*, the overproducer mutant (DE442) could be used for further investigation.

4.6 References

Casjens S, Hatfull G & Hendrix RW (1992) Evolution of dsDNA tailed-bacteriophage genomes. In *Seminars in Virology* **3**: 383-383.

Chen F, Spano A, Goodman BE, Blaster KR, Sabat A, Jeffery E, Norris A, Shabanowitz J, Hunt DE & Lebedev N (2008) Proteomic analysis and identification of the structural and regulatory proteins of the *Rhodobacter capsulatus* gene transfer agent. *Journal of Proteome Research* **8**: 967-973.

Deprez RHL, Fijnvandraat AC, Ruijter JM & Moorman AFM (2002) Sensitivity and accuracy of quantitative real-time polymerase chain reaction using SYBR green I depends on cDNA synthesis conditions. *Analytical Biochemistry* **307**: 63-69.

Fogg PC, Westbye AB & Beatty JT (2012) One for all or all for one: heterogeneous expression and host cell lysis are key to gene transfer agent activity in *Rhodobacter capsulatus*. *PloS One* **7**: e43772.

Gomelsky L, Sram J, Moskvin OV, Horne IM, Dodd HN, Pemberton JM, McEwan AG, Kaplan S & Gomelsky M (2003) Identification and in vivo characterization of PpaA, a regulator of photosystem formation in *Rhodobacter sphaeroides*. *Microbiology* **149**: 377-388.

Grüll MP, Pena-Castillo L, Mulligan ME & Lang AS (2017) Genome-wide identification and characterization of small RNAs in *Rhodobacter capsulatus* and identification of small RNAs affected by loss of the response regulator CtrA. *RNA Biology* 1-12.

Hynes A, Shakya M, Mercer R, *et al.* (2016) Functional and evolutionary characterization of a gene transfer agent's multi-locus "genome". *Molecular Biology and Evolution* **33**: 2530-2543.

Hynes AP, Mercer RG, Watton DE, Buckley CB & Lang AS (2012) DNA packaging bias and differential expression of gene transfer agent genes within a population during production and release of the *Rhodobacter capsulatus* gene transfer agent, RcGTA. *Molecular Microbiology* **85**: 314-325.

Jahn CE, Charkowski AO & Willis DK (2008) Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *Journal of Microbiological Methods* **75**: 318-324.

Kordes E, Jock S, Fritsch J, Bosch F & Klug G (1994) Cloning of a gene involved in rRNA precursor processing and 23S rRNA cleavage in Rhodobacter capsulatus. *Journal of Bacteriology* **176**: 1121-1127.

Koressaar T & Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**: 1289-1291.

Lang AS, Zhaxybayeva O & Beatty JT (2012) Gene transfer agents: phage-like elements of genetic exchange. *Nature Reviews Microbiology* **10**: 472-482.

Lang AS & Beatty JT (2007) Importance of widespread gene transfer agent genes in alpha-proteobacteria. *Trends in Microbiology* **15**: 54-62.

Lang AS, Taylor TA & Beatty JT (2002) Evolutionary implications of phylogenetic analyses of the gene transfer agent (GTA) of *Rhodobacter capsulatus*. *Journal of Molecular Evolution* **55**: 534-543.

Lang AS & Beatty JT (2000) Genetic analysis of a bacterial genetic exchange element: The gene transfer agent of *Rhodobacter capsulatus*. *Proceedings of the National Academy of Sciences USA* **97**: 859-864.

Livak KJ & Schmittgen TD (2001) Analysis of relative gene expression data using real-time

quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-408.
Marrs B (1974) Genetic recombination in *Rhodopseudomonas capsulata*. *Proceedings of the National Academy of Sciences USA* **71**: 971-973.
Marrs B, Wall JD & Gest H (1977) Emergence of the biochemical genetics and molecular biology of photosynthetic bacteria. *Trends in Biochemical Sciences* **2**: 105-108.
Matson EG, Thompson MG, Humphrey SB, Zuerner RL & Stanton TB (2005) Identification of genes of VSH-1, a prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. *Journal of Bacteriology* **187**: 5885-5892.
Mercer R & Lang AS (2014) Identification of a predicted partner-switching system that affects production of the gene transfer agent RcGTA and stationary phase viability in *Rhodobacter capsulatus*. *BMC Microbiology* **14**.
Mercer RG, Quinlan M, Rose AR, Noll S, Beatty JT & Lang AS (2012) Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. *FEMS Microbiology Letters* **331**: 53-62.
Mercer RG, Callister SJ, Lipton MS, Pasa-Tolic L, Strnad H, Paces V, Beatty JT & Lang AS (2010) Loss of the response regulator CtrA causes pleiotropic effects on gene expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *Journal of Bacteriology* **192**: 2701-2710.
Quon KC, Marczynski GT & Shapiro L (1996) Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell* **84**: 83-93.
Riede I, Schwarz H & Jahnig F (1987) Predicted structure of tail-fiber proteins of T-even type phages. *FEBS Letters* **215**: 145-150.
Ririe KM, Rasmussen RP & Wittwer CT (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Analytical Biochemistry* **245**: 154-160.
Ritz M, Garenaux A, Berge M & Federighi M (2009) Determination of rpoA as the most suitable internal control to study stress response in *C. jejuni* by RT-qPCR and application to oxidative stress. *Journal of Microbiological Methods* **76**: 196-200.
Schulz EC & Ficner R (2011) Knitting and snipping: chaperones in beta-helix folding. *Current Opinion in Structural Biology* **21**: 232-239.
Skerker JM & Laub MT (2004) Cell-cycle progression and the generation of asymmetry in *Caulobacter crescentus*. *Nature Reviews in Microbiology* **2**: 325-337.
Solioz M & Marrs B (1977) The gene transfer agent of *Rhodopseudomonas capsulata*,: Purification and characterization of its nucleic acid. *Archives of Biochemistry and Biophysics* **181**: 300-307.
Strnad H, Lapidus A, Paces J, Ulbrich P, Vlcek C, Paces V & Haselkorn R (2010) Complete genome sequence of the photosynthetic purple nonsulfur bacterium Rhodobacter capsulatus SB 1003. *Journal of Bacteriology* **192**: 3545-3546.
Tremblay PL, Zhang T, Dar SA, Leang C & Lovley DR (2012) The Rnf complex of *Clostridium ljungdahlii* is a proton-translocating ferredoxin:NAD+ oxidoreductase essential for autotrophic growth. *mBio* **4**: e00406-00412.
Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M & Rozen SG (2012) Primer3--new capabilities and interfaces. *Nucleic Acids Research* **40**: e115.
Wagner MA, Zahrl D, Rieser G & Koraimann G (2009) Growth phase- and cell division-dependent activation and inactivation of the σ32 regulon in *Escherichia coli*. *Journal of Bacteriology* **191**: 1695-1702.

Weaver PF, Wall JD & Gest H (1975) Characterization of *Rhodopseudomonas capsulata*. *Archives of Microbiology* **105**: 207-216.

Westbye AB, Kuchinski K, Yip CK & Beatty JT (2016) The gene transfer agent RcGTA contains head spikes needed for binding to the *Rhodobacter capsulatus* polysaccharide cell capsule. *Journal of Molecular Biology* **428**: 477-491.

Westbye AB, Leung MM, Florizone SM, Taylor TA, Johnson JA, Fogg PC & Beatty JT (2013) Phosphate concentration and the putative sensor kinase protein CckA modulate cell lysis and release of the *Rhodobacter capsulatus* gene transfer agent. *Journal of Bacteriology* **195**: 5025-5040.

Wilhelm J, Pingoud A & Hahn M (2003) Real-time PCR-based method for the estimation of genome sizes. *Nucleic Acids Research* **31**.

Yen HC, Hu NT & Marrs BL (1979) Characterization of the gene transfer agent made by an overproducer mutant of *Rhodopseudomonas capsulata*. *Journal of Molecular Biology* **131**: 157-168.

5. Summary

Currently, our understanding of the global regulation of gene expression in *R. capsulatus* is limited to a few studies. The first study to investigate this on a genome-wide scale was performed in 2010 using microarrays based on a preliminary version of the complete genome sequence of *R. capsulatus* (Mercer *et al*., 2010). The microarrays contained oligonucleotide probes for 3,635 ORFs and 1,452 intergenic regions greater than 90 bps. Because of the size restrictions on microarrays, in terms of the number of genes that can be investigated, the results gave an insight into gene expression patterns in *R. capsulatus* but were not able to capture a complete transcription profile of all genes in the genome. Mercer's studies focused on the response regulator CtrA, and the resulting hypothesis was that CtrA might exert some of its effects through sRNAs. No sRNA work had previously been done for *R. capsulatus*, so I decided to characterize sRNAs in this organism and to investigate the possibility that CtrA regulated the expression of some sRNAs.

To identify and characterize sRNAs in *R. capsulatus*, I had to selectively isolate sRNAs from a pool of total RNAs. I performed the sequencing on the high-throughput instrument that was available in our group, an Ion Torrent PGM, which generated a total of 3,966,797 reads for the wild-type strain that were successfully mapped to the reference genome. For the *ctrA* mutant strain, I was able to generate a total of 3,461,348 mapped reads. After data processing using bioinformatic tools that were developed in-house, my study yielded a total of 422 putative sRNAs from the RNA-seq data. Of these, 124 were conserved in at least one out of 23 bacterial species that were investigated, but only 19 out of the 422 putative sRNAs were assigned a predicted function, reflecting a general lack of knowledge regarding sRNA functions in bacteria. Putative *cis*-targets for antisense and partially overlapping sRNAs were found to be enriched

with protein-coding genes involved in primary metabolic processes, photosynthesis, compound binding, and with genes forming part of macromolecular complexes. A differential expression analysis was performed to compare the wild-type strain to the strain lacking CtrA. This resulted in the discovery of 18 putative sRNAs with differing expression levels in the two strains. The use of Northern blot analysis on several predicted putative sRNAs validated these sRNAs as true sRNAs with differential expression between the strains. Further experimental work will be required to determine the potential role of these sRNAs in affecting CtrA-regulated processes. Based on my results, there are still 298 putative sRNAs remaining for which we have no indication of possible function in *R. capsulatus*. More work is needed to validate the existence and investigate the potential function of these putative sRNAs in *R. capsulatus*, including those regulated by CtrA.

After conducting the sRNA sequencing experiments, the next step was to investigate the overall transcriptome, especially the remaining RNAs that were excluded from the sRNA analysis. To be able to investigate the full array of RNAs expressed by *R. capsulatus*, I performed total RNA sequencing experiments with the wild-type strain SB1003. These resulted in good coverage for parts of the *R. capsulatus* genome, but this did not include the GTA structural gene cluster region. Because the RcGTA genes are only expressed in a sub-population (~3%) of cells in stationary phase cultures (Hynes *et al.*, 2012), the amount of sequencing reads generated in wild-type cultures were presumably not high enough for the GTA cluster region to surpass the signal threshold in this analysis. Therefore, I decided to use an RcGTA overproducer mutant (DE442) to obtain more GTA gene cluster reads to enable investigation of transcription in this region. I implemented techniques that allowed for the prediction of TSS and TTS genome-wide using RNA-seq. I was able to alter an existing protocol designed for the Illumina

sequencing platform, dRNA-seq, for use on the Ion Torrent sequencing platform. I used this for not only the traditional 5' end-targeted sequencing, but also developed a custom workflow for targeting RNA 3' ends.

This work revealed a high complexity in the *R. capsulatus* transcriptome, as has been observed in other bacterial species (Sharma *et al.*, 2010, Thomason *et al.*, 2015, Ettwiller *et al.*, 2016). The hypothesis is that more complex operons, in terms of numbers of TSS and TTS, are likely to be longer and more highly expressed than those with less TSS and TTS. I found that approximately half of the operons in the *R. capsulatus* genome are simple, with one identified TSS and one identified TTS. Most of the remaining operons were found to have multiple TSS and multiple TTS, or a single TSS and multiple TTS. Approximately 80% of all identified TSS and TTS were within 500 nt of the 5' and 3' end of an annotated gene or putative sRNA. For the 415 putative sRNAs that we had previously identified, 307 (74%) had either a TSS or a TTS within 50 nts of their 5' or 3' end in at least one of the studied strains. Out of all putative sRNAs, 124 (30%) have both a TSS and a TTS in at least one of the sequenced strains and 50 (12%) of all putative sRNAs have both a TSS and TTS in both sequenced strains. From the data it became apparent that, when comparing putative sRNAs to the remaining sequencing reads, putative sRNAs have fewer internal TSS and TTS than annotated genes. These results provided confidence in the predicted putative sRNAs from the previous study because the majority of these putative sRNAs were found to have predicted TSS or TTS. Using 5' RACE for experimental validation of predicted TSS, I was able to validate transcript ends in close proximity to the predicted sites. Utilizing all the data obtained from this study it was then possible to predict a conserved promoter motif in *R. capsulatus*. Only one study had previously attempted to identify a promoter motif in *R. capsulatus*, but that prediction was based on only 14

promoter sequences from nine different studies (Leung *et al.*, 2013). Based on my results, it was possible to identify a single strong motif in *R. capsulatus* promoter sequences (Figure 5.1).

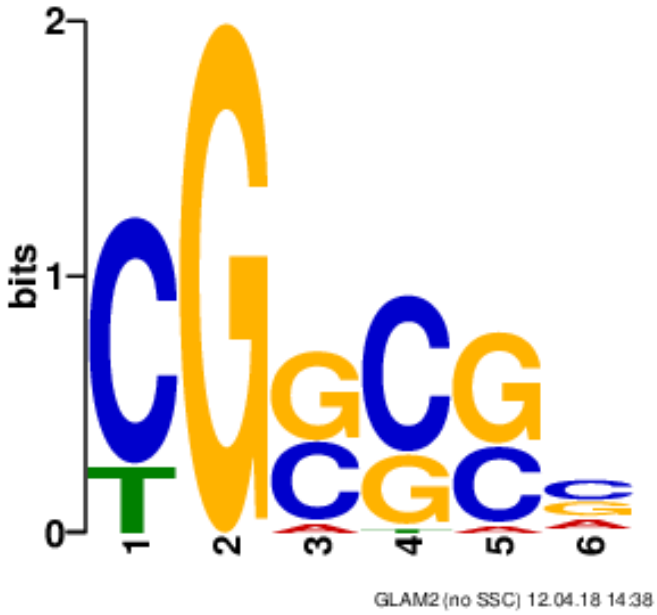

Figure 5.1: DNA sequence promoter motif identified in *R. capsulatus* based on predicted TSS and total RNA sequencing data.

In this study I was able to successfully integrate dRNA-seq and total RNA-seq data to describe the operon complexity in *R. capsulatus* and to identify a single conserved promoter motif. Using sequencing data from an RcGTA overproducer mutant I was able to gain an insight into the expression patterns of the genes within the RcGTA gene cluster and experimentally validate the promoter location upstream of the gene cluster. The limited number of 5' RACE results that I was able to obtain do confirm the predicted TSS locations for the chosen genes, but it is hard to make a prediction about the overall accuracy of all the predictions made. Comparing my results to those of other studies (Sharma *et al.*, 2010, Thomason *et al.*, 2015, Ettwiller *et al.*,

2016), the number of predicted TSS and TTS are similar. However, more experimental validation is necessary to verify predicted sites. Overall this study gives confident suggestions about the presence and location of potential TSS and TTS. The predicted operon structures and promoter motif can be of value for other investigators in future investigations.

After identifying total RNAs and sRNAs present during early stationary phase in *R. capsulatus* and gaining more insight into the gene expression of the RcGTA gene cluster, based on the results from total RNA-seq using an RcGTA overproducer, I studied the transcript levels of different GTA-associated in different mutant strains. I chose eight genes to investigate in my study using qPCR and the goal was to determine the relative expression of these genes in strains lacking genes known or predicted to affect RcGTA production at a regulatory level. Three of the genes that I chose to investigate are part of the GTA structural cluster and the remaining five are located elsewhere in the genome. The genes outside of the GTA gene cluster had been identified and given predicted functions in previous studies (Hynes *et al.*, 2012, Westbye *et al.*, 2013, Hynes *et al.*, 2016). I used existing RcGTA production phenotypes from the previous studies to compare to my results. I was able to demonstrate that the loss of known and putative GTA regulators can have a great effect on the transcription of GTA loci, with the expression patterns of the investigated genes being in accordance with the predictions from previous studies. Loss of gene *rcc01865* showed similar effects to the loss of *ctrA*, therefore the mechanism of Rc01865 regulation of GTA genes needs to be investigated further.

It might be of value to repeat the experiments presented in this work to investigate different growth conditions to get a better coverage of additional regions of the *R. capsulatus* genome. The expression of genes varies under different growth conditions, therefore further investigation will give us a more detailed view of the whole transcriptome of this organism.

While there are still many important questions unanswered, the work elaborated here has helped to gain a greater insight into the whole transcriptome and operon complexity of *R. capsulatus*. The discovery of novel sRNAs and the identification of a conserved promoter motif will serve as starting point for new studies to further investigate the *R. capsulatus* transcriptome and cellular processes.

5.1 References

Ettwiller L, Buswell J, Yigit E & Schildkraut I (2016) A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics* **17**: 199.

Hynes A, Shakya M, Mercer R, *et al.* (2016) Functional and evolutionary characterization of a gene transfer agent's multi-locus "genome". *Molecular Biology and Evolution* **33**: 2530-2543.

Hynes AP, Mercer RG, Watton DE, Buckley CB & Lang AS (2012) DNA packaging bias and differential expression of gene transfer agent genes within a population during production and release of the *Rhodobacter capsulatus* gene transfer agent, RcGTA. *Molecular Microbiology* **85**: 314-325.

Leung MM, Brimacombe CA & Beatty JT (2013) Transcriptional regulation of the *Rhodobacter capsulatus* response regulator CtrA. *Microbiology* **159**: 96-106.

Mercer RG, Callister SJ, Lipton MS, Pasa-Tolic L, Strnad H, Paces V, Beatty JT & Lang AS (2010) Loss of the response regulator CtrA causes pleiotropic effects on gene expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *Journal of Bacteriology* **192**: 2701-2710.

Sharma CM, Hoffmann S, Darfeuille F, *et al.* (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**: 250-255.

Thomason MK, Bischler T, Eisenbart SK, Forstner KU, Zhang A, Herbig A, Nieselt K, Sharma CM & Storz G (2015) Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *Journal of Bacteriology* **197**: 18-28.

Westbye AB, Leung MM, Florizone SM, Taylor TA, Johnson JA, Fogg PC & Beatty JT (2013) Phosphate concentration and the putative sensor kinase protein CckA modulate cell lysis and release of the *Rhodobacter capsulatus* gene transfer agent. *Journal of Bacteriology* **195**: 5025-5040.

Appendix 1 – Supplementary Figures and Tables for Chapter 2

Table S1:: Homologous sRNAs in other bacterial species (pink marks in Figure 1). References to corresponding studies are given in Table 2.1.

A. tumefaciens												
Query	Subject	PID	Length	QLen	Mismatch	Gaps	QStart	QEnd	SStart	SEnd	Evalue	Score
sRNA00598	NC_003062.2|:101444-101546(-)	80.77	52	52	9	1	1	51	49	100	0.00000001	50
sRNA01156	NC_003062.2|:2007230-2007565(+)	88.71	62	121	7	0	60	121	270	331	2E-18	88
N. gonorrhoeae												
Query	Subject	PID	Length	QLen	Mismatch	Gaps	QStart	QEnd	SStart	SEnd	Evalue	Score
sRNA01158	NC_002946.2|:863130-863310(+)	90.62	32	65	3	0	32	63	149	180	0.00000002	48
P_acnes												
Query	Subject	PID	Length	QLen	Mismatch	Gaps	QStart	QEnd	SStart	SEnd	Evalue	Score
sRNA01140	AE017283.1|:718284-718670(+)	81.82	55	52	5	2	1	52	104	156	5E-10	54
P. aeruginosa												
Query	Subject	PID	Length	QLen	Mismatch	Gaps	QStart	QEnd	SStart	SEnd	Evalue	Score
sRNA01158	NC_008463.1|:4749730-4750083(+)	85	40	65	4	1	26	65	316	353	0.0000002	46

R. sphaeroides												
Query	Subject	PID	Length	QLen	Mismatch	Gaps	QStart	QEnd	SStart	SEnd	Evalue	Score
sRNA00470	NC_007493.2|:1444045-1444119(-)	100	20	68	0	0	1	20	55	74	0.000001	40
sRNA00648	NC_007493.2|:2495804-2495967(+)	82.19	73	160	11	2	87	158	5	76	6E-15	72
sRNA01077	NC_007493.2|:694642-694848(+)	84.03	144	194	17	6	53	192	53	194	2E-35	148
S. avermitilis												
Query	Subject	PID	Length	QLen	Mismatch	Gaps	QStart	QEnd	SStart	SEnd	Evalue	Score
sRNA00627	NC_003155.4|:8583426-8583516(-)	81.03	58	79	8	2	19	73	21	78	0.000000001	54
S. meliloti												
Query	Subject	PID	Length	QLen	Mismatch	Gaps	QStart	QEnd	SStart	SEnd	Evalue	Score
sRNA00197	NC_003047.1|:3473144-3473210(+)	100	19	145	0	0	61	79	47	29	0.00004	38
sRNA00598	NC_003047.1|:259927-260032(+)	81.13	53	52	8	2	1	51	48	100	0.00000002	48
sRNA00822	NC_003047.1|:3522269-3522383(+)	86.21	29	85	4	0	51	79	75	103	0.00002	38
sRNA01156	NC_003047.1|:2291226-2291466(+)	79.03	62	121	13	0	60	121	174	235	1E-10	58
V. cholerae												

Query	Subject	PID	Length	QLen	Mismatch	Gaps	QStart	QEnd	SStart	SEnd	Evalue	Score
sRNA00237	NC_002505.1|:603874-603969(+)	86.67	30	180	4	0	148	177	62	91	0.00004	40
sRNA00627	NC_002505.1|:2722409-2722532(-)	96.77	31	79	1	0	41	71	83	113	7E-10	56
sRNA01158	NC_002505.1|:914289-914427(-)	88.24	34	65	4	0	32	65	104	137	0.00000008	48
sRNA01158	NC_002505.1|:914290-914371(-)	88.24	34	65	4	0	32	65	48	81	0.00000008	48
sRNA01158	NC_002505.1|:914290-914382(-)	88.24	34	65	4	0	32	65	59	92	0.00000008	48
sRNA01158	NC_002505.1|:914290-914405(-)	88.24	34	65	4	0	32	65	82	115	0.00000008	48

Table S2: List of putative sRNAs and their features.

sRNA Features													
sRNA ID	**Replicon**	**Start**	**End**	**Read Depth Coverage**	**Strand**	**Free Energy SS**	**Predicted -10 Promoter Position**	**Predicted -10 Promoter Sequence**	**Predicted -35 Promoter Position**	**Predicted -35 Promoter Sequence**	**Distance (nt) to closest predicted Rho-independent terminator**	**Type (I = Intergenic, AS = antisense, S = sense, PO = Partial overlap, POb = Partial overlap both ends)**	**Probability of being a sRNA (logistic regression)**
sRNA00627	Chromosome	3040756	3040834	84.41314386	+	-25.8						I	0.450323972
sRNA00822	Chromosome	397556	397640	9.920967056	-	-43.8	-36	ACGTTGACT	-56	TTGCGG		I	0.265373175
sRNA00687	Chromosome	3552725	3552827	18.88194351	+	-35.2						I	0.970380953
sRNA00526	Chromosome	2547301	2547384	10.13352946	+	-36.2	19	GGGTCGGAT	-6	TTGACG		I	0.720077833
sRNA00688	Chromosome	3554054	3554135	9.860754143	+	-30.4					42	I	0.709411666
sRNA00508	Chromosome	2274456	2274545	19.23337121	+	-33.2						I	0.512307618
sRNA01035	Chromosome	2197241	2197301	12.8390625	-	-13.8						I	0.251174671
sRNA01179	Chromosome	3161732	3161881	19.46585499	-	-45.5						POb	0.085629943
sRNA00798	Chromosome	260229	260433	149.0104329	-	-78.5					0	PO	0.762034684

sRNA01135	Chromosome	2939279	2939479	58.3216969	-	-75.1					0	PO	0.663962408
sRNA01109	Chromosome	2797053	2797137	9.85399701	-	-22.5						PO	0.068798749
sRNA00123	Chromosome	354470	354612	12.61973042	+	-34.1						PO	0.124161317
sRNA00598	Chromosome	2945375	2945426	146.6827896	+	-9.3						I	0.485176636
sRNA01208	Chromosome	3338581	3338766	18.85227873	-	-72.4	-102	GGCTACAGT	-126	TTTTCG	15	PO	0.951885483
sRNA01158	Chromosome	3017606	3017670	261.6568042	-	-18.7						I	0.841633809
sRNA01157	Chromosome	3017496	3017582	5886.197036	-	-6.7						I	0.61807136
sRNA01156	Chromosome	3017375	3017495	21880.0309	-	-42.6	-18	TGCTCAAAT	-35	TGGACG		I	0.821436335
sRNA01077	Chromosome	2555180	2555373	57.04507555	-	-68.9	-95	AGTTAACAT	-112	TTATCC	0	I	0.985539684
sRNA00648	Chromosome	3265930	3266089	265.3022446	+	-38.3	-19	GCCTATCTT	-39	TTGCCG	33	I	0.647510061
sRNA00470	Chromosome	1964312	1964379	42.22965279	+	-21.3	-20	GGCTATAAA	-40	TTGGAA		I	0.224665506
sRNA01140	Chromosome	2940745	2940796	12.53167631	-	-10.5						I	0.899098222
sRNA01139	Chromosome	2940515	2940610	24.35123337	-	-22.1						I	0.742355797
sRNA01141	Chromosome	2940843	2940921	23.74825725	-	-37.4						I	0.973604108

sRNA00489	Chromosome	2066779	2066982	30.61411979	+	-79.1						PO	0.275114403
sRNA00001	Chromosome	114	198	16.18580688	+	-34.1	-73	TGCTAGCGT	-94	TTGCAG		I	3.89E-006
sRNA00004	Chromosome	8074	8126	9.753193485	+	-8.8						PO	0.03700954
sRNA00005	Chromosome	9590	9689	10.81044694	+	-42.4						PO	0.089503145
sRNA00007	Chromosome	38487	38588	41.79395696	+	-44.6						POb	0.083077741
sRNA00011	Chromosome	56827	56878	17.64831187	+	-5.9	-50	TGGAAAAAT	-68	TAGCCT		I	0.065284331
sRNA00012	Chromosome	56957	57063	47.09162884	+	-57						AS	0.070938373
sRNA00014	Chromosome	91337	91431	9.909705445	+	-28.2	-18	GGGCAATTT	-38	TTGCCA		PO	0.395519597
sRNA00017	Chromosome	97656	97770	72.6041861	+	-49.8	-51	CCATACTAT	-73	TTGACG		PO	0.443141324
sRNA00019	Chromosome	97846	97958	10.59608779	+	-27.3						POb	0.045805879
sRNA00029	Chromosome	100790	100872	116.0685542	+	-23.1						AS	0.021518706
sRNA00030	Chromosome	116089	116140	10.06131407	+	-6.6						PO	0.015497588
sRNA00031	Chromosome	123487	123578	9.79167829	+	-36.2	-72	CGTTCTGAT	-92	CTGAAA		PO	0.279043802
sRNA00032	Chromosome	123764	123860	12.74089869	+	-17.7						PO	0.205104638

sRNA00069	Chromosome	231142	231244	12.72072157	+	-50.9					0	PO	0.416975978
sRNA00071	Chromosome	245524	245569	49.73314093	+	-14.5	-14	GCCTATATG	-34	TTGCAA	4	I	0.472240204
sRNA00081	Chromosome	265408	265515	40.61225374	+	-43.2						POb	0.055764361
sRNA00082	Chromosome	276808	276942	21.3864047	+	-42.2	-39	CGTTATCAC	-60	TTTCCC		PO	0.264239524
sRNA00084	Chromosome	285482	285602	13.80571509	+	-45.4	-99	CTGCACAAT	-116	TTGCGA		AS	0.086944784
sRNA00090	Chromosome	297435	297476	9.719512195	+	-14.4	-31	CGCCAACAT	-50	TGGACG	0	I	0.3623966
sRNA00091	Chromosome	300709	300780	15.00582742	+	-29.4	-20	TGGGATATT	-40	TTCACA		I	0.118201205
sRNA00093	Chromosome	317819	317991	9.600238884	+	-32.1					403	PO	0.255843492
sRNA00105	Chromosome	342677	342749	9.809744564	+	-34.7	-16	TGATAGAAG	-36	TCGACG	0	I	0.584759916
sRNA00110	Chromosome	346270	346346	10.71861664	+	-39						PO	0.114092988
sRNA00111	Chromosome	346376	346444	38.84185036	+	-23.5					434	I	0.287903228
sRNA00112	Chromosome	346874	346990	41.5557049	+	-48.1					0	PO	0.450916529
sRNA00118	Chromosome	351407	351489	10.14832728	+	-5	-35	CTGTAGAAT	-56	TTGCCT		I	0.6733938
sRNA00129	Chromosome	369655	369731	33.56535396	+	-25.9	-27	TGATTACAT	-45	TTGCCA		PO	0.195730995

sRNA00131	Chromosome	377404	377527	32.61743903	+	-36	-92	AAGTCAATT	-112	TTTAGG		PO	0.441311386
sRNA00133	Chromosome	398708	398755	11.65334361	+	-11.7					16	I	0.295407704
sRNA00137	Chromosome	426050	426127	9.747479802	+	-12.7	-42	TGGGATTAT	-63	GTGCTG		POb	0.034359893
sRNA00139	Chromosome	443687	443797	17.70995564	+	-54						POb	0.080713331
sRNA00141	Chromosome	445665	445750	40.90189179	+	-51					51	PO	0.567400334
sRNA00142	Chromosome	450573	450717	14.74926543	+	-28.1						PO	0.14633058
sRNA00143	Chromosome	451859	451989	10.27229107	+	-53.8	-48	CGGAAATAT	-65	TTTCCG		POb	0.19051001
sRNA00145	Chromosome	466182	466255	24.7265271	+	-34					0	I	0.450256481
sRNA00146	Chromosome	469113	469200	13.00027697	+	-28.2						PO	0.21505441
sRNA00153	Chromosome	533987	534024	22.85751351	+	-9.6	0	CGTTCTGAT	-24	TCGACT		PO	0.319005791
sRNA00154	Chromosome	534316	534404	10.04524501	+	-35.7						I	0.768300546
sRNA00156	Chromosome	543129	543311	10.65188837	+	-62	-23	GCCTATGAT	-43	TCGCAA		POb	0.238607232
sRNA00160	Chromosome	577032	577100	13.87829615	+	-31.1	-61	TGGTCCAAT	-82	TAGCCG		I	0.386905551
sRNA00161	Chromosome	581122	581205	21.39953827	+	-32.1	-14	AGTTTTCCT	-34	TGGTCG		I	0.643782922

sRNA00164	Chromosome	591406	591466	9.723246334	+	-13.4						POb	0.019390036
sRNA00165	Chromosome	593824	593876	9.781691713	+	-11.8	-4	CGTTAACCT	-25	CTGAGC		PO	0.059197831
sRNA00172	Chromosome	646749	646850	20.98656554	+	-48.5					24	PO	0.5148417
sRNA00173	Chromosome	660988	661058	71.43267523	+	-38.3						PO	0.117111145
sRNA00175	Chromosome	735334	735389	10.15194753	+	-14.5	-90	GGCTAAACC	-111	TTGATG		I	0.607305191
sRNA00178	Chromosome	736299	736413	31.26089774	+	-39.6	-70	TGGTATGAC	-91	TCGTCA		PO	0.220521966
sRNA00182	Chromosome	737598	737791	10.72662831	+	-74.9					83	PO	0.954604742
sRNA00187	Chromosome	740008	740126	72.84011519	+	-48.3						PO	0.096832046
sRNA00196	Chromosome	751911	751936	9.96	+	-0.2						I	0.024818098
sRNA00197	Chromosome	752359	752503	26.84886585	+	-56.1					32	PO	0.635968508
sRNA00198	Chromosome	755197	755355	37.35326464	+	-63.6	-48	GTGTAAGTT	-65	TTGTCC		PO	0.479632322
sRNA00201	Chromosome	757144	757311	10.03406905	+	-79.4					0	PO	0.825595896
sRNA00213	Chromosome	762679	762796	496.711228	+	-38.1						PO	0.123660041
sRNA00215	Chromosome	762923	762944	43.94592593	+	0	-27	GGTCATGAT	-44	GTGCTC		PO	0.031565171

sRNA00217	Chromosome	763035	763142	136.2480774	+	-45.3	-87	CTACAAAAT	-104	CTGAAA		PO	0.228157598
sRNA00223	Chromosome	765016	765166	61.30999194	+	-30.1	-87	CGGCATCCT	-107	GTGACG	100	PO	0.410095947
sRNA00225	Chromosome	765215	765260	9.837590099	+	-6.8					6	PO	0.301078934
sRNA00226	Chromosome	765305	765409	87.31823348	+	-48.9					0	I	0.616925542
sRNA00231	Chromosome	767299	767382	30.26901376	+	-37.6					4	PO	0.502926371
sRNA00234	Chromosome	793306	793403	11.19413024	+	-45.1						PO	0.128309397
sRNA00237	Chromosome	812910	813089	10.76193706	+	-64					0	PO	0.600334076
sRNA00241	Chromosome	828085	828153	13.67986781	+	-33.3	-100	GATCACAAT	-121	CTGACG		I	0.145057779
sRNA00242	Chromosome	834148	834298	64.0527127	+	-70.6						I	0.999511427
sRNA00243	Chromosome	842743	842850	12.9749904	+	-30.2	-91	AGGAAAAAT	-112	ATCACA		PO	0.382039248
sRNA00245	Chromosome	906010	906091	56.97689759	+	-27.7	-34	CGTTAGAAT	-53	TTGCAG		PO	0.090890453
sRNA00251	Chromosome	912264	912392	10.70298947	+	-60.1					0	PO	0.677139867
sRNA00254	Chromosome	935531	935619	14.47902415	+	-33.5	-15	GGTTAAGAT	-39	TTTTCG		I	0.271967756
sRNA00256	Chromosome	995379	995440	30.15777979	+	-24.2	-44	GGGTAGGTT	-68	TTGCTT	159	I	0.868352222

sRNA00257	Chromosome	995526	995630	71.16842427	+	-54.5					0	I	0.966071885
sRNA00258	Chromosome	996171	996275	11.28026855	+	-54.2					0	I	0.857067591
sRNA00260	Chromosome	996517	996609	10.95151226	+	-39.1	-27	GGCGATCAT	-46	TTGCCG	32	PO	0.670148815
sRNA00291	Chromosome	1009417	1009438	150.183156	+	0	-19	AACCATAAT	-37	TTCAAA	0	I	0.612152712
sRNA00292	Chromosome	1011051	1011123	9.708764508	+	-22.4	-75	GCGTATTGT	-100	TTGCTT		I	0.99885724
sRNA00293	Chromosome	1030090	1030187	9.908035965	+	-44.5	-7	ATCGAAAAT	-27	CTGACG		AS	0.080268051
sRNA00294	Chromosome	1040851	1040901	15.82780952	+	-20.7	-32	TATGACAAT	-49	TTGATT		PO	0.368007826
sRNA00295	Chromosome	1053192	1053232	889.8385856	+	-9.9	-21	TGTTCTCCT	-40	TTGAAA	31	I	0.999970957
sRNA00296	Chromosome	1090755	1090923	47.80920903	+	-61	-59	TTTTTTCCT	-79	TTGTCA		AS	0.14157452
sRNA00304	Chromosome	1123660	1123734	24.78031464	+	-28.6						AS	0.026207974
sRNA00306	Chromosome	1151033	1151115	26.21533675	+	-33.2	-20	TGCTAAATG	-40	TTGTTT		I	0.641636877
sRNA00307	Chromosome	1151220	1151321	13.71343117	+	-35	-89	GCTGAAAAT	-111	TTGGTA		PO	0.881816665
sRNA00310	Chromosome	1185594	1185628	15.51858274	+	-18					0	I	0.977789773
sRNA00320	Chromosome	1226623	1226691	190.4162552	+	-18.2	-30	GGGTGTCAT	-47	TTGATC		I	0.257974782

sRNA00324	Chromosome	1229193	1229278	9.721263081	+	-33.6						POb	0.057047273
sRNA00325	Chromosome	1262589	1262643	13.59290568	+	-26.9						I	0.050938629
sRNA00326	Chromosome	1285270	1285349	304.3835106	+	-41.5	-80	ACGCAAAAT	-105	TTCTGG		I	0.359785049
sRNA00331	Chromosome	1317044	1317106	31.81383828	+	-32.2						I	0.07736509
sRNA00332	Chromosome	1317116	1317203	9.794312232	+	-18.4						PO	0.086516218
sRNA00336	Chromosome	1318224	1318297	36.23020834	+	-44.7						I	0.102939548
sRNA00338	Chromosome	1340949	1341045	365.2779248	+	-25.4						I	0.999994831
sRNA00339	Chromosome	1349361	1349439	30.77376785	+	-21.1	-14	TGCCAATCT	-34	TGGACA	445	I	0.856259351
sRNA00340	Chromosome	1349725	1349860	23.35308533	+	-64.2					24	I	0.995776936
sRNA00342	Chromosome	1356292	1356365	23.49686864	+	-26.6						I	0.689691856
sRNA00343	Chromosome	1356459	1356531	9.794630082	+	-20.4						PO	0.827750501
sRNA00347	Chromosome	1418092	1418174	12.68986916	+	-28.6	-38	TCTTACAAT	-57	TTGACC	359	I	0.991937281
sRNA00352	Chromosome	1489672	1489785	11.02148809	+	-51.7	-5	CTATATCGT	-22	TTCACG	7	PO	0.700260149
sRNA00385	Chromosome	1573542	1573731	182.2220823	+	-46.3	-51	TGTCATCAT	-68	TTGTCG	33	I	0.934227526

sRNA00386	Chromosome	1573738	1573794	51.24208767	+	-29.4					0	I	0.96398069
sRNA00389	Chromosome	1579065	1579182	11.89057547	+	-43.3						POb	0.07951329 3
sRNA00390	Chromosome	1584428	1584599	14.93127641	+	-72.7					0	I	0.90977606
sRNA00392	Chromosome	1585164	1585286	327.4131765	+	-55.34					0	I	0.48574181 4
sRNA00394	Chromosome	1616725	1616843	21.25843882	+	-26.9	8	GGGTATTCT	-14	GTGCAA	438	PO	0.39065849 5
sRNA00395	Chromosome	1617406	1617486	13.18543646	+	-34.2						AS	0.03199934 5
sRNA00397	Chromosome	1645930	1646074	14.97261861	+	-53.9						PO	0.12137760 2
sRNA00402	Chromosome	1648508	1648656	10.76122451	+	-43.5	-3	CGGAATACT	-24	AAGACA		I	0.17307482 7
sRNA00403	Chromosome	1649088	1649171	12.59235671	+	-43.1						I	0.16204719 3
sRNA00411	Chromosome	1666039	1666130	56.16138227	+	-50.1					0	PO	0.63627684 3
sRNA00414	Chromosome	1710546	1710663	10.82133909	+	-36.8	-15	GCCTATATC	-35	TTGTCG	164	PO	0.67139130 9
sRNA00418	Chromosome	1721369	1721411	30.57782683	+	-6.3					20	I	0.75860112 7
sRNA00419	Chromosome	1723069	1723160	44.22770837	+	-31.6	-59	GGCGATAGT	-79	TTTCCT		I	0.81812189 7
sRNA00420	Chromosome	1723474	1723521	10.2571809	+	-9						S	0.02393048 2

sRNA00421	Chromosome	1723733	1723789	9.954523155	+	-24.1					289	S	0.155935299
sRNA00424	Chromosome	1751614	1751707	11.83937707	+	-27.7						PO	0.196992418
sRNA00425	Chromosome	1752021	1752045	10.07319444	+	4.5						PO	0.397259787
sRNA00427	Chromosome	1753536	1753582	40.68856948	+	-7					0	PO	0.311667466
sRNA00437	Chromosome	1792719	1792828	20.04888713	+	-48.9						PO	0.303089433
sRNA00438	Chromosome	1793068	1793153	11.15301626	+	-11.5						PO	0.453510829
sRNA00443	Chromosome	1794820	1794910	15.37535854	+	-35.1					0	PO	0.484553932
sRNA00445	Chromosome	1795790	1795845	9.796002861	+	-25.9	-40	CCGTAGGAT	-61	TTGCCG		PO	0.110437281
sRNA00446	Chromosome	1801470	1801620	43.43408167	+	-60.9						PO	0.142127132
sRNA00450	Chromosome	1812907	1812986	53.09822175	+	-18.7	19	CGGTCGAAT	-5	ATGCCT	0	PO	0.55715802
sRNA00452	Chromosome	1846001	1846059	15.77186088	+	-23.2						I	0.088296719
sRNA00458	Chromosome	1891632	1891727	9.679969314	+	-46.1						POb	0.061643584
sRNA00459	Chromosome	1891820	1891885	9.911597638	+	-19.5						AS	0.018904659
sRNA00463	Chromosome	1897509	1897693	23.15948308	+	-59.9						POb	0.137113793

sRNA00471	Chromosome	1968156	1968228	12.78040966	+	-14.4	-47	GTCTATAGG	-67	TTGCCA		I	0.366051529
sRNA00472	Chromosome	1998051	1998073	9.956818182	+	0	-67	GGAAACGAT	-86	TTGACC		I	0.0923676
sRNA00476	Chromosome	2005569	2005625	10.08392663	+	-30.7					0	PO	0.365937276
sRNA00477	Chromosome	2022493	2022550	12.83969298	+	-25.2	-75	GACCAAAAT	-100	TTTTCG		I	0.679921601
sRNA00478	Chromosome	2027832	2027900	33.68641863	+	-15.2	-25	CCATAAGAT	-47	TTGAAA		PO	0.499943959
sRNA00479	Chromosome	2027911	2028103	17.69855738	+	-65.2	-104	CCATAAGAT	-126	TTGAAA	377	S	0.571666341
sRNA00480	Chromosome	2028255	2028455	11.28526635	+	-62.52					25	PO	0.659814322
sRNA00485	Chromosome	2048984	2049050	343.0878985	+	-33.6					0	I	0.253491314
sRNA00487	Chromosome	2064274	2064447	15.81002459	+	-62.9					15	PO	0.618033558
sRNA00490	Chromosome	2066994	2067020	11.83799534	+	-4.6	-44	ATGCAAAAT	-69	ATGAAG		PO	0.053311068
sRNA00505	Chromosome	2162456	2162493	87.28225617	+	-13.3	-25	TGGTAAAGG	-45	TTGATT		I	0.564021283
sRNA00506	Chromosome	2180710	2180762	10.06458428	+	-27.4					0	I	0.515778028
sRNA00507	Chromosome	2209782	2209855	12.00605795	+	-31.8						I	0.228644916
sRNA00509	Chromosome	2279772	2279846	13.28055181	+	-24.4	-20	ACTTGTTAT	-39	GTGATT		I	0.380393114

sRNA00515	Chromosome	2322539	2322679	10.57586591	+	-53.4	-94	GGCTATGAC	-115	TTGCGG		AS	0.112972913
sRNA00517	Chromosome	2389741	2389841	9.931734041	+	-38.7	-42	TGCCACGCT	-62	TTGAAG		PO	0.520910601
sRNA00518	Chromosome	2468938	2469026	14.725127118	+	-41.9					3	I	0.816183217
sRNA00519	Chromosome	2497353	2497490	128.7589997	+	-57.3	-77	CTGAACACT	-97	ATGCGA	0	I	0.921919754
sRNA00520	Chromosome	2497538	2497607	10.45435964	+	-22.4	-4	AAGGAAACT	-25	ATGACG		PO	0.653299568
sRNA00521	Chromosome	2498788	2498926	46.98758737	+	-43.6	-25	TCCTATACT	-45	TTGTCG	216	PO	0.544940806
sRNA00522	Chromosome	2498975	2499174	14.312444452	+	-68.6					0	PO	0.737433949
sRNA00534	Chromosome	2575917	2576032	11.1629946	+	-34.2					0	PO	0.250871474
sRNA00539	Chromosome	2591933	2591966	9.775009353	+	-18.8						I	0.192983384
sRNA00540	Chromosome	2646061	2646128	9.910447761	+	-43.7					0	I	0.94579457
sRNA00541	Chromosome	2650234	2650349	15.34867438	+	-55.3						POb	0.11832681
sRNA00544	Chromosome	2660125	2660250	21.49764666	+	-46.7						PO	0.097198097
sRNA00547	Chromosome	2671417	2671578	10.90202887	+	-51.9	-66	GCGGACAAT	-86	TTAACC		PO	0.43741893
sRNA00554	Chromosome	2677469	2677641	18.191759555	+	-67.7					0	PO	0.816322429

sRNA00555	Chromosome	2677699	2677817	105.9759643	+	-55	-21	GCGTAAACT	-40	TGGTAT		I	0.47886618
sRNA00558	Chromosome	2716288	2716385	9.824258065	+	-27.4	-85	GGCTAGCCT	-105	TTGATC	304	PO	0.870621759
sRNA00559	Chromosome	2720229	2720377	165.5475997	+	-52.7	-58	CTTTACACT	-78	TTGCCG	318	PO	0.999508034
sRNA00561	Chromosome	2720625	2720708	96.93575454	+	-24.6					0	PO	0.327737282
sRNA00567	Chromosome	2722085	2722194	40.71843019	+	-46.4	-6	CCTTATAAT	-28	TTTCGG		PO	0.173303959
sRNA00568	Chromosome	2722211	2722361	30.30774925	+	-39.3						PO	0.107192348
sRNA00575	Chromosome	2770866	2770989	13.01434859	+	-28.9						I	0.864233734
sRNA00582	Chromosome	2789670	2789752	11.71859921	+	-26.6	-80	GGGTAAAAC	-105	TTGCCG	459	PO	0.618801475
sRNA00586	Chromosome	2825149	2825303	26.03093582	+	-62.5	-40	AGCCATGAT	-56	TTGTTA		PO	0.389069023
sRNA00589	Chromosome	2848652	2848825	10.09939173	+	-53.07	9	GGCGCAGAT	-11	CTGCAA		PO	0.210603264
sRNA00592	Chromosome	2878742	2878779	13.91814672	+	-9.1						I	0.181286289
sRNA00593	Chromosome	2893190	2893340	9.931587432	+	-44					284	POb	0.279325311
sRNA00599	Chromosome	2953225	2953317	114.0935812	+	-29.2					0	I	0.693060578
sRNA00603	Chromosome	2957800	2957882	12.87159944	+	-28.4						POb	0.047603929

sRNA00604	Chromosome	2958553	2958685	50.29255906	+	-42.1					12	PO	0.423966781
sRNA00606	Chromosome	2961685	2961766	39.07977871	+	-51						I	0.109227995
sRNA00614	Chromosome	2986973	2987096	16.50612716	+	-52.4	-99	ATTCATCAT	-116	TTTTCG		POb	0.187233041
sRNA00616	Chromosome	2994124	2994270	11.09669371	+	-24.47						PO	0.044571787
sRNA00617	Chromosome	2995453	2995499	10.28347332	+	-16.3	-34	CGGCAACAT	-54	GTGATC		PO	0.091144943
sRNA00619	Chromosome	3000278	3000327	10.18252011	+	-19.6	-18	AAGTAAGCT	-35	TCGACG	0	I	0.570837742
sRNA00621	Chromosome	3002907	3002998	23.17540937	+	-40.4						POb	0.072060638
sRNA00623	Chromosome	3005760	3005844	13.26624981	+	-45.4					0	PO	0.344235694
sRNA00630	Chromosome	3124914	3125002	24.44948143	+	-34.7	-14	GGGTATATT	-35	ATACCA		PO	0.954253692
sRNA00641	Chromosome	3232955	3233020	11.6123348	+	-32.7						POb	0.05529546
sRNA00649	Chromosome	3270994	3271074	25.46340295	+	-26.7	-31	CGCTATTCT	-51	TGGAAA		PO	0.140235679
sRNA00650	Chromosome	3292307	3292351	22.45813809	+	-5.5	-28	TGTTAAGAT	-48	TCGCAC		PO	0.226159545
sRNA00652	Chromosome	3296761	3296783	10.90151515	+	0	-14	TGCCAAGCT	-34	TTGACT		PO	0.055970615
sRNA00653	Chromosome	3307974	3308087	29.86066059	+	-41.6					329	I	0.681649797

sRNA00654	Chromosome	3308099	3308201	27.39491152	+	-29.2					215	PO	0.855459762
sRNA00656	Chromosome	3308391	3308454	10.08965617	+	-26.6					0	PO	0.28267608
sRNA00659	Chromosome	3321235	3321330	83.02120677	+	-52.9						PO	0.128897995
sRNA00661	Chromosome	3321694	3321779	12.78925475	+	-14.5	-25	CGGTCAAAT	-46	TTGTCG		PO	0.197946715
sRNA00663	Chromosome	3322000	3322095	10.44259083	+	-40.9	-45	GATGAAGAT	-62	TTTCCG		POb	0.091229964
sRNA00664	Chromosome	3323221	3323288	9.764053123	+	-27.8	-20	CCCTATTCC	-40	TTGACG	330	I	0.382633059
sRNA00665	Chromosome	3323520	3323598	9.723026637	+	-22.6					20	PO	0.216703324
sRNA00670	Chromosome	3379902	3380015	11.48201744	+	-54.5	-83	CGGTAGGAT	-106	TGGCAG		PO	0.825737924
sRNA00673	Chromosome	3428233	3428442	10.95956956	+	-83.8						PO	0.28966616
sRNA00677	Chromosome	3429329	3429422	26.19517664	+	-39.3						POb	0.069406152
sRNA00678	Chromosome	3436146	3436285	35.10504456	+	-56.3						PO	0.198106926
sRNA00679	Chromosome	3443989	3444060	9.979543779	+	-25.6	-31	CGTCATGCT	-51	TTGACC		PO	0.279931274
sRNA00680	Chromosome	3499192	3499290	10.64841406	+	-30.9	-22	ACTTACGGT	-43	TTTCCA		I	0.324461715
sRNA00684	Chromosome	3515945	3516046	10.51494231	+	-21.8	-16	CCTTAAACT	-36	TTTCGT		PO	0.246801081

sRNA00686	Chromosome	3516459	3516592	18.81923129	+	-45.5	-9	AGCTACAAT	-34	GGGCCG		POb	0.104238262
sRNA00690	Chromosome	3583152	3583297	11.00582036	+	-53.5						PO	0.133120394
sRNA00691	Chromosome	3583857	3583878	11	+	0	-14	GTCTAAGCA	-34	TTGCCC		PO	0.050520475
sRNA00693	Chromosome	3599072	3599123	10.88235294	+	-13.3	-54	TGTTGAAAT	-72	TCACCA		AS	0.027743179
sRNA00696	Chromosome	3606380	3606507	122.5780015	+	-41.8	18	TGCCATGGT	-6	TTGCAG		PO	0.338066551
sRNA00702	Chromosome	3611616	3611797	13.07405007	+	-52.7	-53	TGGTTCGAT	-72	TTGACA	3	S	0.638850328
sRNA00714	Chromosome	3631589	3631727	9.800122871	+	-39.8						PO	0.134168718
sRNA00715	Chromosome	3631772	3631853	20.57238151	+	-21.6	-71	GTGTAAACT	-94	TCGGCG		I	0.140618579
sRNA00717	Chromosome	3632085	3632190	12.92531716	+	-28.4						PO	0.065336801
sRNA00726	Chromosome	3670500	3670576	43.53780793	+	-28.2	-71	GTCTAGAGT	-91	ATGAAA	0	I	0.874913969
sRNA00732	Chromosome	3706449	3706528	18.92429792	+	-41.3	-24	TGCCACAGT	-44	GTGCAG		PO	0.137709867
sRNA00733	Chromosome	3714509	3714574	9.649274448	+	-35	-35	GGCCAGTAT	-56	TTCACG		I	0.376450647
sRNA00734	Chromosome	3718146	3718188	13.58627623	+	-9.5	-13	CAATAGACT	-33	TTGAGC		I	0.111726192
sRNA00741	Chromosome	28297	28384	9.870744115	-	-27.7						AS	0.025377689

sRNA00745	Chromosome	36613	36687	18.86648098	-	-7.2	8	TGATATTTT	-16	TTGGTG		I	0.058910396
sRNA00747	Chromosome	40221	40387	15.25122289	-	-56.6	-4	CTTTATACG	-27	TTGAAA		PO	0.235949401
sRNA00752	Chromosome	42203	42293	10.02387982	-	-35.2					14	PO	0.603870432
sRNA00754	Chromosome	74642	74688	9.832969616	-	-7.9	-21	AGGTATCGT	-40	TCGATT		PO	0.045761824
sRNA00755	Chromosome	83200	83260	26.9628132	-	-16.6						AS	0.01702767
sRNA00756	Chromosome	88327	88405	10.25788278	-	-31.8						POb	0.053594384
sRNA00762	Chromosome	111959	112044	10.26248779	-	-15.1						I	0.124789076
sRNA00763	Chromosome	130119	130238	11.53926624	-	-32.4						POb	0.054722861
sRNA00769	Chromosome	140939	141037	17.13026206	-	-32.2					4	PO	0.384712803
sRNA00771	Chromosome	141470	141500	10.08941799	-	-19.2						I	0.065933555
sRNA00778	Chromosome	182716	182788	9.949704227	-	-30.5	-32	CGGTAGTCT	-52	TAGACA		S	0.090164496
sRNA00780	Chromosome	191374	191447	9.935053919	-	-24.6	-17	ATCTGAAAT	-36	ATCCAG		PO	0.076074082
sRNA00782	Chromosome	213287	213404	11.08519354	-	-49.8						POb	0.078700389
sRNA00783	Chromosome	219880	219975	20.84346258	-	-26.5						POb	0.04453901

sRNA00786	Chromosome	222620	222708	89.90595605	-	-22	2	AGTTAGAAG	-19	TTGAGG		PO	0.107977215
sRNA00788	Chromosome	235136	235262	10.12798144	-	-54	-18	CGGCAATAT	-38	TTGCGC		AS	0.110718487
sRNA00789	Chromosome	235700	235768	12.87665036	-	-37					0	I	0.644904457
sRNA00790	Chromosome	251044	251125	10.06380671	-	-27.5	-105	AGCTGTAAT	-128	TTAAGT	18	PO	0.527254069
sRNA00794	Chromosome	254780	254868	10.97204885	-	-27.7	-87	TGATACATT	-108	ATGGCG		AS	0.047044206
sRNA00795	Chromosome	259049	259168	12.75985057	-	-42.3					19	PO	0.657093035
sRNA00802	Chromosome	262231	262262	13.96509615	-	-8.9	-90	TCGCAAACT	-110	TTGGCG		I	0.076270028
sRNA00803	Chromosome	262264	262295	12.03225806	-	-15.9	-57	TCGCAAACT	-77	TTGGCG		I	0.112531816
sRNA00806	Chromosome	296132	296228	9.780008687	-	-32.3	-23	GGCTAGAGA	-43	TTGTCG	460	PO	0.441656753
sRNA00808	Chromosome	304207	304370	16.36982097	-	-67.5	-56	CGGCAACAT	-73	TTGCGT		PO	0.308078201
sRNA00813	Chromosome	377245	377278	11	-	-6.4						I	0.060187528
sRNA00827	Chromosome	403682	403770	10.28687502	-	-22.3	-45	ACGCAAACT	-59	TTGACA		I	0.231997067
sRNA00828	Chromosome	453842	453967	10.19293874	-	-37.3						AS	0.035719213
sRNA00829	Chromosome	454002	454098	17.46952199	-	-48.5						AS	0.052927111

sRNA00830	Chromosome	454114	454140	10.08653846	-	-5.3						AS	0.011310526
sRNA00833	Chromosome	460574	460655	10.40736502	-	-34.3	-42	TGGCATGCT	-62	TTGCCA		I	0.363179395
sRNA00838	Chromosome	495907	496012	10.38503496	-	-31.8						POb	0.053594384
sRNA00839	Chromosome	497307	497366	11	-	-21.8						I	0.036576961
sRNA00840	Chromosome	519942	520023	10.02501992	-	-16.02						PO	0.03769628
sRNA00841	Chromosome	524969	525088	38.61250762	-	-40.9					385	PO	0.332637483
sRNA00845	Chromosome	526173	526314	170.037602	-	-48.7	-31	GGTGAAAAT	-48	TTTCTA		PO	0.18491892
sRNA00846	Chromosome	533354	533411	9.710933471	-	-6.2						PO	0.096227321
sRNA00848	Chromosome	552629	552700	27.17235458	-	-26.1						I	0.155858291
sRNA00850	Chromosome	579123	579223	14.69838879	-	-32.7						AS	0.030336625
sRNA00852	Chromosome	579915	579945	10.93333333	-	-0.2						POb	0.017438325
sRNA00853	Chromosome	586124	586279	11.04517403	-	-16.8					11	PO	0.247054134
sRNA00855	Chromosome	589329	589387	9.755257556	-	-20.1	-12	TCTTATTCT	-31	ATGATT	9	I	0.898333823
sRNA00856	Chromosome	596978	597078	12.88153736	-	-46						AS	0.048510262

sRNA00857	Chromosome	644526	644612	33.02765873	-	-25.8	15	ATTTAAAAAA	-7	CTGCAA	0	I	0.801250792
sRNA00858	Chromosome	644717	644832	42.63517482	-	-41.07					163	I	0.880950252
sRNA00863	Chromosome	648549	648673	89.34664654	-	-62.4	-5	AGGCATCCT	-22	TTCGCG	0	PO	0.853120382
sRNA00866	Chromosome	649597	649720	102.6266188	-	-55	-47	GCCCAAAAT	-66	TTGAAC		I	0.41884529
sRNA00868	Chromosome	702053	702206	9.534001449	-	-42.6	-28	GTATATCTT	-49	TTTACA		PO	0.301425055
sRNA00871	Chromosome	722977	723033	105.5466506	-	-33.4						I	0.113436403
sRNA00872	Chromosome	723049	723116	22.85384004	-	-21.1						I	0.1047664
sRNA00873	Chromosome	724289	724384	17.47126401	-	-46.9						PO	0.120026234
sRNA00876	Chromosome	725081	725165	39.99463564	-	-27.8						POb	0.046615017
sRNA00880	Chromosome	728113	728145	10.45845647	-	-6.7	-24	CGGCAGAAT	-48	TCGAAG		PO	0.042496766
sRNA00888	Chromosome	733201	733331	39.07341964	-	-36.2						POb	0.062405101
sRNA00891	Chromosome	745621	745682	20.3729519	-	-33.4	-72	CATTACATT	-92	TTGCCC		PO	0.102332978
sRNA00898	Chromosome	778421	778504	10.26281477	-	-22.1	-38	TCTAACAAT	-58	TTGCAC		PO	0.107209868
sRNA00900	Chromosome	814352	814502	10.2628073	-	-46.4	-14	CCCTATCCT	-34	TTGTCA		PO	0.122451675

sRNA00901	Chromosome	823087	823187	108.06487	-	-47.7					4	PO	0.46811721
sRNA00902	Chromosome	825306	825512	12.52409837	-	-75.8	-72	CGCTAAAAT	-90	AAGCCA		PO	0.39795348
sRNA00911	Chromosome	935529	935619	13.98476371	-	-34.9	-89	CGCGATGAT	-113	TTTCTT		I	0.429243621
sRNA00912	Chromosome	935627	935768	14.2926005	-	-49.7						PO	0.607551075
sRNA00913	Chromosome	954369	954467	14.32017694	-	-36.6						AS	0.034844434
sRNA00916	Chromosome	961239	961279	87.71395665	-	-6.7	-104	CGGCAAGAT	-124	TTGCGC		I	0.062488057
sRNA00918	Chromosome	1053744	1053781	179.9385225	-	-4.5	-49	AAGTATGCT	-74	ATGCCT	463	I	0.169274527
sRNA00922	Chromosome	1111620	1111775	22.82931742	-	-35.4						PO	0.078093803
sRNA00929	Chromosome	1116043	1116117	9.773975404	-	-34.4					115	S	0.278475301
sRNA00936	Chromosome	1217470	1217540	9.990896046	-	-20.2					60	PO	0.617381302
sRNA00938	Chromosome	1222626	1222704	14.27161777	-	-45.9						POb	0.068924122
sRNA00939	Chromosome	1230106	1230205	106.1817908	-	-52.8	-55	AGATATGTT	-77	TTGATC		PO	0.28375416
sRNA00944	Chromosome	1305159	1305278	26.10933964	-	-48.1	-11	GGGCATGCT	-31	CTGATG		POb	0.157391966
sRNA00945	Chromosome	1306068	1306143	10.73480951	-	-19.2	-103	GGTTAAAAT	-127	CTGTCT	179	S	0.270337651

sRNA00947	Chromosome	1330043	1330175	15.44032161	-	-35.7					0	I	0.998999537
sRNA00948	Chromosome	1346194	1346330	141.5848578	-	-30.6					0	I	0.968607275
sRNA00949	Chromosome	1346331	1346360	56.25733588	-	0	-106	CGTTCCAAT	-123	TTGAAT	123	I	0.982620674
sRNA00950	Chromosome	1445520	1445560	9.797368421	-	-14.9						AS	0.016013805
sRNA00952	Chromosome	1469753	1469867	10.22271336	-	-35	-69	CGGTAGAAT	-90	TTGGCA		S	0.106726256
sRNA00953	Chromosome	1473899	1474006	16.96994056	-	-48.5					470	PO	0.2610797
sRNA00955	Chromosome	1489876	1489986	41.59862177	-	-49.7					0	PO	0.703264787
sRNA00956	Chromosome	1490099	1490163	11.83375509	-	-14.1	-18	TCATATGCT	-39	TTCCCC	196	PO	0.425223452
sRNA00958	Chromosome	1501916	1501990	10.97684124	-	-14.7	-57	TGCCAAAGT	-76	GTGCCA		PO	0.073737863
sRNA00959	Chromosome	1584435	1584508	49.07754871	-	-37.8					0	I	0.7182218
sRNA00960	Chromosome	1586516	1586607	10.3496143	-	-32.1						I	0.331000095
sRNA00961	Chromosome	1614036	1614117	11.95046101	-	-30.5					63	PO	0.714099112
sRNA00962	Chromosome	1617291	1617386	36.9711847	-	-50.4						PO	0.09499135
sRNA00970	Chromosome	1724168	1724274	10.74776431	-	-34.9					11	POb	0.27504272

sRNA00973	Chromosome	1730522	1730680	12.45675347	-	-37.71						POb	0.065728904
sRNA00977	Chromosome	1744099	1744179	10.27635524	-	-47.6					0	PO	0.464105796
sRNA00984	Chromosome	1839788	1839892	9.944054768	-	-32.7	-32	GGGCATGAT	-54	ATGCGA	11	PO	0.738989467
sRNA00986	Chromosome	1854582	1854650	82.89550404	-	-27						PO	0.055640653
sRNA00987	Chromosome	1854682	1854800	83.11419101	-	-52.2						PO	0.256967997
sRNA00988	Chromosome	1861269	1861351	10.31036228	-	-17.7						PO	0.045534333
sRNA00991	Chromosome	1868786	1868856	9.566207881	-	-2.3						PO	0.019373907
sRNA00993	Chromosome	1906294	1906396	10.6657647	-	-38.3					18	PO	0.641148248
sRNA00995	Chromosome	1952992	1953080	12.01620715	-	-27.3	-13	CGCTAATCT	-33	GTGACA		PO	0.109202067
sRNA00996	Chromosome	1970115	1970192	35.25484482	-	-22.5					52	I	0.650915251
sRNA00998	Chromosome	1970996	1971088	50.3397641	-	-19.1						PO	0.067886393
sRNA00999	Chromosome	1974332	1974444	10.40161463	-	-45						PO	0.107546896
sRNA01010	Chromosome	2035198	2035304	10.22082531	-	-34.2						POb	0.058243979
sRNA01012	Chromosome	2040374	2040500	13.1472039	-	-49.7	-25	GTGCATTTT	-45	TTGGCG	390	I	0.737482976

sRNA01014	Chromosome	2057477	2057543	22.20818877	-	-14.5	-19	TGGTATCAC	-39	TTCCCC		I	0.910774493
sRNA01015	Chromosome	2083565	2083636	10.89970777	-	-30.3	-66	GCTTATTCT	-86	GTCATA		I	0.981002802
sRNA01018	Chromosome	2116847	2116944	12.09368269	-	-20.9	-64	CGTCAGCAT	-84	TTGAGC		AS	0.036555513
sRNA01019	Chromosome	2117036	2117117	17.31022677	-	-40.6					0	PO	0.395694801
sRNA01023	Chromosome	2136162	2136265	10.19411184	-	-25.3	-75	TGCGATGCT	-95	TTTTCG		I	0.819197251
sRNA01024	Chromosome	2156849	2156915	46.99178194	-	-37.2						I	0.076812836
sRNA01025	Chromosome	2157784	2157810	9.883012821	-	-1.1	-36	TGGTATGCT	-55	CTGATG		I	0.718064778
sRNA01026	Chromosome	2157829	2157939	114.0723263	-	-23.7						I	0.729935991
sRNA01027	Chromosome	2157992	2158023	10.29406986	-	0.5						I	0.809654092
sRNA01028	Chromosome	2158037	2158148	11.76610139	-	-28.3						I	0.914087199
sRNA01029	Chromosome	2167726	2167778	35.65158438	-	-26	-83	GCACAAAAT	-101	TTGCGC	269	I	0.954998399
sRNA01030	Chromosome	2169155	2169297	10.11514012	-	-43.5	-23	GGCTATCAG	-43	TTGCAT		PO	0.412519011
sRNA01034	Chromosome	2182669	2182731	10.01612903	-	-33.6	-22	GCATAAGAT	-46	TTCACC		AS	0.055719065
sRNA01039	Chromosome	2246036	2246166	11.48678075	-	-64.6						POb	0.158837865

sRNA01040	Chromosome	2258669	2258706	178.9385225	-	-4.5	-49	AAGTATGCT	-74	ATGCCT		I	0.050950481
sRNA01042	Chromosome	2268622	2268733	11.4233425	-	-52.5					63	PO	0.994258055
sRNA01050	Chromosome	2287527	2287623	10.71375333	-	-23.9						PO	0.110867806
sRNA01052	Chromosome	2296013	2296102	9.675655008	-	-31	-97	CTTTATGAG	-118	TTGAAA		PO	0.21144073
sRNA01053	Chromosome	2316656	2316749	12.29363242	-	-30.1						I	0.509690908
sRNA01060	Chromosome	2337039	2337099	15.23622909	-	-29.2						I	0.418297868
sRNA01062	Chromosome	2352621	2352699	10.94381373	-	-26.8	-8	CGTTGAAAT	-29	TCGACA		PO	0.063578751
sRNA01064	Chromosome	2472987	2473119	13.1208855	-	-61.7	-2	CAAGAAACT	-22	GTGCCG		POb	0.234386826
sRNA01067	Chromosome	2473527	2473681	15.80215883	-	-48.9						PO	0.205914453
sRNA01069	Chromosome	2474598	2474686	52.7739808	-	-37.6	-54	CCGCATAAG	-74	TTGACA		I	0.38618206
sRNA01070	Chromosome	2486449	2486478	9.895274585	-	-8.8					8	I	0.402145933
sRNA01071	Chromosome	2495370	2495419	10.05342552	-	-13.9	-94	GCGTAGAAT	-116	TTGATC		PO	0.089691847
sRNA01072	Chromosome	2516172	2516226	10.53715273	-	-26.2	-34	TGCTATGAG	-59	TTGTTG		I	0.068830018
sRNA01074	Chromosome	2534847	2534930	39.98875689	-	-30.8					101	AS	0.156859731

sRNA01078	Chromosome	2629296	2629392	10.86472006	-	-49.3	-58	CGACAACAT	-78	ATGAAG		PO	0.210035755
sRNA01079	Chromosome	2637447	2637580	681.7259625	-	-53.9					0	PO	0.486472869
sRNA01082	Chromosome	2637704	2637726	20.77045455	-	-5.1	-102	TGGCACGCT	-121	TTGATT	233	I	0.544172538
sRNA01083	Chromosome	2637734	2637804	44.99927799	-	-25.2	-92	TGCAAAAAT	-112	TTCCAT	263	I	0.685784885
sRNA01084	Chromosome	2641207	2641282	10.10349359	-	-11.2					126	PO	0.765225988
sRNA01094	Chromosome	2755099	2755197	38.61691808	-	-39.7	-72	TGCCAATAT	-91	TTGACA	311	I	0.862651447
sRNA01095	Chromosome	2761480	2761546	10.99627746	-	-17.6						AS	0.017653257
sRNA01100	Chromosome	2779945	2780064	47.58099595	-	-31.1						PO	0.072391462
sRNA01102	Chromosome	2786403	2786487	10.12635155	-	-27.5						I	0.350299735
sRNA01103	Chromosome	2787356	2787438	57.41189458	-	-37	-66	AGGCATCCT	-86	TTCCTG		PO	0.202415288
sRNA01110	Chromosome	2808461	2808557	16.02584877	-	-55						I	0.131790112
sRNA01114	Chromosome	2827155	2827227	14.83690716	-	-18.1						PO	0.035780111
sRNA01115	Chromosome	2827541	2827627	10.19990867	-	-35.2						S	0.060290953
sRNA01119	Chromosome	2841983	2842058	43.08587508	-	-24.4	-45	AGGGAACCT	-66	TTGACA		I	0.45243731

sRNA01125	Chromosome	2872705	2872785	21.54537868	-	-36.5						POb	0.063052722
sRNA01128	Chromosome	2877320	2877426	10.03095519	-	-33.4						AS	0.031101881
sRNA01129	Chromosome	2894220	2894289	24.73506294	-	-35.8	-87	ATTCATCCT	-106	TTGAAA		I	0.462172427
sRNA01131	Chromosome	2910836	2910917	12.52066714	-	-10.1	15	TGGGAAACT	-6	TTCCCT		PO	0.999986254
sRNA01132	Chromosome	2911027	2911108	10.83258906	-	-8.5	-13	TGGTGCACT	-31	TCGACA		PO	0.073908752
sRNA01136	Chromosome	2939525	2939648	20.78179935	-	-36.7					0	PO	0.421452787
sRNA01138	Chromosome	2940280	2940397	10.08506083	-	-35.7						PO	0.105208863
sRNA01143	Chromosome	2970400	2970489	115.3783117	-	-34.3	-56	CGGGAATCT	-76	ATCACG	0	I	0.649954253
sRNA01147	Chromosome	2981897	2981985	58.03876037	-	-34.8	-74	CGCGAAGCT	-94	TTCACG	0	I	0.830199518
sRNA01149	Chromosome	2982108	2982185	9.961038961	-	-33.5					0	I	0.899012707
sRNA01150	Chromosome	2992095	2992195	10.18689126	-	-24.3						PO	0.073155481
sRNA01151	Chromosome	3005800	3005968	15.84254854	-	-66.7					10	PO	0.662976027
sRNA01155	Chromosome	3017133	3017254	16.91027857	-	-25.09	-13	CCCTATGCT	-32	TTTTCT		PO	0.124648867
sRNA01159	Chromosome	3020072	3020154	13.10075352	-	-22.1						POb	0.038148102

sRNA01162	Chromosome	3072914	3073031	18.30069129	-	-22.7					23	PO	0.502019706
sRNA01164	Chromosome	3090687	3090778	19.33389116	-	-29.2	-58	TTGCAAAAT	-79	TTGAAA	227	I	0.549448042
sRNA01165	Chromosome	3113069	3113158	10	-	-27.9						I	0.061491667
sRNA01166	Chromosome	3116435	3116533	12.93589856	-	-43.4						POb	0.079782446
sRNA01167	Chromosome	3116714	3116837	10.90771042	-	-30						PO	0.052046138
sRNA01168	Chromosome	3135679	3135755	17.50745038	-	-29.1						I	0.845953402
sRNA01175	Chromosome	3159386	3159518	10.18982236	-	-62.2					305	POb	0.419985909
sRNA01183	Chromosome	3179546	3179644	50.34017413	-	-21.5	-24	TGTTAGGTT	-44	TTCACA		PO	0.102490509
sRNA01188	Chromosome	3192617	3192718	10.02646531	-	-37.7					136	PO	0.428043249
sRNA01206	Chromosome	3310280	3310387	16.5439211	-	-32.9	-72	AGGTAGCAT	-93	TTTCGG		PO	0.290967756
sRNA01207	Chromosome	3311632	3311749	19.94404507	-	-48.8	-19	GGCTCATAT	-39	TTGCCT		PO	0.191310682
sRNA01209	Chromosome	3338805	3338874	9.921634247	-	-30.4	-45	TCATATCCT	-67	TTGCAA	239	I	0.479590202
sRNA01214	Chromosome	3360664	3360703	13.10000341	-	-6.9						PO	0.029046939
sRNA01215	Chromosome	3379503	3379576	222.670663	-	-29.7						AS	0.027258771

sRNA01216	Chromosome	3379598	3379679	34.82147795	-	-18.6	-95	CGCTGAACT	-120	CTGCCA		I	0.152758013
sRNA01224	Chromosome	3421554	3421644	14.74475003	-	-23.4						PO	0.046349784
sRNA01226	Chromosome	3424323	3424459	15.46481751	-	-52.8	-37	GGCTATGCT	-56	TTGACC		PO	0.296850599
sRNA01228	Chromosome	3507826	3507924	10.04756922	-	-29.9						PO	0.057306695
sRNA01231	Chromosome	3511824	3511920	54.00036228	-	-39.3					0	PO	0.394858076
sRNA01232	Chromosome	3515451	3515526	10.42938302	-	-21.8						POb	0.02964925
sRNA01234	Chromosome	3525098	3525255	12.89510748	-	-57.2						POb	0.125800832
sRNA01237	Chromosome	3533238	3533295	10.98277512	-	-8						I	0.876270554
sRNA01238	Chromosome	3533438	3533474	9.916666667	-	-17.4	-16	TGCTTTCAT	-33	TTTACA		I	0.984724019
sRNA01280	Chromosome	3594731	3594840	16.03296971	-	-38.3						PO	0.124023155
sRNA01281	Chromosome	3601331	3601456	25.86935391	-	-58.3						POb	0.130310151
sRNA01285	Chromosome	3650881	3650982	10.24012359	-	-27						AS	0.024749678
sRNA01286	Chromosome	3655525	3655669	11.28567734	-	-72.8						POb	0.203296216
sRNA01297	Chromosome	3735839	3735932	17.2193372	-	-41	18	GTGTAATAT	1	TGCACT		POb	0.123952521

sRNA01299	pRCB133	104798	104945	75.92543021	+	-42.6	-63	GGCGACAAT	-81	TTGCAA		PO	0.216655401
sRNA01300	pRCB133	120827	120980	23.19821539	+	-49.8						S	0.098827919
sRNA01301	pRCB133	121129	121195	10.84077116	+	-24.4						S	0.041370572
sRNA01302	pRCB133	32125	32146	10	-	-3.7	-86	GGGTAACCA	-106	TTGCCG		I	0.677766945
sRNA01303	pRCB133	46729	46791	17.74011513	-	-20.5						I	0.36390787
sRNA01304	pRCB133	72168	72233	30.82004841	-	-17.5						S	0.032411803
sRNA01306	pRCB133	99509	99547	11.05785124	-	-13.5	-18	TGATACATC	-37	TTGACT		I	0.090144357

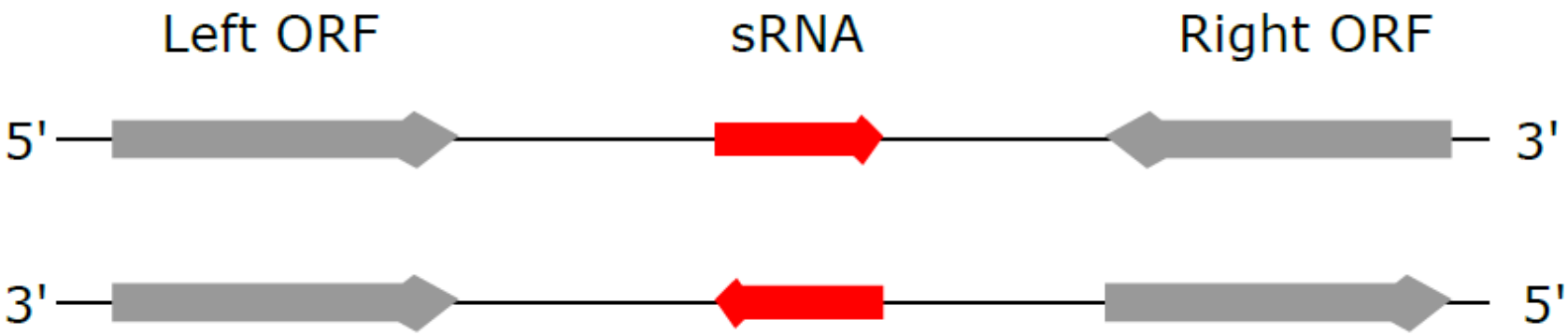

Figure S1: Neighboring ORFs. Left ORFs are located at the 5’ end of a sRNA on the forward strand or at the 3’ end of a sRNA on the reverse strand. Right ORFs are located at the 3’ end of a sRNA on the forward strand or at the 5’ end of a sRNA on the reverse strand. The first base in a replicon is numbered 1.

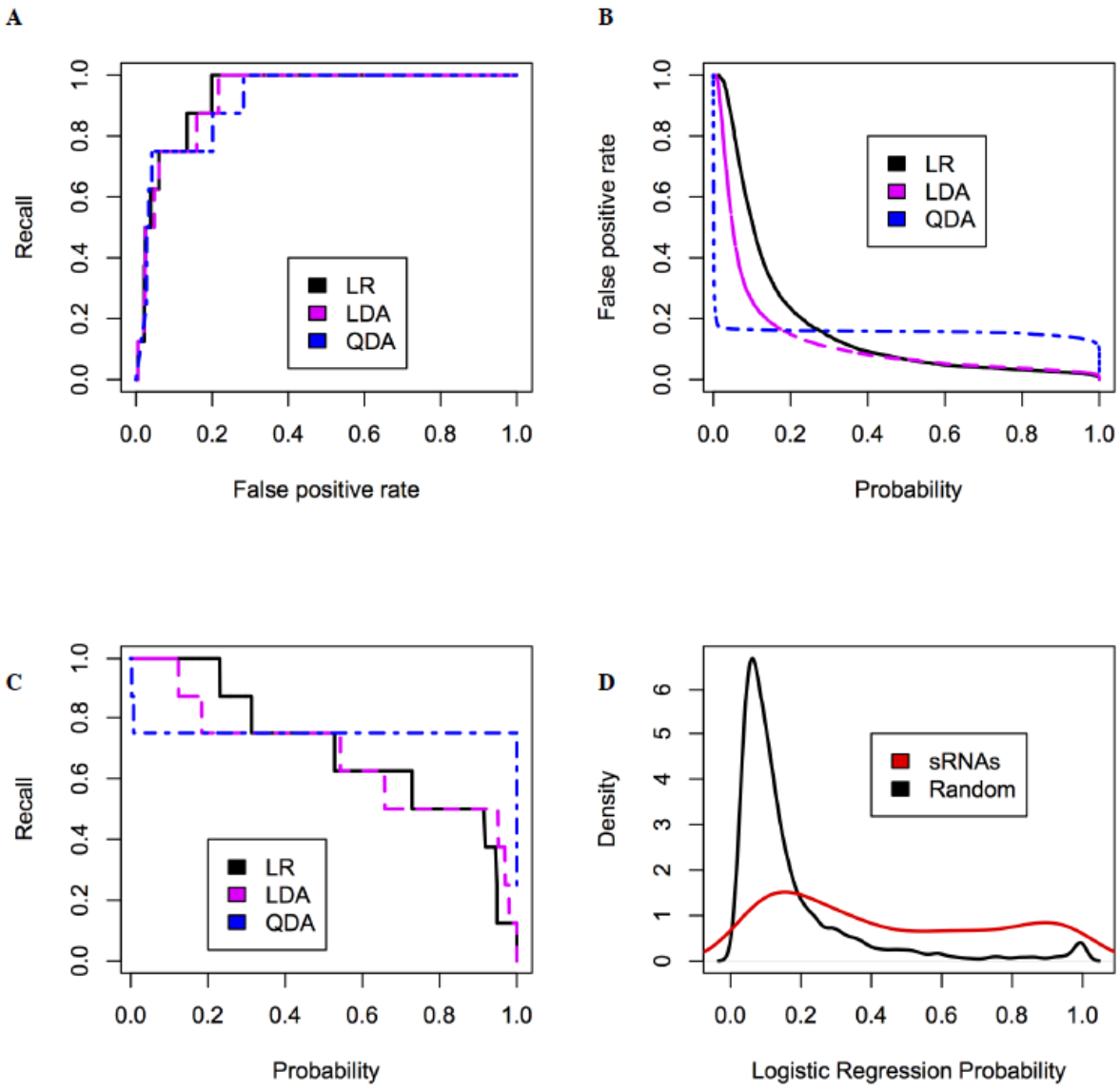

Figure S2: Performance comparison of three classifiers to quantify the probability of a genomic sequence being a sRNA. A. Classifiers' recall as a function of their false positive rate. B. Classifiers' false positive rate as a function of the probability threshold used to classify genomic sequences as "actual sRNAs". C. as b but showing the classifiers' recall. D. Distribution of the probabilities assigned to random genomic sequences and putative sRNAs.

Table S3: CopraRNA's predicted targets of sRNA00295.

FDR	p-value	NC_014034	NC_000913	NC_000962	NC_000964	NC_002505	NC_003028	NC_003047	NC_003062	NC_003450
0.37302072939	0.00004878339	rcap_rcc01474(N/A|-1.90736|0.9445494|43|60|21|36|GeneID:9004297)	b0112(aroP|-10.28490|0.05733184|282|298|24|37|GeneID:946018)	rv2127(ansP1|-8.50468|0.1570987|240|273|2|33|GeneID:887715)	bsu09460(ctrA|-8.42508|0.06274043|234|260|5|30|GeneID:939746)	vca1062(potE|-11.45910|0.01708510|62|87|15|39|GeneID:2612040)	sp_1001(N/A|-4.55148|0.5769226|105|116|13|22|GeneID:931514)	smc02616(N/A|-7.35686|0.229062|16|52|10|37|GeneID:12317724)	atu2143(N/A|-10.56380|0.05308207|278|287|29|38|GeneID:1134181)	ncgl1062(N/A|-10.58690|0.03540691|224|251|9|35|GeneID:1019092)
0.37302072929	0.00005461775	rcap_rcp00009(N/A|-4.75044|0.6774198|1|7|2|8|GeneID:9006433)	b1320(ycjW|-11.14130|0.03937623|101|110|29|38|GeneID:945875)	rv3575c(N/A|-0.75462|0.9905982|202|208|14|20|GeneID:888084)	bsu30260(msmR|-7.59708|0.08672543|288|295|33|40|GeneID:937257)	vc1557(N/A|-9.64276|0.04278154|173|187|4|18|GeneID:2613936)	sp_1854(N/A|-9.25142|0.07008954|250|261|2|13|GeneID:930980)	sm_b20967(N/A|-13.73650|0.01077097|91|113|5|37|GeneID:1237423)	atu5078(N/A|-10.78360|0.04791822|66|74|33|41|GeneID:1136851)	ncgl1203(N/A|-5.29413|0.3115785|100|106|31|37|GeneID:1019233)
0.37302072929	0.00007795515	rcap_rcc00101(N/A|-10.00750|0.1673472|217|229|11|23|GeneID:9002930)	b3542(dppC|-7.90360|0.1543839|232|251|11|37|GeneID:948064)	rv3664c(dppC|-5.79043|0.4260534|230|236|10|16|GeneID:885483)	bsu11450(oppC|-3.20011|0.494876|12|20|5|13|GeneID:936396)	vca0589(N/A|-8.23460|0.08463534|128|136|1|9|GeneID:2612571)	sp_1889(N/A|-9.98262|0.04557807|74|99|13|39|GeneID:930933)	sma1374(N/A|-10.90040|0.04784883|45|73|5|37|GeneID:1235789)	atu6075(dfpC|-9.14242|0.1006535|33|41|6|14|GeneID:1137398)	ncgl2240(N/A|-3.71286|0.5338234|288|295|19|26|GeneID:1020273)
0.37302072929	0.00008387 2		b2587(kgtP|-12.37750|0.02256852|271|281|29|39|GeneID:947069)	rv3331(sugI|-12.97340|0.02310004|92|119|2|31|GeneID:887504)	bsu40840(yyaJ|-10.29380|0.03109372|190|215|8|35|GeneID:937884)	vca0669(N/A|-5.76582|0.25444|249|273|2|31|GeneID:2612471)		smc04362(N/A|-8.52886|0.1421740|70|78|26|34|GeneID:1233743)	atu2154(tphA|-11.03980|0.04248412|244|273|5|37|GeneID:1134192)	ncgl2922(N/A|-7.05200|0.1564781|116|124|30|38|GeneID:1020966)
0.41394271	0.00011634318	rcap_rcc02606(N/A|-13.68460|0.04653805|273|283|26|37|GeneID:9005425)	b3193(mlaD|-18.88780|0.0009497701|149|190|5|40|GeneID:947712)	rv3498c(mce4B|-11.36850|0.04696510|23|48|1|30|GeneID:888349)		vc2518(N/A|-13.24550|0.006696776|157|192|5|33|GeneID:2615182)				
0.447	0.001	rcap_rcc00024(gcdH|-	b0039(caiA|-10.14600|0.060	rv0400c(fadE7|-	bsu24150(mmgC|-			smc01639(N/A|-	atu4418(gcdH|-15.17480|0.005	ncgl0283(N/A|-

32 59 27	50 86 88	9.90725|0.1728917|93|107|25|37|GeneID:9002853)	88265|240|263|11|31|GeneID:949064)	13.16770|0.02117566|188|214|2|27|GeneID:886427)	3.41599|0.4577109|187|195|28|36|GeneID:938664)			10.72170|0.05221026|97|107|28|37|GeneID:1233917)	247319|70|82|26|37|GeneID:1136292)	3.66556|0.5414947|285|291|6|12|GeneID:1021351)
0.5 60 61 45 58	0.0 02 20 59 03	rcap_rcc01400 (N/A|-16.16440|0.01835227|172|196|14|40|GeneID:9004223)	b2469(narQ|-1.75975|0.8775566|66|73|15|22|GeneID:946948)	rv0845(N/A|-7.75491|0.2111492|173|200|2|30|GeneID:885218)	bsu08290(yfiJ|-7.71947|0.08263623|59|66|2|9|GeneID:936171)	vca0675(N/A|-4.76034|0.3768172|125|147|18|35|GeneID:2612440)	sp_0386(N/A|-6.47416|0.2850912|220|233|5|19|GeneID:930303)	sma0204(N/A|-7.25296|0.2384100|135|170|2|37|GeneID:1235333)	atu5266(N/A|-4.43126|0.560873|32|38|2|8|GeneID:1137039)	ncgl1935(N/A|-10.88550|0.03117655|109|116|30|37|GeneID:1019967)
0.7 24 02 62 24	0.0 03 38 39 22	rcap_rcc00616 (acrB|-10.56890|0.1390497|260|289|10|41|GeneID:9003445)	b0462(acrB|-8.07576|0.1441835|268|276|27|35|GeneID:945108)			vc0914(N/A|-7.02303|0.1480470|57|64|2|9|GeneID:2614134)		smc01457(N/A|-13.31760|0.01359738|232|249|20|37|GeneID:1233806)	atu2483(mexF|-8.30737|0.1436961|125|136|8|22|GeneID:1139913)	
0.7 24 02 62 24	0.0 04 23 80 48	rcap_rcc00505 (N/A|-5.31645|0.6074442|34|44|11|22|GeneID:9003334)	b0107(hofB|-10.15960|0.06052614|279|291|28|40|GeneID:947481)	rv3659c(N/A|-4.73086|0.5828222|140|146|14|20|GeneID:885627)		vc2424(N/A|-7.97107|0.09582078|286|292|5|11|GeneID:2612966)	sp_2053(N/A|-6.66222|0.2626085|199|210|15|25|GeneID:930328)	smc02820(cpaF1|-7.43872|0.2218944|95|105|26|37|GeneID:1231775)	atu6177(virB11|-18.50620|0.0007912066|136|148|25|37|GeneID:1137500)	ncgl0297(N/A|-10.00350|0.04539199|26|45|19|37|GeneID:1021365)
0.7 24 02 62 24	0.0 05 89 67 51	rcap_rcc01291 (N/A|-1.65701|0.9565823|203|209|7|13|GeneID:9004118)	b1372(stfR|-8.40210|0.1264592|188|195|32|39|GeneID:945207)	rv1452c(PE_PGRS28|-5.44425|0.4749327|233|240|24|31|GeneID:886595)		vc1764(N/A|-5.37767|0.2975771|195|215|9|26|GeneID:2613644)		smc01910(N/A|-11.96470|0.02799426|138|144|32|38|GeneID:1232907)	atu1348(N/A|-3.10835|0.7553027|12|31|5|25|GeneID:1138836)	
0.7 24 02 62 24	0.0 07 51 47 94	rcap_rcc02771 (N/A|-11.96320|0.08625916|89|97|29|37|GeneID:9005590)	b1618(uidR|-4.74701|0.4684992|229|245|6|21|GeneID:946150)	rv0144(N/A|-13.44270|0.01871629|203|224|1|30|GeneID:886854)	bsu08370(yfiR|-7.94145|0.07573423|150|157|2|9|GeneID:936176)	vc1746(N/A|-5.74020|0.2571288|153|182|13|37|GeneID:2613751)		smc04348(N/A|-5.05769|0.4993243|44|53|26|37|GeneID:1233735)	atu1624(N/A|-5.03222|0.4735046|26|46|4|34|GeneID:1133662)	ncgl1483(N/A|-6.56217|0.1907272|113|121|32|40|GeneID:1019513)
0.7 24 02	0.0 08 27	rcap_rcc00194 (N/A|-10.33290|0.15091|122|130|19|	b1990(ldtA|-5.99481|0.313991|122|130|19|		bsu23940(yqjB|-4.82760|0.264	vca0058(N/A|-7.28365|0.1316158|202|242|11		smc01575(N/A|-11.36800|0.037	atu3331(N/A|-9.27758|0.0948 7738|243|274|5|	

6224	822	3922|127|147|12|32|GeneID:9003023)	27|GeneID:945273)		7284|27|35|11|19|GeneID:938684)	|38|GeneID:2612134)		93983|73|89|25|40|GeneID:1233874)	37|GeneID:1135205)	
0.72402624	0.009264176			rv0938(ligD|-10.17830|0.07847485|7|35|1|31|GeneID:885561)	bsu20500(ligB|-3.95748|0.3732528|160|166|7|13|GeneID:939992)	vc1542(N/A|-7.07337|0.1447390|33|39|4|10|GeneID:2613921)		sm_b21044(N/A|-7.84377|0.1889374|5|13|6|14|GeneID:1236947)	atu5055(ligD|-7.18237|0.2256674|167|186|22|41|GeneID:1136828)	
0.72402624	0.009551294	rcap_rcc02677(aspC2|-5.24831|0.6158966|178|209|8|37|GeneID:9005496)	b0600(ybdL|-15.20900|0.005965756|55|75|6|31|GeneID:945211)	rv0858c(dapC|-6.09945|0.3849553|162|187|2|30|GeneID:885784)	bsu31400(alaT|-1.84573|0.7538462|155|161|20|26|GeneID:938846)		sp_0035(N/A|-11.39660|0.01851741|13|31|6|23|GeneID:929782)	smc02262(N/A|-7.74648|0.1964785|157|165|6|14|GeneID:1232203)	atu0529(aspB|-14.33550|0.008202494|279|289|26|36|GeneID:1132567)	ncgl0780(N/A|-5.97130|0.2408359|116|122|29|35|GeneID:1018809)
0.72402624	0.009915022	rcap_rcc01546(N/A|-4.87232|0.6624453|72|82|8|18|GeneID:9004369)	b1701(fadK|-8.25193|0.1343610|255|261|1|7|GeneID:946213)	rv0119(fadD7|-8.72169|0.143919|231|239|10|18|GeneID:886896)	bsu10270(yhfL|-9.72661|0.03830910|190|208|17|34|GeneID:939308)	vc1985(N/A|-4.60078|0.3993573|284|291|19|26|GeneID:2613489)		sm_b20650(N/A|-12.73250|0.01868688|184|192|29|37|GeneID:1237690)	atu0405(fadD|-9.41820|0.0891828|5|278|285|6|13|GeneID:1132443)	ncgl0279(N/A|-4.10685|0.4717496|30|36|30|36|GeneID:1021346)
0.72402624	0.010787739		b2128(yehW|-5.16499|0.4126237|104|134|13|40|GeneID:949028)		bsu33720(opuBB|-5.66891|0.1883557|286|299|4|18|GeneID:936225)		sp_1860(N/A|-7.54209|0.1741185|126|164|1|23|GeneID:930974)	sm_b21147(N/A|-6.98592|0.2637349|251|262|4|14|GeneID:1237144)	atu0198(N/A|-14.29700|0.008370022|168|183|21|35|GeneID:1132236)	
0.72402624	0.011695966		b1175(minD|-9.19911|0.09107019|4|25|5|37|GeneID:945741)		bsu27990(minD|-9.37503|0.04368393|235|271|6|37|GeneID:937499)	vc1960(N/A|-8.70306|0.0676|7636|253|262|27|37|GeneID:2613464)		sm_b21523(minD|-12.52980|0.02082082|107|123|23|41|GeneID:1237685)	atu3248(minD|-6.91663|0.2496795|229|241|25|37|GeneID:1135122)	
0.7240	0.01172	rcap_rcc02781(N/A|-7.31137|0.3767248|11|33|7|33|	b1301(puuB|-10.21240|0.0591604|18|42|10|					smc00761(ordL2|-8.62127|0.1366696|70|78|6|14|	atu0601(ordL|-7.45027|0.2033373|294|300|2|8|	

62 24	23 48	GeneID:90056 00)	37|GeneID:945 072)					GeneID:12323 44)	GeneID:113263 9)	
0.7 24 02 62 24	0.0 11 88 70 35	rcap_rcc03109 (N/A|- 12.27370|0.077 33092|23|30|28 |35|GeneID:90 05925)				vc0392(N/A|- 11.68530|0.015 20107|227|258| 6|37|GeneID:2 614989)		sma1495(N/A|- 13.65330|0.011 28528|1|26|11| 37|GeneID:123 5850)	atu3217(aspC|- 7.30703|0.2150 459|109|115|30| 36|GeneID:113 5091)	
0.7 24 02 62 24	0.0 12 80 20 55	rcap_rcc02197 (N/A|- 2.70047|0.8927 805|128|141|12 |32|GeneID:90 05016)	b0064(araC|- 8.18847|0.1378 292|101|108|2| 9|GeneID:9447 80)		bsu07010(yes S|- 12.07380|0.01 656096|111|12 0|26|35|GeneI D:936075)	vc1825(N/A|- 7.86523|0.1006 823|262|271|33 |41|GeneID:26 13579)			atu4297(N/A|- 3.32630|0.7247 146|45|51|19|25| GeneID:113617 1)	
0.7 24 02 62 24	0.0 12 88 99 77	rcap_rcc00228 (N/A|- 5.94546|0.5302 069|9|27|8|36|G eneID:9003057)		rv2969c(N/A|- 8.26990|0.172 5522|209|215| 14|20|GeneID: 888481)	bsu33480(bdb D|- 5.39041|0.210 8953|248|255| 34|41|GeneID: 936031)			smc01342(N/A |- 3.28956|0.7533 96|127|133|7|1 3|GeneID:1232 970)	atu1333(N/A|- 9.55701|0.0838 6322|162|187|6| 27|GeneID:113 3371)	
0.7 24 02 62 24	0.0 13 50 71 28			rv3370c(dnaE 2|- 6.13339|0.380 5987|12|18|14| 20|GeneID:88 7259)				sma0892(dnaE 3|- 6.02564|0.3702 496|72|79|34|4 1|GeneID:1235 525)	atu3228(dnaE|- 6.12010|0.3331 186|225|232|4|1 1|GeneID:1135 102)	ncgl0611(dnaE 2|- 8.92122|0.0718 4476|227|235|3 1|39|GeneID:1 018640)
0.7 24 02 62 24	0.0 13 90 43 03	rcap_rcc03388 (N/A|- 11.20220|0.112 2585|192|209|2 6|40|GeneID:9 006204)	b0149(mrcB|- 8.33779|0.1297 912|186|214|21 |39|GeneID:94 4843)	rv3682(ponA2| - 10.74210|0.06 163442|90|116 |1|30|GeneID:8 85751)	bsu37510(pbp G|- 4.85423|0.261 9108|105|144| 8|37|GeneID:9 37200)	vc0602(N/A|- 3.64661|0.5486 637|284|300|20 |37|GeneID:26 15390)	sp_2010(N/A| - 7.30166|0.195 6258|292|300| 2|10|GeneID: 932011)	smc02856(N/A |- 6.83663|0.2787 140|246|286|6| 37|GeneID:123 1812)	atu0103(pbpC|- 7.20345|0.2238 427|223|248|5|3 7|GeneID:1132 141)	ncgl0274(N/A| - 5.21777|0.3205 015|179|185|1| 7|GeneID:1021 340)
0.7 24 02	0.0 14 03	rcap_rcc02182 (N/A|- 11.10080|0.116 2111|28|35|34|		rv3329(N/A|- 3.81202|0.725 4407|102|115|	bsu19740(yod T|- 8.72159|0.055 97393|132|139			sm_b20277(N/ A|- 8.48847|0.1446 370|282|290|27	atu3407(N/A|- 5.18804|0.4517 698|73|84|9|22|	ncgl2355(N/A| - 3.40459|0.5844 472|245|251|20

62 24	51 63	41|GeneID:900 5001)		8|21|GeneID:8 88028)	|32|39|GeneID: 940050)			|35|GeneID:12 36598)	GeneID:113528 1)	|26|GeneID:10 20388)
0.7 24 02 62 24	0.0 14 75 04 02	rcap_rcc01573 (N/A|- 5.64099|0.5673 371|250|256|33 |39|GeneID:90 04396)	b3194(mlaE|- 7.97343|0.1501 731|265|274|24 |33|GeneID:94 7732)	rv1964(yrbE3 A|- 8.66950|0.146 9951|176|183| 14|21|GeneID: 885948)		vc2519(N/A|- 6.53141|0.1839 766|25|31|3|9|G eneID:2615183)		smc00174(N/A |- 7.82987|0.1900 005|47|78|11|3 7|GeneID:1233 498)	atu1647(N/A|- 4.75962|0.5125 399|14|47|4|37| GeneID:113368 5)	
0.7 24 02 62 24	0.0 15 45 00 35	rcap_rcc01904 (N/A|- 10.95510|0.122 1088|85|97|27| 37|GeneID:900 4727)	b3743(asnC|- 3.28663|0.6814 874|93|99|15|2 1|GeneID:9482 59)	rv3291c(lrpA|- 8.45225|0.160 4395|60|69|1|1 0|GeneID:888 728)	bsu06540(yez C|- 6.66064|0.126 0488|213|219| 8|14|GeneID:9 36047)	vc0071(N/A|- 1.96640|0.8204 871|228|235|17 |24|GeneID:26 12951)		smc00425(N/A |- 8.23505|0.1609 094|101|118|5| 23|GeneID:123 1978)		
0.7 24 02 62 24	0.0 15 52 09 85		b2131(osmF|- 8.02516|0.1471 186|22|34|25|3 9|GeneID:9466 81)	rv3759c(proX| - 2.96739|0.842 4065|269|284| 7|31|GeneID:8 86270)	bsu33810(opu CC|- 12.47280|0.01 445316|4|22|5| 36|GeneID:93 6248)		sp_1860(N/A| - 7.54209|0.174 1185|126|164| 1|23|GeneID: 930974)	sma1462(N/A|- 4.24165|0.6180 148|290|300|26 |36|GeneID:12 35834)	atu0199(N/A|- 10.42810|0.056 51982|218|254| 3|40|GeneID:11 32237)	
0.7 24 02 62 24	0.0 15 80 09 9					vc1416(N/A|- 11.57230|0.016 11580|228|248| 20|37|GeneID: 2614048)			atu4348(vgrG|- 7.41724|0.2059 906|39|46|7|14| GeneID:113622 2)	
0.7 24 02 62 24	0.0 15 87 30 22	rcap_rcc02456 (lolE|- 5.07166|0.6378 056|256|266|11 |20|GeneID:90 05275)	b1116(lolC|- 11.50290|0.033 51574|94|123|1 1|36|GeneID:9 45673)			vc1882(N/A|- 11.58780|0.015 98727|1|31|5|2 9|GeneID:2613 636)		smc01935(N/A |- 9.01401|0.1152 284|283|296|26 |38|GeneID:12 32935)	atu1289(N/A|- 4.14838|0.6031 837|259|265|8|1 4|GeneID:1133 327)	ncgl0606(N/A| - 5.49405|0.2891 364|189|214|15 |38|GeneID:10 18635)
0.7 24 02 62 24	0.0 16 75 27 71	rcap_rcc02494 (hemN2|- 0.17775|0.9935 564|143|161|20	b3867(hemN|- 10.02790|0.064 06165|2|34|8|3 4|GeneID:9483 62)			vc0116(N/A|- 1.11025|0.9199 528|228|234|16 |22|GeneID:26 14871)		sma1266(hem N|- 8.58874|0.1385 866|16|32|22|3	atu1601(hemN| - 4.75080|0.5138 217|227|233|4|1	

		|39|GeneID:9005313)						7|GeneID:1235727)	0|GeneID:1133639)	
0.724026224	0.0168248463		b4124(dcuR|-7.04936|0.2146487|183|206|11|36|GeneID:948640)		bsu04460(dctR|-10.05070|0.03398685|142|153|26|38|GeneID:940148)	vc1604(N/A|-5.79658|0.251241|7|14|34|41|GeneID:2613860)	sp_0156(N/A|-7.65695|0.1645206|284|292|24|32|GeneID:929947)			ncgl0068(N/A|-6.11306|0.2278723|125|137|4|16|GeneID:1021127)
0.724026224	0.0172363745		b3437(gntK|-17.50690|0.001920926|185|216|11|39|GeneID:947937)			vc0287(N/A|-4.21339|0.4572776|223|233|29|38|GeneID:2615010)		sm_b21119(gntK|-10.29290|0.06415433|109|115|29|35|GeneID:1237029)	atu0195(gntK|-4.36449|0.5708234|240|273|7|37|GeneID:1132233)	
0.724026224	0.0175878416	rcap_rcc00532(aspS|-17.98950|0.009011323|96|120|11|37|GeneID:9003361)	b1866(aspS|-5.98191|0.315388|259|293|4|38|GeneID:946385)	rv2572c(aspS|-6.24656|0.3663050|222|232|20|30|GeneID:888532)	bsu27550(aspS|-4.51083|0.3004623|42|73|4|37|GeneID:937549)	vc1166(aspS|-7.45983|0.1214473|267|273|2|8|GeneID:2614599)	sp_2114(aspS|-2.70793|0.8606541|54|61|15|22|GeneID:931956)	smc01756(aspS|-7.76298|0.1951831|171|180|7|16|GeneID:1232838)	atu1153(N/A|-7.78221|0.1781824|20|26|33|39|GeneID:1133191)	ncgl1573(aspS|-3.55392|0.5597508|72|78|33|39|GeneID:1019604)
0.724026224	0.0185509099	rcap_rcc00831(dld|-10.70950|0.1326557|5|34|7|32|GeneID:9003660)	b2979(glcD|-1.78631|0.8748529|70|76|16|22|GeneID:947353)	rv3353c(N/A|-11.57130|0.04297927|74|89|15|30|GeneID:888095)	bsu08800(ygaK|-3.86015|0.3874716|2|13|3|13|GeneID:939245)	vca0985(N/A|-5.59648|0.2726253|32|38|2|8|GeneID:2612794)		sm_b21415(N/A|-10.99780|0.04561163|54|67|26|39|GeneID:1237574)	atu0665(glcD|-7.91121|0.1691267|195|204|26|34|GeneID:1132703)	ncgl0865(N/A|-8.17019|0.0985388|211|232|9|30|GeneID:1018894)
0.724026224	0.0190113993		b2064(asmA|-7.67271|0.1690424|252|258|1|7|GeneID:946582)					smc01454(N/A|-8.49460|0.1442610|129|139|29|38|GeneID:1233803)	atu2052(N/A|-6.85875|0.2551606|16|22|26|32|GeneID:1139500)	
0.72402	0.001938		b4284(insI1|-8.18771|0.1378712|182|209|13	rv0678(N/A|-10.74770|0.06148577|92|127			sp_0915(N/A|-10.23870|0.03898005|18|37|	smc01195(N/A|-3.13220|0.7741542|125|138|25	atu1488(tnp|-9.90116|0.0718827|31|45|26|40|	ncgl0179(N/A|-4.92569|0.35643|83|90|32|39|G

62 24	44 32		|41|GeneID:94 8815)	|1|32|GeneID:8 88235)			3|19|GeneID: 930868)	|39|GeneID:12 33252)	GeneID:113352 6)	eneID:1021244)
0.7 24 02 62 24	0.0 19 44 11 99	rcap_rcc02432 (N/A|- 4.79299|0.6722 021|172|180|6| 14|GeneID:900 5251)	b2119(yehL|- 4.14793|0.5540 143|252|265|15 |31|GeneID:94 6656)	rv0370c(N/A|- 4.24058|0.659 5247|86|106|1| 31|GeneID:88 6465)				smc00515(N/A |- 11.47100|0.036 02295|256|273| 24|38|GeneID: 1233417)		
0.7 24 02 62 24	0.0 19 56 34 02	rcap_rcc01091 (gsiC|- 11.37000|0.105 9823|188|219|5 |38|GeneID:90 03920)	b1486(ddpB|- 10.67040|0.048 46384|231|243| 27|38|GeneID: 946044)	rv3665c(dppB| - 10.30030|0.07 449869|65|75| 10|20|GeneID: 885474)	bsu12930(dpp B|- 6.17291|0.153 5147|154|169| 6|23|GeneID:9 39881)	vca0590(N/A|- 7.98385|0.0952 485|286|300|29 |39|GeneID:26 12563)	sp_1890(N/A| - 1.05124|0.980 5382|3|9|14|20 |GeneID:9309 41)	smc02831(N/A |- 7.51601|0.2152 854|104|125|20 |37|GeneID:12 31786)	atu6142(accD|- 9.33195|0.0926 3828|280|287|2| 9|GeneID:1137 465)	ncgl2239(N/A| - 3.34126|0.5949 964|279|285|20 |26|GeneID:10 20272)
0.7 24 02 62 24	0.0 19 65 87 89	rcap_rcc00711 (N/A|- 6.56426|0.4574 694|151|188|7| 39|GeneID:900 3540)	b0607(uspG|- 14.53410|0.008 246858|35|66|5 |37|GeneID:94 5229)	rv2028c(N/A|- 8.45691|0.160 1401|153|180| 1|27|GeneID:8 88494)	bsu09690(nha X|- 7.51225|0.089 68547|158|164 |4|10|GeneID:9 39286)	vc0076(N/A|- 5.47440|0.2863 434|221|228|34 |41|GeneID:26 15765)	sp_1996(N/A| - 10.89350|0.02 578304|55|70| 5|20|GeneID: 931917)	smc00048(N/A |- 9.07156|0.1123 402|3|18|28|41| GeneID:12326 30)	atu0496(N/A|- 3.17216|0.7464 604|170|176|35| 41|GeneID:113 2534)	ncgl2842(N/A| - 5.57325|0.2806 113|167|175|23 |31|GeneID:10 20885)
0.7 24 02 62 24	0.0 20 39 07 67	rcap_rcc03465 (N/A|- 5.70060|0.5600 19|261|269|28| 37|GeneID:900 6281)	b2372(yfdV|- 6.56245|0.2569 025|252|260|1| 9|GeneID:9491 10)		bsu37040(ywk B|- 7.04671|0.107 9482|256|262| 2|8|GeneID:93 6875)	vca0024(N/A|- 3.24077|0.6169 663|51|58|20|2 7|GeneID:2612 525)		smc01358(N/A |- 4.84023|0.5304 75|224|236|8|1 9|GeneID:1232 954)	atu0080(N/A|- 10.35030|0.058 58141|11|31|12| 34|GeneID:113 2118)	ncgl2303(N/A| - 6.49042|0.1962 757|197|224|9| 40|GeneID:102 0336)
0.7 24 02 62 24	0.0 20 54 36 41	rcap_rcc01180 (copA2|- 11.77430|0.092 1431|282|289|3 0|37|GeneID:9 004009)	b0484(copA|- 3.64996|0.6277 578|79|85|16|2 2|GeneID:9461 06)	rv0092(ctpA|- 8.19558|0.177 7143|166|178| 22|34|GeneID: 886946)		vc2215(N/A|- 4.44209|0.4225 585|119|129|6| 16|GeneID:261 3254)	sp_0729(N/A| - 7.51860|0.176 1345|255|266| 1|13|GeneID: 930679)	sma1013(actP|- 5.14215|0.4873 876|266|282|10 |29|GeneID:12 35584)	atu1195(N/A|- 6.55077|0.2858 564|60|66|7|13| GeneID:113323 3)	ncgl0375(N/A| - 3.88111|0.5068 885|267|284|8| 32|GeneID:102 1421)
0.7 24 02	0.0 20 90	rcap_rcc02138 (N/A|- 8.37729|0.2783 162|272|283|25	b4363(yjjB|- 7.90987|0.1540 017|263|284|8|	rv3737(N/A|- 6.00176|0.397 6718|166|172|		vc0439(N/A|- 10.40500|0.029 23940|271|285|	sp_1730(N/A| - 8.60056|0.100 7381|130|145|			ncgl2533(N/A| - 6.00079|0.2380 883|123|163|7|

62 24	94 51	|35|GeneID:90 04957)	32|GeneID:948 811)	14|20|GeneID: 885794)		22|36|GeneID: 2615700)	7|21|GeneID: 931102)			34|GeneID:102 0569)
0.7 24 02 62 24	0.0 21 62 11 71		b1657(ydhP|- 8.37257|0.1279 797|218|247|6| 40|GeneID:947 340)		bsu29040(ytb D|- 8.31729|0.065 41402|114|130 |15|32|GeneID: 936655)			sm_b20071(N/ A|- 12.50490|0.021 09774|181|212| 11|37|GeneID: 1236362)	atu4292(N/A|- 5.54904|0.4033 441|270|276|8|1 4|GeneID:1136 166)	
0.7 24 02 62 24	0.0 22 20 34 05	rcap_rcc00502 (N/A|- 8.82616|0.2431 306|178|196|25 |40|GeneID:90 03331)	b3391(hofQ|- 6.33657|0.2785 946|53|75|4|33| GeneID:94790 1)			vc2630(N/A|- 13.35690|0.006 310387|6|13|30 |37|GeneID:26 15647)		smc04111(cpa C1|- 6.29580|0.3378 89|121|130|4|1 3|GeneID:1231 772)	atu0221(ctpD|- 6.73320|0.2673 618|149|155|4|1 0|GeneID:1132 259)	
0.7 24 02 62 24	0.0 22 23 49 89	rcap_rcc02024 (N/A|- 3.94483|0.7725 342|191|206|11 |27|GeneID:90 04847)	b4032(malG|- 12.87690|0.017 94513|77|104|1 2|37|GeneID:9 48530)	rv1237(sugB|- 10.02360|0.08 380386|269|29 6|6|32|GeneID: 887121)	bsu34590(mdx G|- 6.35336|0.142 6978|99|108|3 2|40|GeneID:9 36499)	vca0943(malG| - 7.87976|0.1000 019|26|32|2|8|G eneID:2612568)	sp_2110(N/A| - 7.74724|0.157 2745|35|57|5|2 2|GeneID:930 236)	smc01626(N/A |- 9.60562|0.0883 6898|3|41|4|37| GeneID:12339 04)	atu5081(N/A|- 6.41689|0.3000 079|3|10|7|14|G eneID:1136854)	ncgl0694(N/A| - 5.41713|0.2976 141|237|264|5| 37|GeneID:101 8723)
0.7 24 02 62 24	0.0 22 61 60 45	rcap_rcc02199 (pflA1|- 4.77292|0.6746 646|233|240|7| 14|GeneID:900 5018)	b3952(pflC|- 8.42369|0.1253 577|100|107|33 |40|GeneID:94 8453)	rv3138(pflA|- 1.90938|0.944 0093|255|261| 14|20|GeneID: 887973)		vca0512(nrdG| - 7.76922|0.1052 857|231|247|26 |37|GeneID:26 12588)	sp_1976(N/A| - 9.04652|0.078 73258|6|22|1|1 9|GeneID:929 784)			
0.7 24 02 62 24	0.0 23 17 08 19	rcap_rcc02243 (tauC|- 9.81718|0.1780 061|238|259|10 |37|GeneID:90 05062)	b0934(ssuC|- 5.54052|0.3658 239|61|68|2|9|G eneID:947216)					sm_b20571(N/ A|- 7.73092|0.1977 062|95|112|13| 35|GeneID:123 7815)	atu0158(N/A|- 6.85180|0.2558 248|84|90|32|38| GeneID:113219 6)	ncgl1174(N/A| - 5.61881|0.2758 004|36|42|1|7|G eneID:1019204)
0.7 24 02	0.0 23 37	rcap_rcc02394 (sdaA|- 6.87971|0.4223 032|276|282|31	b2797(sdaB|- 10.38300|0.054 94145|273|298|	rv0069c(sdaA| - 8.96082|0.130 5516|14|41|2|2		vc1300(N/A|- 11.17830|0.019 73734|256|278|	sp_0105(N/A| - 6.39585|0.294 8271|33|39|3|9	smc01256(sda| - 7.43166|0.2225 058|184|196|26	atu1759(sdaA|- 6.52643|0.2883 928|186|218|8|3	ncgl1583(N/A| - 4.06996|0.4774 059|227|244|18

62 24	59 91	|37|GeneID:90 05213)	10|37|GeneID: 947262)	9|GeneID:886 986)		4|29|GeneID:2 614754)	|GeneID:9298 82)	|37|GeneID:12 33061)	8|GeneID:1133 797)	|39|GeneID:10 19614)
0.7 24 02 62 24	0.0 24 00 18 65		b2286(nuoC|- 8.89519|0.1033 442|283|296|25 |37|GeneID:94 6759)	rv3147(nuoC|- 7.46089|0.236 2969|75|81|12| 18|GeneID:88 8816)				smc01914(nuo C1|- 8.26877|0.1586 613|177|191|26 |37|GeneID:12 32914)		
0.7 24 02 62 24	0.0 24 27 90 12	rcap_rcc00139 (N/A|- 10.30470|0.151 8001|32|42|27| 37|GeneID:900 2968)		rv3394c(N/A|- 6.19280|0.373 0501|119|129| 1|11|GeneID:8 87945)				sma0890(N/A|- 5.93318|0.3817 318|243|273|8| 37|GeneID:123 5524)	atu5099(N/A|- 10.70850|0.049 62826|19|33|27| 41|GeneID:113 6872)	ncgl0591(N/A| - 5.01637|0.3449 71|44|51|33|40| GeneID:10186 20)
0.7 24 02 62 24	0.0 24 94 23 8	rcap_rcc03529 (flgD|- 9.02334|0.2288 261|160|182|14 |37|GeneID:90 06345)	b1075(flgD|- 5.13480|0.4165 323|18|25|16|2 3|GeneID:9458 13)		bsu16280(flgD |- 4.41530|0.312 0668|111|120| 26|34|GeneID: 936501)	vc2198(flgD|- 6.86533|0.1588 428|278|290|18 |31|GeneID:26 13238)		smc03052(flg D|- 8.30318|0.1563 938|112|118|30 |36|GeneID:12 32323)	atu0579(flgD|- 7.32984|0.2131 461|255|282|10| 37|GeneID:113 2617)	
0.7 24 02 62 24	0.0 24 94 99 59	rcap_rcc02184 (potH2|- 9.93134|0.1715 452|5|12|30|37| GeneID:90050 03)	b1125(potB|- 9.38824|0.0841 1842|28|50|16| 37|GeneID:945 692)			vc1427(potB|- 8.59743|0.0711 9858|172|179|2 |9|GeneID:261 4059)	sp_1388(N/A| - 7.60469|0.168 8343|52|82|13| 37|GeneID:93 1436)	smc01633(N/A |- 10.94220|0.046 87695|147|161| 24|38|GeneID: 1233911)	atu0611(potB|- 5.29932|0.4365 412|252|258|34| 40|GeneID:113 2649)	
0.7 24 02 62 24	0.0 25 79 21 15	rcap_rcc02545 (N/A|- 8.42387|0.2744 932|6|19|15|28| GeneID:90053 64)		rv2435c(N/A|- 5.39453|0.482 1738|245|254| 1|10|GeneID:8 85891)		vca1056(N/A|- 9.32997|0.0499 0632|33|45|25| 35|GeneID:261 2073)		smc01104(mcp X|- 10.44080|0.059 78564|159|186| 6|37|GeneID:1 232050)	atu0387(mcpA| - 11.18440|0.039 67393|229|255| 10|39|GeneID:1 132425)	
0.7 24 02	0.0 25 96	rcap_rcc01948 (N/A|- 9.42803|0.2015 984|54|63|24|3	b0981(etk|- 4.78393|0.4634 193|174|187|20		bsu34360(eps B|- 5.88206|0.172 7395|170|211|	vc0937(N/A|- 5.97864|0.2329 55|93|116|11|3		smc01795(N/A |- 12.36460|0.022 72171|214|232|	atu1239(exoP|- 10.17430|0.063 50074|274|284|	ncgl0337(N/A| - 5.95941|0.2419 513|87|101|7|2

62 24	26 94	3|GeneID:9004 771)	|31|GeneID:94 7409)		6|37|GeneID:9 38640)	0|GeneID:2614 157)		20|37|GeneID: 1232879)	28|37|GeneID:1 133277)	2|GeneID:1021 402)
0.7 24 02 62 24	0.0 26 25 71 37		b3020(ygiS|- 8.90408|0.1029 649|80|87|30|3 7|GeneID:9471 40)	rv3666c(dppA| - 14.84800|0.00 9901222|276|2 99|2|31|GeneI D:885315)	bsu12960(dpp E|- 7.18582|0.102 1122|236|269| 7|37|GeneID:9 38099)		sp_0366(N/A| - 7.45408|0.181 7661|281|289| 15|23|GeneID :930256)	sm_b21196(op pA|- 2.25627|0.8747 954|119|134|14 |33|GeneID:12 37192)	atu4259(N/A|- 5.28304|0.4387 531|49|59|11|22| GeneID:113613 3)	ncgl2294(N/A| - 5.28904|0.3121 672|64|71|28|3 5|GeneID:1020 327)
0.7 24 02 62 24	0.0 27 38 48 04		b1234(rssA|- 8.23508|0.1352 743|190|196|33 |39|GeneID:94 5725)	rv3728(N/A|- 7.81362|0.206 4064|124|130| 14|20|GeneID: 885271)	bsu15040(ylb K|- 9.96087|0.035 12907|217|224 |2|9|GeneID:93 6382)	vc0603(N/A|- 3.75036|0.5314 795|293|299|10 |16|GeneID:26 15391)		smc00930(N/A |- 3.86267|0.6732 859|189|195|25 |31|GeneID:12 32444)	atu1751(N/A|- 20.77250|0.000 1934701|156|18 9|5|36|GeneID: 1133789)	
0.7 24 02 62 24	0.0 27 62 74 03	rcap_rcc01535 (nuoL|- 6.22877|0.4963 655|278|284|29 |35|GeneID:90 04358)	b2278(nuoL|- 12.80850|0.018 52014|265|300| 7|37|GeneID:9 45540)	rv3156(nuoL|- 3.72067|0.739 1084|26|32|6|1 2|GeneID:888 063)				smc01925(nuo L|- 5.78410|0.4006 629|258|264|8| 14|GeneID:123 2925)	atu1281(nuoL|- 5.27705|0.4395 683|40|46|8|14| GeneID:113331 9)	
0.7 24 02 62 24	0.0 27 72 30 68	rcap_rcc02555 (N/A|- 9.40370|0.2031 562|49|57|6|14| GeneID:90053 74)	b4212(ytfH|- 7.48226|0.1820 158|6|19|19|31| GeneID:94873 0)		bsu03470(hxl R|- 8.42584|0.062 722|187|207|1 8|34|GeneID:9 38311)			smc01493(N/A |- 12.27120|0.023 86505|233|254| 12|37|GeneID: 1233748)	atu5151(attZ|- 10.27030|0.060 77269|142|151| 5|14|GeneID:11 36924)	ncgl2034(N/A| - 2.46106|0.7412 052|32|38|32|3 8|GeneID:1020 066)
0.7 24 02 62 24	0.0 29 12 56 94	rcap_rcc03382 (tyrB|- 9.54832|0.1940 414|61|74|5|20| GeneID:90061 98)	b4054(tyrB|- 8.08096|0.1438 847|101|143|7| 40|GeneID:948 563)			vca0513(N/A|- 6.07909|0.2233 341|187|193|1| 7|GeneID:2612 575)		smc00387(tatA |- 5.21572|0.4770 752|82|105|5|3 7|GeneID:1231 941)		
0.7 24 02	0.0 29 96	rcap_rcc03119 (N/A|- 8.78495|0.2462 081|154|163|5|			bsu31322(yug O|- 6.78630|0.119 8328|259|265|	vca0194(N/A|- 5.87855|0.2428 714|164|202|1|		sm_b20361(N/ A|- 8.44127|0.1475 602|273|284|25		

62 24	78 53	14|GeneID:900 5935)			1|7|GeneID:93 8842)	37|GeneID:261 2372)		|36|GeneID:12 36678)		
0.7 24 02 62 24	0.0 30 29 91 97					vc1584(N/A|- 8.51027|0.0742 3254|125|142|1 5|32|GeneID:2 613838)		smc01158(N/A |- 3.21187|0.7637 268|121|141|12 |36|GeneID:12 31995)	atu1798(N/A|- 11.56130|0.033 13691|102|109| 30|37|GeneID:1 133836)	ncgl2263(N/A| - 7.50447|0.1299 656|234|262|6| 33|GeneID:102 0296)
0.7 24 02 62 24	0.0 30 67 32 57	rcap_rcc02524 (N/A|- 10.81930|0.127 8445|232|260|6 |39|GeneID:90 05343)	b0366(tauB|- 9.16244|0.0924 7728|17|24|27| 34|GeneID:945 027)		bsu08830(ssu B|- 12.41360|0.01 474638|172|21 2|11|37|GeneI D:939730)		sp_2196(N/A| - 4.26416|0.625 8802|38|44|13| 19|GeneID:93 0180)	sm_b20569(N/ A|- 6.34226|0.3325 077|108|125|12 |37|GeneID:12 37813)	atu1047(N/A|- 9.07349|0.1037 177|244|264|5|2 3|GeneID:1133 085)	ncgl1175(N/A| - 6.33341|0.2089 207|63|87|9|31| GeneID:10192 05)
0.7 24 02 62 24	0.0 30 94 56 66	rcap_rcc01902 (N/A|- 11.38810|0.105 3246|179|188|2 9|38|GeneID:9 004725)	b3963(fabR|- 8.20341|0.1370 057|15|35|8|28| GeneID:94846 0)	rv0681(N/A|- 13.23610|0.02 053607|93|119 |1|27|GeneID:8 88239)	bsu19070(yob S|- 7.39549|0.093 93598|124|152 |8|32|GeneID:9 39639)	vc0152(N/A|- 13.19520|0.006 878643|89|128| 7|41|GeneID:2 614851)			atu4199(N/A|- 4.03067|0.6208 505|54|60|3|9|G eneID:1136073)	
0.7 24 02 62 24	0.0 31 10 67 65		b1840(yebZ|- 6.04515|0.3085 819|20|57|13|3 4|GeneID:9470 78)	rv0102(N/A|- 7.77472|0.209 5387|247|256| 1|10|GeneID:8 86926)	bsu14930(cta G|- 6.15772|0.154 4627|219|226| 3|10|GeneID:9 39909)			sma1198(N/A|- 4.01761|0.6508 098|121|127|26 |32|GeneID:12 35690)		ncgl2370(N/A| - 5.58305|0.2795 707|171|177|4| 10|GeneID:102 0403)
0.7 24 02 62 24	0.0 31 14 00 07	rcap_rcc01719 (N/A|- 4.90641|0.6582 424|99|106|26| 33|GeneID:900 4542)	b0820(ybiT|- 8.28413|0.1326 308|8|20|2|14|G eneID:945440)	rv1473(N/A|- 8.95523|0.130 8508|262|270| 1|9|GeneID:88 6549)	bsu07420(yfm M|- 5.26767|0.221 6515|220|242| 5|29|GeneID:9 36103)	vc2608(N/A|- 7.45689|0.1216 110|5|11|34|40| GeneID:26156 25)	sp_2230(N/A| - 4.06433|0.659 675|70|83|5|19 |GeneID:9298 40)	sma0036(N/A|- 8.97211|0.1173 707|258|281|12 |37|GeneID:12 35242)	atu3869(N/A|- 6.08218|0.3375 201|126|132|26| 32|GeneID:113 5743)	ncgl1493(N/A| - 8.94101|0.0712 4642|136|153|2 0|40|GeneID:1 019523)
0.7 24 02	0.0 31 15		b3486(rbbA|- 13.86570|0.011 31870|139|188|	rv1747(N/A|- 5.38098|0.484 156|221|230|1|	bsu13220(thi W|- 6.36796|0.141 8576|182|213|		sp_0720(N/A| - 7.85144|0.149 2316|61|87|3|2		atu4667(N/A|- 5.01127|0.4764 611|290|296|35|	ncgl0510(N/A| - 6.31383|0.2105

62 24	06 21		4|37|GeneID:947998)	10|GeneID:885311)	7|36|GeneID:936472)		4|GeneID:930670)		41|GeneID:1136541)	47|42|48|3|9|GeneID:1021527)
0.7 24 02 62 24	0.0 31 29 08 31	rcap_rcc02050 (cobO|-11.52870|0.1003398|55|81|3|41|GeneID:9004873)	b1270(btuR|-7.19901|0.2028594|140|166|7|31|GeneID:945839)	rv2849c(cobO|-6.46426|0.3398357|203|245|1|30|GeneID:888236)		vc1040(N/A|-2.33213|0.7665507|5|17|11|23|GeneID:2614310)		smc04302(cobO|-9.85577|0.07876074|75|103|5|38|GeneID:1233621)	atu2807(cobO|-9.01704|0.1062882|156|164|29|37|GeneID:1134845)	
0.7 24 02 62 24	0.0 31 56 70 72	rcap_rcc01402 (N/A|-10.76190|0.13034|215|235|20|37|GeneID:9004225)	b0936(ssuA|-6.65300|0.2485832|217|232|5|21|GeneID:945560)		bsu08840(ssuA|-12.34560|0.01509126|5|31|5|30|GeneID:936215)	vc1369(N/A|-3.12120|0.6372133|285|291|6|12|GeneID:2614823)		sma2087(N/A|-5.40581|0.4508385|31|42|8|19|GeneID:1236175)	atu1884(ssuA|-12.34180|0.02265370|237|252|28|41|GeneID:1133922)	ncgl1176(N/A|-2.22304|0.7785987|264|271|6|13|GeneID:1019206)
0.7 24 02 62 24	0.0 32 26 14 25		b3667(uhpC|-10.45290|0.05329537|19|52|6|41|GeneID:948184)	rv1672c(N/A|-6.40720|0.3466411|82|88|14|20|GeneID:885700)	bsu12360(exuT|-9.42286|0.04290688|8|14|33|39|GeneID:939405)	vca0137(glpT|-3.37604|0.5940776|274|280|7|13|GeneID:2612732)			atu2708(N/A|-4.46856|0.5553287|258|264|8|14|GeneID:1134746)	ncgl2144(N/A|-6.71406|0.1794428|185|211|3|35|GeneID:1020177)
0.7 24 02 62 24	0.0 33 45 93 11	rcap_rcc01680 (fabF1|-12.50660|0.07120139|62|71|28|37|GeneID:9004503)	b2323(fabB|-0.38641|0.9723225|63|78|9|37|GeneID:946799)	rv2246(kasB|-7.00914|0.2796500|15|25|1|12|GeneID:887539)	bsu17140(pksF|-6.64027|0.1270876|38|59|16|37|GeneID:939500)		sp_0422(N/A|-5.64744|0.3984933|94|106|12|23|GeneID:930358)	smc00574(fabF|-8.06995|0.1722945|246|253|34|41|GeneID:1232788)	atu1596(fabF|-7.03771|0.2385053|2|17|6|23|GeneID:1133634)	ncgl0802(N/A|-5.32850|0.3076257|1|20|8|32|GeneID:1018831)
0.7 24 02 62 24	0.0 34 17 70 96				bsu33200(yvrE|-6.18511|0.1527576|281|288|2|9|GeneID:935979)			sma0717(N/A|-13.92400|0.009689755|9|33|8|35|GeneID:1235201)	atu4190(N/A|-7.41363|0.2062823|134|162|8|37|GeneID:1136064)	
0.7 24 02	0.0 34 29	rcap_rcc00626 (N/A|-11.83060|0.09035221|186|209|	b2307(hisM|-6.26449|0.2858011|233|256|7|			vca0757(artM|-8.98724|0.0589 9083|240|269|7		smc00138(N/A|-6.56991|0.3069424|37|46|5|14|	atu2363(N/A|-7.19834|0.2242840|52|59|30|37|	

62 24	99 27	16|38|GeneID: 9003455)	31|GeneID:946 790)			|32|GeneID:26 11843)		GeneID:12335 37)	GeneID:113440 1)	
0.7 24 02 62 24	0.0 34 69 17 85	rcap_rcc02522 (N/A|- 13.20260|0.055 46339|34|74|5| 37|GeneID:900 5341)					sp_2198(N/A| - 8.01656|0.137 1686|7|28|1|19 |GeneID:9319 18)	smc01826(N/A |- 12.03190|0.027 03711|189|218| 8|35|GeneID:1 234126)	atu2390(N/A|- 10.01310|0.068 3332|254|261|7| 14|GeneID:113 4428)	
0.7 24 02 62 24	0.0 34 81 87 46	rcap_rcc02898 (N/A|- 6.44283|0.4713 827|8|21|3|13|G eneID:9005717)	b0914(msbA|- 1.89467|0.8634 967|188|202|19 |31|GeneID:94 5530)		bsu08690(yga D|- 9.15085|0.047 53498|158|166 |32|40|GeneID: 936200)	vc1878(N/A|- 4.12880|0.4704 731|225|239|8| 21|GeneID:261 3632)		sm_b20813(ms bA1|- 12.92230|0.016 87165|183|206| 19|39|GeneID: 1236884)	atu4600(N/A|- 6.73259|0.2674 221|5|11|26|32| GeneID:113647 4)	
0.7 24 02 62 24	0.0 35 22 21 05	rcap_rcc01881 (sufB|- 4.56534|0.6999 582|67|77|25|3 7|GeneID:9004 704)	b1683(sufB|- 8.22280|0.1359 433|240|252|11 |23|GeneID:94 5753)	rv1461(N/A|- 14.20210|0.01 328285|216|24 8|1|33|GeneID: 886609)	bsu32700(suf D|- 7.14997|0.103 5837|284|292| 14|22|GeneID: 938871)		sp_0868(N/A| - 5.38078|0.439 6283|63|72|23| 32|GeneID:93 0820)		atu1824(sufB|- 6.15734|0.3288 335|227|233|3|9| GeneID:113386 2)	
0.7 24 02 62 24	0.0 36 66 46 89	rcap_rcc02505 (tadA|- 10.62900|0.136 2840|56|63|30| 37|GeneID:900 5324)		rv3752c(N/A|- 2.99908|0.838 4916|65|86|2|2 4|GeneID:885 586)	bsu00180(tad A|- 1.48269|0.820 2497|107|113| 19|25|GeneID: 937989)	vc0864(N/A|- 3.87898|0.5104 211|187|213|10 |32|GeneID:26 14531)	sp_0020(N/A| - 3.14216|0.804 213|262|270|4| 12|GeneID:92 9767)	smc00905(N/A |- 8.72774|0.1305 498|230|240|5| 14|GeneID:123 2441)	atu0688(codA|- 9.53895|0.0845 3885|177|212|5| 33|GeneID:113 2726)	ncgl0225(N/A| - 11.52240|0.023 76644|182|197| 21|35|GeneID: 1021295)
0.7 24 02 62 24	0.0 36 68 68 7	rcap_rcc02052 (bluB|- 10.02160|0.166 5798|1|31|5|41| GeneID:90048 75)		rv3368c(N/A|- 6.89709|0.291 3152|79|102|1| 30|GeneID:88 7641)				smc00166(N/A |- 5.76172|0.4035 48|161|173|26| 37|GeneID:123 3506)	atu1654(N/A|- 13.35490|0.013 62647|20|64|5|3 7|GeneID:1133 692)	
0.7 24 02	0.0 36 85	rcap_rcc01379 (N/A|- 6.36329|0.4806 005|2|12|25|35|	b2799(fucO|- 6.81688|0.2340 674|255|271|13		bsu31050(gbs B|- 4.90320|0.256 8027|10|35|8|3	vc1147(N/A|- 6.31305|0.2021 83|93|131|2|38|	sp_2026(N/A| - 8.35146|0.115 1433|269|291|	sma0263(N/A|- 6.22062|0.3467 125|186|215|12	atu5074(attL|- 8.57400|0.1284 834|1|31|7|33|G	ncgl1112(N/A| - 11.12960|0.028 09637|146|170|

62 24	81 16	GeneID:90042 02)	|34|GeneID:94 7273)		3|GeneID:938 832)	GeneID:26145 80)	5|22|GeneID: 932025)	|36|GeneID:12 35365)	eneID:1136847)	9|31|GeneID:1 019142)
0.7 24 02 62 24	0.0 37 05 57 71	rcap_rcc00896 (betI|- 7.64784|0.3434 217|219|244|4| 22|GeneID:900 3725)	b0313(betI|- 0.67701|0.9594 672|27|35|21|3 0|GeneID:9449 81)					sma1726(N/A|- 11.44060|0.036 57941|228|246| 12|36|GeneID: 1235985)	atu2059(N/A|- 2.43748|0.8407 808|158|165|5|12 |GeneID:1134 097)	
0.7 24 02 62 24	0.0 37 15 34 88		b1968(yedV|- 12.25690|0.023 84437|258|298| 10|40|GeneID: 946487)	rv0902c(prrB|- 4.41031|0.632 9673|207|216| 8|17|GeneID:8 85647)				smc02367(N/A |- 3.63562|0.7057 134|199|211|25 |37|GeneID:12 32666)	atu0979(ragB|- 0.60663|0.9760 215|234|249|6|3 7|GeneID:1133 017)	ncgl0840(N/A| - 4.81868|0.3703 027|139|169|15 |38|GeneID:10 18869)
0.7 24 02 62 24	0.0 37 56 87 78	rcap_rcc01655 (N/A|- 9.07085|0.2254 826|160|167|30 |37|GeneID:90 04478)	b0644(ybeQ|- 2.70512|0.7640 37|292|298|3|9| GeneID:94525 1)			vca1026(N/A|- 6.46663|0.1892 273|211|236|5| 30|GeneID:261 2428)		smc02078(exo R|- 8.58601|0.1387 485|230|237|34 |41|GeneID:12 33171)	atu0499(N/A|- 8.40740|0.1378 165|88|94|29|35| GeneID:113253 7)	
0.7 24 02 62 24	0.0 37 59 89 6				bsu30750(mnt C|- 3.87368|0.385 4695|85|92|6|1 3|GeneID:937 203)		sp_1649(N/A| - 3.67599|0.723 5587|84|90|13| 19|GeneID:93 1187)	smc02507(sitC |- 7.78887|0.1931 641|27|48|8|36| GeneID:12347 12)	atu4469(sitC|- 11.75170|0.030 22864|266|291| 8|37|GeneID:11 36343)	
0.7 24 02 62 24	0.0 38 14 83 04	rcap_rcc03048 (lldD|- 3.83180|0.7850 65|124|130|8|1 4|GeneID:9005 864)	b3605(lldD|- 3.68482|0.6225 741|49|64|14|3 1|GeneID:9481 21)	rv1872c(lldD2| - 9.01606|0.127 6279|261|281| 14|32|GeneID: 885754)		vca0984(lldD|- 5.54388|0.2784 732|68|74|17|2 3|GeneID:2612 817)	sp_0715(N/A| - 8.48054|0.107 4778|2|21|1|21 |GeneID:9306 65)	sm_b20858(N/ A|- 15.14360|0.004 757681|255|29 0|4|35|GeneID: 1237482)	atu2318(lldA|- 3.62923|0.6807 042|239|255|5|2 2|GeneID:1134 356)	ncgl2817(N/A| - 4.37970|0.4310 651|52|58|20|2 6|GeneID:1020 860)
0.7 24 02	0.0 38 84	rcap_rcc01502 (gph2|- 14.05230|0.040 65629|188|219|	b3385(gph|- 6.00136|0.3132 834|97|140|7|3	rv2232(ptkA|- 5.09571|0.526 6875|231|259|		vc2624(N/A|- 10.88190|0.022 96397|77|117|6	sp_0104(N/A| - 3.90589|0.686 104|97|109|1|1	smc01276(gph 2|- 4.40179|0.5944 868|156|163|15	atu1472(N/A|- 8.42518|0.1367 934|137|165|7|3	

62 24	19 4	7|36|GeneID:9 004325)	8|GeneID:9478 95)	1|29|GeneID:8 87597)		|37|GeneID:26 15641)	2|GeneID:929 880)	|22|GeneID:12 33040)	4|GeneID:1133 510)	
0.7 24 02 62 24	0.0 39 08 08 58	rcap_rcc02897 (N/A|- 5.14495|0.6287 194|247|272|8| 38|GeneID:900 5716)	b2078(baeS|- 5.32215|0.3926 209|180|220|6| 37|GeneID:946 611)		bsu03760(ycl K|- 10.63770|0.02 744988|180|19 4|25|38|GeneI D:938283)		sp_0084(N/A| - 6.06915|0.337 7755|276|297| 1|21|GeneID: 929854)			ncgl2862(N/A| - 3.94358|0.4970 421|189|195|32 |38|GeneID:10 20905)
0.7 24 02 62 24	0.0 39 75 03 22	rcap_rcc00056 (flhA|- 2.22809|0.9261 452|108|114|34 |40|GeneID:90 02885)	b1879(flhA|- 3.41446|0.6626 796|265|295|9| 39|GeneID:946 390)			vc2069(flhA|- 4.13462|0.4695 595|92|112|11| 34|GeneID:261 3449)		smc03054(flh A|- 7.07688|0.2548 975|150|158|25 |33|GeneID:12 32325)	atu0581(flhA|- 3.07389|0.7600 353|244|254|25| 37|GeneID:113 2619)	
0.7 24 02 62 24	0.0 40 02 15 27	rcap_rcc02368 (gph4|- 10.95410|0.122 1502|37|60|11| 37|GeneID:900 5187)	b1317(ycjU|- 4.99785|0.4345 207|144|150|4| 10|GeneID:945 891)	rv3400(N/A|- 10.28770|0.07 490038|263|29 3|2|34|GeneID: 887918)	bsu34550(pgc M|- 2.92815|0.544 1818|177|191| 5|23|GeneID:9 38624)	vca0798(N/A|- 4.64133|0.3935 526|58|65|18|2 5|GeneID:2611 825)		sm_b21214(orf 24|- 8.78278|0.1274 774|26|33|26|3 3|GeneID:1237 212)	atu3251(N/A|- 6.26805|0.3163 147|249|257|8|1 6|GeneID:1135 125)	ncgl1449(N/A| - 6.67495|0.1822 894|256|277|13 |38|GeneID:10 19479)
0.7 24 02 62 24	0.0 40 14 25 72	rcap_rcc01211 (folP|- 7.14202|0.3942 364|46|58|11|2 3|GeneID:9004 039)	b3177(folP|- 4.60179|0.4887 236|44|78|4|31| GeneID:94769 1)	rv3608c(folP1| - 5.67315|0.442 2962|160|187| 1|30|GeneID:8 85831)		vc0638(N/A|- 2.74043|0.7010 532|275|296|12 |31|GeneID:26 15426)	sp_0289(N/A| - 10.12910|0.04 16932|276|29 7|4|22|GeneID :930100)	smc00462(folP |- 0.17648|0.9882 404|245|258|7| 23|GeneID:123 3476)	atu1352(folP|- 13.30840|0.013 95311|2|39|2|37| GeneID:113339 0)	ncgl2601(N/A| - 7.02244|0.1583 738|174|214|5| 36|GeneID:102 0641)
0.7 24 02 62 24	0.0 40 15 21 29	rcap_rcc00248 (N/A|- 16.91680|0.013 7233|190|218|8 |35|GeneID:90 03077)	b1533(eamA|- 6.48075|0.2645 93|247|253|29| 35|GeneID:946 081)		bsu18770(cye A|- 7.94508|0.075 6266|245|251| 1|7|GeneID:94 0126)				atu1863(N/A|- 6.36093|0.3060 681|244|250|32| 38|GeneID:113 3901)	
0.7 24 02	0.0 40 47	rcap_rcp00098 (N/A|- 7.80469|0.3285 942|180|192|5|			bsu10260(yhf K|- 5.11411|0.235 8600|182|191|	vca1054(N/A|- 5.58660|0.2737 165|3|35|8|38|G	sp_1627(N/A| - 3.11767|0.807 6413|27|33|15|	sma2099(N/A|- 2.52945|0.8464 944|213|220|6|	atu1999(N/A|- 6.89060|0.2521 333|71|77|7|13|	ncgl2943(N/A| - 8.45898|0.0872 984|225|249|9|

62 24	75 56	16|GeneID:900 6522)			26|35|GeneID: 936309)	eneID:2612081)	21|GeneID:93 1213)	13|GeneID:123 6181)	GeneID:113403 7)	37|GeneID:102 0990)
0.7 24 02 62 24	0.0 41 65 12 24	rcap_rcc02308 (rarD|- 12.17170|0.080 16556|159|177| 11|34|GeneID: 9005127)			bsu19480(yoj E|- 6.45765|0.136 8064|256|262| 32|38|GeneID: 939447)	vc0195(N/A|- 5.43860|0.2904 634|268|274|32 |38|GeneID:26 14559)		smc02545(N/A |- 10.56460|0.056 33413|8|18|32| 41|GeneID:123 2761)	atu1068(N/A|- 0.27729|0.9853 82|250|257|9|16| GeneID:113310 6)	ncgl2050(N/A| - 5.15263|0.3282 675|269|284|20 |32|GeneID:10 20082)
0.7 24 02 62 24	0.0 42 44 33 62	rcap_rcc01121 (trmB|- 6.54478|0.4596 88|5|11|25|31|G eneID:9003950)	b2960(trmI|- 9.01213|0.0984 5414|57|95|4|4 0|GeneID:9474 48)	rv0208c(N/A|- 5.70410|0.437 9769|60|82|1|3 1|GeneID:886 740)	bsu29900(trm B|- 8.10555|0.071 02854|22|28|2| 8|GeneID:936 447)	vc0453(trmB|- 5.25450|0.3123 484|85|107|20| 37|GeneID:261 5115)	sp_0550(trmB |- 5.39542|0.437 3209|241|250| 12|23|GeneID :930494)	smc01108(trm B|- 5.94368|0.3804 178|20|33|10|2 2|GeneID:1232 046)	atu0363(N/A|- 5.27216|0.4402 344|168|174|2|8| GeneID:113240 1)	ncgl2767(trmB |- 4.87668|0.3627 367|267|284|9| 27|GeneID:102 0808)
0.7 24 02 62 24	0.0 42 66 78 82	rcap_rcc00892 (N/A|- 12.22420|0.078 69483|177|189| 26|37|GeneID: 9003721)				vca0963(N/A|- 9.21697|0.0527 4505|83|89|29| 35|GeneID:261 2745)		sm_b20112(N/ A|- 8.03198|0.1750 03|242|254|29| 40|GeneID:123 6403)	atu4150(N/A|- 9.27079|0.0951 6036|154|192|5| 37|GeneID:113 6024)	
0.7 24 02 62 24	0.0 42 70 59 46			rv1679(fadE16 |- 5.78250|0.427 1409|252|287| 1|30|GeneID:8 85688)	bsu04520(ydb M|- 10.55530|0.02 827838|130|14 9|26|40|GeneI D:939949)			sma2073(N/A|- 5.15108|0.4861 315|76|87|28|3 7|GeneID:1236 168)	atu2572(acd|- 5.38762|0.4246 435|216|227|25| 37|GeneID:113 4610)	ncgl0974(N/A| - 5.54804|0.2833 026|92|109|6|3 9|GeneID:1019 003)
0.7 24 02 62 24	0.0 42 75 44	rcap_rcc00197 (comF|- 11.18040|0.113 0979|118|149|5 |41|GeneID:90 03026)	b3413(gntX|- 4.05170|0.5681 62|133|141|6|1 4|GeneID:9479 15)	rv3242c(N/A|- 2.96297|0.842 949|107|113|1 5|21|GeneID:8 88757)			sp_2207(N/A| - 4.64325|0.561 3086|150|179| 14|37|GeneID :929909)	smc02444(N/A |- 12.70060|0.019 00890|178|190| 26|37|GeneID: 1234293)	atu3512(comF|- 6.29991|0.3127 734|65|80|7|20| GeneID:113538 6)	ncgl0724(N/A| - 4.47357|0.4175 665|6|26|9|31|G eneID:1018753)
0.7 24 02	0.0 43 07	rcap_rcc03078 (N/A|- 11.77580|0.092 09497|7|32|12|	b2062(wza|- 4.74566|0.4686 855|258|298|5|					smc02274(rkp U|- 8.24539|0.1602 173|128|135|33	atu3272(pssN|- 7.35767|0.2108 464|249|274|6|4	

62 24	77 33	37|GeneID:900 5894)	41|GeneID:946 558)					|40|GeneID:12 32215)	1|GeneID:1135 146)	
0.7 24 02 62 24	0.0 43 22 17 61	rcap_rcc01961 (gluQ|- 7.74392|0.3342 856|109|119|28 |37|GeneID:90 04784)	b0144(gluQ|- 4.28505|0.5340 089|256|285|8| 37|GeneID:944 846)			vc0595(N/A|- 2.55426|0.7314 517|205|211|20 |26|GeneID:26 15383)			atu3589(gltX|- 8.07296|0.1583 19|71|91|25|38| GeneID:113546 3)	ncgl0233(N/A| - 7.91849|0.1094 586|57|87|6|38| GeneID:10213 02)
0.7 24 02 62 24	0.0 43 32 79 3	rcap_rcc02463 (N/A|- 10.36570|0.148 7689|1|7|29|35| GeneID:90052 82)				vca0849(N/A|- 10.19700|0.032 46204|154|160| 29|35|GeneID: 2612210)		sm_b21543(N/ A|- 9.31717|0.1006 994|53|61|30|3 8|GeneID:1237 315)	atu2707(rzcA|- 0.53088|0.9784 917|112|118|8|1 4|GeneID:1134 745)	
0.7 24 02 62 24	0.0 43 41 97 21		b2430(yfeW|- 8.73424|0.1104 335|10|30|9|37| GeneID:94690 7)	rv2463(lipP|- 12.73220|0.02 572636|132|15 4|6|28|GeneID: 888572)	bsu34440(pbp E|- 9.64651|0.039 46719|89|116| 2|37|GeneID:9 38615)		sp_1448(N/A| - 3.77784|0.707 119|188|195|1 5|22|GeneID: 931391)	sm_b21600(am pC|- 0.82054|0.9706 348|235|242|11 |18|GeneID:12 37371)	atu3077(ampC|- 4.68665|0.5231 754|178|184|3|9| GeneID:113487 9)	ncgl2331(N/A| - 4.49680|0.4142 673|95|113|4|2 4|GeneID:1020 364)
0.7 24 02 62 24	0.0 44 26 43 24	rcap_rcc00580 (cbbR1|- 16.57050|0.015 69478|118|150| 5|37|GeneID:9 003409)		rv2282c(N/A|- 2.65895|0.878 0573|241|249| 10|19|GeneID: 887253)	bsu37650(cys L|- 3.77326|0.400 5256|189|200| 23|34|GeneID: 936502)	vc2324(N/A|- 6.61616|0.1772 948|47|71|7|34| GeneID:26131 20)		sm_b21535(N/ A|- 9.03337|0.1142 498|234|240|34 |40|GeneID:12 37306)	atu4411(N/A|- 4.06601|0.6155 467|221|227|8|1 4|GeneID:1136 285)	ncgl0015(N/A| - 7.86979|0.1117 008|210|225|15 |35|GeneID:10 21314)

FDR	**NC_006085**	**NC_007494**	**NC_008463**	**NC_010338**	**NC_013971**	**NC_016810**	**NC_018750**	**Annotation**	**Additional**	**Amou**

									homologs	nt sampled
0.3730207 29	ppa1643(N/A|-6.93556|0.1263353|250|257|6|13|GeneID:2931302)		pa14_61250(lysP|-18.69280|0.03192169|164|181|1|15|GeneID:4382995)	caul_1399(N/A|-10.96740|0.3375374|27|43|1|15|GeneID:5898854)	eam_2410(ansP|-13.36310|0.1402427|239|279|1|39|GeneID:8950029)	sl1344_4265(yjeH|-13.95950|0.009812732|1811|230|2|38|N/A)	sven_4322(N/A|-10.79490|0.03668777|172|198|12|37|GeneID:13819480)	amino acid permease		1
0.3730207 29	ppa0088(N/A|-7.67357|0.08947679|43|110|4|41|GeneID:2932390)	rsp_3939(N/A|-9.37788|0.1663119|2|22|16|37|GeneID:3712064)	pa14_35370(ptxS|-17.20640|0.05788242|206|231|1|25|GeneID:4384424)	caul_3995(N/A|-18.37940|0.01584733|131|145|1|15|GeneID:5901457)	eam_2770(galR|-14.26120|0.09307622|250|260|5|15|GeneID:8950731)	sl1344_3853(rbsR|-7.42181|0.1901317|231|265|1|37|N/A)	sven_3916(N/A|-11.68310|0.02570970|126|138|26|37|GeneID:13821741)	LacI family transcriptional regulator	rcap_rcc01770 rcap_rcc00129	0

0.3 730 207 29	ppa1275(N/A|- 4.27909|0.39225 78|255|270|7|37| GeneID:2932819)	rsp_3233(N/A|- 11.47100|0.07276 126|2|9|30|37|Gen eID:3721840)	pa14_58450(dpp C|- 15.11380|0.1276 053|248|258|5|15 |GeneID:438262 5)		eam_1083(ophC| - 11.84220|0.2578 128|270|283|1|15 |GeneID:895160 2)	sl1344_1624(sapC|- 8.89585|0.10 39556|135|16 6|4|39|N/A)	sven_4761(N/A|- 16.00580|0.00469 8862|269|300|5|3 8|GeneID:138242 16)	ABC transport er permeas e	rcap_rcc 02498 rcap_rcc 01090 rcap_rcc 02277 rcap_rcc 00704 rcap_rcc 00848	1
0.3 730 207 29	ppa0987(N/A|- 9.83058|0.03150 663|213|252|4|41| GeneID:2932775)		pa14_72960(N/ A|- 18.58970|0.0332 935|233|245|1|15 |GeneID:438239 1)	caul_1092(N/A|- 12.23240|0.2273 394|102|125|2|26| GeneID:5898547)	eam_1281(iolT|- 20.40050|0.0015 9083|236|254|2|1 7|GeneID:89494 41)	sl1344_1070(N/A|- 12.21450|0.0 2307250|267| 297|7|37|N/A)	sven_0301(N/A|- 5.75919|0.266669 1|181|188|31|38|G eneID:13817407)			3
0.4 139 442 71		rsp_0194(N/A|- 7.98027|0.271116 7|272|282|27|38|G eneID:3719483)	pa14_57850(N/ A|- 7.34773|0.89867 57|255|262|8|15|	caul_1576(N/A|- 9.00808|0.55416 61|246|262|2|17|	eam_3095(N/A|- 6.96830|0.83714 53|67|88|1|23|Ge neID:8950379)	sl1344_3283(N/A|- 18.18350|0.0 01045845|14	sven_0808(N/A|- 6.61518|0.193160 3|187|196|32|41|G eneID:13822527)	mammal ian cell entry domain-		6

			GeneID:4383003)	GeneID:5899031)		9|190|5|40|N/A)		containing protein		
0.473259	ppa2216(N/A|-0.37738|0.962989|39|45|34|40|GeneID:2933071)	rsp_0156(N/A|-10.3602 0|0.1143090|191|202|25|35|GeneID:3719540)	pa14_43420(N/A|-15.54790|0.1088684|202|228|1|25|GeneID:4381648)	caul_3157(N/A|-14.72770|0.08935224|226|233|1|8|GeneID:5900612)		sl1344_0074(caiA|-11.25800|0.03622612|240|263|11|31|N/A)	sven_2560(N/A|-12.30240|0.02008377|103|133|4|41|GeneID:13819016)	glutaryl-CoA dehydrogenase	rcap_rcc03306 rcap_rcc01510 rcap_rcc01564	3
0.560614558	ppa0925(N/A|-9.56326|0.03593175|71|81|30|40|GeneID:2932143)		pa14_38910(N/A|-16.84720|0.06657498|107|133|1|24|GeneID:4382777)	caul_0622(N/A|-13.20160|0.1619469|66|80|2|15|GeneID:5898077)	eam_2412(N/A|-9.58418|0.5187247|10|54|1|41|GeneID:8950030)	sl1344_2443(narQ|-8.56027|0.1197804|185|191|32|38|N/A)	sven_6117(N/A|-13.70300|0.01153373|16|39|9|38|GeneID:13820050)	signal transduction histidine kinase		1
0.724026224		rsp_3318(acrB|-8.48453|0.2287842|176|193|26|37|GeneID:3722020)	pa14_18780(N/A|-15.80550|0.09894157|284|300|5|2	caul_5122(N/A|-17.39900|0.02638257|155|172|1|1	eam_2185(mdtC|-13.02060|0.1623645|250|285|1|40	sl1344_0468(acrB|-6.48598|0.27	sven_1821(N/A|-12.16650|0.02120049|41|71|5|41|GeneID:13820945)	acriflavine resistanc	rcap_rcc02895 rcap_rcc00432	5

			3|GeneID:4381487)	6|GeneID:5897408)	|GeneID:8951419)	09226|268|284|8|35|N/A)		e protein B		
0.724026224	ppa0230(N/A|-9.01651|0.04694055|118|150|8|35|GeneID:2932178)	rsp_1905(N/A|-7.78618|0.2888122|182|188|34|40|GeneID:3719211)	pa14_58750(pilB|-14.66430|0.1499209|173|193|1|20|GeneID:4382648)	caul_5330(N/A|-12.65260|0.1969893|25|33|1|9|GeneID:5897166)	eam_2838(N/A|-10.05110|0.4584423|136|153|5|23|GeneID:8950230)	sl1344_0143(hofB|-2.34203|0.823676|108|114|26|32|N/A)	sven_4680(N/A|-5.08571|0.339939|189|211|14|35|GeneID:13824135)	type II secretion system protein E		1
0.724026224		rsp_1938(N/A|-14.27690|0.02062000|16|39|4|40|GeneID:3719248)		caul_1751(N/A|-12.26000|0.2252501|3|16|2|14|GeneID:5899206)		sl1344_0988(N/A|-9.59640|0.07679535|24|42|11|30|N/A)	sven_5048(N/A|-19.18820|0.001401717|131|168|2|40|GeneID:13824503)	kinetochore Spc7 domain-containing protein		6
0.724026224		rsp_3317(N/A|-16.63450|0.006258299|21|31|27|37|GeneID:3722019)	pa14_16790(N/A|-12.57180|0.3010704|290|299|2|11	caul_1312(N/A|-18.06920|0.01869403|71|89|1|21|GeneID:5898767)	eam_1729(N/A|-10.70250|0.3787321|78|90|9|21|GeneID:8949660)		sven_3803(N/A|-17.08590|0.003103842|95|127|9|41|GeneID:13821628)	TetR family transcriptional regulator		3

			|GeneID:4381602)							
0.724026224		rsp_2218(N/A|-14.87160|0.01544566|138|146|27|35|GeneID:3719747)	pa14_27180(N/A|-14.29570|0.1706464|279|292|1|14|GeneID:4380901)	caul_1274(N/A|-16.35970|0.04361131|269|278|1|10|GeneID:5898729)	eam_1228(N/A|-14.78840|0.07187974|261|271|5|15|GeneID:8950922)	sl1344_0812(ybiS|-4.13045|0.5709006|59|74|17|33|N/A)		ErfK/YbiS/YcfS/YnhG family protein/ Tat domain-containing protein	rcap_rcc02469 rcap_rcc00430 rcap_rcc01589	5
0.724026224		rsp_2413(lig2|-9.10013|0.1840824|184|190|31|37|GeneID:3720010)	pa14_36910(ligD|-11.80090|0.3786781|5|17|9|21|GeneID:4385768)	caul_1769(ligD|-11.36750|0.2996747|103|113|1|11|GeneID:5899224)			sven_7275(N/A|-6.48645|0.2029174|179|194|25|41|GeneID:13823803)			7

0.7 240 262 24		rsp_3439(N/A|- 9.81950|0.140933 2|226|234|5|13|Ge neID:3721725)	pa14_14940(N/ A|- 11.88350|0.3698 014|94|121|2|24| GeneID:438456 4)	caul_0857(N/A|- 12.82690|0.1853 070|190|200|2|12| GeneID:5898312)	eam_0882(N/A|- 10.52880|0.3993 566|158|169|7|20 |GeneID:894923 4)	sl1344_0591(N/A|- 7.02268|0.22 18321|37|46| 27|37|N/A)	sven_4337(N/A|- 4.04776|0.480386 6|62|68|5|11|Gene ID:13819495)	aspartate aminotra nsferase	rcap_rcc 01471	2
0.7 240 262 24		rsp_1041(N/A|- 10.77660|0.09682 305|105|132|15|37 |GeneID:3720951)	pa14_14060(N/ A|- 12.41480|0.3159 062|202|221|1|20 |GeneID:438534 8)	caul_4604(N/A|- 16.32620|0.0442 9653|120|143|3|2 4|GeneID:59020 66)	eam_1956(fadD| - 9.42168|0.54006 95|80|92|1|12|Ge neID:8951878)	sl1344_0072(caiC|- 9.52451|0.07 9252|249|285 |7|37|N/A)	sven_4199(N/A|- 13.43350|0.01282 679|152|161|32|4 1|GeneID:138193 57)	AMP- depende nt syntheta se and ligase	rcap_rcc 03140	2
0.7 240 262 24			pa14_13610(N/ A|- 19.62010|0.0217 5369|280|290|1|1 1|GeneID:43819 33)	caul_1376(N/A|- 13.79100|0.1298 355|107|116|1|10| GeneID:5898831)		sl1344_2142(yehY|- 8.43206|0.12 63657|138|15 1|20|33|N/A)	sven_3246(N/A|- 11.12070|0.03219 651|137|146|32|4 1|GeneID:138236 66)			7
0.7 240			pa14_22020(min D|-		eam_1953(minD |-	sl1344_1743(minD|-				8

26224			10.53840|0.5284824|243|266|1|24|GeneID:4381242)		12.08900|0.2352757|79|90|10|20|GeneID:8951448)	8.93401|0.1022793|3|25|5|32|N/A)				
0.724026224		rsp_0378(N/A|-8.22741|0.2497140|92|114|12|38|GeneID:3718886)	pa14_70100(N/A|-15.42820|0.1137752|92|104|2|14|GeneID:4384011)	caul_3943(N/A|-12.98540|0.1751367|286|293|2|9|GeneID:5901405)		sl1344_3102(ordL|-1.03005|0.9447407|95|101|16|22|N/A)		FAD dependent oxidoreductase		8
0.724026224		rsp_0245(N/A|-8.55580|0.2232219|87|101|28|41|GeneID:3719387)	pa14_47300(phnW|-8.83355|0.7466459|156|189|3|39|GeneID:4381782)	caul_3706(N/A|-10.14310|0.4234135|67|73|3|9|GeneID:5901162)	eam_0876(N/A|-9.48478|0.5317665|231|240|2|11|GeneID:8949228)		sven_1506(N/A|-2.53699|0.7224968|111|124|26|40|GeneID:13820630)	serine--glyoxylate aminotransferase/ alanine--glyoxylate aminotransferase/		7

								serine--pyruvate aminotransferase		
0.7240262424			pa14_34440(mtlR|-14.66650|0.149804|133|149|1|15|GeneID:4380398)		eam_1701(araC|-11.60270|0.2809705|202|226|2|25|GeneID:8950647)	sl1344_2012(pocR|-10.71400|0.04654916|249|262|26|38|N/A)		two component AraC family transcriptional regulator		8
0.7240262424		rsp_3906(N/A|-13.03140|0.03684418|253|262|30|39|GeneID:4796440)		caul_2570(N/A|-14.83850|0.08533354|250|260|6|16|GeneID:5900025)		sl1344_1054(scsC|-7.12185|0.2135830|20|26|32|38|N/A)	sven_1688(N/A|-11.53440|0.02728388|144|164|5|35|GeneID:13820812)	DSBA family oxidoreductase	rcap_rcc01743	7
0.7240	ppa1650(dnaE2|-4.28476|0.39142		pa14_55610(dnaE2|-11.08350|0.4608994|108|130|1|25	caul_4461(N/A|-10.31730|0.4044667|262|272|1|10|			sven_1356(N/A|-9.05794|0.073669			8

26224	03|40|50|27|37|GeneID:2931181)		|GeneID:4384917)	GeneID:5901922)			94|140|184|5|41|GeneID:13823075)			
0.724026224	ppa0126(N/A|-3.48541|0.5196791|19|25|4|10|GeneID:2933151)		pa14_62200(mrcB|-15.10440|0.1280405|220|233|2|15|GeneID:4384639)	caul_1078(N/A|-15.68150|0.0593682|168|180|1|16|GeneID:5898533)	eam_0797(mrcB|-7.96052|0.728593|19|27|33|41|GeneID:8951257)	sl1344_0191(mrcB|-10.43990|0.05272819|204|238|5|37|N/A)	sven_4705(N/A|-8.17791|0.1047022|43|51|4|12|GeneID:13824160)	1A family penicillin-binding protein		1
0.724026224		rsp_3534(N/A|-15.08840|0.01387461|144|155|30|41|GeneID:3721948)	pa14_01620(aptA|-14.04220|0.1862613|273|288|1|16|GeneID:4384249)	caul_3957(N/A|-11.70150|0.270171|185|200|3|18|GeneID:5901419)			sven_5316(N/A|-7.39507|0.1426791|183|216|8|37|GeneID:13821843)	class III aminotransferase	rcap_rcc02240 rcap_rcc02598	6
0.724026224		rsp_2809(N/A|-7.04792|0.3631266|17|32|26|41|GeneID:3720583)	pa14_22680(N/A|-16.66760|0.0713556|248|260|5|15	caul_1329(N/A|-16.17280|0.04755181|237|254|1|18|GeneID:5898784)	eam_3096(N/A|-12.26920|0.2196776|63|71|6|15|GeneID:8950380)	sl1344_3284(N/A|-7.40625|0.1912939|257|274|24|40|N/A)		ABC transporter permease		5

			|GeneID:4381205)							
0.724026224	ppa1220(N/A|-11.92630|0.01109911|176|196|18|38|GeneID:2931311)	rsp_2719(N/A|-8.47330|0.2296701|275|294|19|41|GeneID:3720449)	pa14_30770(N/A|-15.51660|0.1101330|114|124|7|16|GeneID:4380607)		eam_0002(asnC|-13.16290|0.1528792|117|127|1|11|GeneID:8951451)	sl1344_3843(asnC|-2.80225|0.7644538|226|236|12|23|N/A)	sven_2685(N/A|-12.30280|0.02008058|64|93|6|36|GeneID:13823105)	AsnC/Lrp family transcriptional regulator		4
0.724026224			pa14_13600(N/A|-8.89377|0.7393182|2|13|5|15|GeneID:4381934)	caul_1376(N/A|-13.79100|0.1298355|107|116|1|10|GeneID:5898831)		sl1344_1423(N/A|-6.60978|0.2589218|129|154|9|38|N/A)	sven_3244(N/A|-4.13553|0.4673202|189|212|9|36|GeneID:13823664)			6
0.724026224		rsp_3489(N/A|-8.08794|0.2616368|33|40|34|41|GeneID:3721905)	pa14_67230(N/A|-19.19920|0.02591887|197|219|1|23|GeneID:4383402)	caul_1564(N/A|-5.45907|0.9096026|266|275|1|10|GeneID:5899019)	eam_0594(N/A|-16.27060|0.03225042|137|167|1|21|GeneID:8949097)	sl1344_0284(vrgS|-7.56158|0.1799520|240|250|12|22|N/A)	sven_5841(N/A|-1.42474|0.8798246|94|100|33|39|GeneID:13822368)			8

0.7 240 262 24	ppa0053(N/A|- 7.69558|0.08855 017|14|27|1|13|G eneID:2932665)	rsp_0777(LolE|- 7.71866|0.295150 7|161|170|27|37|G eneID:3718143)	pa14_25430(N/ A|- 12.28430|0.3286 204|200|216|5|20 |GeneID:438101 5)	caul_2978(N/A|- 9.69059|0.47427 53|254|261|1|8|G eneID:5900433)	eam_1505(lolC|- 12.99490|0.1641 229|160|170|1|11 |GeneID:894954 2)	sl1344_1156(N/A|- 7.82874|0.16 17605|6|17|1 5|28|N/A)	sven_3306(N/A|- 4.73386|0.383919 8|91|97|31|37|Gen eID:13823726)	lipoprote in- releasing system transme mbrane protein LolE		3
0.7 240 262 24		rsp_0317(hemN|- 3.85397|0.757351 4|281|287|8|14|Ge neID:3719077)	pa14_44470(he mN|- 16.14750|0.0870 1628|252|269|1|1 6|GeneID:43830 37)	caul_3872(N/A|- 12.95570|0.1770 099|233|240|2|9| GeneID:5901334)	eam_0029(hem N|- 25.99900|8.3007 05e- 08|254|266|5|17| GeneID:894877 1)			oxygen- indepen dent copropor phyrinog en-III oxidase		7
0.7 240 262 24					eam_1541(dcuR| - 9.50656|0.52890 46|4|17|2|15|Gen eID:8949562)	sl1344_0614(citB|- 8.02041|0.14 96916|84|92| 1|9|N/A)	sven_1352(N/A|- 10.74820|0.03738 141|85|105|20|40| GeneID:1382307 1)			8

0.7 240 262 24	ppa0739(N/A|- 6.93046|0.12663 29|158|192|5|37| GeneID:2931458)		pa14_34640(N/ A|- 10.78560|0.4973 922|252|263|8|19 |GeneID:438436 5)	caul_2070(N/A|- 10.99470|0.3348 693|161|177|1|17| GeneID:5899525)	eam_3275(gntK| - 5.99451|0.91564 73|248|260|2|17| GeneID:895116 8)	sl1344_3508(gntK|- 17.01100|0.0 01996073|18 8|211|15|37|N /A)	sven_1278(N/A|- 4.75733|0.380865 2|76|82|31|37|Gen eID:13822997)			6
0.7 240 262 24	ppa1175(aspS|- 3.93402|0.44530 78|107|113|4|10| GeneID:2932005)	rsp_0815(aspS|- 7.15197|0.351999 5|98|115|11|37|Ge neID:3718393)	pa14_51820(asp S|- 17.98410|0.0425 2817|272|291|1|2 4|GeneID:43844 51)	caul_2557(aspS|- 14.43470|0.1007 282|212|219|1|8| GeneID:5900012)	eam_2008(aspS|- 15.42800|0.0515 807|269|288|1|24 |GeneID:895156 9)	sl1344_1836(aspS|- 5.13118|0.42 87243|259|29 2|6|38|N/A)	sven_3561(N/A|- 8.97046|0.076297 72|55|81|6|37|Gen eID:13821386)	aspartyl- tRNA syntheta se		0
0.7 240 262 24	ppa0776(N/A|- 3.15180|0.57788 38|34|56|5|38|Ge neID:2931878)	rsp_1020(glcD|- 8.18966|0.252901 4|96|108|27|37|Ge neID:3720987)	pa14_04140(N/ A|- 12.87800|0.2735 895|127|156|1|23 |GeneID:438372 2)	caul_4478(N/A|- 23.37440|0.0005 73618|182|199|1| 20|GeneID:5901 939)	eam_1235(dld|- 12.48130|0.2022 447|253|293|1|29 |GeneID:894941 1)	sl1344_2145(dld|- 2.48495|0.80 6031|123|129 |20|26|N/A)	sven_1412(N/A|- 6.31814|0.216340 5|137|149|26|37|G eneID:13820536)	D- lactate dehydro genase	rcap_rcc 02871 rcap_rcc 01174 rcap_rcc 03093	1
0.7 240		rsp_2006(amsA|- 6.44171|0.431842|	pa14_67975(N/ A|-	caul_4644(N/A|- 9.78912|0.46301	eam_3377(N/A|- 18.02600|0.0105	sl1344_2097(asmA|-				8

26224		24|30|35|41|GeneID:3719339)	10.56540|0.5250563|55|72|1|15|GeneID:4384257)	99|135|147|1|15|GeneID:5902106)	9071|53|67|1|15|GeneID:8950491)	9.35130|0.08546624|241|256|24|38|N/A)				
0.7240262 24				caul_5446(N/A|-13.87220|0.1258339|18|42|2|24|GeneID:5897118)			sven_3410(N/A|-6.53911|0.1988736|103|114|11|22|GeneID:13821235)			8
0.7240262 24		rsp_2653(N/A|-9.32041|0.1698710|29|64|6|40|GeneID:3720344)	pa14_29130(N/A|-13.23390|0.2440542|265|279|2|15|GeneID:4384678)	caul_2894(N/A|-15.48420|0.06476044|61|77|2|16|GeneID:5900349)			sven_6130(N/A|-9.69737|0.05699969|183|193|32|41|GeneID:13820063)	ATPase AAA		8
0.7240262 24	ppa2064(N/A|-3.37622|0.5385191|64|92|7|31|GeneID:2931073)	rsp_3030(N/A|-10.43800|0.1108520|122|138|7|22|GeneID:3721615)	pa14_41140(N/A|-11.28560|0.4368646|274|294|6|21		eam_1288(N/A|-15.38870|0.05267461|149|179|5|39|GeneID:8949445)	sl1344_3595(dppB|-10.45850|0.05228603|68|83|25|40|N/A)	sven_4822(N/A|-6.06752|0.2377837|12|50|3|40|GeneID:13824277)	glutathione import ABC transporter	rcap_rcc02499 rcap_rcc00153 rcap_rcc00102	1

			|GeneID:4383091)					permease GsiC	rcap_rcc02276 rcap_rcc00705 rcap_rcc00847	
0.724026224	ppa1028(N/A|-3.77851|0.4704649|84|103|7|37|GeneID:2931757)	rsp_3980(N/A|-6.06144|0.4779095|126|132|33|39|GeneID:3711817)	pa14_21220(N/A|-13.58340|0.2175204|38|48|5|15|GeneID:4385422)	caul_1934(N/A|-7.21066|0.7606924|14|20|33|39|GeneID:5899389)	eam_3325(uspA|--14.60000|0.07895542|9|21|1|13|GeneID:8952287)	sl1344_0602(N/A|-3.09235|0.7240773|44|67|6|31|N/A)	sven_2578(N/A|-11.33150|0.02959059|28|62|5|37|GeneID:13819034)	universal stress family protein		0
0.724026224		rsp_0851(N/A|-18.87630|0.00177398|14|36|12|39|GeneID:3718317)	pa14_72140(N/A|-11.81050|0.3776398|168|190|1|20|GeneID:4384999)	caul_3529(N/A|-15.44150|0.06597963|185|207|1|23|GeneID:5900984)				membrane transport family protein	rcap_rcc01247	6
0.724	ppa2240(N/A|-4.69679|0.3337104|114|123|5|14|	rsp_2890(N/A|-4.57809|0.668090	pa14_13170(N/A|-19.07580|0.0272	caul_2307(N/A|-13.59030|0.1401559|205|213|1|9|	eam_1044(copA|--13.51760|0.1310	sl1344_0491(ybaR|-6.55589|0.26	sven_2534(N/A|-8.08171|0.108783	copper-transporting P-		1

26224	GeneID:2931784)	5|161|168|33|40|GeneID:3720630)	7534|38|51|2|15|GeneID:4381968)	GeneID:5899762)	432|67|88|1|24|GeneID:8949319)	40953|167|179|11|23|N/A)	3|237|253|9|28|GeneID:13818990)	type ATPase		
0.724026224	ppa0210(N/A|-5.79832|0.2106601|226|232|34|40|GeneID:2932885)			caul_2678(N/A|-11.68050|0.2719683|159|175|1|19|GeneID:5900133)		sl1344_4476(N/A|-11.42610|0.03349607|188|219|8|35|N/A)	sven_3262(N/A|-3.02919|0.642462|203|210|34|41|GeneID:13823682)		rcap_rcc02137	6
0.724026224			pa14_56640(N/A|-14.10680|0.1821732|26|48|2|24|GeneID:4382459)	caul_5131(N/A|-18.82540|0.0124133|99|115|1|15|GeneID:5897357)	eam_2773(N/A|-9.29188|0.5571907|152|162|5|15|GeneID:8951403)	sl1344_0252(N/A|-2.29950|0.8287822|88|94|8|14|N/A)	sven_0177(N/A|-4.23223|0.4531548|270|276|1|7|GeneID:13817284)			7
0.724026224		rsp_1909(N/A|-8.70008|0.2122759|54|73|3|22|GeneID:3719214)	pa14_23970(xcpQ|-14.19470|0.1767310|195|233|1|39	caul_4193(N/A|-12.11490|0.2363861|148|162|2|16|GeneID:5901655)	eam_3241(hofQ|-13.70950|0.1202669|16|31|1|15|GeneID:8950671)	sl1344_3455(hofQ|-2.30866|0.8276882|91|107|26|37|N/A)		type II and III secretion system protein		6

			|GeneID:4381105)							
0.724026224	ppa1658(N/A|-6.20065|0.176401|132|138|4|10|GeneID:2932659)	rsp_0093(smoG|-6.89505|0.379844|104|112|27|35|GeneID:3719924)	pa14_34390(N/A|-8.18486|0.820749|8|19|28|5|14|GeneID:4380401)			sl1344_4163(malG|-5.51903|0.3787692|100|126|8|37|N/A)	sven_2711(N/A|-6.22939|0.2237312|189|206|11|29|GeneID:13823131)	polyols ABC transporter permease		2
0.724026224			pa14_28910(N/A|-12.15100|0.3419627|64|77|1|14|GeneID:4380737)		eam_3060(nrdG|-11.51920|0.2893381|239|248|5|14|GeneID:8951559)	sl1344_4064(N/A|-12.65060|0.01870471|236|247|30|41|N/A)		[pyruvate formate-lyase]-activating enzyme	rcap_rcc03344 rcap_rcc02865	8
0.724026224		rsp_2598(N/A|-11.88110|0.0611557|243|255|26|37|GeneID:3720270)	pa14_02340(N/A|-15.87920|0.09625437|239|249|5|1	caul_2657(N/A|-9.16029|0.5361353|64|70|1|7|GeneID:5900112)	eam_1362(ssuC|-17.81390|0.01224879|27|47|1|22|		sven_0779(N/A|-4.71586|0.3862741|292|299|9|16|GeneID:13822498)	taurine ABC transporter	rcap_rcc01403 rcap_rcc02223	6

			5|GeneID:4383585)		GeneID:8952073)			permease TauC		
0.724026224	ppa0396(N/A|-7.93383|0.07908238|184|190|3|9|GeneID:2931066)	rsp_0361(sda|-3.46076|0.8018218|83|89|8|14|GeneID:3718984)	pa14_33030(sdaA|-17.47270|0.05212607|240|260|1|24|GeneID:4380531)	caul_0654(N/A|-7.40125|0.74039|09|210|223|1|16|GeneID:5898109)	eam_1964(sdaA|-9.12402|0.579372|251|265|9|21|GeneID:8949797)	sl1344_3212(tdcG|-5.68624|0.3583614|29|41|11|22|N/A)		L-serine ammonia-lyase		2
0.724026224	ppa1934(N/A|-5.17614|0.2746268|245|251|5|11|GeneID:2931326)	rsp_2514(nuoC|-8.17084|0.2545015|212|219|5|12|GeneID:3720131)	pa14_29990(nuoD|-14.59750|0.1535089|92|106|1|15|GeneID:4380670)	caul_2834(N/A|-10.10790|0.4272882|227|236|1|10|GeneID:5900289)	eam_2290(nuoC|-15.36670|0.05329523|141|149|7|15|GeneID:8951577)	sl1344_2295(nuoC|-2.07605|0.8544407|215|227|11|21|N/A)	sven_4267(N/A|-5.13146|0.3345081|262|268|31|37|GeneID:13819425)			6
0.724026224	ppa1651(N/A|-5.27403|0.2636283|232|238|4|10|GeneID:2932956)	rsp_1458(N/A|-6.61538|0.4115143|239|264|12|37|GeneID:3718752)	pa14_55600(N/A|-10.14190|0.5794357|52|62|5|15|GeneID:4384918)	caul_5202(N/A|-14.67500|0.0913172|94|132|1|39|GeneID:5897254)				nucleotidyltransferase/DNA polymerase		7

								involved in DNA repair		
0.7 240 262 24		rsp_1336(N/A|-8.13221|0.257808 7|76|86|28|37|Gen eID:3720862)	pa14_50460(flg D|- 6.49202|0.95295 24|29|35|10|16|G eneID:4380044)	caul_1009(N/A|- 19.71110|0.0074 53776|120|136|1| 15|GeneID:5898 464)	eam_2553(flgD|- 16.78490|0.0237 4567|105|120|1|1 6|GeneID:89509 91)	sl1344_1113(flgD|- 6.84867|0.23 6918|264|275 |11|22|N/A)		flagellar hook capping protein		5
0.7 240 262 24		rsp_0017(N/A|- 9.47525|0.160419 2|191|207|19|36|G eneID:3720152)	pa14_03950(spu G|- 13.75510|0.2053 574|44|58|6|21|G eneID:4383707)	caul_3950(N/A|- 9.03985|0.55039 65|224|242|2|20| GeneID:5901412)	eam_1303(potH| - 9.05225|0.58885 39|229|235|8|14| GeneID:895169 2)	sl1344_1162(potB|- 9.22392|0.09 031362|58|80 |16|37|N/A)		polyamine ABC transporter permease PotH	rcap_rcc 02267 rcap_rcc 02450 rcap_rcc 01895 rcap_rcc 01390	5
0.7 240		rsp_0218(mcpM|- 4.62436|0.662173	pa14_38970(N/ A|- 12.98550|0.2643	caul_0236(N/A|- 11.03710|0.3307 496|269|281|2|15|				sensor histidine kinase/response	rcap_rcc 02887 rcap_rcc 03014	8

26224		4|3|24|6|39|GeneID:3719377)	97|6|21|1|15|GeneID:4382773)	GeneID:5897510)				regulator receiver protein	rcap_rcc03425 rcap_rcc01364 rcap_rcc01749	
0.724026224		rsp_2561(exoP|-13.70170|0.02707024|130|152|8|32|GeneID:3720195)		caul_4816(N/A|-7.11062|0.771093|65|87|1|23|GeneID:5902278)	eam_2171(amsA|-7.83711|0.743359|17|27|1|10|GeneID:8950622)	sl1344_2093(wzc|-6.69353|0.2510356|228|260|5|37|N/A)		lipopolysaccharide biosynthesis family protein	rcap_rcc01958	5
0.724026224	ppa2152(N/A|-0.31315|0.9665639|240|246|21|27|GeneID:2932358)	rsp_1481(OppA|-12.01330|0.05777922|156|189|5|38|GeneID:3718717)			eam_1831(mppA|-12.13560|0.2311725|69|81|6|19|GeneID:8949733)	sl1344_1677(oppA|-4.53446|0.5117175|15|46|13|37|N/A)	sven_4759(N/A|-10.15230|0.04748216|268|276|32|40|GeneID:13824214)			4
0.724	ppa1953(N/A|-0.50291|0.95504		pa14_20870(N/A|-	caul_3155(N/A|-11.62200|0.2770		sl1344_1685(N/A|-				6

26224	55|189|210|8|30|GeneID:2933166)		16.11830|0.08798144|242|253|5|17|GeneID:4384598)	161|179|210|1|24|GeneID:5900610)		4.18810|0.5623606|219|225|27|33|N/A)				
0.724026224	ppa1924(N/A|-4.14011|0.4131399|13|19|16|22|GeneID:2931952)	rsp_2527(nuoL|-9.49549|0.1592158|271|283|27|37|GeneID:3720144)	pa14_29880(nuoL|-10.51700|0.5312026|26|35|5|15|GeneID:4380678)	caul_2821(N/A|-17.71540|0.02246901|219|235|2|17|GeneID:5900276)	eam_2282(nuoL|-8.05599|0.7169743|222|234|5|17|GeneID:8951585)	sl1344_2287(nuoL|-8.80207|0.1081794|265|296|11|37|N/A)	sven_4276(N/A|-3.42505|0.5778314|233|239|7|13|GeneID:13819434)	NADH-quinone oxidoreductase subunit L		4
0.724026224		rsp_0611(N/A|-6.61452|0.4116138|221|242|11|37|GeneID:3718239)		caul_0523(N/A|-9.66126|0.4776428|44|51|2|9|GeneID:5897978)		sl1344_4335(N/A|-9.04300|0.09762381|7|15|2|10|N/A)	sven_3283(N/A|-8.72815|0.08406797|106|130|5|35|GeneID:13823703)	HxlR family transcriptional regulator		6
0.724026224		rsp_0886(TyrB|-8.36600|0.2382641|59|72|5|20|GeneID:3718353)	pa14_23500(tyrB|-13.74330|0.2061753|144|151|5|12		eam_1356(aspC|-18.02750|0.0105797|222|234|1|12	sl1344_0935(aspC|-7.88116|0.15		aromatic-amino-acid		8

			|GeneID:4381139)		|GeneID:8949474)	83801|242|281|5|35|N/A)		aminotransferase		
0.724026224			pa14_45070(N/A|-12.84410|0.2765373|241|266|2|21|GeneID:4383162)	caul_4869(N/A|-8.13208|0.6578857|84|100|2|19|GeneID:5902331)		sl1344_1672(N/A|-6.56998|0.2627351|231|237|1|7|N/A)	sven_1132(N/A|-7.37423|0.1438505|239|245|32|38|GeneID:13822851)	ion transport 2 family protein		8
0.724026224			pa14_21530(N/A|-12.94160|0.2681224|209|218|5|14|GeneID:4381306)	caul_2929(N/A|-12.56540|0.203032|216|225|5|14|GeneID:5900384)	eam_3344(N/A|-13.19140|0.1510301|36|62|2|27|GeneID:8950470)		sven_7088(N/A|-11.67460|0.02579715|127|160|10|41|GeneID:13818354)			8
0.724026224		rsp_3167(N/A|-8.08810|0.2616229|35|57|8|31|GeneID:3721459)	pa14_34780(N/A|-12.92490|0.2695499|208|223|1|15	caul_2658(N/A|-13.05400|0.1708669|225|245|1|17|GeneID:5900113)	eam_1361(ssuB|-8.10167|0.7113584|18|31|6|16|GeneID:8951652)		sven_6762(N/A|-8.14118|0.1062426|131|160|8|38|GeneID:13818028)	pyrimidine ABC transporter ATP-	rcap_rcc02225 rcap_rcc01404	4

			|GeneID:4384353)					binding protein	rcap_rcc02242	
0.724026224	ppa2077(N/A|-0.89855|0.9230532|85|95|22|33|GeneID:2933086)	rsp_0014(N/A|-3.30433|0.8184565|278|292|23|35|GeneID:3720149)	pa14_25800(N/A|-14.69560|0.1482646|78|108|2|36|GeneID:4384338)	caul_3929(N/A|-17.48260|0.02529564|197|236|1|39|GeneID:5901391)	eam_0144(N/A|-10.45360|0.4084389|247|268|1|21|GeneID:8952077)		sven_3912(N/A|-9.99117|0.05065621|80|125|5|41|GeneID:13821737)	TetR family transcriptional regulator	rcap_rcc03032	4
0.724026224				caul_2346(N/A|-9.75906|0.4664441|219|231|1|12|GeneID:5899801)	eam_1986(N/A|-18.78230|0.006129386|89|113|1|26|GeneID:8949814)		sven_0824(N/A|-10.83580|0.03609094|248|282|8|37|GeneID:13822543)			8
0.724026224	ppa1538(N/A|-2.37782|0.7150806|269|285|29|41|GeneID:2931203)	rsp_2895(N/A|-17.71000|0.003475886|19|56|5|41|GeneID:3720634)	pa14_39130(N/A|-14.54050|0.1566284|233|261|1|28|GeneID:4385662)	caul_2537(N/A|-15.28220|0.0706979|163|179|1|16|GeneID:5899992)	eam_1282(N/A|-11.48490|0.2928189|265|285|1|21|GeneID:8951322)	sl1344_0813(ybiT|-4.87454|0.4636108|274|300|7|37|N/A)	sven_1486(N/A|-2.96046|0.6537365|22|28|8|14|GeneID:13820610)	ABC transporter ATP-binding protein		0

0.7 240 262 24	ppa0157(N/A|-5.60405|0.2291449|81|108|9|41|GeneID:2932462)		pa14_69070(N/A|-12.68020|0.2911218|154|189|1|39|GeneID:4382439)	caul_1950(N/A|-4.77427|0.9459706|67|73|27|33|GeneID:5899405)		sl1344_0793(ybhF|-6.71269|0.2492577|244|250|33|39|N/A)	sven_1441(N/A|-8.90908|0.0781964|5|198|206|32|40|GeneID:13820565)			5
0.7 240 262 24	ppa0437(N/A|-2.67109|0.6635097|26|32|32|38|GeneID:2933179)	rsp_2831(cobO|-13.90570|0.0245931|60|72|30|41|GeneID:3720558)	pa14_47790(cobO|-10.16490|0.5764546|281|293|2|15|GeneID:4384692)	caul_2737(N/A|-14.88560|0.0836706|96|111|1|18|GeneID:5900192)		sl1344_1650(btuR|-6.66062|0.2541123|245|252|2|9|N/A)	sven_1497(N/A|-2.75791|0.6868506|206|215|28|37|GeneID:13820621)	cob(I)yrinic acid a c-diamide adenosyltransferase		4
0.7 240 262 24		rsp_3152(N/A|-5.74284|0.517863|62|70|26|34|GeneID:3721444)	pa14_19500(N/A|-15.83130|0.0979933|280|292|5|15|GeneID:4385235)	caul_4454(N/A|-9.83007|0.4583694|97|107|3|15|GeneID:5901915)	eam_1765(N/A|-14.96340|0.0657742|133|153|1|27|GeneID:8951500)		sven_6755(N/A|-6.38146|0.2112004|166|174|15|23|GeneID:13818021)	ABC transporter substrate-binding protein	rcap_rcc02224 rcap_rcc02241	4

0.7 240 262 24	ppa0295(N/A|- 5.19982|0.27193 34|257|264|6|13| GeneID:2932258)	rsp_1373(N/A|- 0.39633|0.987372 2|202|215|24|38|G eneID:3720819)	pa14_44190(N/ A|- 14.99620|0.1331 449|33|43|5|15|G eneID:4382788)	caul_2075(N/A|- 16.58510|0.0392 3026|133|152|1|2 0|GeneID:58995 30)	eam_2451(N/A|- 15.45620|0.0508 0729|34|46|1|13| GeneID:895005 0)	sl1344_4345(N/A|- 10.06660|0.0 6236464|180| 208|8|37|N/A)	sven_6558(N/A|- 4.08948|0.474151 5|271|278|34|41|G eneID:13820491)			3
0.7 240 262 24	ppa0980(N/A|- 3.53539|0.51113 87|126|132|8|14| GeneID:2932293)	rsp_2464(fabF|- 14.22420|0.02114 692|125|138|28|40 |GeneID:3720079)	pa14_43690(fab B|- 12.58090|0.3002 26|148|156|5|13| GeneID:438176 2)	caul_5061(N/A|- 15.64090|0.0604 4599|90|115|5|39| GeneID:5902523)	eam_1460(fabF|- 8.19634|0.69961 11|226|236|5|15| GeneID:895059 2)	sl1344_2347(fabB|- 8.10017|0.14 49017|17|24| 1|8|N/A)	sven_6796(N/A|- 8.79342|0.081901 51|246|273|5|38|G eneID:13818062)	3- oxoacyl- ACP synthase II	rcap_rcc 01870 rcap_rcc 02668	1
0.7 240 262 24		rsp_3138(N/A|- 5.46199|0.553815 7|230|237|11|18|G eneID:3721430)	pa14_47490(N/ A|- 9.14708|0.70787 76|268|282|1|15| GeneID:438416 0)	caul_3389(N/A|- 15.29580|0.0702 8445|112|124|2|1 5|GeneID:59008 44)	eam_2366(N/A|- 14.73990|0.0736 5039|115|146|1|3 8|GeneID:89500 07)	sl1344_2626(N/A|- 9.94914|0.06 571613|4|11| 2|9|N/A)	sven_6642(N/A|- 7.45442|0.139392 2|287|295|4|12|Ge neID:13817908)			7

0.7 240 262 24		rsp_0374(N/A|- 8.18655|0.253165 3|171|181|27|36|G eneID:3718882)	pa14_26220(his M|- 13.67910|0.2106 715|29|43|1|15|G eneID:4380971)		eam_1308(artM| - 11.06780|0.3371 168|273|283|5|15 |GeneID:895067 6)	sl1344_0864(artM|- 10.99800|0.0 4086037|186| 197|26|37|N/ A)		amino acid ABC transport er permeas e	rcap_rcc 02776	7
0.7 240 262 24		rsp_0180(N/A|- 7.19166|0.347809 9|272|278|8|14|Ge neID:3719506)	pa14_02340(N/ A|- 15.87920|0.0962 5437|239|249|5|1 5|GeneID:43835 85)	caul_1260(N/A|- 14.94460|0.0816 2533|22|31|1|10| GeneID:5898715)			sven_7015(N/A|- 6.71922|0.185587 1|102|109|33|40|G eneID:13818281)	pyrimidi ne ABC transport er permeas e	rcap_rcc 02523 rcap_rcc 02855	8
0.7 240 262 24		rsp_2254(N/A|- 11.19320|0.08167 781|101|116|10|35 |GeneID:3719783)	pa14_66080(ms bA|- 11.19660|0.4473 713|241|254|5|19 |GeneID:438549 9)	caul_4538(N/A|- 21.40680|0.0025 23240|163|200|1| 39|GeneID:5901 999)	eam_1343(msbA |- 11.19400|0.3233 37|48|54|10|16|G eneID:8949468)	sl1344_0921(msbA|- 10.71720|0.0 4648116|77|1 05|5|37|N/A)		ABC transport er ATP- binding/ permeas e	rcap_rcc 03139 rcap_rcp 00108	5

0.7 240 262 24		rsp_0440(sufB|- 13.75690|0.02638 044|85|101|25|41| GeneID:3718814)		caul_2588(N/A|- 6.61547|0.81959 98|89|96|3|10|Ge neID:5900043)	eam_1651(sufB|- 12.76980|0.1801 264|54|71|1|17|G eneID:8952082)	sl1344_1304(N/A|- 11.09170|0.0 3913008|5|30 |3|35|N/A)		FeS assembl y protein SufB		6
0.7 240 262 24	ppa2182(N/A|- 10.38840|0.0239 1668|5|38|2|35|G eneID:2932241)	rsp_2267(N/A|- 14.66150|0.01712 071|25|58|13|37|G eneID:3719796)	pa14_15680(N/ A|- 11.27350|0.4382 856|238|279|1|38 |GeneID:438130 2)	caul_4636(N/A|- 13.38680|0.1512 617|157|181|1|19| GeneID:5902098)	eam_2504(tadA| - 14.12970|0.0990 6808|122|137|5|1 8|GeneID:89500 85)		sven_3789(N/A|- 8.12544|0.106909 4|183|194|11|22|G eneID:13824037)	tRNA- specific adenosin e deamina se		2
0.7 240 262 24	ppa0953(N/A|- 2.60570|0.67511 25|176|210|8|37| GeneID:2932693)	rsp_3218(N/A|- 5.58974|0.537395 1|1|9|5|13|GeneID :3721825)	pa14_47750(N/ A|- 9.17013|0.70497 34|146|160|1|16| GeneID:438469 4)	caul_2980(N/A|- 13.15500|0.1647 240|220|228|1|9| GeneID:5900435)			sven_1153(N/A|- 4.16428|0.463082 9|81|107|4|38|Gen eID:13822872)	cob(II)yr inic acid a c- diamide reductas e		7
0.7 240		rsp_2661(N/A|- 6.58262|0.415311	pa14_49580(N/ A|- 11.70600|0.3890	caul_3781(N/A|- 11.79020|0.2626	eam_1900(adhE| - 11.88710|0.2536	sl1344_2953(fucO|- 9.34842|0.08	sven_1359(N/A|- 1.96630|0.809397	iron- containi ng	rcap_rcc 02210	2

26224		6|52|69|26|37|GeneID:3720352)	358|65|72|8|15|GeneID:4380114)	657|28|35|5|12|GeneID:5901243)	120|132|138|10|16|GeneID:8949761)	55732|245|271|13|41|N/A)	6|62|68|9|15|GeneID:13823078)	alcohol dehydrogenase		
0.724026224	ppa1034(N/A|-6.10567|0.1840182|172|178|5|11|GeneID:2931496)	rsp_2182(betI|-6.37525|0.4397457|268|277|27|35|GeneID:3719652)	pa14_70970(betI|-17.75650|0.04657704|203|241|1|38|GeneID:4385078)	caul_3936(N/A|-17.02830|0.03169823|197|213|1|17|GeneID:5901398)	eam_1685(betI|-12.90520|0.1703701|75|92|3|20|GeneID:8952144)	sl1344_1506(N/A|-6.95398|0.2276944|190|196|32|38|N/A)	sven_2214(N/A|-9.07428|0.0731891|161|178|19|38|GeneID:13818670)	transcriptional repressor BetI		5
0.724026224	ppa2113(N/A|-4.06622|0.4245124|60|66|31|37|GeneID:2932826)	rsp_3240(N/A|-6.12567|0.4699921|122|132|11|20|GeneID:3721845)	pa14_45870(N/A|-18.34950|0.03670612|132|169|1|39|GeneID:4382872)	caul_3885(N/A|-11.79450|0.2623055|249|261|1|14|GeneID:5901347)		sl1344_1034(copS|-3.48220|0.6674056|169|175|1|7|N/A)	sven_7375(N/A|-9.14612|0.07111158|43|82|5|37|GeneID:13823903)			5
0.724026224		rsp_3497(N/A|-22.13900|0.0002148921|275|286|26|	pa14_58600(N/A|-13.88500|0.1965	caul_1450(N/A|-5.22329|0.9235403|80|86|1|7|GeneID:5898905)		sl1344_1688(N/A|-6.57423|0.26	sven_4858(N/A|-7.73924|0.1245929|213|221|27|35|GeneID:13824313)			6

		37|GeneID:3721913)	275|1|18|2|18|GeneID:4382636)			23259|53|59|3|9|N/A)				
0.724026224	ppa1504(N/A|-6.69723|0.1409491|128|137|5|14|GeneID:2931946)	rsp_0908(sitD|-9.81429|0.1412128|123|136|7|19|GeneID:3717970)			eam_2090(sitD|-13.54910|0.1292254|284|294|1|11|GeneID:8952024)	sl1344_2844(sitD|-5.09130|0.4340556|176|200|15|37|N/A)				8
0.724026224		rsp_0829(lctB|-6.46032|0.4296409|83|96|6|19|GeneID:3718369)	pa14_33860(lldA|-12.81200|0.2793503|172|183|7|17|GeneID:4380444)	caul_3608(lldD|-6.37092|0.8415159|252|259|1|8|GeneID:5901063)		sl1344_1550(N/A|-9.50407|0.07996338|108|121|24|37|N/A)	sven_0183(N/A|-8.13489|0.1065086|123|132|26|35|GeneID:13817290)	L-lactate dehydrogenase		3
0.724026224		rsp_1278(cbbZ|-10.36240|0.1142100|121|130|28|37|GeneID:3718192)	pa14_07930(N/A|-16.74600|0.0692314|268|288|6|28|GeneID:4385484)	caul_3547(N/A|-12.69580|0.1940449|27|42|1|15|GeneID:5901002)	eam_3235(gph|-13.21780|0.1493323|125|152|1|24|GeneID:8952240)	sl1344_3449(gph|-9.02866|0.09822531|223|248|12|35|N/A)	sven_6699(N/A|-4.67149|0.3921199|270|276|33|39|GeneID:13817965)	phosphoglycolate phosphatase	rcap_rcc01826 rcap_rcc00333	3

0.7 240 262 24					eam_2186(baeS| - 12.34910|0.2129 931|154|168|6|20 |GeneID:895142 1)	sl1344_2107(baeS|- 6.19135|0.30 11416|235|25 2|7|23|N/A)	sven_4967(N/A|- 9.93659|0.051778 81|208|231|5|38|G eneID:13824422)	sensor histidine kinase		8
0.7 240 262 24		rsp_0034(flhA|- 9.07911|0.185486 4|86|94|28|36|Gen eID:3720112)	pa14_45680(flh A|- 12.14350|0.3427 240|99|117|2|21| GeneID:438017 9)	caul_1017(flhA|- 12.36470|0.2174 473|3|41|1|39|Ge neID:5898472)	eam_2587(flhA|- 13.94730|0.1078 773|257|276|5|24 |GeneID:895101 3)	sl1344_1848(flhA|- 4.95911|0.45 19671|74|126 |5|40|N/A)		flagellar biosynth esis protein FlhA		6
0.7 240 262 24	ppa1107(N/A|- 4.28024|0.39208 78|36|46|26|35|G eneID:2931530)	rsp_0902(N/A|- 8.50898|0.226864 4|236|263|8|37|Ge neID:3717965)	pa14_37770(N/ A|- 17.55360|0.0504 8498|217|233|1|1 5|GeneID:43803 11)	caul_0320(N/A|- 14.43610|0.1006 711|184|217|1|37| GeneID:5897594)	eam_2636(N/A|- 14.51530|0.0823 1335|14|40|1|24| GeneID:895012 2)	sl1344_1257(N/A|- 3.23788|0.70 31772|194|20 0|35|41|N/A)	sven_2056(N/A|- 5.34854|0.309645 7|185|191|6|12|Ge neID:13818512)	phospho glycolat e phosphat ase	rcap_rcc 03410 rcap_rcc 02020 rcap_rcc 01825	1

0.7 240 262 24	ppa0271(N/A|- 6.26040|0.17175 25|87|94|1|8|Gen eID:2931909)	rsp_1862(N/A|- 5.61597|0.534036 3|46|72|6|36|Gene ID:3719129)	pa14_62850(fol P|- 17.46660|0.0522 5179|98|125|1|25 |GeneID:438319 9)	caul_4421(N/A|- 12.08790|0.2384 997|94|101|2|9|G eneID:5901882)	eam_3077(folP|- 11.73100|0.2684 084|128|156|10|4 1|GeneID:89514 94)	sl1344_3266(folP|- 10.17350|0.0 5945155|249| 293|12|40|N/ A)	sven_5577(N/A|- 5.73511|0.269045 2|133|147|7|22|Ge neID:13822104)	dihydrop teroate synthase		1
0.7 240 262 24		rsp_1390(N/A|- 6.52867|0.421603 2|254|260|8|14|Ge neID:3720795)	pa14_63190(N/ A|- 11.53450|0.4081 738|255|288|1|39 |GeneID:438322 8)	caul_4262(N/A|- 15.27620|0.0708 8095|215|228|2|1 5|GeneID:59017 23)	eam_2049(N/A|- 11.52270|0.2889 844|127|133|9|15 |GeneID:894984 0)	sl1344_1447(N/A|- 8.65393|0.11 51625|194|20 0|28|34|N/A)				7
0.7 240 262 24	ppa0707(N/A|- 7.25345|0.10900 39|117|126|8|17| GeneID:2933161)		pa14_54690(N/ A|- 12.49860|0.3079 248|250|267|2|21 |GeneID:438528 8)				sven_4326(N/A|- 10.63860|0.03906 164|284|292|33|4 1|GeneID:138240 52)	NAD- dependent epimerase/dehydratase family protein		6

0.7 240 262 24		rsp_3340(rarD|- 9.21863|0.176324 2|69|77|6|14|Gene ID:3722042)	pa14_06320(N/ A|- 8.89718|0.73890 15|266|273|5|12| GeneID:438385 4)	caul_4619(N/A|- 6.05802|0.86737 6|36|42|33|39|Ge neID:5902081)			sven_4296(N/A|- 9.80867|0.054508 28|224|264|5|38|G eneID:13819454)	RarD protein		6
0.7 240 262 24	ppa1808(N/A|- 5.43805|0.24599 54|151|158|7|14| GeneID:2931436)	rsp_3593(N/A|- 16.83620|0.00561 7299|134|156|12|3 7|GeneID:372211 0)	pa14_05000(trm B|- 16.12200|0.0878 586|125|157|1|24 |GeneID:438377 1)	caul_0028(trmB|- 13.98630|0.1203 764|6|13|2|9|Gen eID:5897740)	eam_2845(trmB| - 9.29703|0.55651 06|259|267|1|9|G eneID:8950237)	sl1344_3084(N/A|- 3.56127|0.65 56976|67|73| 30|36|N/A)	sven_3866(N/A|- 3.72636|0.529766 7|158|164|31|37|G eneID:13821691)	tRNA (guanine -N(7)-)- methyltr ansferas e		0
0.7 240 262 24	ppa0633(N/A|- 1.89078|0.79591 26|23|34|13|29|G eneID:2931243)	rsp_0858(N/A|- 8.69839|0.212401 7|218|248|8|37|Ge neID:3718325)	pa14_02970(N/ A|- 14.24760|0.1735 218|206|226|1|20 |GeneID:438365 0)	caul_4826(N/A|- 9.97873|0.44163 22|5|12|2|9|GeneI D:5902288)	eam_1032(N/A|- 10.57910|0.3933 323|26|36|1|11|G eneID:8949312)				rcap_rcc 00450 rcap_rcc 00086 rcap_rcc 03090 rcap_rcc 01716	7

									rcap_rcc00756 rcap_rcc00381 rcap_rcc02147 rcap_rcp00047 rcap_rcc02763	
0.724026224			pa14_34200(N/A|-11.34470|0.429957|9|2|2|1|21|GeneID:4385114)	caul_2948(N/A|-19.23290|0.009858998|171|180|1|10|GeneID:5900403)	eam_1759(N/A|-12.69860|0.185415|8|106|115|7|16|GeneID:8952043)		sven_0431(N/A|-2.88525|0.66606|281|287|9|15|GeneID:13817537)			7
0.724026224	ppa1335(N/A|-6.14912|0.1804985|70|83|25|38|GeneID:2931065)	rsp_1195(comF|-18.00610|0.002940726|134|157|5|37|GeneID:3718859)	pa14_06360(N/A|-10.64150|0.515437|8|259|275|2|16	caul_0939(N/A|-14.06650|0.116654|1|152|170|1|18|GeneID:5898394)	eam_3258(gntX|-10.98090|0.346788|4|78|90|2|12|GeneID:8950729)	sl1344_3477(N/A|-3.39106|0.6808316|259|268|22|35|N/A)	sven_2753(N/A|-8.97363|0.0762009|178|186|33|41|GeneID:13823173)	competence protein F		2

			|GeneID:4383858)							
0.724026224		rsp_4086(N/A|-10.36840|0.1139404|43|83|9|41|GeneID:3711961)		caul_2079(N/A|-13.24480|0.1594044|186|216|5|37|GeneID:5899534)	eam_2173(amsH|-6.10251|0.9084634|264|275|7|16|GeneID:8950629)	sl1344_2095(wza|-1.36604|0.9213592|278|286|11|20|N/A)		polysaccharide biosynthesis/export family protein	rcap_rcc01960 rcap_rcc03226	8
0.724026224		rsp_2628(gltX/glnS|-6.24773|0.4550925|190|199|5|13|GeneID:3720300)	pa14_62510(N/A|-12.54790|0.3032963|163|188|2|24|GeneID:4383172)	caul_3205(N/A|-14.99590|0.07988049|169|182|1|15|GeneID:5900660)	eam_0792(gluQ|-13.30330|0.1439317|255|264|7|16|GeneID:8951640)	sl1344_0186(yadB|-9.82226|0.06952106|70|76|32|38|N/A)		glutamyl-Q tRNA(Asp) synthetase		6
0.724026224		rsp_3993(N/A|-10.73640|0.09840478|12|48|7|37|GeneID:3711793)	pa14_48060(aprA|-6.94873|0.92726	caul_4884(N/A|-10.31730|0.4044667|26|33|2|9|GeneID:5902346)	eam_3368(prtA|-8.04195|0.7186932|278|287|16|25			hemolysin-type calcium-binding repeat	rcap_rcc02063 rcap_rcc00158 rcap_rcc	8

			8|203|221|4|17|GeneID:4385542)		|GeneID:8951317)			family protein	01362 rcap_rcc 02665 rcap_rcc 02520 rcap_rcc 01242 rcap_rcc 02060 rcap_rcc 01900 rcap_rcc 02786 rcap_rcc 00178 rcap_rcc 03191 rcap_rcc 00280 rcap_rcc 00964	

0.7 240 262 24	ppa1137(N/A|- 3.57473|0.50445 64|242|258|16|31| GeneID:2933040)	rsp_3749(ampC|- 3.71965|0.772936 1|77|90|25|37|Gen eID:3721508)	pa14_10790(am pC|- 13.38670|0.2321 568|211|224|5|15 |GeneID:438209 7)	caul_4601(N/A|- 17.00100|0.0321 2351|150|176|1|1 5|GeneID:59020 63)	eam_2914(ampC |- 14.46000|0.0845 6664|238|254|1|1 5|GeneID:89513 32)	sl1344_0370(ampH|- 5.54697|0.37 53101|2|20|1| 20|N/A)	sven_2995(N/A|- 7.97728|0.113387 4|233|243|32|41|G eneID:13823415)			2
0.7 240 262 24		rsp_1286(cbbR|- 8.47942|0.229187 0|87|99|26|37|Gen eID:3718200)	pa14_71750(N/ A|- 15.25270|0.1213 229|84|94|5|15|G eneID:4385026)		eam_2218(N/A|- 8.75319|0.62820 08|63|74|2|16|Ge neID:8951402)		sven_1212(N/A|- 7.64866|0.129129 2|210|216|32|38|G eneID:13822931)	RuBisC O operon transcrip tional regulator CbbR	rcap_rcc 01835	5

Table S4: Functional enrichment of potential cis-targets of putative antisense and partially overlapping sRNAs.

pathway ID	pathway description	observed gene count	false discovery rate	matching proteins in your network (IDs)	matching proteins in your network (labels)
GO.0044238	primary metabolic process	28	1.97E-005	RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00659,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc01909,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchN,cbbM,ccoP,crtE,crtF,crtJ,crtK,dnaJ,dnaK,dxs1,glnA,groS,ihfA,petB,pucA,pucB,pufA,pufB,puhA,rpmF,rpsF
GO.0015979	photosynthesis	16	3.98E-005	RCAP_rcc00659,RCAP_rcc00660,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00695,RCAP_rcc01829,RCAP_rcc02530,RCAP_rcc02531	bchB,bchJ,bchN,bchO,cbbM,crtE,crtF,crtJ,crtK,pucA,pucB,pucC1,pufA,pufB,pufX,puhA
GO.0044	cellular metabolic	32	0.00001	RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00659,RCAP_rcc00660,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00695,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc01909,RCAP_rcc01912,R	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,cbbM,ccoP,crtE,crtF,crtJ,crtK,dnaJ,dnaK,dxs1,glnA,groS,ihfA,petB,pucA,pucB,pucC1,pufA,pufB,pufX,puhA,rpmF,rpsF

237	process		13	CAP_rcc01913,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	
GO.0071704	organic substance metabolic process	30	0.000113	RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00659,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc01909,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,cbbM,ccoP,crtE,crtF,crtJ,crtK,dnaJ,dnaK,dxs1,glnA,groS,ihfA,petB,pucA,pucB,pufA,pufB,puhA,rpmF,rpsF
GO.0009987	cellular process	34	0.000269	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00659,RCAP_rcc00660,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00695,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc01909,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,cbbM,ccoP,crtE,crtF,crtJ,crtK,dnaJ,dnaK,dxs1,glnA,groS,ihfA,petB,pucA,pucB,pucC1,pufA,pufB,pufX,puhA,rpmF,rpsF,secA,tatC
GO.0008150	biological_process	37	0.000137	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00659,RCAP_rcc00660,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00695,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc00767,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc01519,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc01909,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,cbbM,ccoP,crtE,crtF,crtJ,crtK,dnaJ,dnaK,dxs1,glnA,groS,hupA,ihfA,nuoC,petB,petC,pucA,pucB,pucC1,pufA,pufB,pufX,puhA,rpmF,rpsF,secA,tatC
GO.0008152	metabolic process	35	0.000137	RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00659,RCAP_rcc00660,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00695,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc00767,RCAP_rcc01160,RCAP_rcc01519,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc01909,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,cbbM,ccoP,crtE,crtF,crtJ,crtK,dnaJ,dnaK,dxs1,glnA,groS,hupA,ihfA,nuoC,petB,petC,pucA,pucB,pucC1,pufA,pufB,pufX,puhA,rpmF,rpsF
GO.0044	cellular protein met	10	0.0001	RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc01909,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531	dnaJ,dnaK,groS,pucA,pucB,pufA,pufB,puhA,rpmF,rpsF

267	abolic process		37		
GO.1901566	organonitrogen compound biosynthetic process	17	0.00137	RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01674,RCAP_rcc01909,RCAP_rcc02011,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,crtE,crtF,crtJ,crtK,dxs1,glnA,rpmF,rpsF
GO.1901564	organonitrogen compound metabolic process	19	0.00169	RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01674,RCAP_rcc01909,RCAP_rcc02011,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,ccoP,crtE,crtF,crtJ,crtK,dxs1,glnA,petB,rpmF,rpsF
GO.0019684	photosynthesis, light reaction	5	0.00196	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc02530,RCAP_rcc02531	pucA,pucB,pufA,pufB,puhA

GO.0044271	cellular nitrogen compound biosynthetic process	18	0.00196	RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01909,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,crtE,crtF,crtJ,crtK,dxs1,ihfA,rpmF,rpsF
GO.1901576	organic substance biosynthetic process	21	0.00196	RCAP_rcc00223,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc01909,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,cbbM,crtE,crtF,crtJ,crtK,dnaJ,dxs1,glnA,ihfA,rpmF,rpsF
GO.0015992	proton transport	7	0.00263	RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,petB
GO.0018298	protein-chromophore linkage	5	0.00263	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc02530,RCAP_rcc02531	pucA,pucB,pufA,pufB,puhA

GO.0034641	cellular nitrogen compound metabolic process	21	0.00263	RCAP_rcc00223,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01909,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,ccoP,crtE,crtF,crtJ,crtK,dnaJ,dxs1,ihfA,petB,rpmF,rpsF
GO.0042777	plasma membrane ATP synthesis coupled proton transport	5	0.00263	RCAP_rcc00744,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF
GO.0044249	cellular biosynthetic process	20	0.00263	RCAP_rcc00223,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01674,RCAP_rcc01909,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,crtE,crtF,crtJ,crtK,dnaJ,dxs1,glnA,ihfA,rpmF,rpsF

GO.1901137	carbohydrate derivative biosynthetic process	6	0.00263	RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,dxs1
GO.1902600	hydrogen ion transmembrane transport	6	0.00263	RCAP_rcc00744,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,petB
GO.0006725	cellular aromatic compound metabolic process	19	0.00354	RCAP_rcc00223,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,ccoP,crtE,crtF,crtJ,crtK,dnaJ,dxs1,ihfA,petB

GO.0019438	aromatic compound biosynthetic process	16	0.00354	RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,crtE,crtF,crtJ,crtK,dxs1,ihfA
GO.0044711	single-organism biosynthetic process	16	0.00354	RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,cbbM,crtE,crtF,crtJ,crtK,dxs1,glnA
GO.0046034	ATP metabolic process	7	0.00354	RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,petB
GO.0046483	heterocycle metabolic process	19	0.00354	RCAP_rcc00223,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,ccoP,crtE,crtF,crtJ,crtK,dnaJ,dxs1,ihfA,petB

GO.0190135	carbohydrate derivative metabolic process	8	0.00354	RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,dxs1,petB
GO.0018130	heterocyclebiosynthetic process	16	0.00355	RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,crtE,crtF,crtJ,crtK,dxs1,ihfA
GO.0019637	organophosphate metabolic process	9	0.00355	RCAP_rcc00684,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,crtE,dxs1,petB
GO.1901360	organic cyclic compound metabol	19	0.00355	RCAP_rcc00223,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,ccoP,crtE,crtF,crtJ,crtK,dnaJ,dxs1,ihfA,petB

	ic process				
GO.1901362	organic cyclic compound biosynthetic process	16	0.00355	RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,crtE,crtF,crtJ,crtK,dxs1,ihfA
GO.0090407	organophosphate biosynthetic process	7	0.00386	RCAP_rcc00684,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,crtE,dxs1
GO.0015995	chlorophyll biosynthetic process	8	0.00503	RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685	bchB,bchJ,bchN,bchO,crtE,crtF,crtJ,crtK
GO.00	terpenoid	5	0.00	RCAP_rcc00668,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696	crtE,crtF,crtJ,crtK,dxs1

16114	biosynthetic process		567		
GO.0044260	cellular macromolecule metabolic process	12	0.00567	RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc01909,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531	RCAP_rcc01913,dnaJ,dnaK,groS,ihfA,pucA,pucB,pufA,pufB,puhA,rpmF,rpsF
GO.0006091	generation of precursor metabolites and energy	7	0.00616	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc01160,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02769	ccoP,petB,pucA,pucB,pufA,pufB,puhA
GO.0044710	single-organism metabolic	22	0.00907	RCAP_rcc00659,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc00767,RCAP_rcc01160,RCAP_rcc01519,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,cbbM,ccoP,crtE,crtF,crtJ,crtK,dxs1,glnA,hupA,nuoC,petB,petC,puhA

	process				
GO.0055085	transmembrane transport	8	0.00909	RCAP_rcc00220,RCAP_rcc00744,RCAP_rcc01460,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,petB,secA,tatC
GO.0006807	nitrogen compound metabolic process	22	0.00103	RCAP_rcc00223,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01674,RCAP_rcc01909,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,ccoP,crtE,crtF,crtJ,crtK,dnaJ,dxs1,glnA,ihfA,petB,rpmF,rpsF
GO.0006796	phosphate-containing compound metabolic process	9	0.00108	RCAP_rcc00684,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,crtE,dxs1,petB
GO.0016	carotenoid biosynt	4	0.001	RCAP_rcc00668,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685	crtE,crtF,crtJ,crtK

117	hetic process		73		
GO.0044763	single-organism cellular process	19	0.02833	RCAP_rcc00220,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc01674,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,ccoP,crtE,crtF,crtJ,crtK,dxs1,glnA,petB,secA,tatC
GO.0044699	single-organism process	24	0.02844	RCAP_rcc00220,RCAP_rcc00659,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00670,RCAP_rcc00675,RCAP_rcc00681,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc00767,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc01519,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchJ,bchN,bchO,cbbM,ccoP,crtE,crtF,crtJ,crtK,dxs1,glnA,hupA,nuoC,petB,petC,puhA,secA,tatC
GO.0006457	protein folding	3	0.0039	RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc02477	dnaJ,dnaK,groS
GO.0044765	single-organism transport	9	0.03988	RCAP_rcc00220,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,petB,secA,tatC
GO.0006	nucleobase-cont	10	0.0	RCAP_rcc00223,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,dnaJ,ihfA,petB

139	aining compound metabolic process		43		
GO.0006810	transport	10	0.00472	RCAP_rcc00220,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc01519,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,nuoC,petB,secA,tatC
pathway ID	**pathway description**	**observed gene count**	**false discovery rate**	**matching proteins in your network (IDs)**	**matching proteins in your network (labels)**
GO.0097159	organic cyclic compound	24	0.00004	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00659,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00675,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00696,RCAP_rcc01160,RCAP_rcc01674,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,bchB,bchN,bchO,ccoP,crtJ,dnaJ,dnaK,dxs1,glnA,groS,ihfA,petC,pucA,pucB,pufA,pufB,puhA,rpsF,secA

	binding				
GO.1901363	heterocyclic compound binding	24	0.00004	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00659,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00675,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00696,RCAP_rcc01160,RCAP_rcc01674,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,bchB,bchN,bchO,ccoP,crtJ,dnaJ,dnaK,dxs1,glnA,groS,ihfA,petC,pucA,pucB,pufA,pufB,puhA,rpsF,secA
GO.0005488	binding	29	0.000285	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00659,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00675,RCAP_rcc00684,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00696,RCAP_rcc00767,RCAP_rcc01160,RCAP_rcc01519,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,bchB,bchN,bchO,cbbM,ccoP,crtE,crtJ,dnaJ,dnaK,dxs1,glnA,groS,hupA,ihfA,nuoC,petB,petC,pucA,pucB,pufA,pufB,puhA,rpsF,secA
GO.0042314	bacteriochlorophyll binding	5	0.000285	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc02530,RCAP_rcc02531	pucA,pucB,pufA,pufB,puhA
GO.0045156	electron transporter, transferring electrons wit	5	0.000285	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc02530,RCAP_rcc02531	pucA,pucB,pufA,pufB,puhA

	hin the cyclic electron transport pathway of photosynthesis activity				
GO.0005524	ATP binding	12	0.00307	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00675,RCAP_rcc01674,RCAP_rcc02477,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,bchB,bchN,bchO,dnaJ,dnaK,glnA,groS,secA
GO.0015078	hydrogen ion transmembrane transporter activity	6	0.00307	RCAP_rcc00744,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,petB
GO.00	proton-tran	5	0.00	RCAP_rcc00744,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF

46933	sporting ATP synthase activity, rotational mechanism		307		
GO.0043167	ion binding	23	0.00318	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00675,RCAP_rcc00684,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00696,RCAP_rcc00767,RCAP_rcc01160,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,bchB,bchN,bchO,cbbM,ccoP,crtE,dnaJ,dnaK,dxs1,glnA,groS,hupA,petB,petC,pucA,pucB,pufA,pufB,secA
GO.0046906	tetrapyrrole binding	7	0.00318	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc01160,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02770	ccoP,petC,pucA,pucB,pufA,pufB,puhA
GO.0003674	molecular_function	33	0.00334	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00659,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00668,RCAP_rcc00675,RCAP_rcc00684,RCAP_rcc00685,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00696,RCAP_rcc00744,RCAP_rcc00767,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc01519,RCAP_rcc01674,RCAP_rcc01829,RCAP_rcc01909,RCAP_rcc01912,RCAP_rcc01913,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc01913,RCAP_rcc02972,atpA,atpC,atpD,atpF,bchB,bchN,bchO,cbbM,ccoP,crtE,crtF,crtJ,dnaJ,dnaK,dxs1,glnA,groS,hupA,ihfA,nuoC,petB,petC,pucA,pucB,pufA,pufB,puhA,rpmF,rpsF,secA,tatC
GO.0009	electron carrier	8	0.006	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc01160,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02769,RCAP_rcc02770	ccoP,petB,petC,pucA,pucB,pufA,pufB,puhA

055	activity		25		
GO.0046961	proton-transporting ATPase activity, rotational mechanism	3	0.00625	RCAP_rcc02970,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC
GO.0043168	anion binding	13	0.00124	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00675,RCAP_rcc00696,RCAP_rcc01674,RCAP_rcc02477,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,bchB,bchN,bchO,dnaJ,dnaK,dxs1,glnA,groS,secA
GO.0036094	small molecule binding	13	0.00161	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00224,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00675,RCAP_rcc00696,RCAP_rcc01674,RCAP_rcc02477,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,bchB,bchN,bchO,dnaJ,dnaK,dxs1,glnA,groS,secA
GO.00022	substrate-specific	7	0.00207	RCAP_rcc00744,RCAP_rcc01460,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,petB,tatC

891	transmembrane transporter activity				
GO.0046872	metal ion binding	15	0.03511	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00664,RCAP_rcc00665,RCAP_rcc00684,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00696,RCAP_rcc00767,RCAP_rcc01160,RCAP_rcc01829,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02769,RCAP_rcc02770	bchB,bchN,cbbM,ccoP,crtE,dnaJ,dxs1,hupA,petB,petC,pucA,pucB,pufA,pufB,secA
pathway ID	**pathway description**	**observed gene count**	**false discovery rate**	**matching proteins in your network (IDs)**	**matching proteins in your network (labels)**
GO.0032991	macromolecular complex	17	3.20E-0	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00744,RCAP_rcc00767,RCAP_rcc01460,RCAP_rcc01909,RCAP_rcc02011,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02532,RCAP_rcc02533,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,hupA,petB,pucA,pucB,pucC,pucD,pufA,pufB,puhA,rpmF,rpsF,tatC

			07		
GO.0043234	protein complex	15	1.42E-006	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00744,RCAP_rcc00767,RCAP_rcc01460,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02532,RCAP_rcc02533,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,hupA,petB,pucA,pucB,pucC,pucD,pufA,pufB,puhA,tatC
GO.0005886	plasma membrane	20	1.61E-006	RCAP_rcc00220,RCAP_rcc00660,RCAP_rcc00681,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00695,RCAP_rcc00744,RCAP_rcc00767,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc01519,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02532,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,crtK,hupA,nuoC,petB,petC,pucA,pucB,pucC,pucC1,pufA,pufB,pufX,secA,tatC
GO.0071944	cell periphery	20	3.47E-006	RCAP_rcc00220,RCAP_rcc00660,RCAP_rcc00681,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00695,RCAP_rcc00744,RCAP_rcc00767,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc01519,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02532,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,crtK,hupA,nuoC,petB,petC,pucA,pucB,pucC,pucC1,pufA,pufB,pufX,secA,tatC
GO.0030076	light-harvesting complex	7	7.25E-006	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02532,RCAP_rcc02533	pucA,pucB,pucC,pucD,pufA,pufB,puhA
GO.0044	intracellula	19	2.83E	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00744,RCAP_rcc00767,RCAP_rcc01674,RCAP_rcc01909,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,R	RCAP_rcc02972,atpA,atpC,atpD,atpF,dnaJ,glnA,groS,hupA,petB,pucA,pucB,pucC,pucD,pufA,pufB,rpmF,rpsF,secA

424	r part		-005	CAP_rcc02531,RCAP_rcc02532,RCAP_rcc02533,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	
GO.0016020	membrane	19	3.00E-005	RCAP_rcc00220,RCAP_rcc00659,RCAP_rcc00660,RCAP_rcc00681,RCAP_rcc00691,RCAP_rcc00695,RCAP_rcc00744,RCAP_rcc00767,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc01519,RCAP_rcc02530,RCAP_rcc02532,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,crtK,hupA,nuoC,petB,petC,pucB,pucC,pucC1,pufB,pufX,puhA,secA,tatC
GO.0044444	cytoplasmic part	11	4.17E-005	RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00744,RCAP_rcc01909,RCAP_rcc02011,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,pucA,pucB,pufA,pufB,rpmF,rpsF
GO.0044425	membrane part	17	5.37E-005	RCAP_rcc00659,RCAP_rcc00660,RCAP_rcc00681,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02532,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,crtK,petB,petC,pucA,pucB,pucC,pucC1,pufA,pufB,puhA,tatC
GO.0044446	intracellular organelle part	6	9.76E-005	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc01909,RCAP_rcc02530,RCAP_rcc02531	pucA,pucB,pufA,pufB,puhA,rpmF

GO.0005622	intracellular	19	0.000101	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00744,RCAP_rcc00767,RCAP_rcc01674,RCAP_rcc01909,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02532,RCAP_rcc02533,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,dnaJ,glnA,groS,hupA,petB,pucA,pucB,pucC,pucD,pufA,pufB,rpmF,rpsF,secA
GO.0043229	intracellular organelle	7	0.000177	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc01909,RCAP_rcc02011,RCAP_rcc02530,RCAP_rcc02531	pucA,pucB,pufA,pufB,puhA,rpmF,rpsF
GO.0042716	plasma membrane-derived chromatophore	9	0.000196	RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00744,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,pucA,pucB,pufA,pufB
GO.0030077	plasma membrane light-harvesting complex	5	0.000228	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc02530,RCAP_rcc02531	pucA,pucB,pufA,pufB,puhA

GO.0005737	cytoplasm	15	0.00050 9	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00744,RCAP_rcc01674,RCAP_rcc01909,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,dnaJ,glnA,groS,pucA,pucB,pufA,pufB,rpmF,rpsF,secA
GO.0031410	cytoplasmic vesicle	5	0.00050 9	RCAP_rcc00659,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc02530,RCAP_rcc02531	pucA,pucB,pufA,pufB,puhA
GO.0045261	proton-transporting ATP synthase complex, catalytic core F(1)	4	0.000601	RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD
GO.0045259	proton-transporting ATP synt	5	0.0016	RCAP_rcc00744,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF

	hase complex				
GO.0044464	cell part	21	0.00168	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00660,RCAP_rcc00681,RCAP_rcc00695,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc01519,RCAP_rcc01674,RCAP_rcc01909,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02532,RCAP_rcc02533,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,crtK,dnaJ,glnA,groS,nuoC,petB,petC,pucC,pucC1,pucD,pufX,rpmF,rpsF,secA,tatC
GO.0005623	cell	21	0.00176	RCAP_rcc00220,RCAP_rcc00223,RCAP_rcc00660,RCAP_rcc00681,RCAP_rcc00695,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc01519,RCAP_rcc01674,RCAP_rcc01909,RCAP_rcc02011,RCAP_rcc02477,RCAP_rcc02532,RCAP_rcc02533,RCAP_rcc02769,RCAP_rcc02770,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,ccoP,crtK,dnaJ,glnA,groS,nuoC,petB,petC,pucC,pucC1,pucD,pufX,rpmF,rpsF,secA,tatC
GO.0016021	integral component of membrane	13	0.00176	RCAP_rcc00659,RCAP_rcc00660,RCAP_rcc00681,RCAP_rcc00691,RCAP_rcc00692,RCAP_rcc00744,RCAP_rcc01160,RCAP_rcc01460,RCAP_rcc02530,RCAP_rcc02531,RCAP_rcc02532,RCAP_rcc02769,RCAP_rcc02770	atpF,ccoP,crtK,petB,petC,pucA,pucB,pucC,pucC1,pufA,pufB,puhA,tatC
GO.0098796	membrane protein complex	7	0.00431	RCAP_rcc00744,RCAP_rcc01460,RCAP_rcc02769,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,petB,tatC
GO.0070	respiratory chain	3	0.004	RCAP_rcc01160,RCAP_rcc02769,RCAP_rcc02770	ccoP,petB,petC

469			58		
GO.0009579	thylakoid	6	0.01337	RCAP_rcc00659,RCAP_rcc00744,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,puhA
GO.0019866	organelle inner membrane	2	0.01337	RCAP_rcc00692,RCAP_rcc02531	pucA,pufA
GO.0031967	organelle envelope	2	0.01337	RCAP_rcc00692,RCAP_rcc02531	pucA,pufA
GO.0042717	plasma membrane-derived chromatophore membrane	6	0.01337	RCAP_rcc00659,RCAP_rcc00744,RCAP_rcc02970,RCAP_rcc02971,RCAP_rcc02972,RCAP_rcc02973	RCAP_rcc02972,atpA,atpC,atpD,atpF,puhA

GO.0043227	membrane-bounded organelle	2	0.0366	RCAP_rcc00692,RCAP_rcc02531	pucA,pufA

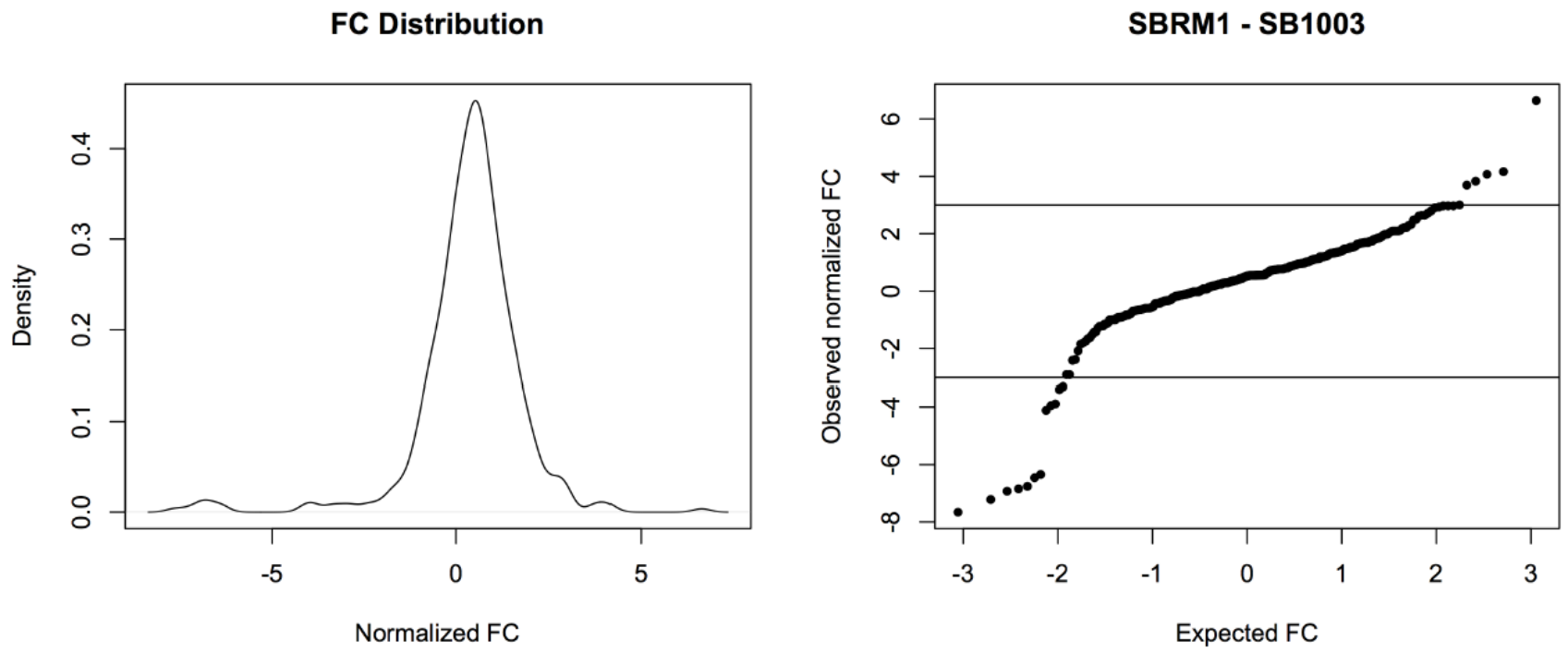

Figure S3: Differential expression (normalized log2 fold change) between the genome-sequenced strain, SB1003, and its derived *ctrA* mutant strain, SBRM1. Most putative sRNAs have similar transcription level between both strains. 18 sRNAs had an absolute log2 fold change >= 3. These 18 sRNAs are indicated on the right plot as the points above the top horizontal line or below the bottom horizontal line.

Appendix 2 - Supplementary Figures and Tables for Chapter 3

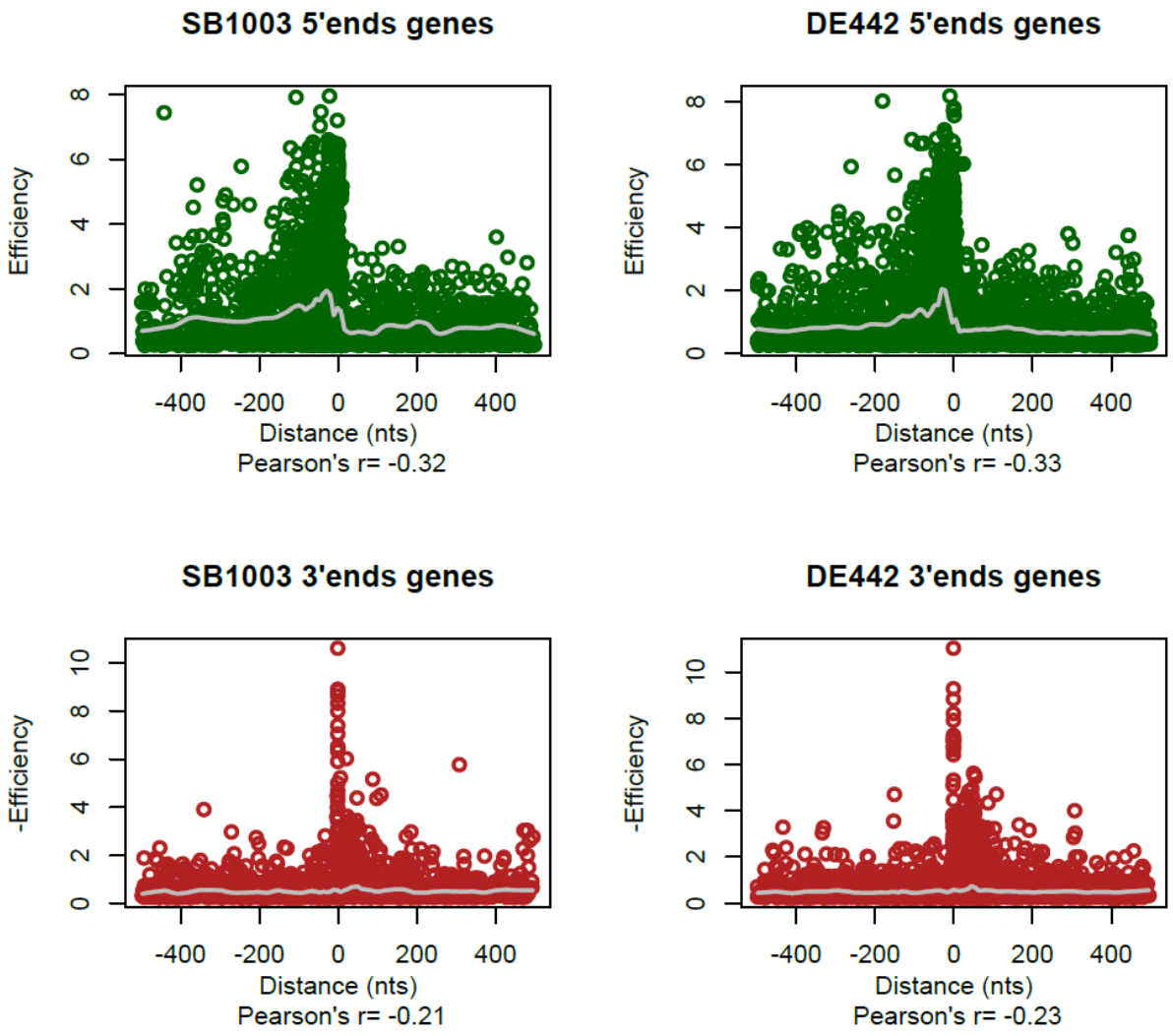

Figure S4: Scatter plot showing differences between TSS and TTS associated with annotated genes.

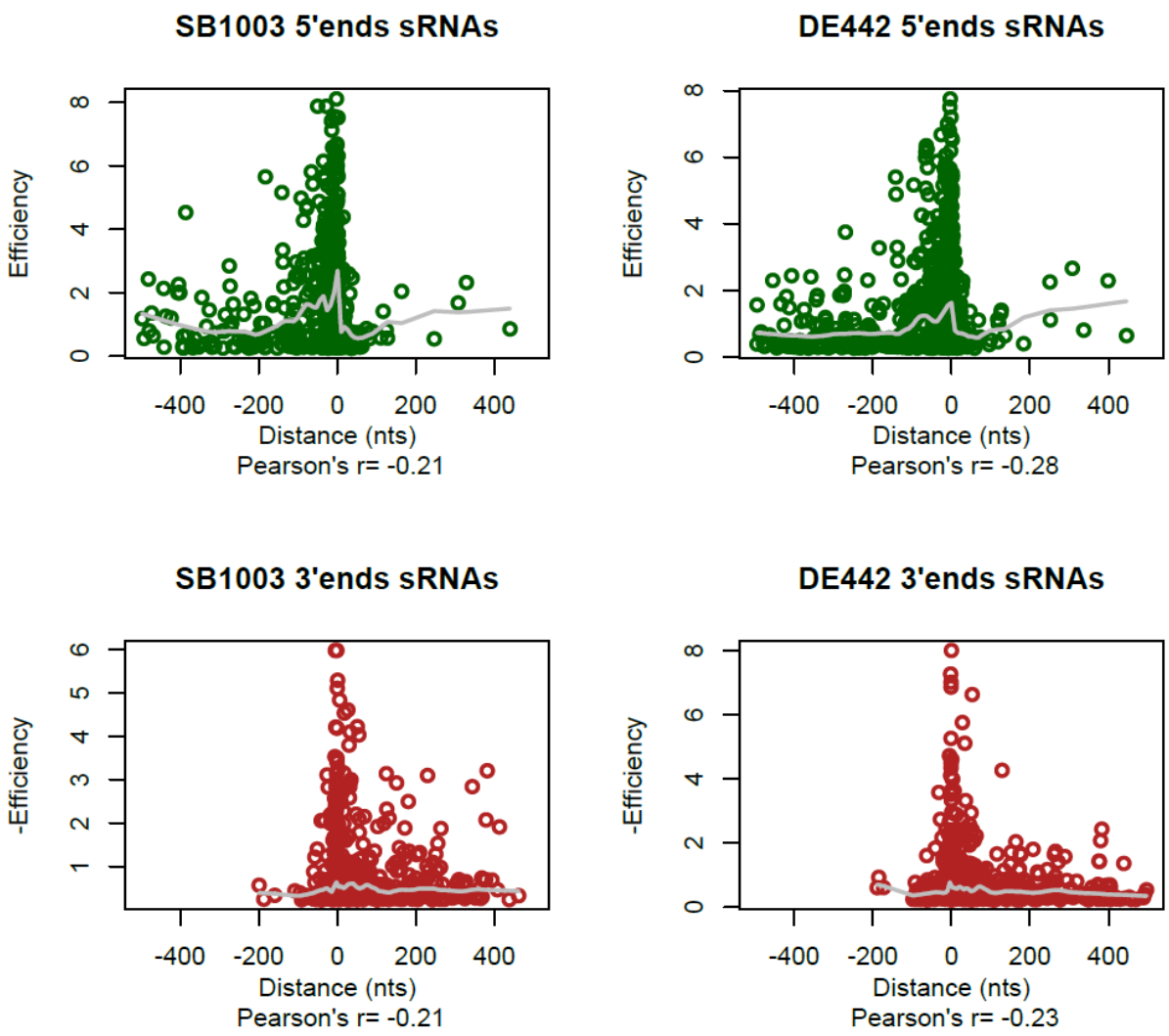

Figure S5: Scatter plot showing differences between TSS and TTS associated with the putative sRNAs

Appendix 3 – Supplementary Figures and Tables for Chapter 4

Table S5: Mean difference and significance of relative gene expression analysis. Each strain was compared to the relative expression values obtained from the *R. capsulatus* wild-type strain SB1003.

rpoZ								
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	A-?		
SB1003 rpoZ vs. SB1865 rpoZ	0.2565	-0.05068 to 0.5636	No	ns	0.1082	B	SB1865 rpoZ	
SB1003 rpoZ vs. SB1866 rpoZ	0.2565	-0.05068 to 0.5636	No	ns	0.1082	C	SB1866 rpoZ	
SB1003 rpoZ vs. SBcckA rpoZ	0.2565	-0.05068 to 0.5636	No	ns	0.1082	D	SBcckA rpoZ	
SB1003 rpoZ vs. SBRM1 rpoZ	0.2565	-0.05068 to 0.5636	No	ns	0.1082	E	SBRM1 rpoZ	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
SB1003 rpoZ vs. SB1865 rpoZ	1.001	0.7444	0.2565	0.1063	3	3	2.414	10
SB1003 rpoZ vs. SB1866 rpoZ	1.001	0.7444	0.2565	0.1063	3	3	2.414	10
SB1003 rpoZ vs. SBcckA rpoZ	1.001	0.7444	0.2565	0.1063	3	3	2.414	10
SB1003 rpoZ vs. SBRM1 rpoZ	1.001	0.7444	0.2565	0.1063	3	3	2.414	10
rpoA								
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	A-?		

SB1003 rpoA vs. SB1865 rpoA	-0.3662	-1.034 to 0.3011	No	ns	0.3704	B	SB1865 rpoA	
SB1003 rpoA vs. SB1866 rpoA	-0.3662	-1.034 to 0.3011	No	ns	0.3704	C	SB1866 rpoA	
SB1003 rpoA vs. SBcckA rpoA	-0.3662	-1.034 to 0.3011	No	ns	0.3704	D	SBcckA rpoA	
SB1003 rpoA vs. SBRM1 rpoA	-0.3662	-1.034 to 0.3011	No	ns	0.3704	E	SBRM1 rpoA	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
SB1003 rpoA vs. SB1865 rpoA	1.012	1.378	-0.3662	0.2309	3	3	1.586	10
SB1003 rpoA vs. SB1866 rpoA	1.012	1.378	-0.3662	0.2309	3	3	1.586	10
SB1003 rpoA vs. SBcckA rpoA	1.012	1.378	-0.3662	0.2309	3	3	1.586	10
SB1003 rpoA vs. SBRM1 rpoA	1.012	1.378	-0.3662	0.2309	3	3	1.586	10
171								
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	A-?		
SB1003 171 vs. SB1865 171	0.9522	0.7609 to 1.143	Yes	****	<0.0001	B	SB1865 171	
SB1003 171 vs. SB1866 171	-1.613	-1.804 to -1.422	Yes	****	<0.0001	C	SB1866 171	
SB1003 171 vs. SBcckA 171	0.9763	0.7625 to 1.19	Yes	****	<0.0001	D	SBcckA 171	
SB1003 171 vs. SBRM1 171	0.9756	0.7844 to 1.167	Yes	****	<0.0001	E	SBRM1 171	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF

SB1003 171 vs. SB1865 171	1.007	0.05439	0.9522	0.06487	3	3	14.68	9
SB1003 171 vs. SB1866 171	1.007	2.62	-1.613	0.06487	3	3	24.87	9
SB1003 171 vs. SBcckA 171	1.007	0.03029	0.9763	0.07253	3	2	13.46	9
SB1003 171 vs. SBRM1 171	1.007	0.03093	0.9756	0.06487	3	3	15.04	9
555								
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	A-?		
SB1003 555 vs. SB1865 555	0.8415	0.558 to 1.125	Yes	****	<0.0001	B	SB1865 555	
SB1003 555 vs. SB1866 555	-0.6886	-0.972 to -0.4051	Yes	***	0.0001	C	SB1866 555	
SB1003 555 vs. SBcckA 555	0.9169	0.6334 to 1.2	Yes	****	<0.0001	D	SBcckA 555	
SB1003 555 vs. SBRM1 555	0.9386	0.6552 to 1.222	Yes	****	<0.0001	E	SBRM1 555	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
SB1003 555 vs. SB1865 555	1.017	0.1757	0.8415	0.09806	3	3	8.581	10
SB1003 555 vs. SB1866 555	1.017	1.706	-0.6886	0.09806	3	3	7.022	10
SB1003 555 vs. SBcckA 555	1.017	0.1003	0.9169	0.09806	3	3	9.351	10
SB1003 555 vs. SBRM1 555	1.017	0.07858	0.9386	0.09806	3	3	9.572	10
1079								
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	A-?		

SB1003 1079 vs. SB1865 1079	0.9975	0.6947 to 1.3	Yes	****	<0.0001	B	SB1865 1079	
SB1003 1079 vs. SB1866 1079	-0.1336	-0.4364 to 0.1692	No	ns	0.5442	C	SB1866 1079	
SB1003 1079 vs. SBcckA 1079	1.017	0.7143 to 1.32	Yes	****	<0.0001	D	SBcckA 1079	
SB1003 1079 vs. SBRM1 1079	1.004	0.7011 to 1.307	Yes	****	<0.0001	E	SBRM1 1079	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
SB1003 1079 vs. SB1865 1079	1.027	0.02932	0.9975	0.1047	3	3	9.522	10
SB1003 1079 vs. SB1866 1079	1.027	1.16	-0.1336	0.1047	3	3	1.275	10
SB1003 1079 vs. SBcckA 1079	1.027	0.009699	1.017	0.1047	3	3	9.71	10
SB1003 1079 vs. SBRM1 1079	1.027	0.02287	1.004	0.1047	3	3	9.584	10
1683								
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	A-?		
SB1003 1683 vs. SB1865 1683	0.4991	0.223 to 0.7752	Yes	**	0.0013	B	SB1865 1683	
SB1003 1683 vs. SB1866 1683	0.06873	-0.2074 to 0.3448	No	ns	0.8793	C	SB1866 1683	
SB1003 1683 vs. SBcckA 1683	0.669	0.3929 to 0.9451	Yes	***	0.0001	D	SBcckA 1683	
SB1003 1683 vs. SBRM1 1683	0.9796	0.7035 to 1.256	Yes	****	<0.0001	E	SBRM1 1683	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF

SB1003 1683 vs. SB1865 1683	1.01	0.5109	0.4991	0.09552	3	3	5.225	10
SB1003 1683 vs. SB1866 1683	1.01	0.9412	0.06873	0.09552	3	3	0.7195	10
SB1003 1683 vs. SBcckA 1683	1.01	0.3409	0.669	0.09552	3	3	7.004	10
SB1003 1683 vs. SBRM1 1683	1.01	0.03033	0.9796	0.09552	3	3	10.26	10
1684								
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	A-?		
SB1003 1684 vs. SB1865 1684	1.025	0.6205 to 1.43	Yes	***	0.0001	B	SB1865 1684	
SB1003 1684 vs. SB1866 1684	-0.6468	-1.052 to -0.2418	Yes	**	0.0032	C	SB1866 1684	
SB1003 1684 vs. SBcckA 1684	0.6952	0.2902 to 1.1	Yes	**	0.0019	D	SBcckA 1684	
SB1003 1684 vs. SBRM1 1684	1.033	0.6279 to 1.438	Yes	****	<0.0001	E	SBRM1 1684	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
SB1003 1684 vs. SB1865 1684	1.05	0.02428	1.025	0.1401	3	3	7.32	10
SB1003 1684 vs. SB1866 1684	1.05	1.697	-0.6468	0.1401	3	3	4.617	10
SB1003 1684 vs. SBcckA 1684	1.05	0.3546	0.6952	0.1401	3	3	4.962	10
SB1003 1684 vs. SBRM1 1684	1.05	0.0169	1.033	0.1401	3	3	7.372	10
1686								
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	A-?		

SB1003 1686 vs. SB1865 1686	1.007	0.7495 to 1.264	Yes	****	<0.0001	B	SB1865 1686	
SB1003 1686 vs. SB1866 1686	0.2448	-0.01267 to 0.5022	No	ns	0.0631	C	SB1866 1686	
SB1003 1686 vs. SBcckA 1686	0.5512	0.2938 to 0.8087	Yes	***	0.0004	D	SBcckA 1686	
SB1003 1686 vs. SBRM1 1686	1.006	0.749 to 1.264	Yes	****	<0.0001	E	SBRM1 1686	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
SB1003 1686 vs. SB1865 1686	1.012	0.004558	1.007	0.08906	3	3	11.31	10
SB1003 1686 vs. SB1866 1686	1.012	0.7668	0.2448	0.08906	3	3	2.748	10
SB1003 1686 vs. SBcckA 1686	1.012	0.4603	0.5512	0.08906	3	3	6.19	10
SB1003 1686 vs. SBRM1 1686	1.012	0.00514	1.006	0.08906	3	3	11.3	10
1865								
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	A-?		
SB1003 1865 vs. SB1865 1865	-4.27E+08	-427019905 to -427019893	Yes	****	<0.0001	B	SB1865 1865	
SB1003 1865 vs. SB1866 1865	-12.92	-19.16 to -6.676	Yes	***	0.0006	C	SB1866 1865	
SB1003 1865 vs. SBcckA 1865	0.4345	-6.545 to 7.414	No	ns	0.999	D	SBcckA 1865	
SB1003 1865 vs. SBRM1 1865	0.8246	-5.418 to 7.067	No	ns	0.9837	E	SBRM1 1865	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF

SB1003 1865 vs. SB1865 1865	1.016	427019900	-427019899	2.117	3	3	201665841	9
SB1003 1865 vs. SB1866 1865	1.016	13.93	-12.92	2.117	3	3	6.101	9
SB1003 1865 vs. SBcckA 1865	1.016	0.5812	0.4345	2.367	3	2	0.1835	9
SB1003 1865 vs. SBRM1 1865	1.016	0.1912	0.8246	2.117	3	3	0.3894	9
1866								
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	A-?		
SB1003 1866 vs. SB1865 1866	0.216	-0.03505 to 0.467	No	ns	0.0962	B	SB1865 1866	
SB1003 1866 vs. SB1866 1866	0.6469	0.3958 to 0.8979	Yes	****	<0.0001	C	SB1866 1866	
SB1003 1866 vs. SBcckA 1866	0.7802	0.5292 to 1.031	Yes	****	<0.0001	D	SBcckA 1866	
SB1003 1866 vs. SBRM1 1866	0.8111	0.56 to 1.062	Yes	****	<0.0001	E	SBRM1 1866	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
SB1003 1866 vs. SB1865 1866	1.009	0.7929	0.216	0.08685	3	3	2.487	10
SB1003 1866 vs. SB1866 1866	1.009	0.362	0.6469	0.08685	3	3	7.448	10
SB1003 1866 vs. SBcckA 1866	1.009	0.2287	0.7802	0.08685	3	3	8.983	10
SB1003 1866 vs. SBRM1 1866	1.009	0.1978	0.8111	0.08685	3	3	9.339	10

Appendix 4 – Other Scientific Contributions

Marc P. Grüll, Martin E. Mulligan, Andrew S. Lang, (2018) Small extracellular particles with big potential for horizontal gene transfer: membrane vesicles and gene transfer agents. *FEMS Microbiology Letters*, **365**(19): fny192, doi: 10.1093/femsle/fny192.

Lourdes Peña-Castillo, Marc P. Grüll, Martin E. Mulligan, Andrew S. Lang, (2016) Detection of bacterial small transcripts from RNA-Seq data: a comparative assessment. In *Biocomputing*: *Proceedings of the Pacific Symposium:* 456-467.

Hynes, A., Shakya, M., Mercer, R., Grüll, M., Bown, L., Davidson, F., Steffen, E., Matchem, E., Peach, M., Berger, T., Grebe, K., Zhaxybayeva, O., Lang, A.S., (2016). Functional and evolutionary characterization of a gene transfer agent's multilocus "genome". *Molecular Biology and Evolution* **33**(10): 2530-2543, doi: 10.1093/molbev /msw125.

Ayre, D.C., Chute, I.C., Joy, A.P., Barnett, D.A., Hogan, A.M., Grüll, M.P., Peña-Castillo, L., Lang, A.S., Lewis, S.M., Christian, S.L., (2017). CD24 induces changes to the surface receptors of B cell microvesicles with variable effects on their RNA and protein cargo. *Scientific Reports*, *7*(1), 8642, doi: 10.1038/s41598-017-08094-8.

FEMS Microbiology Letters, 365, 2018, fny192

doi: 10.1093/femsle/fny192
Advance Access Publication Date: 6 August 2018
Minireview

MINIREVIEW – Environmental Microbiology

Small extracellular particles with big potential for horizontal gene transfer: membrane vesicles and gene transfer agents

M.P. Grüll[1], M.E. Mulligan[2] and A.S. Lang[1,*]

[1]Departments of Biology and [2]Biochemistry, Memorial University of Newfoundland, St John's, NL, Canada

*Corresponding author: Department of Biology, Memorial University of Newfoundland, 232 Elizabeth Ave., St John's, NL A1B 3X9, Canada. Tel/Fax: +1-709-864-7517; E-mail: aslang@mun.ca

One sentence summary: Prokaryotic cells release various molecules and supramolecular structures into the extracellular environment including membrane vesicles and virus-like particles, a comparison of which is discussed in this minireview.

Editor: Olga Ozoline

ABSTRACT

Bacteria are known to release different types of particles that serve various purposes such as the processing of metabolites, communication, and the transfer of genetic material. One of the most interesting aspects of the production of such particles is the biogenesis and trafficking of complex particles that can carry DNA, RNA, proteins or toxins into the surrounding environment to aid in bacterial survival or lead to gene transfer. Two important bacterial extracellular complexes are membrane vesicles and gene transfer agents. In this review, we will discuss the production, contents and functions of these two types of particles as related to their abilities to facilitate horizontal gene transfer.

Keywords: gene transfer; extracellular environment; vesicle; virus

INTRODUCTION

Gene transfer agents (GTAs) were first discovered in 1974 in the bacterium *Rhodobacter capsulatus* as virus-sized, DNase-resistant and protease-sensitive mediators of cell contact-independent genetic exchange (Marrs 1974; Solioz, Yen and Marrs 1975). They are tailed phage-like particles that contain DNA from the producing cell's genome (Lang and Beatty 2007; Stanton 2007) which are released into the extracellular environment via lysis of the producing cell (Matson *et al.* 2005; Fogg, Westbye and Beatty 2012; Hynes *et al.* 2012). GTAs have now been documented in a wide range of prokaryotes, including both bacteria and archaea (Marrs 1974; Rapp and Wall 1987; Humphrey *et al.* 1997; Bertani 1999; Biers *et al.* 2008; Berglund *et al.* 2009; Nagao *et al.* 2015; Tomasch *et al.* 2018). The factors controlling GTA production appear to differ between different organisms but, to date, the sole known function for GTAs is to mediate gene transfer.

Membrane vesicles (MVs) associated with prokaryotes were first discovered in 1965 in cell-free culture supernatants of *Escherichia coli* (Bishop and Work 1965) and their production has subsequently been documented from additional Gram-negative and Gram-positive bacteria and archaea (Beveridge 1999; Ellen *et al.* 2009; Rivera *et al.* 2010; Thay, Wai and Oscarsson 2013). The production of MVs can occur during all stages of bacterial growth and the amount of MVs produced can be quite substantial. In *E. coli* it has been found that MVs typically represent 1–5% of total outer membrane material in growing cultures (Gankema *et al.* 1980; Wensink and Witholt 1981; Wai *et al.* 2003); in log-phase cultures of *Neisseria meningitidis*, the total amount of released MVs corresponds to 12% of the total lipopolysaccharides (LPS) (Devoe 1973). Unlike GTAs, MVs can contain and transport a variety of cargo (Table 1) and, in addition to mediating horizontal gene transfer (HGT), their known activities include trafficking of signals for quorum sensing, defence against viral infection,

Received: 5 February 2018; **Accepted:** 4 August 2018

Table 1. Examples of different cargo carried by MVs.

Organism	Cargo	References
Aggregatibacter actinomycetemcomitans[a]	LPS, endotoxin, leukotoxin, Gro-EL homolog	Nowotny *et al.* 1982; Goulhen *et al.* 1998; Kato, Kowashi and Demuth 2002
Porphyromonas gingivalis[b]	Proteases	Grenier and Mayrand 1987
Escherichia coli	Shiga toxins 1 and 2, DNA	Kolling and Matthews 1999
Bartonella henselae	Hemin-binding proteins	Roden *et al.* 2012
Pseudomonas aeruginosa	PaAP aminopeptidase, DNA	Bauman and Kuehn 2006
Prochlorococcus	Proteins, DNA, RNA	Biller *et al.* 2014
Streptomyces coelicolor	Proteins	Schrempf *et al.* 2011
Pseudomonas fragi	Protease	Thompson, Naidu and Pestka 1985
Lysobacter sp.	Proteins (endopeptidase L5)	Vasilyeva *et al.* 2008

[a]Formerly *Actinobacillus actinomycetemcomitans*
[b]Formerly *Bacteroides gingivalis*

transport of toxic compounds, and they can also play a role in pathogenicity (Devoe 1973; Loeb and Kilner 1978; Nowotny *et al.* 1982; Kadurugamuwa and Beveridge 1997; Horstman and Kuehn 2000; Yaron *et al.* 2000; Renelli *et al.* 2004; Mashburn and Whiteley 2005; Rivera *et al.* 2010; Thay, Wai and Oscarsson 2013; Deo *et al.* 2018). Across different microorganisms, some MVs' functions are conserved (Deatherage and Cookson 2012), and similar mechanistic release strategies based on membrane blebbing have been observed (Rivera *et al.* 2010).

It is possible that MVs and GTAs exist in abundance in all environments on Earth (McDaniel *et al.* 2010; Biller *et al.* 2014; Nagao *et al.* 2015), but identifying them and distinguishing them from true viral particles represents a challenge. MVs, just like viruses and virus-like particles, can carry enough DNA to be detected using DNA stains and fluorescence microscopy (Brussaard 2009; Ortmann and Suttle 2009). The morphology, dimensions and molecular composition of MVs can be very similar to those of virions (Forterre *et al.* 2013); likewise, GTAs have the structure of bacteriophage particles and contain DNA (Stanton 2007; Lang, Zhaxybayeva and Beatty 2012; Lang, Westbye and Beatty 2017). Recent studies have therefore suggested that earlier estimates of the abundance of viruses or virus-like particles in the environment might be overestimates due to the unrecognized occurrence of MVs and GTAs (Forterre *et al.* 2013; Soler *et al.* 2015). Correspondingly, it is therefore very likely that the numbers of MVs and GTAs in environmental samples have been greatly underestimated, and that their roles in genetic transfer in nature could be much greater than previously appreciated.

HGT is one of the most powerful forces in the evolution of bacteria and archaea. There are three canonical mechanisms by which HGT can occur: (i) conjugation (direct cell-to-cell transfer of elements such as plasmids), (ii) natural transformation (uptake of free DNA from the environment), and (iii) transduction (transfer by phage particles). GTAs incorporate genomic DNA into particles that resemble phages and their mechanism of gene transfer is most similar to transduction. MVs resemble enveloped viruses and, similar to GTAs, their DNA cargo is protected inside after release from the producing cell, and they can then attach and introduce the DNA content into recipients. However, while both GTAs and MVs facilitate HGT in ways similar to transduction, the formation of GTAs and MVs, their selection of DNA to be incorporated into the particles, and the mechanism of their release from the host cells, are very different. In this review we attempt to highlight the similarities and differences between MVs and GTAs with respect to their activities in HGT.

We refer readers to previous reviews that cover many additional details concerning GTAs (Lang and Beatty 2001; Stanton 2007; Lang, Zhaxybayeva and Beatty 2012; Lang, Westbye and Beatty 2017; Westbye, Beatty and Lang 2017) and MVs (McBroom and Kuehn 2005; Ellis and Kuehn 2010; Kulp and Kuehn 2010; Haurat, Elhenawy and Feldman 2015).

PRODUCTION AND RELEASE OF DNA-CONTAINING PARTICLES

MVs are small spherical structures that vary in size from approximately 10 to 300 nm in diameter (Ellis and Kuehn 2010; Kulp and Kuehn 2010; Haurat, Elhenawy and Feldman 2015; Tsatsaronis *et al.* 2018). For a given organism, the sizes of MVs produced are not uniform, and different sizes of MVs can be produced depending on environmental factors and cell stressors (McBroom and Kuehn 2007). In Gram-negative bacteria, MVs can be produced from live cells by a mechanism called blebbing, which is an outward bulging of the outer membrane in an area that lacks membrane-peptidoglycan bonds (Chatterjee and Das 1967; Grenier and Mayrand 1987; Mayrand 1989; Kadurugamuwa and Beveridge 1995; Patrick *et al.* 1996; Beveridge 1999; Kolling and Matthews 1999). Blebbing is then followed by membrane fission that does not require direct energy from ATP/GTP (McBroom and Kuehn 2005; Kulp and Kuehn 2010). The process is comparable to viral budding from the membrane of wall-less eukaryotic cells. In Gram-positive bacteria, the process of MV production is similar to Gram-negative bacteria in that these vesicles are also created by an outward bulging of the membrane (Prados-Rosales *et al.* 2011). However, since Gram-positive bacteria lack an outer membrane, the MVs are derived from the cytoplasmic membrane in areas of the cell where membrane microdomains protrude (Mayer and Gottschalk 2003). In archaean organisms, it has been found that vesicles are also released in a budding-off process similar to that found in Gram-negative bacteria, but the details of this process have not been well characterized (Näther and Rachel 2001). Overall, mechanisms that control the formation of MVs are not well understood (Kulp and Kuehn 2010).

Another mechanism for the biogenesis of MVs, which was recently identified in *Pseudomonas aeruginosa* biofilms using dynamic super-resolution microscopy, is by means of 'explosive cell lysis' (Turnbull *et al.* 2016). The cell lysis is due to the activity of a prophage endolysin, and membrane fragments and cellular components are released into the extracellular milieu.

The cellular components can then be captured by MVs that form from the released membrane fragments. This raises the possibility that MV formation might accompany cell lysis events associated with phage or GTA release in other organisms.

Different types of nucleic acids, such as chromosomal, plasmid or phage DNA and RNA, can be found associated with MVs (Biller *et al.* 2014; Gaudin *et al.* 2014; Erdmann *et al.* 2017; Tsatsaronis *et al.* 2018), and can be carried either on the outer surface (Perez-Cruz *et al.* 2013; Biller *et al.* 2014; Perez-Cruz *et al.* 2015; Bitto *et al.* 2017) or in the lumen (Renelli *et al.* 2004; Fulsundar *et al.* 2014; Gaudin *et al.* 2014; Ho *et al.* 2015). Nucleic acid cargo that is carried inside the particles is protected against nucleases in the extracellular milieu (Rumbo *et al.* 2011; Biller *et al.* 2014; Fulsundar *et al.* 2014; Jiang *et al.* 2014), which makes it more likely they will survive and be delivered to other cells. In most cases it is unclear if nucleic acids are actively transported to MVs or if their presence in these particles is the result of random processes during MV formation (Domingues and Nielsen 2017). Transformation studies investigating competence in *Acinetobacter baylyi* identified genes involved in DNA transport that could in addition be involved in the packaging of DNA into MVs (Link *et al.* 1998; Busch, Rosenpläntер and Averhoff 1999; Herzberg, Friedrich and Averhoff 2000; Averhoff and Graf 2008). It was also found that antibiotic stress increases the DNA content in the MVs released by this bacterium (Fulsundar *et al.* 2014). The recent findings of production of MVs through explosive cell lysis gives another mechanism to explain how DNA, RNA and cytoplasmic material can be found in MVs, but it seems unlikely that the cargo of MVs can be precisely controlled with this mechanism (Turnbull *et al.* 2016).

One example for specific selection of cargo and function are the plasmid-mobilizing MVs produced by the archaeon *Halorubrum lacusprofundi* where a plasmid encodes the genes for the production of a specialized MV, referred to as a PV (plasmid vesicle), which is responsible for packaging the plasmid DNA into the PV and for its transmission to other cells (Erdmann *et al.* 2017). This plasmid MV virus-like hybrid element is the latest example of the blurred distinction between what should be called a plasmid or a virus. Such blurred boundaries seem common within the spectrum of mobile genetic elements (e.g. integrative and conjugative elements that have properties of both plasmids and transposons (Burrus 2017), and transposable phages that have properties of both phages and transposons (Toussaint and Rice 2017)).

GTAs resemble small-headed tailed bacteriophages and several are known to be released from the producing cell by lysis (Matson *et al.* 2005; Fogg, Westbye and Beatty 2012; Hynes *et al.* 2012). Their capsid sizes vary from 30 to 80 nm and they contain 4–14 kbp of DNA packaged in the protein capsid shell (Yen, Hu and Marrs 1979; Rapp and Wall 1987; Humphrey, Stanton and Jensen 1995; Eiserling, Pushkin and Bertani 1999; Barbian and Minnick 2000; Nagao *et al.* 2015; Tomasch *et al.* 2018). Bacteriophages usually package their entire genomes and are able to replicate in suitable infected host cells. By contrast, GTA particles do not package sufficient DNA to encode their own production and no replication occurs after the encapsidated DNA is injected into a cell by a GTA particle. GTAs contain fragments of DNA from across the entire genome of the host cell and most GTA particles do not contain any GTA-encoding genes (Lang and Beatty 2007; Stanton 2007; Lang, Zhaxybayeva and Beatty 2012). This means that GTAs are non-replicative.

The genes that encode GTA structural proteins are contained within the hosts' genomes and are known for several of the GTAs. For example, a cluster of ~15 open reading frames encodes most of the *R. capsulatus* GTA (RcGTA) particle proteins (Lang and Beatty 2000; Chen *et al.* 2008), with at least three additional structural proteins encoded at two distant loci (Hynes *et al.* 2016; Westbye *et al.* 2016). The *Bartonella* GTA (BaGTA) genes are organized within a 80 kb-long genome segment (Quebatte *et al.* 2017) with a cluster of 11 genes at the 5' end of this BaGTA genome region that belong to the bona fide GTA locus that had previously been inferred by sequence conservation analysis (Guy *et al.* 2013). The GTA of *Brachyspira hyodysenteriae* is encoded in two co-regulated gene clusters (Matson *et al.* 2005; Stanton *et al.* 2009). Unlike *R. capsulatus* and *Bartonella* spp., the *B. hyodysenteriae* GTA structural genes show no sequence homology with known phage genes. The organizations of the GTA structural genes are not yet known for some other organisms such as *Desulfovibrio desulfuricans* and *Methanococcus voltae*.

As mentioned above, GTA particles appear to contain DNA from throughout the producing cells' genomes when examined using low-resolution techniques such as agarose gel electrophoresis (Yen, Hu and Marrs 1979; Rapp and Wall 1987; Anderson *et al.* 1994; Humphrey *et al.* 1997; Bertani 1999). Hybridization of DNA from *R. capsulatus* GTA particles to a *R. capsulatus* whole-genome microarray showed that DNA packaging is essentially random in this organism (Hynes *et al.* 2012). In *Bartonella*, microarray hybridization of the GTA-packaged DNA showed that a phage-derived region of the genome associated with run-off replication is packaged more frequently (Berglund *et al.* 2009), presumably due to its higher copy number. Subsequent gene transfer experiments showed that this GTA transfers chromosomal DNA, but not episomal DNA (Quebatte *et al.* 2017). The first documented clear deviation from random packaging comes from the bacterium *Dinoroseobacter shibae* where high-throughput sequencing of GTA-packaged DNA showed that, although all genomic DNA could be detected in the particles, chromosomal position and GC content are associated with differences in packaging, and DNA from different replicons in the cells are packaged at very different rates (Tomasch *et al.* 2018). DNA packaging into tailed phage particles is performed by the terminase enzyme (Casjens 2011), and the *Bartonella, R. capsulatus* and *D. shibae* GTA clusters contain predicted terminase-encoding genes (Lang and Beatty 2000; Berglund *et al.* 2009; Tomasch *et al.* 2018). The sequence specificity of this enzyme, or lack thereof, possibly combined with other features of the genomic DNA, are expected to determine exactly what DNA is packaged into the GTA particles.

The production of GTA particles is dependent on the host's physiology. For example, in *R. capsulatus* the production of GTA particles varies according to the growth phase of lab cultures (Solioz, Yen and Marrs 1975) and is affected by phosphate concentration (Westbye *et al.* 2013); production of GTAs is increased in response to certain DNA-damaging antibiotics in *B. hyodysenteriae* (Humphrey, Stanton and Jensen 1995; Stanton *et al.* 2008), and the stringent response controls GTA production in *Bartonella* (Quebatte *et al.* 2017). Regulation of GTA production has been most extensively studied in *R. capsulatus*, and expression of its GTA genes and release of the particles into the extracellular milieu are regulated by several cellular regulatory systems and two phage-related genes located outside of the main structural gene cluster (reviewed in (Lang, Zhaxybayeva and Beatty 2012; Lang, Westbye and Beatty 2017; Westbye, Beatty and Lang 2017)). The cellular regulators include multiple regulatory systems (Lang and Beatty 2000; Mercer *et al.* 2012; Westbye *et al.* 2013; Mercer and Lang 2014; Kuchinski *et al.* 2016; Westbye *et al.* 2018), with the growth-phase regulation occurring via a long-chain acyl-homoserine lactone quorum-sensing system that was

found to induce five- to seven-fold higher GTA gene expression in the stationary phase (Schaefer *et al.* 2002). The mechanisms by which the phage-related regulatory genes affect the production of GTAs is not yet known.

EFFECTS OF RECEIVING DNA FROM MVS AND GTAS

The effects that MVs can have when released into their environment obviously depends on their content. Interactions with a recipient cell can be via MV lysis, which allows the content of an MV to subsequently diffuse through the cell membrane into target cells (Dorward and Garon 1990; Schooling, Hubley and Beveridge 2009), or could lead to DNA transfer if the recipient cell is competent for transformation. Indeed, competence proteins seem to play a role in the uptake of DNA delivered by MVs in *Acinetobacter baylyi* (Fulsundar *et al.* 2014). Another type of interaction with a target organism is via direct attachment followed by internalization or membrane fusion (Kadurugamuwa and Beveridge 1996; Kadurugamuwa and Beveridge 1999), which can deliver DNA directly into the recipient cell. A recent study showed that internal DNA has the potential for genetic exchange (Bitto *et al.* 2017), while surface-bound DNA appears to be important for biofilm formation and protection of the biofilm (Rumbo *et al.* 2011; Gloag *et al.* 2013; Liao *et al.* 2014). However, treatment of MVs with DNase was found to lower the efficiency of gene transfer in another study with *Porphyromonas gingivalis* (Ho *et al.* 2015), suggesting that DNA on the surface can be important for HGT.

MVs carrying DNA have been known to mediate HGT for some time, but the mechanisms by which they achieve the transfer of genetic material have only been identified in recent years. The first demonstration of DNA and RNA content in MVs was from *Neisseria gonorrhoeae* almost 30 years ago (Dorward, Garon and Judd 1989). In that study, MV-mediated transfer of plasmid DNA was observed in the presence of DNase and resulted in the efficient exchange of penicillinase-specifying R plasmids. Transfer of plasmids seems to be a common theme for MVs. Another example is *Acinetobacter baumannii*, where the MVs it produces are able to transfer plasmid DNA as well as the OXA-24 carbapenemase gene (Rumbo *et al.* 2011), showing the potential importance of MVs in the dissemination of antimicrobial resistance. Plasmid transfer by MVs has also been demonstrated in archaea, where *Thermococcus kodakaraensis* released MVs harbouring a plasmid that were able to transfer the plasmid into plasmid-free cells (Gaudin *et al.* 2013). Also, as mentioned in the previous section, a plasmid was shown to direct the formation of MVs for its packaging and transmission in the archaeon *H. lacusprofundi* (Erdmann *et al.* 2017). MV-mediated DNA transfer has also been documented in Gram-positive bacteria, with the MV-containing fraction of a culture supernatant of *Ruminococcus* sp. able to restore a specific metabolic activity to mutant strains (Klieve *et al.* 2005).

Another membrane-derived particle that slightly differs from any other MV was identified by adding environmental samples to *E. coli* amino acid auxotrophs and selecting for prototrophs (Chiura *et al.* 2011). The recipient cells were able to produce new MVs that contained up to 370 kbp of DNA and were capable of gene transfer, a phenomenon termed serial transduction. This DNA content is much higher than any other known MVs that we are aware of, and the process resembled generalized transduction as no markers were preferentially transferred.

Overall, it is now known that MVs produced by bacteria and archaea can move DNA between cells. The fate of any transferred DNA (i.e. maintained versus degraded in the recipient cell) and the genes present on any transferred DNA will obviously be the key determinants of the overall outcome for a recipient cell. MV transfer of plasmids has been repeatedly detected, suggesting that this may be an underappreciated mechanism for plasmid movement among organisms. Membrane-derived particles that are capable of inducing serial transduction in recipient cells have also been detected, indicating that particles very similar to MVs may be capable of more than just HGT but can furthermore transmit the capability for particle production in some instances. More research is required to identify specific underlying mechanisms that select specific cargo for packaging and achieve the transfer of genetic material. At this time, we are unaware of any studies indicating that specific organisms use MV-mediated DNA transfer as a means of recombination within their population, but that seems to be one possibility for MV function.

The potential for GTAs to transfer DNA among different species has not been widely investigated. Due to their tailed phage-like structures, it has been assumed that the GTA particles bind to their recipient cells via specific tail-receptor interactions (Lang, Zhaxybayeva and Beatty 2012), but protein receptors have not yet been identified. A requirement for such specific interactions would be one limitation to GTA DNA transfer. While not all isolates of *R. capsulatus* are capable of producing GTAs, 25 of 33 tested *Rhodobacter* strains were able to act as recipients and successfully take up DNA from GTA particles (Wall, Weaver and Gest 1975). The transfer of genetic material to related species was not observed (Wall, Weaver and Gest 1975). In *R. capsulatus*, the GTAs contain head spikes (Westbye *et al.* 2016) that are needed for the binding to the *R. capsulatus* polysaccharide cell capsule (Brimacombe *et al.* 2013). Although the head spikes are not essential for GTA gene transfer, they greatly enhance the GTA–cell interactions (Hynes *et al.* 2016; Westbye *et al.* 2016). In contrast to RcGTA and its apparently narrow host range, GTA transfer from *Roseovarius nubinhibens* showed broad gene transfer capability and interspecific transfer at very high frequencies (McDaniel *et al.* 2010).

Once DNA carried by the GTA particles is introduced into a recipient cell, the linear DNA molecule then needs to be recombined, via homologous or non-homologous recombination, into the cell's genome in order to be maintained. Homologous recombination involving the recombinase RecA is involved in GTA-transduced DNA incorporation in *R. capsulatus* (Genthner and Wall 1984; Brimacombe, Ding and Beatty 2014), and is dependent on the presence of DprA. DprA was found to non-specifically bind incoming double-stranded DNA, protecting it from endonuclease digestion (Brimacombe, Ding and Beatty 2014). The competence proteins, ComEC and ComF, which are putative DNA transport proteins, are also involved in the process of DNA uptake, and they facilitate the entry of DNA from GTAs into the cytoplasm of recipient cells (Brimacombe *et al.* 2015). The expression of the genes encoding the capsule biosynthesis system and these DNA uptake and processing proteins is co-regulated with genes for the production and release of GTA particles, such that a small proportion of the population becomes activated to make and release GTAs while the remainder of the population becomes active to receive the DNA from GTAs (Lang, Westbye and Beatty 2017; Westbye, Beatty and Lang 2017).

It was recently shown that the induction of GTA production in *Bartonella* is also restricted to a subset of the population (Quebatte *et al.* 2017). In this system, the stringent response signal guanosine-tetraphosphate limits GTA production to the

fast-growing cells in the population. The uptake of DNA from GTAs is only done by actively dividing cells because it requires the presence of the Tol-Pal trans-envelope complex, which is needed to mediate outer-membrane fusion upon cell division. Furthermore, DNA transfer into the cells also requires homologs of the same competence and recombination proteins required in *R. capsulatus*. The outer-membrane TolC and HbpC proteins were also identified as candidate protein receptors on the cell surface for the GTAs.

It has been proposed that the *Bartonella* GTA benefits these bacteria in the expansion of their host range by facilitating transfer of host-adaptation systems (Berglund *et al.* 2009). This was based on the finding that highly dynamic regions of the chromosome, which contain many gene clusters for secretion systems, were extensively amplified and packaged into BaGTA particles (Berglund *et al.* 2009), which were subsequently shown to be functional for gene transfer (Guy *et al.* 2013). Subsequent experiments that revealed details about regulation of its production and limits on which cells can receive the BaGTA DNA indicate this may be a strategy to promote genetic exchange within the fittest members of the population and to disfavour the exchange of detrimental mutations (Quebatte *et al.* 2017).

As described in the previous section, the genes encoding the GTA particles are known for the *B. hyodysenteriae*, *R. capsulatus* and *Bartonella* GTAs. These three GTAs are genetically distinct, indicating they have independent evolutionary origins, and are presumed to be derived from prophage ancestors. Complete sets of homologs of the *R. capsulatus* GTA genes are widely distributed in bacteria belonging to the same taxonomic family within the class α-proteobacteria, the Rhodobacteraceae, with more limited conservation also found in other members of this class (Lang and Beatty 2001; Lang, Taylor and Beatty 2002; Lang and Beatty 2007; Biers *et al.* 2008; Paul 2008; Lang, Zhaxybayeva and Beatty 2012; Shakya, Soucy and Zhaxybayeva 2017). GTA production has been documented for some of these bacteria (Biers *et al.* 2008; McDaniel *et al.* 2010; Nagao *et al.* 2015; Tomasch *et al.* 2018). The *Bartonella* GTA genes are also well-conserved in the Bartonellaceae but show much more limited distribution outside of this family (Berglund *et al.* 2009; Guy *et al.* 2013). For both of these GTAs, the distribution of their genes and evolutionary relationships among the homologs in different genomes argue they have had long evolutionary associations in these bacterial lineages. Furthermore, analysis of nucleotide substitution patterns in their genes indicates they are under similar purifying selection as bona fide cellular genes and different from prophage genes (Lang, Zhaxybayeva and Beatty 2012).

GTAs have evolved independently in divergent organisms, and while possible benefits to recipient cells of receiving DNA from GTA particles have been proposed, these have not been proven. Perhaps there are other functions for GTAs, or the products of their genes, in the producing organisms, though such possible functions remain obscure. GTAs are highly conserved, under selection and specifically regulated in some lineages, which suggests that they probably provide benefits to the producing organisms. Homologous recombination within a prokaryotic population has been predicted to have benefits, which could include the restoration of genes that have been inactivated through mutations (Marrs, Wall and Gest 1977; Takeuchi, Kaneko and Koonin 2014) or the repair of a genome after insertion of a parasitic element (Croucher *et al.* 2016). A mathematical model suggests that competence for transformation in an organism provides an effective defence against parasitic mobile elements (Croucher *et al.* 2016). This could therefore be an important consideration for the long-term maintenance of GTA production in prokaryotes through their prevention of stochastic, irreversible deterioration of genomes.

CONCLUDING REMARKS

Bacteria and archaea release many different compounds and particles from their cells. These include MVs, which might be universally produced, and GTAs, which are currently known from only a few genera. Both of these are capable of transferring DNA from a producing cell to a recipient and thereby mediating HGT. It seems likely that the term MV encompasses a diverse collection of particles that are only unified by the fact that they originate from membranes, with diverse functions carried out by differentiated particles. Some MVs have links with mobile genetic elements such as plasmids and viruses, and more research is needed to fully understand their evolutionary origins, the magnitude of their role in HGT, and to identify the mechanisms by which DNA is selected for packaging. It is a bit simpler for GTAs, which are clearly evolutionarily related to phages. Transfer of DNA by both types of particles has been found to involve proteins associated with natural competence in some instances, indicating an overlap between transduction-like and competence-based DNA transfer and blurring the distinctions typically made among the canonical mechanisms for HGT. Possible benefits of producing and receiving DNA from both types of particles remain to be proven.

We now recognize that viruses and virus-like particles are present and abundant in all environments on Earth, and the same seems to be true for MVs. One of the first techniques used to attempt to quantify viruses and particles of similar size in environmental samples is epifluorescence microscopy. It was recently pointed out that MVs carrying DNA would mimic viral particles under epifluorescence microscopy, as would GTAs if they contained enough DNA to be detected (Forterre *et al.* 2013; Soler *et al.* 2015). Therefore, further work is needed to specifically quantify MVs and GTAs in environmental samples to understand how much of a contribution they make to what we currently view as the viral fraction within environmental samples (Forterre 2013). This will then also allow us to understand better the contributions of MVs and GTAs to gene exchange in natural environments.

ACKNOWLEDGEMENT

Research in ASL's lab is funded by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2017-04636).

Conflict of interest. None declared.

REFERENCES

Anderson B, Goldsmith C, Johnson AP *et al.* Bacteriophage-like particle of *Rochalimaea henselae*. *Mol Microbiol* 1994;**31**:67–73.

Averhoff B, Graf I. The natural transformation system of *Acinetobacter baylyi* ADP1: a unique DNA transport machinery. In: Gerischer U (ed). *Acinetobacter Molecular Biology*. Poole, UK: Caister Academic Press, 2008, 19–39.

Barbian KD, Minnick MF. A bacteriophage-like particle from *Bartonella bacilliformis*. *Microbiology* 2000;**146**:599–609.

Bauman SJ, Kuehn MJ. Purification of outer membrane vesicles from *Pseudomonas aeruginosa* and their activation of an IL-8 response. *Microbes Infect* 2006;**8**:2400–8.

Berglund EC, Frank AC, Calteau A *et al.* Run-off replication of host-adaptability genes is associated with gene transfer

agents in the genome of mouse-infecting *Bartonella grahamii*. *PLoS Genet* 2009;**5**:e1000546.

Bertani G. Transduction-like gene transfer in the methanogen *Methanococcus voltae*. *J Bacteriol* 1999;**181**:2992–3002.

Beveridge TJ. Structures of Gram-negative cell walls and their derived membrane vesicles. *J Bacteriol* 1999;**181**:4725–33.

Biers EJ, Wang K, Pennington C *et al.* Occurrence and expression of gene transfer agent genes in marine bacterioplankton. *Appl Environ Microbiol* 2008;**74**:2933–9.

Biller SJ, Schubotz F, Roggensack SE *et al.* Bacterial vesicles in marine ecosystems. *Science* 2014;**343**:183–6.

Bishop DG, Work E. An extracellular glycolipid produced by *Escherichia coli* grown under lysine-limiting conditions. *Biochem J* 1965;**96**:567–76.

Bitto NJ, Chapman R, Pidot S *et al.* Bacterial membrane vesicles transport their DNA cargo into host cells. *Sci Rep* 2017;**7**:7072.

Brimacombe CA, Ding H, Beatty JT. *Rhodobacter capsulatus* DprA is essential for RecA-mediated gene transfer agent (RcGTA) recipient capability regulated by quorum-sensing and the CtrA response regulator. *Mol Microbiol* 2014;**92**:1260–78.

Brimacombe CA, Ding H, Johnson JA *et al.* Homologues of genetic transformation DNA import genes are required for *Rhodobacter capsulatus* gene transfer agent recipient capability regulated by the response Regulator CtrA. *J Bacteriol* 2015;**197**:2653–63.

Brimacombe CA, Stevens A, Jun D *et al.* Quorum-sensing regulation of a capsular polysaccharide receptor for the *Rhodobacter capsulatus* gene transfer agent (RcGTA). *Mol Microbiol* 2013;**87**:802–17.

Brussaard CPD. Enumeration of bacteriophages using flow cytometry. *Methods Mol Biol* 2009;**501**:97–111.

Burrus V. Mechanisms of stabilization of integrative and conjugative elements. *Curr Opin Microbiol* 2017;**38**:44–50.

Busch S, Rosenplänter C, Averhoff B. Identification and characterization of ComE and ComF, two novel pilin-like competence factors involved in natural transformation of *Acinetobacter* sp. strain BD413. *Appl Environ Microbiol* 1999;**65**:4568–74.

Casjens SR. The DNA-packaging nanomotor of tailed bacteriophages. *Nat Rev Micro* 2011;**9**:647–57.

Chatterjee SN, Das J. Electron microscopic observations on the excretion of cell-wall material by *Vibrio cholerae*. *J Gen Microbiol* 1967;**49**:1–11.

Chen F, Spano A, Goodman BE *et al.* Proteomic Analysis and Identification of the Structural and Regulatory Proteins of the Rhodobacter capsulatus Gene Transfer Agent. *J Proteome Res* 2009;**8**:967–73.

Chiura HX, Kogure K, Hagemann S *et al.* Evidence for particle-induced horizontal gene transfer and serial transduction between bacteria. *FEMS Microbiol Ecol* 2011;**76**:576–91.

Croucher NJ, Mostowy R, Wymant C *et al.* Horizontal DNA transfer mechanisms of bacteria as weapons of intragenomic conflict. *PLoS Biol* 2016;**14**:e1002394.

Deatherage BL, Cookson BT. Membrane vesicle release in bacteria, eukaryotes, and archaea: a conserved yet underappreciated aspect of microbial life. *Infect Immun* 2012;**80**:1948–57.

Deo P, Chow SH, Hay ID *et al.* Outer membrane vesicles from *Neisseria gonorrhoeae* target PorB to mitochondria and induce apoptosis. *PLoS Pathog* 2018;**14**:e1006945.

Devoe IW. Release of endotoxin in the form of cell wall blebs during in vitro growth of *Neisseria meningitidis*. *J Exp Med* 1973;**138**:1156–67.

Domingues S, Nielsen KM. Membrane vesicles and horizontal gene transfer in prokaryotes. *Curr Opin Microbiol* 2017;**38**:16–21.

Dorward DW, Garon CF. DNA Is packaged within membrane-derived vesicles of Gram-negative but not Gram-positive bacteria. *Appl Environ Microbiol* 1990;**56**:1960–2.

Dorward DW, Garon CF, Judd RC. Export and intercellular transfer of DNA via membrane blebs of *Neisseria gonorrhoeae*. *J Bacteriol* 1989;**171**:2499–505.

Eiserling F, Pushkin A, Bertani G. Bacteriophage-like particles associated with the gene transfer agent of *Methanococcus voltae* PS. *J Gen Virol* 1999;**80**:3305–8.

Ellen AF, Albers SV, Huibers W *et al.* Proteomic analysis of secreted membrane vesicles of archaeal *Sulfolobus* species reveals the presence of endosome sorting complex components. *Extremophiles* 2009;**13**:67–79.

Ellis TN, Kuehn MJ. Virulence and immunomodulatory roles of bacterial outer membrane vesicles. *Microbiol Mol Biol Rev* 2010;**74**:81–94.

Erdmann S, Tschitschko B, Zhong L *et al.* A plasmid from an Antarctic haloarchaeon uses specialized membrane vesicles to disseminate and infect plasmid-free cells. *Nat Microbiol* 2017;**2**:1446–55.

Fogg PC, Westbye AB, Beatty JT. One for all or all for one: heterogeneous expression and host cell lysis are key to gene transfer agent activity in *Rhodobacter capsulatus*. *PLoS One* 2012;**7**:e43772.

Forterre P. The virocell concept and environmental microbiology. *ISME J* 2013;**7**:233–6.

Forterre P, Soler N, Krupovic M *et al.* Fake virus particles generated by fluorescence microscopy. *Trends Microbiol* 2013;**21**:1–5.

Fulsundar S, Harms K, Flaten GE *et al.* Gene Transfer Potential of Outer Membrane Vesicles of *Acinetobacter baylyi* and Effects of Stress on Vesiculation. *Appl Environ Microbiol* 2014;**80**:3469–83.

Gankema H, Wensink J, Giuinee PAM *et al.* Some characteristics of the outer membrane material released by growing enterotoxigenic *Escherichia coli*. *Infect Immun* 1980;**29**:704–13.

Gaudin M, Gauliard E, Schouten S *et al.* Hyperthermophilic archaea produce membrane vesicles that can transfer DNA. *Env Microbiol Rep* 2013;**5**:109–16.

Gaudin M, Krupovic M, Marguet E *et al.* Extracellular membrane vesicles harbouring viral genomes. *Environ Microbiol* 2014;**16**:1167–75.

Genthner FJ, Wall JD. Isolation of a recombination deficient mutant of *Rhodopseudomonas capsulata*. *J Bacteriol* 1984;**160**:971–5.

Gloag ES, Turnbull L, Huang A *et al.* Self-organization of bacterial biofilms is facilitated by extracellular DNA. *Proc Natl Acad Sci* 2013;**110**:11541–6.

Goulhen F, Hafezi A, Uitto V-J *et al.* Subcellular localization and cytotoxic activity of the GroEL-like protein isolated from *Actinobacillus actinomycetemcomitans*. *Infect Immun* 1998;**66**:5307–13.

Grenier D, Mayrand D. Functional characterization of extracellular vesicles produced by *Bacteroides gingivalis*. *Infect Immun* 1987;**55**:111–7.

Guy L, Nystedt B, Toft C *et al.* A gene transfer agent and a dynamic repertoire of secretion systems hold the keys to the explosive radiation of the emerging pathogen *Bartonella*. *PLoS Genet* 2013;**9**:e1003393.

Haurat MF, Elhenawy W, Feldman MF. Prokaryotic membrane vesicles: new insights on biogenesis and biological roles. *Biol Chem* 2015;**396**:95–109.

Herzberg C, Friedrich A, Averhoff B. *comB*, a novel competence gene required for natural transformation of *Acinetobacter* sp. BD413: identification, characterization, and analysis of growth-phase-dependent regulation. *Arch Microbiol* 2000;**173**:220–8.

Ho MH, Chen CH, Goodwin JS *et al.* Functional advantages of *Porphyromonas gingivalis* vesicles. *PLoS One* 2015;**10**:e0123448.

Horstman AL, Kuehn MJ. Enterotoxigenic Escherichia coli Secretes Active Heat-labile Enterotoxin via Outer Membrane Vesicles. *J Biol Chem* 2000;**275**:12489–96.

Humphrey SB, Stanton TB, Jensen NS. Mitomycin C induction of bacteriophages from *Serpulina hyodysenteriae* and *Serpulina innocens*. *FEMS Microbiol Lett* 1995;**134**:97–101.

Humphrey SB, Stanton TB, Jensen NS *et al.* Purification and characterization of VSH-1, a generalized transducing bacteriophage of *Serpulina hyodysenteriae*. *J Bacteriol* 1997;**179**:323–9.

Hynes A, Shakya M, Mercer R *et al.* Functional and evolutionary characterization of a gene transfer agent's multi-locus "genome". *Mol Biol Evol* 2016;**33**:2530–43.

Hynes AP, Mercer RG, Watton DE *et al.* DNA packaging bias and differential expression of gene transfer agent genes within a population during production and release of the *Rhodobacter capsulatus* genetransferagent, RcGTA. *Mol Microbiol* 2012;**85**:314–25.

Jiang Y, Kong Q, Roland KL *et al.* Membrane vesicles of *Clostridium perfringens* type A strains induce innate and adaptive immunity. *Int J Med Microbiol* 2014;**304**:431–43.

Kadurugamuwa JL, Beveridge TJ. Virulence factors are released from *Pseudomonas aeruginosa* in association with membrane vesicles during normal growth and exposure to gentamicin: a novel mechanism of enzyme secretion. *J Bacteriol* 1995;**177**:3998–4008.

Kadurugamuwa JL, Beveridge TJ. Bacteriolytic effect of membrane vesicles from *Pseudomonas aeruginosa* on other bacteria including pathogens: conceptually new antibiotics. *J Bacteriol* 1996;**178**:2767–74.

Kadurugamuwa JL, Beveridge TJ. Natural release of virulence factors in membrane vesicles by *Pseudomonas aeruginosa* and the effect of aminoglycoside antibiotics on their release. *J Antimicrob Chemother* 1997;**40**:615–21.

Kadurugamuwa JL, Beveridge TJ. Membrane vesicles derived from *Pseudomonas aeruginosa* and *Shigella flexneri* can be integrated into the surfaces of other Gram-negative bacteria. *Microbiology* 1999;**145**:2051–60.

Kato S, Kowashi Y, Demuth DR. Outer membrane-like vesicles secreted by *Actinobacillus actinomycetemcomitans* are enriched in leukotoxin. *Microb Pathog* 2002;**32**:1–13.

Klieve AV, Yokoyama MT, Forster RJ *et al.* Naturally occurring DNA transfer system associated with membrane vesicles in cellulolytic *Ruminococcus* spp. of ruminal origin. *Appl Environ Microbiol* 2005;**71**:4248–53.

Kolling GL, Matthews KR. Export of virulence genes and shiga toxin by membrane vesicles of *Escherichia coli* O157:H7. *Appl Environ Microbiol* 1999;**65**:1843–8.

Kuchinski KS, Brimacombe CA, Westbye AB *et al.* The SOS response master regulator LexA regulates the gene transfer agent of *Rhodobacter capsulatus* and represses transcription of the signal transduction protein CckA. *J Bacteriol* 2016;**198**:1137–48.

Kulp A, Kuehn MJ. Biological functions and biogenesis of secreted bacterial outer membrane vesicles. *Annu Rev Microbiol* 2010;**64**:163–84.

Lang AS, Beatty JT. Genetic analysis of a bacterial genetic exchange element: The gene transfer agent of *Rhodobacter capsulatus*. *Proc Natl Acad Sci* 2000;**97**:859–64.

Lang AS, Beatty JT. The gene transfer agent of *Rhodobacter capsulatus* and "constitutive transduction" in prokaryotes. *Arch Microbiol* 2001;**175**:241–9.

Lang AS, Beatty JT. Importance of widespread gene transfer agent genes in alpha-proteobacteria. *Trends Microbiol* 2007;**15**:54–62.

Lang AS, Taylor TA, Beatty JT. Evolutionary implications of phylogenetic analyses of the gene transfer agent (GTA) of *Rhodobacter capsulatus*. *J Mol Evol* 2002;**55**:534–43.

Lang AS, Westbye AB, Beatty JT. The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. *Annu Rev Virol* 2017;**4**:87–104.

Lang AS, Zhaxybayeva O, Beatty JT. Gene transfer agents: phage-like elements of genetic exchange. *Nat Rev Micro* 2012;**10**:472–82.

Liao S, Klein MI, Helm KP *et al. Streptococcus* mutans extracellular DNA Is upregulated during growth in biofilms, actively released via membrane vesicles, and influenced by components of the protein secretion machinery. *J Bacteriol* 2014;**196**:2355–66.

Link C, Eickernjäger S, Porstendörfer D *et al.* Identification and characterization of a novel competence gene, *comC*, required for DNA binding and uptake in *Acinetobacter* sp. strain BD413. *J Bacteriol* 1998;**180**:1592–5.

Loeb MR, Kilner J. Release of a special fraction of the outer membrane from both growing and phage T4-infected *Escherichia coli* B. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1978;**514**:117–27.

Marrs B. Genetic recombination in *Rhodopseudomonas capsulata*. *Proc Natl Acad Sci* 1974;**71**:971–3.

Marrs B, Wall JD, Gest H. Emergence of the biochemical genetics and molecular biology of photosynthetic bacteria. *Trends Biochem Sci* 1977;**2**:105–8.

Mashburn LM, Whiteley M. Membrane vesicles traffic signals and facilitate group activities in a prokaryote. *Nature* 2005;**437**:422–5.

Matson EG, Thompson MG, Humphrey SB *et al.* Identification of genes of VSH-1, a prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. *J Bacteriol* 2005;**187**:5885–92.

Mayer F, Gottschalk G. The Bacterial Cytoskeleton and Its PutativeRole in Membrane Vesicle Formation Observedin a Gram-Positive Bacterium Producing Starch-Degrading Enzymes. *J Mol Microbiol Biotechnol* 2003;**6**:127–32.

Mayrand D. Biological activities of outer membrane vesicles. *Can J Microbiol* 1989;**35**:607–13.

McBroom AJ, Kuehn MJ. Outer membrane vesicles. *EcoSal Plus* 2005, DOI: 10.1128/ecosal.2.2.4.

McBroom AJ, Kuehn MJ. Release of outer membrane vesicles by Gram-negative bacteria is a novel envelope stress response. *Mol Microbiol* 2007;**63**:545–58.

McDaniel LD, Young E, Delaney J *et al.* High frequency of horizontal gene transfer in the oceans. *Science* 2010;**330**:50–50.

Mercer R, Lang AS. Identification of a predicted partner-switching system that affects production of the gene transfer agent RcGTA and stationary phase viability in *Rhodobacter capsulatus*. *BMC Microbiol* 2014;**14**:71.

Mercer RG, Quinlan M, Rose AR *et al.* Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. *FEMS Microbiol Lett* 2012;**331**:53–62.

Nagao N, Yamamoto J, Komatsu H *et al.* The gene transfer agent-like particle of the marine phototrophic bacterium

Rhodovulum sulfidophilum. *Biochemistry and Biophysics Reports* 2015;**4**:369–74.

Näther DJ, Rachel R. The outer membrane of the hyperthermophilic archaeon *Ignicoccus*: dynamics, ultrastructure and composition. *Biochm Soc Trans* 2004;**32**:199–203.

Nowotny A, Behling UH, Hammond B *et al.* Release of toxic microvesicles by *Actinobacillus actinomycetemcomitans*. *Infect Immun* 1982;**31**:151–4.

Ortmann AC, Suttle CA. Determination of virus abundance by epifluorescence microscopy. *Methods Mol Biol* 2009;**501**:87–95.

Patrick S, McKenna JP, O'Hagan S *et al.* A comparison of the haemagglutinating and enzymic activities of *Bacteroides fragilis* whole cells and outer membrane vesicles. *Microb Pathog* 1996;**20**:191–202.

Paul JH. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J* 2008;**2**:579–89.

Perez-Cruz C, Carrion O, Delgado L *et al.* New Type of Outer Membrane Vesicle Produced by the Gram-Negative Bacterium *Shewanella vesiculosa* M7 T: Implications for DNA Content. *Appl Environ Microbiol* 2013;**79**:1874–81.

Perez-Cruz C, Delgado L, Lopez-Iglesias C *et al.* Outer-inner membrane vesicles naturally secreted by gram-negative pathogenic bacteria. *PLoS One* 2015;**10**:e0116896.

Prados-Rosales R, Baena A, Martinez LR *et al.* Mycobacteria release active membrane vesicles that modulate immune responses in a TLR2-dependent manner in mice. *J Clin Invest* 2011;**121**:1471–83.

Quebatte M, Christen M, Harms A *et al.* Gene transfer agent promotes evolvability within the fittest subpopulation of a bacterial pathogen. *Cell Systems* 2017;**4**:611–621.e6 e6.

Rapp BJ, Wall JD. Genetic transfer in *Desulfovibrio desulfuricans*. *Proc Natl Acad Sci* 1987;**84**:9128–30.

Renelli M, Matias V, Lo RY *et al.* DNA-containing membrane vesicles of *Pseudomonas aeruginosa* PAO1 and their genetic transformation potential. *Microbiology* 2004;**150**:2161–9.

Rivera J, Cordero RJB, Nakouzi AS *et al. Bacillus anthracis* produces membrane-derived vesicles containing biologically active toxins. *Proc Natl Acad Sci* 2010;**107**:19002–7.

Roden JA, Wells DH, Chomel BB *et al.* Hemin binding protein C is found in outer membrane vesicles and protects *Bartonella henselae* against toxic concentrations of hemin. *Infect Immun* 2012;**80**:929–42.

Rumbo C, Fernandez-Moreira E, Merino M *et al.* Horizontal transfer of the OXA-24 carbapenemase gene via outer membrane vesicles: a new mechanism of dissemination of carbapenem resistance genes in *Acinetobacter baumannii*. *Antimicrob Agents Chemother* 2011;**55**:3084–90.

Schaefer AL, Taylor TA, Beatty JT *et al.* Long-chain acyl-homoserine lactone quorum-sensing regulation of *Rhodobacter capsulatus* gene transfer agent production. *J Bacteriol* 2002;**184**:6515–21.

Schooling SR, Hubley A, Beveridge TJ. Interactions of DNA with biofilm-derived membrane vesicles. *J Bacteriol* 2009;**191**:4097–102.

Schrempf H, Koebsch I, Walter S *et al.* Extracellular *Streptomyces* vesicles: amphorae for survival and defence. *Microb Biotechnol* 2011;**4**:286–99.

Shakya M, Soucy SM, Zhaxybayeva O. Insights into origin and evolution of alpha-proteobacterial gene transfer agents. *Virus Evolution* 2017;**3**:vex036.

Soler N, Krupovic M, Marguet E *et al.* Membrane vesicles in natural environments: a major challenge in viral ecology. *ISME J* 2015;**9**:793–6.

Solioz M, Yen HC, Marrs B. Release and uptake of gene transfer agent by *Rhodopseudomonas capsulata*. *J Bacteriol* 1975;**123**:651–7.

Stanton TB. Prophage-like gene transfer agents-novel mechanisms of gene exchange for *Methanococcus, Desulfovibrio, Brachyspira*, and *Rhodobacter* species. *Anaerobe* 2007;**13**:43–49.

Stanton TB, Humphrey SB, Bayles DO *et al.* Identification of a divided genome for VSH-1, the prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. *J Bacteriol* 2009;**191**:1719–21.

Stanton TB, Humphrey SB, Sharma VK *et al.* Collateral effects of antibiotics: carbadox and metronidazole induce VSH-1 and facilitate gene transfer among *Brachyspira hyodysenteriae* strains. *Appl Environ Microbiol* 2008;**74**:2950–6.

Takeuchi N, Kaneko K, Koonin EV. Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: benefit of DNA from dead cells and population subdivision. *G3: Genes Genom Genet* 2014;**4**:325–39.

Thay B, Wai SN, Oscarsson J. *Staphylococcus aureus* alpha-toxin-dependent induction of host cell death by membrane-derived vesicles. *PLoS One* 2013;**8**:e54661.

Thompson SS, Naidu YM, Pestka JJ. Ultrastructural localization of an extracellular protease in *Pseudomonas fragi* by using the peroxidase-antiperoxidase reaction. *Appl Environ Microbiol* 1985;**50**:1038–42.

Tomasch J, Wang H, Hall ATK *et al.* Packaging of *Dinoroseobacter shibae* DNA into gene transfer agent particles is not random. *Genome Biology and Evolution* 2018;**10**:359–69.

Toussaint A, Rice PA. Transposable phages, DNA reorganization and transfer. *Curr Opin Microbiol* 2017;**38**:88–94.

Tsatsaronis JA, Franch-Arroyo S, Resch U *et al.* Extracellular vesicle RNA: a universal mediator of microbial communication? *Trends Microbiol* 2018;**26**:401–10.

Turnbull L, Toyofuku M, Hynen AL *et al.* Explosive cell lysis as a mechanism for the biogenesis of bacterial membrane vesicles and biofilms. *Nat Comms* 2016;**7**:11220.

Vasilyeva NV, Tsfasman IM, Suzina NE *et al.* Secretion of bacteriolytic endopeptidase L5 of *Lysobacter* sp. XL1 into the medium by means of outer membrane vesicles. *FEBS J* 2008;**275**:3827–35.

Wai SN, Lindmark B, Söderblom T *et al.* Vesicle-mediated export and assembly of pore-forming oligomers of the enterobacterial ClyA cytotoxin. *Cell* 2003;**115**:25–35.

Wall JD, Weaver PF, Gest H. Gene transfer agents, bacteriophages, and bacteriocins of *Rhodopseudomonas capsulata*. *Arch Microbiol* 1975;**105**:217–24.

Wensink J, Witholt B. Outer-membrane vesicles released by normally growing *Escherichia coli* contain very little lipoprotein. *Eur J Biochem* 1981;**116**:331–5.

Westbye AB, Beatty JT, Lang AS. Guaranteeing a captive audience: coordinated regulation of gene transfer agent (GTA) production and recipient capability by cellular regulators. *Curr Opin Microbiol* 2017;**38**:122–9.

Westbye AB, Kater L, Wiesmann C *et al.* The Protease ClpXP and the PAS Domain Protein DivL Regulate CtrA and Gene Transfer Agent Production in *Rhodobacter capsulatus*. *Appl Environ Microbiol* 2018;**84**:e00275–18.

Westbye AB, Kuchinski K, Yip CK *et al.* The gene transfer agent RcGTA contains head spikes needed for binding to the *Rhodobacter capsulatus* polysaccharide cell capsule. *J Mol Biol* 2016;**428**:477–91.

Westbye AB, Leung MM, Florizone SM *et al.* Phosphate concentration and the putative sensor kinase protein CckA modulate cell lysis and release of the *Rhodobacter capsulatus* gene transfer agent. *J Bacteriol* 2013;**195**:5025–40.

Yaron S, Kolling GL, Simon L *et al.* Vesicle-mediated transfer of virulence genes from *Escherichia coli* O157:H7 to other enteric bacteria. *Appl Environ Microbiol* 2000;**66**:4414–20.

Yen HC, Hu NT, Marrs BL. Characterization of the gene transfer agent made by an overproducer mutant of *Rhodopseudomonas capsulata*. *J Mol Biol* 1979;**131**:157–68.