6-28-2021

Elucidating the Evolution and Function of sRNAs that Facilitate Bacterial Stress Tolerance

Madeline Claire Krieger
Portland State University

Elucidating the Evolution and Function of sRNAs that Facilitate Bacterial Stress Tolerance

by

Madeline Claire Krieger

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Biology

Dissertation Committee:
Justin Courcelle, Chair
Rahul Raghavan, Co-Chair
Justin Merritt
Brooke Napier
Suzanne Estes

Portland State University
2021

ABSTRACT

Non-coding small RNAs (sRNAs) are ubiquitous post-transcriptional regulators found in both gram-negative and gram-positive bacteria. Despite their integral role in many regulatory pathways, including those that facilitate stress-tolerance, little is known about the evolutionary forces that drive the emergence of novel sRNAs, how prevalent sRNAs are within bacterial species, or the functions of a majority of these transcripts. In this work, I first describe the evolution of OxyS, a well-characterized sRNA in gram-negative Enterobacteriaceae that is involved in coordinating the bacterial response to oxidative stress. Next, I identify novel sRNAs differentially produced by the gram-positive oral pathogen *Streptococcus mutans* in response to multiple stress conditions. Finally, I investigate the function of one novel sRNA, SmsR4, in fine-tuning sugar-alcohol metabolism in *S. mutans*. These findings reveal protein coding genes to be a new reservoir for sRNA emergence and define a novel evolutionary path through which new sRNAs become incorporated into regulatory networks. Additionally, I show evidence for abundant regulation by sRNAs in *S. mutans* during a variety of stress responses and outline how SmsR4 participates in the regulation of sugar metabolism. Taken together, this work presents a new paradigm for sRNA emergence from protein coding genes and demonstrates the prevalence and utility of the understudied phenomenon of post-transcriptional regulation in oral bacteria.

ACKNOWLEDGEMENTS

Without the guidance of my mentor Dr. Rahul Raghavan, I would have never had the courage to complete this journey. I am departing his lab not only a more confident and careful researcher but also a more resilient and determined human and am eternally grateful for him seeing my potential even when I could not. Additionally, my entire doctoral committee has been incredible in supporting me through my graduate school career and beyond. Through every stage of this program they have challenged me in ways that built both my knowledge and confidence. My labmates and friends Amanda Brenner, Madhur Sachan, and Abe Moses have given me so much support and guidance, and I feel so lucky to have spent the past four years with them. I also received continual encouragement and instruction from Auguste Dutcher, Jess Millar, Dr. Merritt and his lab group at OHSU, and the entire PSU Biology community, with special thanks to all the members of my graduate school cohort.

My amazing and supportive husband Jess has been with me every step of the way through my doctoral program. His love and compassion were essential to my success during my graduate studies, and along with our two canine companions, Avery and Indie, my family provided me with so much joy and comfort during stressful times. My parents have supported me on whatever path I have walked in life, and I have so much gratitude for their unconditional love during the course of my studies and beyond.

I am blessed to be surrounded by a strong community of friends who played an essential role in my success over the last four years. Through endless laughter, horseback adventures, celebrations and frustrations, I could not have completed this journey without the love and support of each and every one of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Introduction to small RNAs and *Streptococcus mutans*

I. SMALL NON-CODING RNAs

A. sRNAs in bacteria

Small RNAs (sRNAs) are ubiquitous regulatory transcripts found throughout the prokaryotic domain, but their importance in the regulation of many metabolic, virulence, and stress response pathways has only recently begun to be recognized. sRNAs are generally non-coding transcripts and act by binding to mRNA targets and regulating their translation (1). Regulation by sRNAs occurs quickly as they are significantly less costly for the cell to produce than protein regulators and can fine-tune regulatory networks by acting downstream of previous genetic responses (2). sRNAs bind to mRNA targets via a stretch of 7-12 complementary nucleotides (nt) known as the "seed region" (3). Typically, sRNAs act through inhibitory binding around the ribosomal binding site (RBS) of mRNAs to block protein synthesis (**Figure 1**), but sRNA binding can also stabilize mRNA transcripts to enhance ribosomal binding (**Figure 2**) (4).

Defining the functions of sRNAs presents a much greater challenge than elucidating the roles of protein coding genes. sRNAs regularly act in refining complex regulatory pathways and interact with multiple mRNA targets, sometimes in concert with RNA-binding proteins. The protein or mRNA targets of sRNAs are often difficult to identify, since sRNAs only need a small functional region of complementary nucleotide

sequence to pair with binding partners (5). Additionally, the abundance of these mRNA targets is generally only affected through subtle fluctuations in transcript availability (6).

sRNAs are able to act downstream of transcription and mark specific mRNAs for RNase degradation. The co-degradation of sRNAs with their mRNA targets offers the ability for the regulator to also be destroyed with the target, a phenomenon that is not common with protein regulators (7). sRNAs can either be classified as *cis*-acting, those that are transcribed from the same genetic loci as the gene they regulate, or *trans*-acting, which are located at a separate location on the genome from their targets. *Cis*-acting sRNAs are commonly located in the 5′ or 3′ untranslated region (UTR) of mRNA transcripts and include riboswitches and metabolite sensing RNAs, which respond to temperature changes or the presence of certain metabolites (8, 9). The genomic location of *trans*-encoded sRNAs can be thousands of nucleotides away from their mRNA targets, and often these sRNAs share only a small region of sequence complementarily to the mRNAs that they regulate (10).

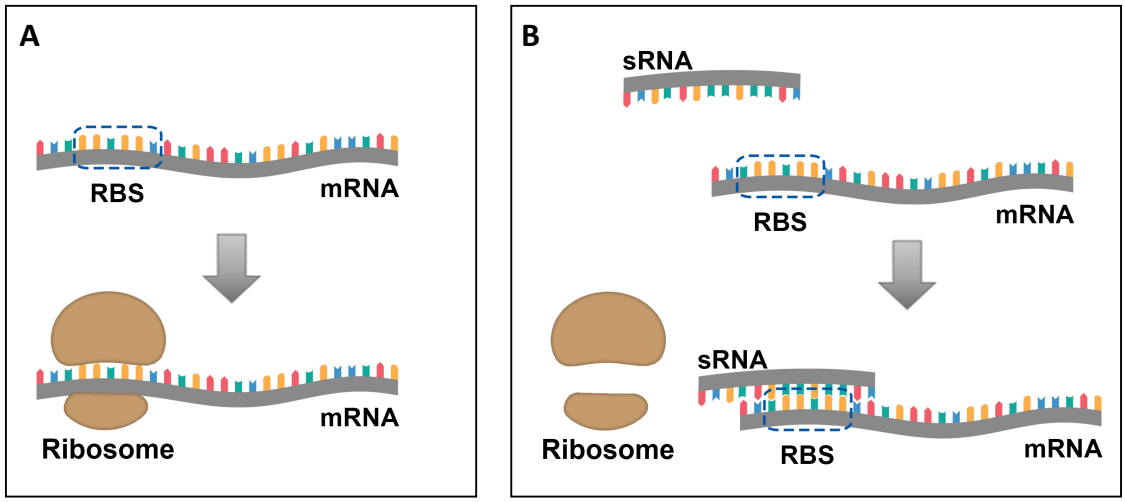

Figure 1. Negative regulation of mRNA translation through sRNA binding around the ribosomal binding site (RBS). A) Without a complementary sRNA present, translation of an mRNA is unimpeded. **B)** A sRNA binds to the RBS of its target mRNA, blocking the ribosome from binding and preventing translation.

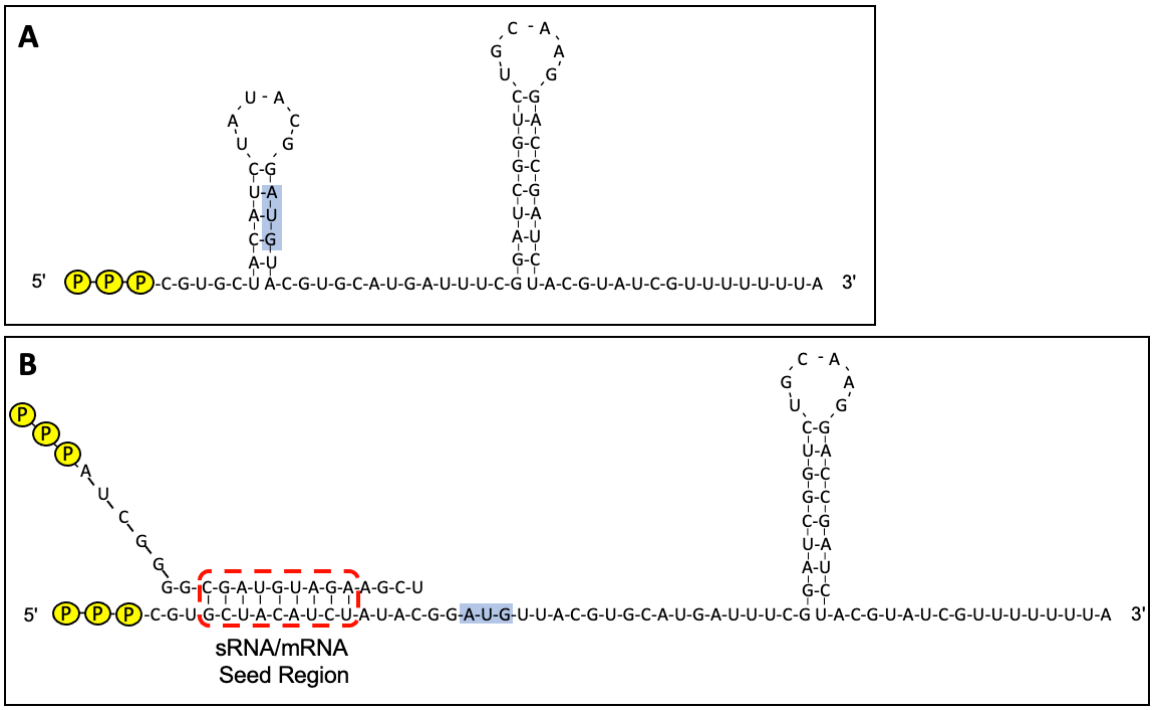

Figure 2. Positive regulation of mRNA translation by sRNA binding to the 5′ untranslated region (UTR). A) Some mRNAs fold into stem-loop structures within their 5′ UTR that obscure the start codon (blue), blocking translation. **B)** The binding of a sRNA to a complementary region on the 5′ UTR just upstream of the start codon (red dotted line) stabilizes the transcript, freeing the start codon for translation.

B. sRNA evolution

sRNAs typically arise in bacterial genomes through one of three mechanisms: *de novo* promoter emergence, horizontal gene transfer, or gene duplication events which allow the redundant gene copy to mutate without deleterious effects to the cell (11). In comparison to protein coding genes, sRNAs evolve rapidly due to several inherent factors (12). sRNAs lack rigid gene structure requirements necessary for translation, and thus landmarks like ribosomal binding sites and standard codon organization are absent from non-coding transcripts (11). Furthermore, their function is largely dependent on secondary structure formation and the base-pairing of a few specific nucleotides that are complementary to the mRNA target, allowing for the possibility of significant sequence divergence between similar transcripts in closely related species that still maintain functional homology (13). The small sRNA-mRNA seed region allows for rapid evolution of sRNAs to find new targets or develop greater or weaker affinity for existing ones (11). This property makes sRNAs uniquely malleable for changing functionality in a short period of evolutionary time, but also creates challenges in defining the evolutionary trajectory of sRNAs through related species. Since homologous sRNAs in different species can diverge significantly in their primary sequences, covariance modeling (cm) is an essential tool for tracking the evolution of sRNAs in related genomes (14, 15). Cm-based scans incorporate both primary and secondary structure to detect similarities in the folding configuration of sRNA homologs, which are essential for stabilizing the functional seed region.

C. RNases

Many sRNAs are effective at physically blocking translation of mRNA targets through base-pairing around the RBS, but the RNase E degradome complex plays an essential role in further degrading mRNA-sRNA duplexes. Although there are multiple RNases that act to degrade mRNAs, is estimated that the action of RNase E accounts for greater than 50% of all mRNA cleavage during exponential growth in *Escherichia coli* (16, 17). RNase E has a substrate preference for single-stranded RNA with 5′ monophosphates and cuts at the consensus sequence A/GAUUA/U (18, 19). This enzyme can associate with the membrane-anchored degradosome complex and work in concert with other RNases and proteins to degrade RNA transcripts; however, RNase E function is not limited solely to the degradosome, and it may act independently on some RNA substrates (20). In addition to assisting with target mRNA degradation, RNases also help produce mature functional sRNAs. Many sRNAs are processed from larger transcripts by 3′ and 5′ cleavage (**Figure 3**) (21). An example of the RNase E mediated production of a mature sRNA is ArcZ, which is produced from the 3′ cleavage of a larger RNA (22).

Although RNase E is essential for sRNA processing in many gram-negative species, gram-positive Firmicutes, including *Streptococcus* species, do not contain RNase E. Instead, two varieties of RNase J (RNase J1 and J2) are present in the gram-positive *Streptococcus mutans*, with RNase J2 likely to be the primary post-transcriptional regulator with a role in regulating RNase J1 (23). Additionally, they encode RNase Y, RNaseIII, PNPase, and a DEAD-box RNA-helicase, which are also found in gram-negative species (2) . This collection of RNase enzymes provides ample potential for RNase mediated sRNA processing in streptococci.

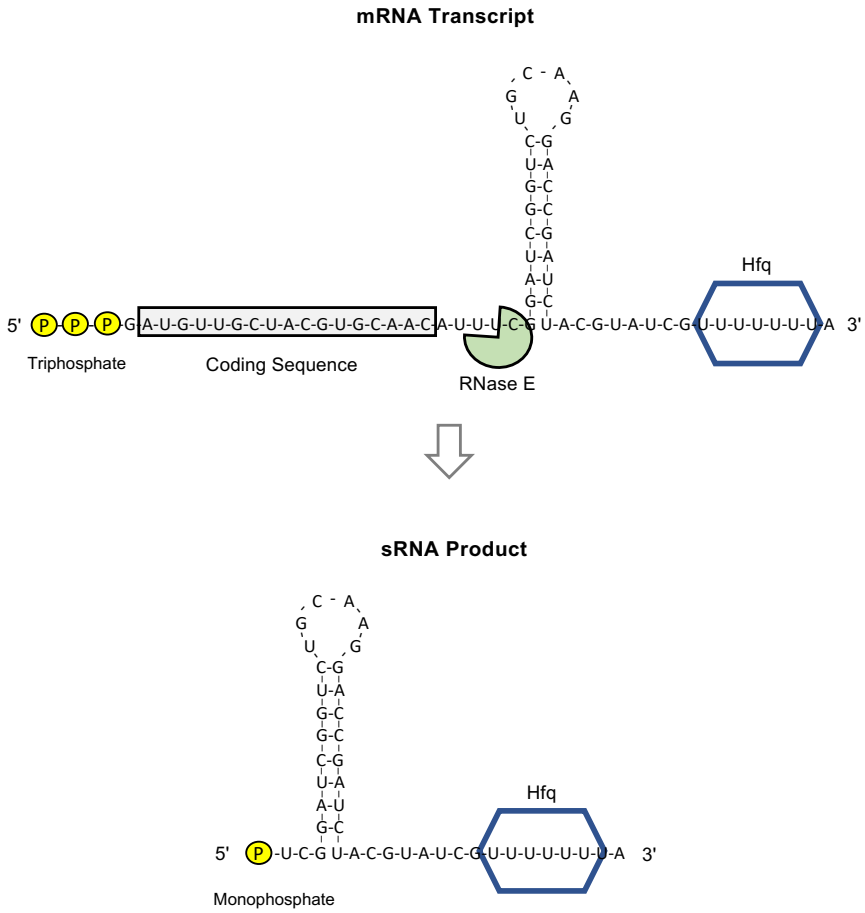

Figure 3. sRNA production through RNase E mediated cleavage. A sRNA is produced from the 3′ UTR of an mRNA through the action of RNase E. The coding sequence is unaffected, while the mature sRNA product, containing a 5′ monophosphate, is released. The RNA-binding protein Hfq often assists in this process by binding to the 3′ end of the transcript, stabilizing the RNA and/or recruiting RNase E.

D. Hfq

The protein Hfq was originally annotated as a host factor necessary for the replication of the bacteriophage Q-beta (24), but further research has illuminated its essential role in sRNA function. In *E. coli* Hfq is an 11.2 kDa homohexameric complex with a proximal and distal face capable of binding U-rich and A-rich regions of RNA simultaneously (25). The binding of U-rich sequences, a Sm1 sequence motif, and multi-subunit structure make Hfq a typical Sm-like protein (26). Although some bacteria lack Hfq entirely, removal of the RNA-binding protein in species such as *E. coli* causes pleiotropic effects throughout the cell, including decreased cell growth and an increase in UV sensitivity (27, 28). A sequence of at least six consecutive U's on the 3′ end of Hfq binding sRNAs is essential for successful Hfq/sRNA pairing (29).

Hfq facilitates mRNA and sRNA base-pairing through stabilizing single stranded RNA molecules and bringing them naturally in close proximity to one another (30). Additionally, if the mRNA and sRNA have the same Hfq binding site, competition for Hfq binding may bring the two RNA transcripts into close proximity. There are often many more sRNA transcripts than Hfq complexes available for binding; due to this discrepancy in availability, RNAs often cycle on and off Hfq (31).

Hfq often works in concert with RNase E either to degrade bound mRNA-sRNA pairs or to generate mature sRNAs from larger transcripts. Research suggests that some sort of RNase E processing is required to produce all functional sRNAs identified in *E. coli* to date (32), and Hfq is often necessary for this essential process of transcript modification (21, 33). In the case of a sRNA that is cleaved from the 3′ end of an mRNA transcript, Hfq can bind to the stretch of U's at the end of the 3′ end and stabilize the

transcript, allowing for cleavage of the mRNA by RNase E (**Figure 3**). Hfq is then already associated with the mature sRNA and is able to further assist in binding mRNA targets. Alternatively, Hfq binding can protect mRNA targets from RNase E degradation. Hfq and RNase E can occupy the same binding site on an RNA transcript, so Hfq may effectively block RNase E binding and degradation (34).

E. OxyS

Oxidative stress is one of the most destructive threats to a bacterial cell. The damage caused by oxidative stress is primarily induced by the formation of reactive oxygen species (ROS) such as superoxide (O_2^-) and hydroxide radicals (OH^-) (35). H_2O_2 exposure can induce ROS formation, and the interaction of H_2O_2 and unincorporated intracellular iron can cause DNA damage and inactivate essential enzymes involved in key metabolic processes (36). Due to its potentially catastrophic impact on survival, bacteria have developed well-organized responses to combat oxidative stress. In *E. coli* and other gram-negative bacteria, the transcription factor OxyR servers as the principal regulator for coordinating the response to oxidative stress. Although its function deviates slightly between species, the constitutively produced OxyR is activated by oxidative stress and turns on the transcription of several dozen genes that are responsible for mitigating the damage caused by ROS (36).

Specifically in *E. coli*, OxyR controls a divergently located sRNA, OxyS (37). OxyS was one of the first sRNAs described in the species and is an essential component of the oxidative stress response. The deletion of OxyS results in higher intracellular levels of H_2O_2 and superoxide, indicating its central role in alleviating damage caused by ROS

(38). OxyS acts to enhance the activity of OxyR and fine-tune *E. coli*'s response to oxidative stress through several mechanisms, including modulating the ratio of the transcriptional activators RpoS and RpoD. RpoS activates genes that mitigate oxidative stress during normal stationary phase growth, but RpoD is necessary to activate additional rescue mechanisms during times of extreme oxidative stress. OxyR expression is dependent on RpoD, and OxyS sequesters the expression of the RpoD competitor RpoS, allowing for increased OxyR transcription and thus greater intensity of the oxidative stress response. In addition, repression of RpoS may help to reduce the activation of genes that may be redundant in the oxidative stress response (39). Unlike the traditional paradigm of sRNA regulation, OxyS interacts with RpoS not through base pairing, but indirectly through competition with other sRNA activators of RpoS or through competition for Hfq (40).

The metabolism of formate from formic acid can contribute to oxidative stress because metal cofactors required for this process increase cellular damage during these stress conditions. The transcriptional activator encoded by *fhlA* is an activator of genes involved in formate metabolism. OxyS reduces levels of the *fhlA* mRNA through direct base pairing around the ribosomal binding site as well as to a second site within the coding sequence to form a kissing-loop structure that inhibits translation (41).

The final indirect mechanism of OxyS involvement in oxidative stress is through negative regulation of NusG mRNA. NusG is a regulator of RNA polymerase, and functions with Rho in transcriptional termination. When OxyS is active, NusG levels are low. NusG acts to silence the *rac* locus, which contains the *kilR* gene. The KilR protein interferes with the action of FtsZ during cell division. Due to reduced levels of NusG

caused by OxyS inhibition, KilR production is increased, which causes arrest of the cell cycle to allow for repair of DNA damage caused by oxidative stress (42).

F. SgrS

The 227-nt sRNA SgrS (sugar transport-related sRNA) has been well characterized in *E. coli* and *Salmonella enterica* for its role in alleviating sugar-phosphate stress. The accumulation of phosphorylated sugars through the uptake of a non-metabolizable glucose-6-phosphate (G6P) analog or a block in glycolysis causes growth defects or cell death in bacteria (43–45). The exact molecular mechanism underlying the toxicity of accumulated phospho-sugars is not well understood, but research suggests that the resulting buildup of toxic byproducts such as methyglyoxal or the depletion of glycolytic intermediates is responsible for sugar-phosphate stress (44–46).

SgrS plays an important role in modulating the accumulation of G6P in the cell and it acts through multiple mechanisms to alleviate sugar-phosphate stress (**Figure 4**) and is required for survival during growth with a non-metabolizable G6P analog (47). Expression of SgrS is regulated by the transcriptional activator SgrR, which responds to rising G6P levels (47). SgrS acts like a traditional base-pairing sRNA and binds around the RBS of *ptsG* mRNA, which encodes a glucose transporter, blocking translation and recruiting RNase E to promote transcript degradation (48). It also regulates glucose transport through inhibitory binding to *manXYZ* operon transcript. Additionally, the 3′ end of the SgrS transcript also encodes a 43 amino acid protein, SgrT, which physically blocks glucose transport through the PTS system (48). Thus, SgrS is a dual-functioning sRNA that acts both as a base-pairing sRNA and a functional mRNA . The effects of

SgrS expression are a decrease in glucose transport, a decrease in intracellular G6P levels, and the switch to the use of alternative carbon sources. Collectively, these efforts help alleviate sugar-phosphate stress caused by excess G6P.

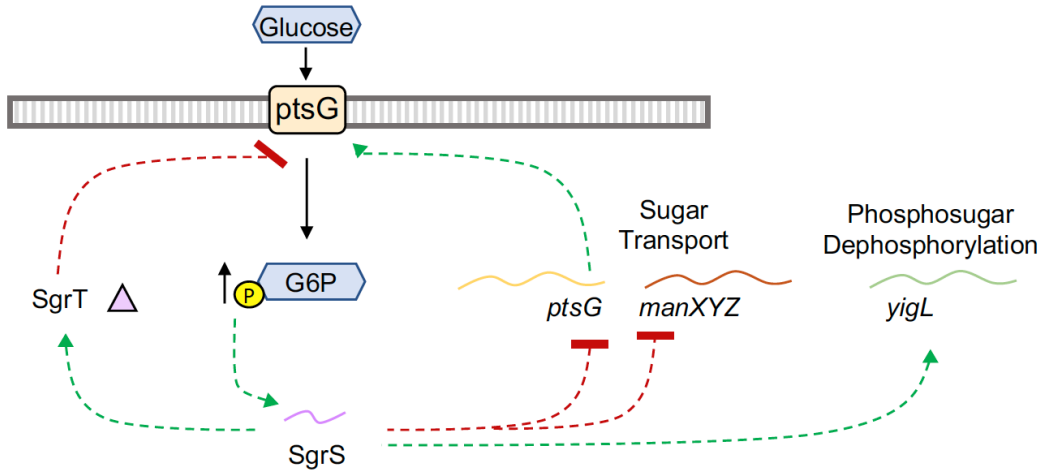

Figure 4. SgrS responds to sugar-phosphate stress through multiple mechanisms. Green arrows indicate activation, while red boxes show repression. Transport of glucose results in elevated glucose-6-phosphate (G6P) levels. Excess G6P causes sugar-phosphate stress, which triggers the synthesis of the sRNA SgrS, which acts to relieve this stress through multiple mechanisms. The sRNA blocks the translation of *ptsG* and *manXYZ* mRNA, which encode components of sugar-transport systems, and enhances the synthesis of yigL, which acts to dephosphorylate phosphosugars. Additionally, a small protein, SgrT, is synthesized from the SgrS sRNA, which physically blocks glucose transport through PtsG.

G. 6S RNA

6S RNA is found in both gram positive and negative bacterial species and exhibits a highly conserved secondary structure (49). The 6S RNA transcript folds back on itself to create a hairpin complex that contains a single-stranded bubble region in the middle which mimics the structure of an open promoter DNA complex (50, 51). This formation acts as a decoy for the sigma-70 RNA polymerase complex (RNAP), which binds to 6S RNA and alleviates competition with alternative sigma factors. RNase processing of both the 5′ and 3′ ends of 6S RNA has been well documented in *E. coli* and Northern blot analysis of 6S RNA transcripts reveals multiple forms (52).

In *E. coli*, 6S RNA is expressed later in growth and assists in transition from exponential to stationary phase (51). *Bacillus subtilis*, like some other Firmicutes, contains two copies of 6S RNA, one of which is highly expressed at the start of growth (53), while the second echoes the expression profile seen in *E. coli* (50). In addition to binding the RNAP, 6S binds with mRNA targets. In *E. coli* 6S RNA binds to the *pspF* mRNA, helping the cell respond to elevated pH levels during stationary growth (54). Surprisingly, an additional function of *E. coli* 6S RNA is to serve as a template for the synthesis of small product RNAs (pRNAs) (55). pRNAs are small, usually <20 nts in length, and have been implicated in reversal of 6S RNA transcription inhibition by disrupting the RNAP/6S RNA complex (56).

In Gammaproteobacteria, the genome location of 6S RNA is highly conserved. In these species, 6S RNA is located next to the gene *ygfA*, which encodes an enzyme that converts the storage form of folate, 5-formyltetrahydrofolate, into an active form (50). The conservation of 6S RNA with *ygfA* may give some clues to its expression profile – in

eukaryotes, orthologs of *ygfA* help with adaptation to nutrient limitation, similar to the conditions that a bacterial cell would be experiencing as it enters stationary phase. No similar locus conservation has been documented in species outside of Gammaproteobacteria.

H. sRNA function in gram-positive bacteria

Studies of sRNAs in gram-positive bacterial species have lagged behind those in gram-negative; however, the essential utility of sRNAs in gram-positive species is apparent from the small body of research that exists. Multiple functional sRNAs have been described in *B. subtilis*, including the RyhB homolog FsrA that responds to iron limitation (57) and the dual-function SR1, a transcript which base-pairs with an arginine-catabolism regulator but also encodes a small peptide (58). RNAIII was the first sRNA to be characterized in *Staphylococcus aureus* (59). This unusually long (514 nt) sRNA is under transcriptional control of a quorum-sensing transcriptional regulator and has both activating and inhibitory functions on a variety of mRNA targets involved in pathogenesis. Inhibition of targets occurs through binding around the RBS to block translation or recruitment of RNase III to promote degradation of bound mRNAs. Interestingly, this sRNA is also a dual-action transcript and contains an open reading frame encoding a delta-hemolysin.

In *Streptococcus* species, few sRNAs have been functionally defined. A whole-genome scan of sRNAs in *Streptococcus pneumoniae* revealed 66 new sRNA candidates, 30 of which were validated by Northern blot and many of which showed evidence of RNase processing from larger transcripts (60). Another study identified 89 putative

sRNAs in *S. pneumoniae* using RNA-seq, and analysis of Tn-seq sRNA mutants predicted that 28 sRNAs participate significantly during lung infection and 18 may play a role in sepsis pathology (61). This work strongly implicated sRNAs in the pathogenesis of *S. pneumoniae* during several routes of infection. In this species, promoters in the two-component regulatory system CiaRH, which controls genes involved in a wide variety of stress response and metabolic networks, were found to drive the production of csRNAs (cia-dependent small RNA), which ranged in size from 87 to 151 nt (62). These csRNAs were shown to impact autolysis in stationary phase, lung infection, and competence development (61, 63). csRNAs homologs have been identified in other *Streptococcus* species as well (64, 65). Work predicting candidate sRNAs in Group A Streptococcus (*Streptococcus pyogenes*) has revealed over 100 putative transcripts that are differentially expressed in various growth conditions (66). Several of these sRNAs have been studied in detail and sRNAs have been shown to play an important role in virulence in this species, including pilus formation and regulation of enzymes that modulate host blood clotting (66). As with other streptococci, *S. mutans* likely contains multiple important sRNAs that play key roles in regulatory networks, but to date none have been functionally defined in this important gram-positive oral pathogen.

II. *Streptococcus mutans*

A. The oral microbiome

The human oral cavity is a unique connector of the exterior world to the interior realm. It contains a multitude of distinct surfaces for bacteria to colonize, including the tongue, cheeks, tonsils, hard and soft palates, and teeth. Behind only the expansive gut, the oral cavity contains the largest and most diverse collection of microbial species in the body (67). An adult mouth is comprised of approximately 20% teeth, 50% keratinized and 30% non-keratinized epithelium, and microbial colonizers inhabit all these surfaces (68), making the composition of such a varied landscape difficult to define. The Human Oral Microbiome Database has been established to characterize the microbial residents of the oral cavity (69), and 16S rDNA sequencing has revealed the presence of over 700 microbial species within the oral cavity, a mere 54% of which have been named (70, 71).

Although there is great variation between individuals in the species compositions of their oral microbiomes, the microbial community develops sequentially throughout early life. The womb is considered to be sterile, but it is estimated that oral cavity associated microorganisms are present in the amniotic fluid of up to 70% of women (67). As with any environment where microbes reside in constant close contact with host tissue, immune tolerance to microbial presence is essential in host maintenance of inflammatory responses. The immune response to these resident microorganisms is dominated by anti-inflammatory T-regulatory cells and influenced greatly by anti-inflammatory cytokines. The development of fetal tolerance to the maternal oral microbiome may be essential in establishing this commensal relationship early in development (72). During birth, the initial colonization of the oral microbiome occurs as

the infant comes in contact with fluids in the birth canal and eventually the breastmilk of their mother. Unsurprisingly, different methods of delivery, such as Cesarian section, can alter the composition of the oral microbiome in infancy (73, 74). In infants born through Cesarian section, *S. mutans* residency is established nearly a year earlier than in those born through vaginal birth (73). After birth, *Streptococcus* and *Veinillonella* species predominate the oral cavity until around 6 weeks of age, but the emergence of teeth allow for a new variety of bacteria within the mouth environment (72). Early colonizers of the clean tooth surface are primarily health associated *Streptococcus* species (75). In this state, gram-positive facultative anaerobic bacteria predominate and a non-inflammatory equilibrium exists between the host's immune response and bacterial residents of the oral cavity (76). However, the bacterial community composition shifts with the development of a biofilm matrix.

B. Biofilm formation by *S. mutans*

With a temperature maintained around 37°C and a stable neutral salivary pH of 6.5-7, the oral cavity can be the ideal place for bacteria to thrive in a well hydrated, nutrient rich environment (67). However, numerous challenges such as fluctuations in temperature, oxygenation, acidity, oxidative stress, and nutrient composition are presented to the microbial communities of the mouth (71). Thus, the oral cavity can be an extremely challenging environmental to inhabit, and bacterial residents must develop methods of shielding themselves from these outside assaults. One primary strategy for protection deployed by oral bacteria is the development of biofilms on the tooth surface.

Biofilms are in no way unique to the oral cavity and form in almost all locations associated with an abundance of surface moisture, including natural environments like rocks in a stream, within the gut, and on foreign implanted medical devices (77, 78). Within the oral cavity, biofilm proliferation occurs when the composition of oral species shifts towards a state of imbalance, or dysbiosis. A diet lacking fermentable carbohydrates favors colonization by commensal organisms, which can easily attach to the smooth surface of a non-biofilm coated tooth, with no deleterious effects on the host. However, with the introduction of fermentable sugars such as sucrose into the diet, the environment of the oral cavity quickly shifts to favor biofilm forming species that can tolerate acidic conditions within the biofilm microenvironment (79).

S. mutans is largely responsible for initiating biofilm formation in the oral cavity and has two methods of biofilm production. A sucrose-independent mechanism that is not associated with its virulence utilizes Antigen I/II (also known as SpaP or Pac1), an adhesin that attaches to glycoprotein-340 and can bind other bacterial species (80). The primary methods of biofilm formation utilize the production of an extrapolymeric substance (EPS), creating a biofilm framework for aciduric species to proliferate and thrive (68). This sucrose-dependent method of biofilm formation relies on the activity of cell-surface glucan binding proteins (Gbps) that attach to glucans synthesized from sucrose by glucosyltransferase (Gtfs) enzymes (81). Gtfs bind to tooth surfaces, building glucans on top of the enamel, or onto other microbial species, transferring the glucan-building properties to other species that lack the genes for such enzymes (81). The EPS layer created by *S. mutans* allows certain bacteria to adhere to surfaces and provides a safety net of protection for the inhabitants, which are primarily acid producers. The acidic

environment favors species that can survive these low pH conditions, self-selecting for communities that contribute increasingly to the acidity of the biofilm.

C. Dental caries pathology

Dental caries is a disease caused almost solely from the byproducts of bacterial metabolism and stems from dysbiosis within the oral cavity (82). Due to the lack of specific virulence factors stimulating disease progression, caries does not have the standard pathology of many diseases. Although caries is polymicrobial in nature, the formation of the disease has been traditionally linked to the presence of *S. mutans*. Over 98% of adults carry *S. mutans*, but in non-cariogenic adults *S. mutans* make up less that 0.01% of plaque bacteria; however, a switch to sugar rich diets may drive *S. mutans* to make up over half of the bacterial species (75). The carcinogenicity of *S. mutans* is driven by three key traits: its ability to survive at a low pH, form biofilms, and produce acidic byproducts during metabolism (82).

In a caries-free oral cavity, *Streptococcus sanguinis, Streptococcus mitis, Streptococcus gordonii*, and *Streptococcus oralis* are the pioneering early colonizers of the clean tooth surface (75). These species promote healthy bacterial community formation through production of alkalis, the bacterial toxins bacterocins, and hydrogen peroxide, all which act against aciduric species. However, when the diet of the host switches to one abundant in fermentable carbohydrates, especially sucrose, caries associated species are able to form a glucan matrix and begin to build a biofilm. The matrix traps nutrients and allows aciduric species, such as *S. mutans*, to thrive. As the population of *S. mutans* expands they produce more and more acidic byproducts, such as

lactic acid, further driving down the pH of their environment and killing health-associated species who are not well adapted to acidic conditions. A positive feedback loop is created as aciduric bacteria such as *S. mutans, Veillonella,* and *Lactobacillus* proliferate (83). Eventually, the acid builds up to a level that causes tooth damage, and caries forms when the rate of demineralization outpaces that of remineralization.

D. *S. mutans* sugar transport and metabolism

Sugar phosphotransferase (PTS) systems in bacteria contain two major components: the sugar-specific membrane transport complexes EII, and the general EI and HPr cytoplasmic machinery (84). The PTS system is a chain of phosphorylation fed by pyruvate, the end product of glycolysis. In all PTS systems, phosphoenol pyruvate (PEP) is generated by the addition of a phosphate group from ATP to pyruvate. PEP passes its phosphate to the EI protein, which phosphorylates HPr. HPr passes its phosphate to the sugar-specific EIIA component, and finally EIIA can phosphorylate the membrane bound EIIB/EIIC complex, which is able to transport the sugar across the plasma membrane while simultaneously attaching a phosphate. The phosphorylated sugar then enters the glycolytic cycle, usually as fructose-6-phosphate (F6P). *S. mutans* UA159 is well equipped to metabolize a variety of different carbohydrates and contains 14 different PTS systems (85). Somewhat counterintuitively, expression of PTS components is negatively correlated with carbohydrate availability, and glucose-limited cells have been shown to have higher rates of PTS activity than those grown with excess glucose (86).

Although sugars are primarily moved across the cell membrane through PTS systems, less specific transport of carbohydrates can occur through other methods (87).

ATP-dependent transport of sugars occurs through ATP-binding cassettes (ABC) (88). ABCs are associated with bacterial virulence (89) and have been connected with bacterial toxin, or bacteriocin, production in *S. mutans* (90).

E. Sorbitol utilization by *S. mutans*

S. mutans can metabolize the sugar-alcohol sorbitol (glucitol), an isomer of mannitol that is found natural occurring in many fruits. Transport of sorbitol across the membrane of *S. mutans* is controlled by an 8.5-kb operon of six sorbitol transport associated genes, designated SMU_308 – SMU_313 (**Chapter 3, Figure 8B**). These genes are sorbitol-6-phosphate dehydrogenase (SMU_308), operon regulator (SMU_309), operon activator (SMU_310), two annotated EIIC PTS components (SMU_311 and SMU_312) – one of which is likely an EIIB protein – and component EIIA (SMU_313) (91). Sorbitol transport occurs through this canonical PTS pathway and the sorbitol operon is repressible in the presence of glucose (**Figure 5**) (91). It has been documented that sorbitol-6-phosphate 2-dehydrogenase, encoded by SMU_308, is essential for conversion of sorbitol-6-phosphate (Sorb6P) to F6P, which can enter the glycolytic cycle; deletion of the SMU_308 gene resulted in a failure to metabolize sorbitol (92).

Although lactic acid is the primary metabolic byproduct when grown in anerobic conditions with an excess of fermentable carbohydrates, ethanol – a volatile end product – is produced when sorbitol is fermented by *S. mutans* (93). Because sugar alcohols contain two additional hydrogen atoms than hexoses, the metabolism of sugar alcohols results in a higher ratio of intracellular NADH to NAD+. This problem becomes compounded under anaerobic conditions when NADH oxidase cannot be utilized to

recycle NAD+. *S. mutans* has developed several strategies to restore NAD+ while fermenting sugar alcohols anaerobically, including the use of pyruvate-formate-lyase (pfl). The gene for pfl has been shown to be essential for the anaerobic fermentation of sorbitol, and a *pfl* knockout strain of *S. mutans* LT11 failed to ferment sorbitol under anaerobic conditions, although it was able to utilize the sugar alcohol aerobically (93).

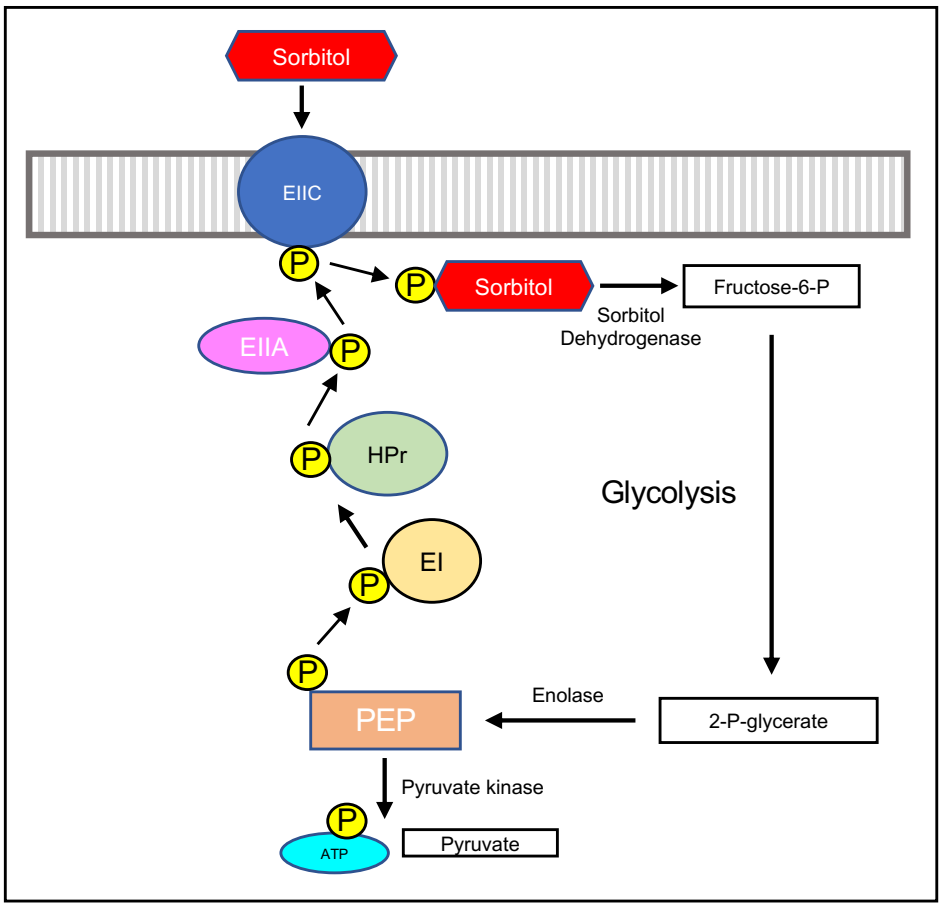

Figure 5. Sorbitol transport and metabolism by *S. mutans*. Sorbitol is transported into the cell via the EIIC component of the phosphotransferase (PTS) system and phosphorylated by EIIA. Sorbitol-6-phosphate can then be converted into fructose-6-phosphate by sorbitol dehydrogenase and enter glycolysis, which produces phosphoenol pyruvate (PEP) along with other byproducts. PEP can either be converted to pyruvate by pyruvate kinase or can act as the initial phosphate donor in the PTS phosphorylation cascade. After receiving a phosphate from PEP, EI will phosphorylate the histidine-containing phosphocarrier protein (HPr), which in turn donates the phosphate to EIIA.

F. Sugar-phosphate stress induced by xylitol

Xylitol is a five-carbon sugar that is found naturally in fruits and vegetables or manufactured artificially from plant materials that are rich in xylan. The oral hygiene benefits of xylitol are well known, and the recommended daily dose of xylitol for the prevention of dental caries is 6-10 grams (94). The anti-caries action of xylitol on *S. mutans* comes from its interference in the cycle of energy production. Transport of xylitol into the cell via the fructose PTS system results in the intracellular formation of xylitol-5-phosphate (X5P), which competes with fructose-6-phosphate for the enzyme phosphofructokinase, an essential metabolite for subsequent steps of glycolysis. The unusable X5P is then dephosphorylated and expelled from the cell, resulting in a futile cycle of energy expenditure with no gain (95). This use of energy eventually causes the cell to starve and results in growth inhibition. Additionally, xylitol has been shown to decrease the synthesis of insoluble polysaccharides used for biofilm formation, promote membrane degradation, and cause the formation of intracellular vacuoles (94). The genome of *S. mutans* contains multiple fructose PTS systems, fruI and fruCD, but only the former engages in the transportation of xylitol, and *S. mutans* can adapt to long-term xylitol exposure by downregulating fruI activity (96). Although the fructose PTS is the only confirmed transport system for xylitol in *S. mutans*, it is likely that there are multiple systems that help move xylitol across the cell membrane.

G. Xylitol and sorbitol co-metabolism

PTS elements are often not specific to only one sugar and alternative routes for sugar transport frequently exist (97). For example, *S. mutans* double mutants for both fructose

PTS systems are still viable when grown with fructose as the sole carbon source, and fructose has been shown to be transported by the sorbitol and mannose PTS systems (98). The same is likely true for the sorbitol-specific PTS system, and evidence exists for mechanistic overlap during the transport of xylitol and sorbitol, which, aside from sorbitol carrying an additional carbon group, are structurally identical sugar-alcohols. The growth of *S. mutans* strain OMZ 176 on sorbitol was shown to be inhibited by xylitol, the uptake of sorbitol in cells pre-treated with xylitol was delayed, and the uptake of xylitol was reduced in the presence of sorbitol (95, 99). These observations indicate that there may be overlap in the transport and/or metabolism of these two sugar-alcohols, but more work is needed to elucidate the exact molecular mechanisms underlying these findings on the effects of concomitant sorbitol/xylitol growth.

H. Stress tolerance in *S. mutans*

The constantly changing conditions of the oral cavity provides many challenges to the microbial residents, and *S. mutans* has developed ways to cope with these environmental assaults. Heat stress poses an obvious threat to bacteria residing in the human mouth as the host ingests both warm and cold substances. The response to heat stress is generally conserved among bacteria, and both *E. coli* and *B. subtilis* rely on alternative sigma factors to activate transcription of heat shock associated genes. However, *S. mutans* lacks these common alternative sigma factors, suggesting that the mechanisms of heat shock response in this bacterium must be different (100). A temperature change from 37°C to 42°C causes 34.8% of genes to be differentially expressed in *S. mutans*, including the typical heat shock associated *dnaK* and *groEL* genes (100). In an effort to enhance the

protective qualities of the biofilm matrix, *S. mutans* upregulates GtfC under heat stress. Sugar-specific transcriptional regulators are also induced, but because only genes in PTS system component EII loci were significantly impacted by heat stress these regulators may be utilized in an alternative manner during temperature induced stress (100).

Bacteria commonly encounter oxidative stress from both their environment and competitive bacterial species. Within the dental plaque, *S. gordinii* and *S. sanguinis* are the primary producers of hydrogen peroxide (101). In an experiment measuring the transcriptome of *S. mutans* under oxidative stress, 7% of genes were found to be altered in response to hydrogen peroxides stress, indicating that multiple mechanisms exist to compensate for ROS (101). The alky-hydroperoxide reductase *aphC*, iron-binding protein *dpr*, superoxide dismutase *sod*, peroxide-sensor *perR*, and transcriptional activator s*px* are some of the more well characterized genes involved in peroxide stress (101).

There is a large overlap between genes induced under acid and oxidative stress (101). Fluctuations in the pH of dental plaque can be sudden and severe, dropping from a neutral pH of 7 to 3 in less than 20 minutes (102). However, *S. mutans* is well equipped to withstand acidic conditions, and in complex media *S. mutans* cells grown at a pH of 5.5 have three times the glycolytic activity of cells grown at pH 6.5 (102). F-ATPases, which are membrane bound H+ translocation machinery, are the primary method by which *S. mutans* copes with acidic conditions. These F-ATPases allow *S. mutans* to maintain a cytoplasm with a higher pH than the outside environment (82). Even brief exposure to acidic conditions can have detrimental effects on the cell, including loss of function for key glycolytic enzymes, damage to proteins and the cell membrane. DNA can also be harmed during acidic fluctuations, and protecting genetic material becomes a

key concern for the bacteria under acid stress. RecA is a key moderator of homologous recombination that also participates in the repair of damaged DNA and the restarting of DNA polymerase activity following a break or altered base. *S. mutans recA* mutants exhibited an increased sensitivity to a drop in pH to 2.5 (102). Abasic sites (AP), which are nucleotides lacking a nitrogenous base, are often produced during growth in acidic conditions. Base excision repair via AP endonucleases and exonuclease III helps to remove and replace the damaged nucleotide. UvrA has also been shown to play a role in DNA damage repair during acid shock; *uvrA* mutants display an increased sensitivity to both UV damage and acidic pH. *S. mutans* cope with a pH drop from neutral to 5 via expulsion of H+, altering the production of acidic metabolic end products, and increasing the production of branched chain amino acids (82).

I. sRNA prevalence in *S. mutans*

Although no functional studies have been conducted thus far on sRNA utility in *S. mutans* (2), several RNA-seq based surveys of putative sRNA expression have been conducted. One such scan on sRNAs produced in response to acid stress identified almost 2,000 candidate sRNA transcripts, the majority of which were less than 25 nt in length (103). An even greater number of sRNA candidates were identified in response to sucrose using the same methods, most belonging to this understudied category of RNAs that are less than 20 nt in length (104). Such studies highlight a major knowledge gap in the field of sRNA biology: when do sRNA scans reveal meaningful sRNAs that fulfil functional roles, and when do sRNA detection experiments yield mostly transcriptional noise? A more robust survey of sRNA expression in *S. mutans* in response to the presence or

absence of the carbon catabolite repression (CCR) protein CcpA identified several hundred predicted sRNA transcripts by a combination of bioinformatical methods, some of which showed differential expression in response to varied growth conditions (105). This study revealed the prevalence and importance of sRNA regulation in metabolic pathways for *S. mutans* and highlighted the need for future work into sRNA regulation in the species.

REFERENCES

1. J. Livny, M. K. Waldor, Identification of small RNAs in diverse bacterial species. *Current Opinion in Microbiology* **10**, 96–101 (2007).

2. J. Merritt, Z. Chen, N. Liu, J. Kreth, Posttranscriptional regulation of oral bacterial adaptive responses. *Curr Oral Health Rep* **1**, 50–58 (2014).

3. K. J. Bandyra, *et al.*, The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. *Mol Cell* **47**, 943–953 (2012).

4. K. Papenfort, C. K. Vanderpool, Target activation by regulatory RNAs in bacteria. *FEMS Microbiol Rev* **39**, 362–378 (2015).

5. L. S. Waters, G. Storz, Regulatory RNAs in bacteria. *Cell* **136**, 615–628 (2009).

6. R. Faigenbaum-Romm, *et al.*, Hierarchy in Hfq chaperon occupancy of small RNA targets plays a major role in their regulation. *Cell Rep* **30**, 3127-3138.e6 (2020).

7. C. L. Beisel, G. Storz, Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol Rev* **34**, 866–882 (2010).

8. A. D. Garst, A. L. Edwards, R. T. Batey, Riboswitches: Structures and mechanisms. *Cold Spring Harb Perspect Biol* **3** (2011).

9. E. Loh, F. Righetti, H. Eichner, C. Twittenhoff, F. Narberhaus, RNA thermometers in bacterial pathogens. *Microbiology Spectrum* **6** (2018).

10. T. B. Updegrove, S. A. Shabalina, G. Storz, How do base-pairing small RNAs evolve? *FEMS Microbiol Rev* **39**, 379–391 (2015).

11. H. A. Dutcher, R. Raghavan, Origin, evolution, and loss of bacterial small RNAs. *Microbiol Spectr* **6** (2018).

12. F. R. Kacharia, J. A. Millar, R. Raghavan, Emergence of new sRNAs in enteric bacteria is associated with low expression and rapid evolution. *J. Mol. Evol.* **84**, 204–213 (2017).

13. A. S. Richter, R. Backofen, Accessibility and conservation. *RNA Biol* **9**, 954–965 (2012).

14. L. Barquist, S. W. Burge, P. P. Gardner, Studying RNA homology and conservation with Infernal: From single sequences to RNA families. *Curr Protoc Bioinformatics* **54**, 12.13.1-12.13.25 (2016).

15. E. P. Nawrocki, S. R. Eddy, Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).

16. J. E. Clarke, L. Kime, D. Romero A, K. J. McDowall, Direct entry by RNase E is a major pathway for the degradation and processing of RNA in *Escherichia coli*. *Nucleic Acids Res* **42**, 11733–11751 (2014).

17. M. B. Stead, *et al.*, Analysis of *Escherichia coli* RNase E and RNase III activity in vivo using tiling microarrays. *Nucleic Acids Res* **39**, 3188–3203 (2011).

18. N. De Lay, S. Gottesman, RNase E finds some sRNAs stimulating. *Mol Cell* **47**, 825–826 (2012).

19. K. J. Bandyra, B. F. Luisi, RNase E and the high-fidelity orchestration of RNA metabolism. *Microbiol Spectr* **6** (2018).

20. J. A. Bernstein, P.-H. Lin, S. N. Cohen, S. Lin-Chao, Global analysis of *Escherichia coli* RNA degradosome function using DNA microarrays. *PNAS* **101**, 2758–2763 (2004).

21. Y. Chao, *et al.*, In vivo cleavage map illuminates the central role of RNase E in coding and non-coding RNA pathways. *Mol Cell* **65**, 39–51 (2017).

22. K. Papenfort, *et al.*, Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Molecular Microbiology* **74**, 139–158 (2009).

23. X. Chen, N. Liu, S. Khajotia, F. Qi, J. Merritt, RNases J1 and J2 are critical pleiotropic regulators in *Streptococcus mutans*. *Microbiology (Reading)* **161**, 797–806 (2015).

24. M. T. Franze de Fernandez, L. Eoyang, J. T. August, Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature* **219**, 588–590 (1968).

25. P. J. Mikulecky, *et al.*, *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat Struct Mol Biol* **11**, 1206–1214 (2004).

26. A. Zhang, K. M. Wassarman, J. Ortega, A. C. Steven, G. Storz, The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol. Cell* **9**, 11–22 (2002).

27. M. Miyakoshi, Y. Chao, J. Vogel, Regulatory small RNAs from the 3′ regions of bacterial mRNAs. *Curr. Opin. Microbiol.* **24**, 132–139 (2015).

28. H. C. Tsui, H. C. Leung, M. E. Winkler, Characterization of broadly pleiotropic phenotypes caused by an *hfq* insertion mutation in *Escherichia coli* K-12. *Mol Microbiol* **13**, 35–49 (1994).

29. H. Otaka, H. Ishikawa, T. Morita, H. Aiba, PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *PNAS* **108**, 13059–13064 (2011).

30. A. Santiago-Frangos, S. A. Woodson, Hfq chaperone brings speed dating to bacterial sRNA. *Wiley Interdiscip Rev RNA* **9**, e1475 (2018).

31. E. G. H. Wagner, Cycling of RNAs on Hfq. *RNA Biol* **10**, 619–626 (2013).

32. B. K. Mohanty, S. R. Kushner, Enzymes involved in post-transcriptional RNA metabolism in gram-negative bacteria. *Microbiol Spectr* **6** (2018).

33. Y. Chao, K. Papenfort, R. Reinhardt, C. M. Sharma, J. Vogel, An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* **31**, 4005–4019 (2012).

34. P. Valentin-Hansen, M. Eriksen, C. Udesen, MicroReview: The bacterial Sm-like protein Hfq: A key player in RNA transactions. *Molecular Microbiology* **51**, 1525–1533 (2004).

35. J. A. Imlay, Cellular defenses against superoxide and hydrogen peroxide. *Annu. Rev. Biochem.* **77**, 755–776 (2008).

36. J. A. Imlay, Transcription factors that defend bacteria against reactive oxygen species. *Annu. Rev. Microbiol.* **69**, 93–108 (2015).

37. S. Altuvia, D. Weinstein-Fischer, A. Zhang, L. Postow, G. Storz, A small, stable RNA induced by oxidative stress: Role as a pleiotropic regulator and antimutator. *Cell* **90**, 43–53 (1997).

38. B. González-Flecha, B. Demple, Role for the oxyS gene in regulation of intracellular hydrogen peroxide in *Escherichia coli*. *J. Bacteriol.* **181**, 3833–3836 (1999).

39. K. S. Fröhlich, S. Gottesman, Small regulatory RNAs in the Enterobacterial response to envelope damage and oxidative stress. *Microbiol Spectr* **6** (2018).

40. A. Zhang, *et al.*, The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein. *EMBO J.* **17**, 6061–6068 (1998).

41. S. Altuvia, A. Zhang, L. Argaman, A. Tiwari, G. Storz, The *Escherichia coli* OxyS regulatory RNA represses fhlA translation by blocking ribosome binding. *EMBO J.* **17**, 6069–6075 (1998).

42. S. Barshishat, *et al.*, OxyS small RNA induces cell cycle arrest to allow DNA damage repair. *EMBO J.* **37**, 413–426 (2018).

43. M. H. Irani, P. K. Maitra, Properties of *Escherichia coli* mutants deficient in enzymes of glycolysis. *J Bacteriol* **132**, 398–410 (1977).

44. K. Kimata, Y. Tanaka, T. Inada, H. Aiba, Expression of the glucose transporter gene, ptsG, is regulated at the mRNA degradation step in response to glycolytic flux in *Escherichia coli*. *EMBO J* **20**, 3587–3595 (2001).

45. T. Morita, W. El-Kazzaz, Y. Tanaka, T. Inada, H. Aiba, Accumulation of glucose 6-phosphate or fructose 6-phosphate is responsible for destabilization of glucose transporter mRNA in *Escherichia coli*. *J Biol Chem* **278**, 15608–15614 (2003).

46. C. K. Vanderpool, Physiological consequences of small RNA-mediated regulation of glucose-phosphate stress. *Current Opinion in Microbiology* **10**, 146–151 (2007).

47. C. K. Vanderpool, S. Gottesman, Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Molecular Microbiology* **54**, 1076–1089 (2004).

48. M. Raina, A. King, C. Bianco, C. K. Vanderpool, Dual-function RNAs. *Microbiol Spectr* **6** (2018).

49. A. E. Trotochaud, K. M. Wassarman, A highly conserved 6S RNA structure is required for regulation of transcription. *Nature Structural & Molecular Biology* **12**, 313–319 (2005).

50. J. E. Barrick, N. Sudarsan, Z. Weinberg, W. L. Ruzzo, R. R. Breaker, 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA* **11**, 774–784 (2005).

51. K. M. Wassarman, 6S RNA: a regulator of transcription. *Molecular Microbiology* **65**, 1425–1431 (2007).

52. K. Kim, Y. Lee, Regulation of 6S RNA biogenesis by switching utilization of both sigma factors and endoribonucleases. *Nucleic Acids Res* **32**, 6057–6068 (2004).

53. Y. Ando, S. Asari, S. Suzuma, K. Yamane, K. Nakamura, Expression of a small RNA, BS203 RNA, from the yocI-yocJ intergenic region of *Bacillus subtilis* genome. *FEMS Microbiol Lett* **207**, 29–33 (2002).

54. A. E. Trotochaud, K. M. Wassarman, 6S RNA regulation of pspF transcription leads to altered cell survival at high pH. *Journal of Bacteriology* **188**, 3936–3943 (2006).

55. N. Gildehaus, T. Neußer, R. Wurm, R. Wagner, Studies on the function of the riboregulator 6S RNA from *Escherichia coli*: RNA polymerase binding, inhibition of in vitro transcription and synthesis of RNA-directed de novo transcripts. *Nucleic Acids Res* **35**, 1885–1896 (2007).

56. B. Steuten, *et al.*, Regulation of transcription by 6S RNAs. *RNA Biol* **11**, 508–521 (2014).

57. A. Gaballa, *et al.*, The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *PNAS* **105**, 11927–11932 (2008).

58. M. Gimpel, N. Heidrich, U. Mäder, H. Krügel, S. Brantl, A dual-function sRNA from *Bacillus subtilis*: SR1 acts as a peptide encoding mRNA on the gapA operon. *Mol Microbiol* **76**, 990–1009 (2010).

59. E. Morfeldt, D. Taylor, A. von Gabain, S. Arvidson, Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense RNA, RNAIII. *EMBO J* **14**, 4569–4577 (1995).

60. D. Sinha, *et al.*, Redefining the small regulatory RNA transcriptome in *Streptococcus pneumoniae* serotype 2 strain D39. *Journal of Bacteriology* **201** (2019).

61. B. Mann, *et al.*, Control of Virulence by Small RNAs in *Streptococcus pneumoniae*. *PLoS Pathogens* **8**, e1002788 (2012).

62. A. Halfmann, M. Kovács, R. Hakenbeck, R. Brückner, Identification of the genes directly controlled by the response regulator CiaR in *Streptococcus pneumoniae*: five out of 15 promoters drive expression of small non-coding RNAs. *Mol Microbiol* **66**, 110–126 (2007).

63. H.-C. T. Tsui, *et al.*, Identification and characterization of noncoding small RNAs in *Streptococcus pneumoniae* serotype 2 strain D39. *J Bacteriol* **192**, 264–279 (2010).

64. P. Marx, M. Nuhn, M. Kovács, R. Hakenbeck, R. Brückner, Identification of genes for small non-coding RNAs that belong to the regulon of the two-component regulatory system CiaRH in *Streptococcus*. *BMC Genomics* **11**, 661 (2010).

65. N. Perez, *et al.*, A genome-wide analysis of small regulatory RNAs in the human pathogen group A Streptococcus. *PLoS One* **4** (2009).

66. E. W. Miller, T. N. Cao, K. J. Pflughoeft, P. Sumby, RNA-mediated regulation in gram-positive pathogens: An overview punctuated with examples from the group A *Streptococcus*. *Mol Microbiol* **94**, 9–20 (2014).

67. P. N. Deo, R. Deshmukh, Oral microbiome: Unveiling the fundamentals. *J Oral Maxillofac Pathol* **23**, 122–128 (2019).

68. W. H. Bowen, R. A. Burne, H. Wu, H. Koo, Oral biofilms: Pathogens, matrix, and polymicrobial interactions in microenvironments. *Trends in Microbiology* **26**, 229–242 (2018).

69. T. Chen, *et al.*, The Human Oral Microbiome Database: A web accessible resource for investigating oral microbe taxonomic and genomic information. *Database (Oxford)* **2010** (2010).

70. L. Gao, *et al.*, Oral microbiomes: More and more importance in oral cavity and whole body. *Protein Cell* **9**, 488–500 (2018).

71. M. Kilian, *et al.*, The oral microbiome – an update for oral healthcare professionals. *British Dental Journal* **221**, 657–666 (2016).

72. M. Kilian, The oral microbiome – friend or foe? *European Journal of Oral Sciences* **126**, 5–12 (2018).

73. Y. Li, P. W. Caufield, A. P. Dasanayake, H. W. Wiener, S. H. Vermund, Mode of delivery and other maternal factors influence the acquisition of *Streptococcus mutans* in infants. *J. Dent. Res.* **84**, 806–811 (2005).

74. P. Lif Holgerson, L. Harnevik, O. Hernell, A. C. R. Tanner, I. Johansson, Mode of birth delivery affects oral microbiota in infants. *J Dent Res* **90**, 1183–1188 (2011).

75. S. M. Colby, R. R. B. Russell, Sugar metabolism by mutans streptococci. *Journal of Applied Microbiology* **83**, 80S-88S (1997).

76. K. Kriebel, C. Hieke, B. Müller-Hilke, M. Nakata, B. Kreikemeyer, Oral biofilms from symbiotic to pathogenic interactions and associated disease –connection of periodontitis and rheumatic arthritis by peptidylarginine deiminase. *Front. Microbiol.* **9** (2018).

77. R. Saini, S. Saini, S. Sharma, Biofilm: A dental microbial infection. *J Nat Sci Biol Med* **2**, 71–75 (2011).

78. R. M. Donlan, Biofilm formation: A clinically relevant microbiological process. *Clin Infect Dis* **33**, 1387–1392 (2001).

79. A. F. P. Leme, H. Koo, C. M. Bellato, G. Bedi, J. A. Cury, The role of sucrose in cariogenic dental biofilm formation—new insight. *J Dent Res* **85**, 878–887 (2006).

80. W. Krzyściak, A. Jurczak, D. Kościelniak, B. Bystrowska, A. Skalniak, The virulence of *Streptococcus mutans* and the ability to form biofilms. *Eur J Clin Microbiol Infect Dis* **33**, 499–515 (2014).

81. M. Matsumoto-Nakano, Role of *Streptococcus mutans* surface proteins for biofilm formation. *Japanese Dental Science Review* **54**, 22–29 (2018).

82. J. A. Lemos, R. A. Burne, A model of efficiency: Stress tolerance by *Streptococcus mutans*. *Microbiology* **154**, 3247–3255 (2008).

83. J. L. Baker, A. Edlund, Exploiting the oral microbiome to prevent tooth decay: Has evolution already provided the best tools? *Front. Microbiol.* **9** (2019).

84. J. Deutscher, C. Francke, P. W. Postma, How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev* **70**, 939–1031 (2006).

85. D. Ajdić, *et al.*, Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. *PNAS* **99**, 14434–14439 (2002).

86. Z. D. Moye, L. Zeng, R. A. Burne, Modification of gene expression and virulence traits in *Streptococcus mutans* in response to carbohydrate availability. *Appl. Environ. Microbiol.* **80**, 972–985 (2014).

87. M. Kawada-Matsuo, Y. Oogai, H. Komatsuzawa, Sugar allocation to metabolic pathways is tightly regulated and affects the virulence of *Streptococcus mutans*. *Genes* **8**, 11 (2017).

88. A. J. Webb, K. A. Homer, A. H. F. Hosie, Two closely related ABC transporters in *Streptococcus mutans* are involved in disaccharide and/or oligosaccharide uptake. *J Bacteriol* **190**, 168–178 (2008).

89. V. G. Lewis, M. P. Ween, C. A. McDevitt, The role of ATP-binding cassette transporters in bacterial pathogenicity. *Protoplasma* **249**, 919–942 (2012).

90. J. D. F. Hale, N. C. K. Heng, R. W. Jack, J. R. Tagg, Identification of nlmTE, the locus encoding the ABC transport system required for export of nonlantibiotic mutacins in *Streptococcus mutans*. *J Bacteriol* **187**, 5036–5039 (2005).

91. D. A. Boyd, T. Thevenot, M. Gumbmann, A. L. Honeyman, I. R. Hamilton, Identification of the operon for the sorbitol (glucitol) phosphoenolpyruvate:sugar phosphotransferase system in *Streptococcus mutans*. *Infect Immun* **68**, 925–930 (2000).

92. E.-M. Decker, C. Klein, D. Schwindt, C. von Ohle, Metabolic activity of *Streptococcus mutans* biofilms and gene expression during exposure to xylitol and sucrose. *Int J Oral Sci* **6**, 195–204 (2014).

93. Y. Yamamoto, *et al.*, Cloning and sequence analysis of the pfl gene encoding pyruvate formate-lyase from *Streptococcus mutans*. *Infect Immun* **64**, 385–391 (1996).

94. P. A. Nayak, U. A. Nayak, V. Khandelwal, The effect of xylitol on dental caries and oral flora. *Clin Cosmet Investig Dent* **6**, 89–94 (2014).

95. S. Assev, G. Rölla, Further studies on the growth inhibition of *Streptococcus mutans* OMZ 176 by xylitol. *Acta Pathol Microbiol Immunol Scand B* **94**, 97–102 (1986).

96. J. M. Tanzer, A. Thompson, Z. T. Wen, R. A. Burne, *Streptococcus mutans*: fructose transport, xylitol resistance, and virulence. *J Dent Res* **85**, 369–373 (2006).

97. B. Erni, The bacterial phosphoenolpyruvate:sugar phosphotransferase system (PTS): An interface between energy and signal transduction. *J Iran Chem Soc* **10**, 593–630 (2013).

98. Z. T. Wen, C. Browngardt, R. A. Burne, Characterization of two operons that encode components of fructose-specific enzyme II of the sugar:phosphotransferase system of *Streptococcus mutans*. *FEMS Microbiol. Lett.* **205**, 337–342 (2001).

99. S. Assev, G. Rølla, Sorbitol increases the growth inhibition of xylitol on *Streptococcus mutans* OMZ 176. *Acta Pathol Microbiol Immunol Scand B* **94**, 231–237 (1986).

100. C. Liu, *et al.*, *Streptococcus mutans* copes with heat stress by multiple transcriptional regulons modulating virulence and energy metabolism. *Sci Rep* **5** (2015).

101. J. K. Kajfasz, T. Ganguly, E. L. Hardin, J. Abranches, J. A. Lemos, Transcriptome responses of *Streptococcus mutans* to peroxide stress: Identification of novel antioxidant pathways regulated by Spx. *Scientific Reports* **7**, 16018 (2017).

102. R. Matsui, D. Cvitkovitch, Acid tolerance mechanisms utilized by *Streptococcus mutans*. *Future Microbiol* **5**, 403–417 (2010).

103. S. Liu, *et al.*, Analysis of small RNAs in *Streptococcus mutans* under acid stress—a new insight for caries research. *International Journal of Molecular Sciences* **17**, 1529 (2016).

104. S. S. Liu, *et al.*, Analysis of sucrose-induced small RNAs in *Streptococcus mutans* in the presence of different sucrose concentrations. *Appl. Microbiol. Biotechnol.* **101**, 5739–5748 (2017).

105. L. Zeng, *et al.*, Gene regulation by CcpA and catabolite repression explored by RNA-Seq in *Streptococcus mutans*. *PLoS One* **8** (2013).

CHAPTER TWO

A peroxide-responding sRNA evolved from a peroxidase mRNA

Madeline C. Krieger[a], H. Auguste Dutcher[a], Andrew J. Ashford[a], Rahul Raghavan[a,b]

[a]Department of Biology, Portland State University, Portland, OR 97201, USA.
[b]Department of Biology, The University of Texas at San Antonio, San Antonio, TX 78249, USA.

Key words: OxyS, sRNA, sRNA evolution, peroxidase, peroxide, RNase E

ABSTRACT

Small RNAs (sRNAs) are critical regulators of gene expression in bacteria, but we lack a clear understanding of how new sRNAs originate and get integrated into regulatory networks. A major obstacle to elucidating their evolution is the difficulty in tracing sRNAs across large phylogenetic distances. To overcome this roadblock, we investigated the prevalence of sRNAs across Enterobacterales, a bacterial order with a rare confluence of factors that allows robust genome-scale sRNA analyses: several well-studied organisms with fairly conserved genome structures, an established phylogeny, and substantial nucleotide diversity within a narrow evolutionary space. Using a covariance modeling-based approach, we analyzed the presence of hundreds of sRNAs in more than a thousand Enterobacterales genomes and showed that a majority of sRNAs arose recently, and uncovered protein-coding genes as a potential source for the generation of new sRNA genes. A detailed investigation of the emergence of OxyS, a peroxide-responding sRNA, demonstrated that it likely evolved from an RNase E-processed, 3′ end fragment of a peroxidase mRNA. Collectively, our data show that the erosion of protein-coding genes can result in the formation of new sRNAs that continue to be part of the original protein's regulon. This novel insight provides a fresh framework for understanding how new sRNAs originate and get incorporated into preexisting regulatory networks.

AUTHOR SUMMARY

Small RNAs (sRNAs) are important gene regulators in bacteria, but it is unclear how new sRNAs originate and become part of regulatory networks that coordinate bacterial response to environmental stimuli. Here, we show that new sRNAs arise from protein-coding genes and are incorporated into the ancestral proteins' regulatory networks. We illustrate this process by defining the origin of OxyS. This peroxide-responding sRNA evolved from and replaced a peroxidase gene but continues to be part of the peroxide-response regulon. In sum, this study has significantly improved our knowledge of sRNA evolution by identifying protein-coding genes as a raw material from which new sRNAs emerge, and by defining a novel evolutionary path through which new sRNAs get incorporated into regulatory networks.

INTRODUCTION

Bacterial small RNAs (sRNAs) control gene expression by modulating translation or by altering the stability of messenger RNAs (mRNAs). sRNAs allow precise and efficient control of gene expression because they are produced quickly, regulate multiple genes simultaneously, and are degraded along with target mRNAs (1). These qualities are especially beneficial under conditions such as oxidative stress that require abrupt reprogramming of regulatory networks (2). In bacteria, oxidative stress caused by hydrogen peroxide (H_2O_2) is mitigated mainly by peroxidases (3, 4). For instance, a peroxidase system encoded by *ahpCF* genes is induced by the regulator OxyR when *Escherichia coli* is exposed to H_2O_2. OxyR simultaneously upregulates the expression of several other genes, including the sRNA OxyS that together assuage H_2O_2 toxicity (5–9).

OxyS is one of the most well-studied sRNAs. More than two decades of research on this sRNA has revealed many of the foundational details about sRNA-mediated gene regulation (10–15). In *E. coli* and *Salmonella enterica*, OxyS is encoded by a gene located in the intergenic region (IGR) between *oxyR* and *argH* genes. Similar to OxyS, most sRNAs in bacteria are transcribed from genes present in IGRs; however, in recent years, numerous sRNAs that are encoded within protein-coding genes have also been identified. In such cases, sRNAs are either transcribed from promoters contained within protein-coding sequences or are generated from the 3′ ends of mRNAs by the endoribonuclease RNase E, which acts in concert with the chaperone protein Hfq (17–22).

Despite the discovery of hundreds of sRNAs, we do not fully understand how new sRNAs originate in bacteria (23). One of the main impediments to elucidating the

evolutionary histories of sRNAs is the difficulty in tracing sRNAs across large phylogenetic distances (24). Unlike proteins that are fairly easy to identify in distant bacteria, sRNAs can only be reliably detected within clusters of related microbes (25). This difficulty is due to a combination of factors, including their small size (50 to 400 nt), rapid turnover, and a lack of open reading frames (ORFs) or other features that serve as signposts (23-27). Given these constraints, an ideal group of bacteria to study sRNA evolution is the order Enterobacterales (25), which has an established phylogeny, substantial nucleotide diversity within a narrow evolutionary space, and contains well-characterized organisms with diverse lifestyles but enough similarity in genome structure that enables meaningful comparative genomics.

Here, by analyzing the prevalence of hundreds of sRNAs in more than a thousand Enterobacterales genomes, we show that most sRNAs arose recently, and that mRNAs are a potentially rich source for the generation of new sRNAs. One sRNA that originated from an mRNA is OxyS, which evolved from an RNase E-processed, 3′ end fragment of a peroxidase mRNA. Interestingly, both the parental peroxidase and OxyS are regulated by OxyR, suggesting a novel paradigm for understanding how new sRNAs arise and are recruited into preexisting regulatory networks: erosion of a protein-coding gene could lead to the formation of a new sRNA gene that continues to function as part of the original protein's regulatory network.

RESULTS

Most sRNAs in enteric bacteria arose recently

We built covariance models for 371 sRNAs described in *Escherichia coli* K-12 MG1655, *S. enterica* Typhimurium SL1344, and *Yersinia pseudotuberculosis* IP32953, and located their homologs across 1105 Enterobacterales genomes. The ensuing phyletic patterns of sRNA presence and absence was used to perform an evolutionary reconstruction of ancestral states using a maximum likelihood approach (**Figure 1, Table S1**). This order-wide analysis showed that 61% of sRNAs (228/371) are in the "young" category, meaning they emerged at the root of a genus or more recently. In comparison, among 148 proteins that function as gene regulators in *E. coli* and *S. enterica*, only 18% fall in this category (**Figure 2, Table S2**). The overrepresentation of recently-evolved sRNAs in our dataset indicates that sRNAs arise rapidly, probably in response to lineage-specific selection pressures.

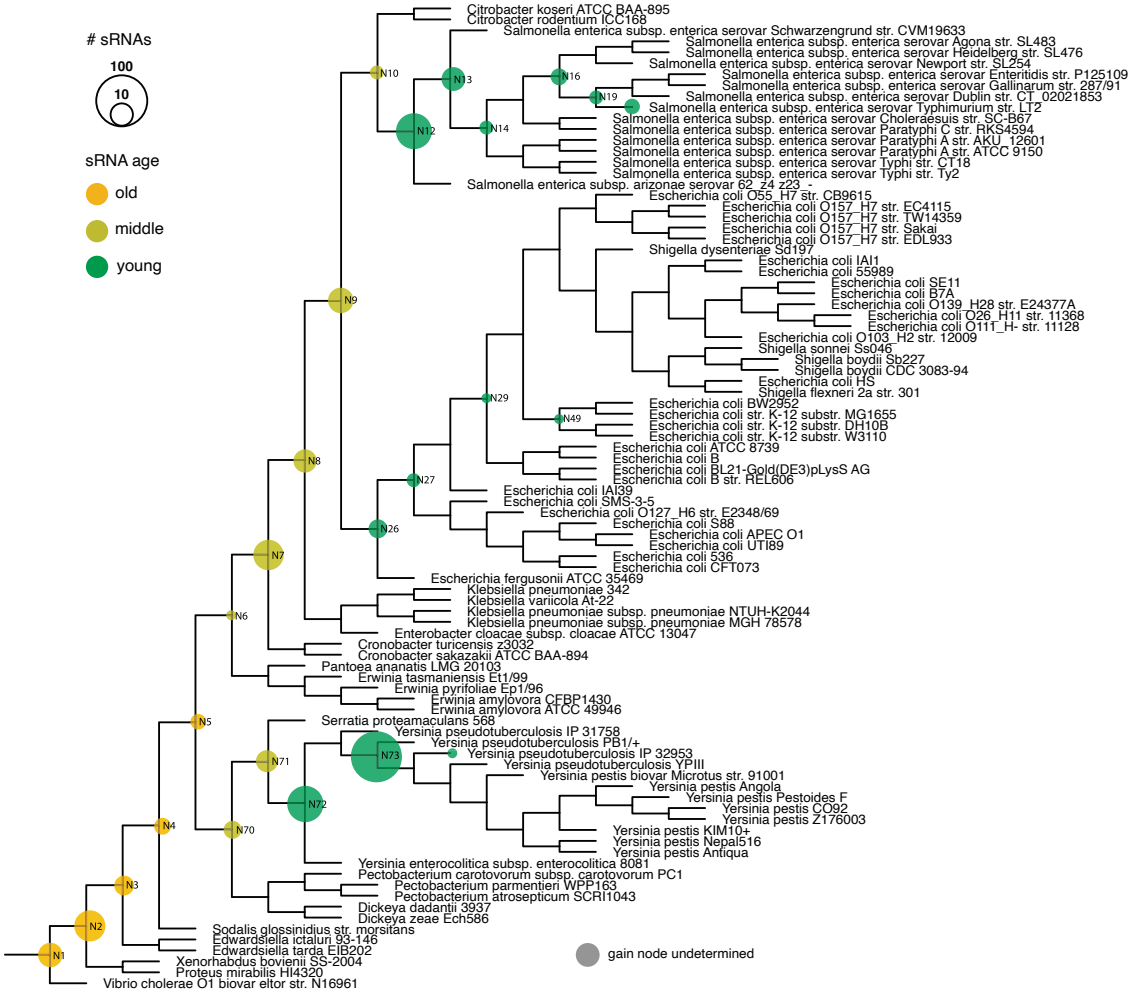

Figure 1. Most sRNAs in enteric bacteria arose recently. sRNAs that arose at each node is depicted by circles. Size and color of each circle corresponds respectively to the number of sRNAs and their ages, as shown in the side panel.

Figure 2. Regulatory proteins in enteric bacteria are older than sRNAs. Regulatory proteins that arose at each node is depicted by circles. Size and color of each circle corresponds respectively to the number of proteins and their ages, as shown in side panel.

sRNA genes overlap with protein-coding genes

Our *cmscan*-based iterative search identified 62 sRNAs that were located in IGRs in the hub genomes (*E. coli* K-12 MG1655, *S. enterica* Typhimurium SL1344, or *Y. pseudotuberculosis* IP32953) but mapped to the coding strands of protein-coding genes in other Enterobacterales members (**Table S3**). These sRNA-ORF overlaps suggest that several of the sRNAs likely started out as segments of mRNAs, and later evolved into independent sRNAs when the original protein-coding genes decayed, leaving behind only the parts that encoded the sRNAs. In the rest of this article, we elucidate this process by tracing the evolution of OxyS, which arose from a 3′ end segment of a peroxidase mRNA, and later evolved into a stand-alone sRNA gene when the rest of the peroxidase gene was deleted from the genome.

A peroxidase gene was replaced by *oxyS* gene in Enterobacteriaceae

In the order Enterobacterales, *oxyS* gene is present only in the family Enterobacteriaceae, where it is located divergently from the *oxyR* gene in the *oxyR-argH* IGR (**Figure 3**). In contrast, a peroxidase (peroxiredoxin-glutaredoxin hybrid) gene occupies the same loci in families Erwiniaceae, Pectobacteriaceae, Yersiniaceae, Hafniaceae, and Budviciaceae. Bacteria belonging to orders Pasteurellales and Vibrionales also contain orthologous peroxidase genes at this location (**Figure 3**). The most parsimonious explanation for this phylogenetic profile is that the peroxidase gene was present in the common ancestor of all Enterobacterales and that it was subsequently replaced by the *oxyS* gene in Enterobacteriaceae. The only family within Enterobacterales that does not have either *oxyS* or peroxidase in this location is Morganellaceae. Bacteria belonging to this family

have intact *oxyR-argH* IGRs, but in place of the peroxidase gene they contain genes for transposases and small hypothetical proteins, suggesting that the ancestral peroxidase gene was lost in this lineage due to the activity of selfish genetic elements.

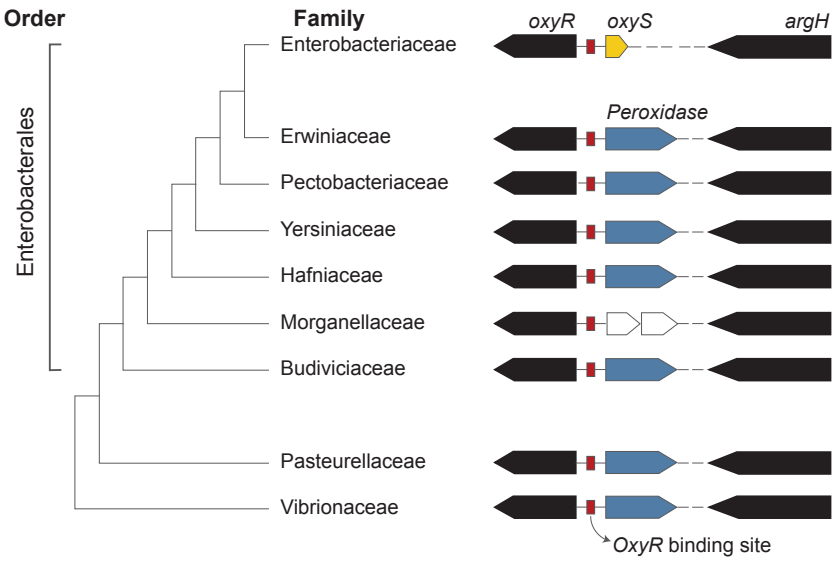

Figure 3. OxyS arose from a peroxidase gene. Arrangement of *oxyR*, peroxidase and *argH* genes in bacterial families within the order Enterobacterales is shown. In Enterobacteriaceae *oxyS* (yellow arrow) is found in place of the peroxidase gene (blue arrow). The intergenic region between peroxidase gene and *argH* varies between families. In Morgenallaceae, the peroxidase gene has been lost due to the activity of selfish genetic elements (white arrows). Orders Pasteurellales and Vibrionales also contain the same gene arrangement as Enterobacterales members. The cladogram is based on Adeolu et al. (16).

Peroxidase genes contain OxyS-like sequences at their 3′ ends

Our *cmscan*-based search identified OxyS-like sequences at the 3′ ends of the peroxidase genes in Pectobacteriaceae and Yersiniaceae but not in other families. We explored this OxyS-like sequence (OLS) in more detail in *Serratia marcescens* (family Yersiniaceae) (**Figure 4**). OLS overlaps the last 56 bp of the peroxidase coding sequence, extends into the 3′ UTR and ends at the downstream intrinsic terminator. Curiously, no matches to OxyS were found in Erwiniaceae even though bacteria in this family contain an intact peroxidase gene in the *oxyR-argH* IGR. A closer examination of *Erwinia amylovora*, a representative member of this family, revealed that the 3′ end of the peroxidase gene has diverged substantially in this bacterium, which likely made it not match our OxyS covariance model.

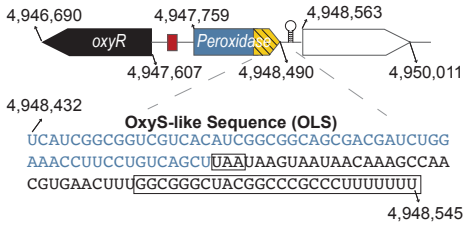

Figure 4. The peroxidase gene in *Serratia marcescens* contains OxyS-like sequence. The peroxidase gene in *Serratia marcescens* ATCC13880 (CP041233) is flanked by *oxyR* (black arrow) and a dihydrolipoyl dehydrogenase gene (white arrow). Covariance-based search detected an OxyS-like sequence (OLS) at the 3′ end of the peroxidase gene. The part of OLS that overlaps the coding sequence is shown as blue nucleotides, and the stop codon and the predicted intrinsic terminator are boxed. The intergenic region between peroxidase and *oxyR* genes contains putative OxyR-binding sites (red square).

Exposure to H_2O_2 induces peroxidase expression and mRNA fragmentation

Similar to *oxyS*, the peroxidase gene is located divergently from *oxyR* and the IGR between the two genes contain putative OxyR-binding sites (**Figure 3, Figure 5**). As observed for OxyS in *E. coli* (10), exposure to H_2O_2 significantly induced the expression of the peroxidase gene in *S. marcescens* (**Figure 6**), indicating that it is also regulated by OxyR. In addition, H_2O_2 exposure caused the peroxidase mRNA to be fragmented into several shorter RNAs, including a ~100 nt 3′ end fragment that contains the OLS (**Figure 7A**). Previous studies have shown that 3′ UTR-derived sRNAs are either transcribed from promoters within protein-coding genes or are cut out of mRNAs by RNase E digestion (17–19). We could not identify any promoter-like sequences within the peroxidase gene, but an *in vitro* digestion produced small 3′ end-containing fragments only when RNase E was present in the reaction (**Figure 7B**).

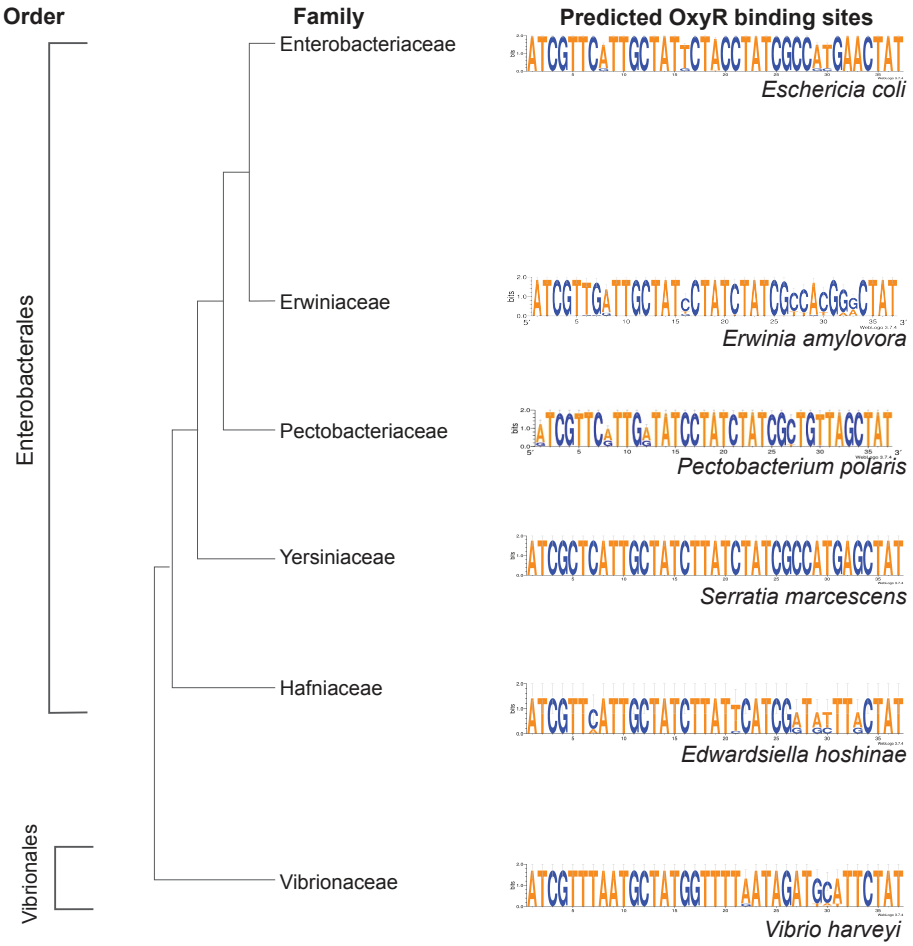

Figure 5. Predicted OxyR binding sites. Putative OxyR binding sites are located between *oxyR* and peroxidase/*oxyS* genes in Enterobacterales members. Predicted OxyR-binding sites in one representative taxa per family is shown.

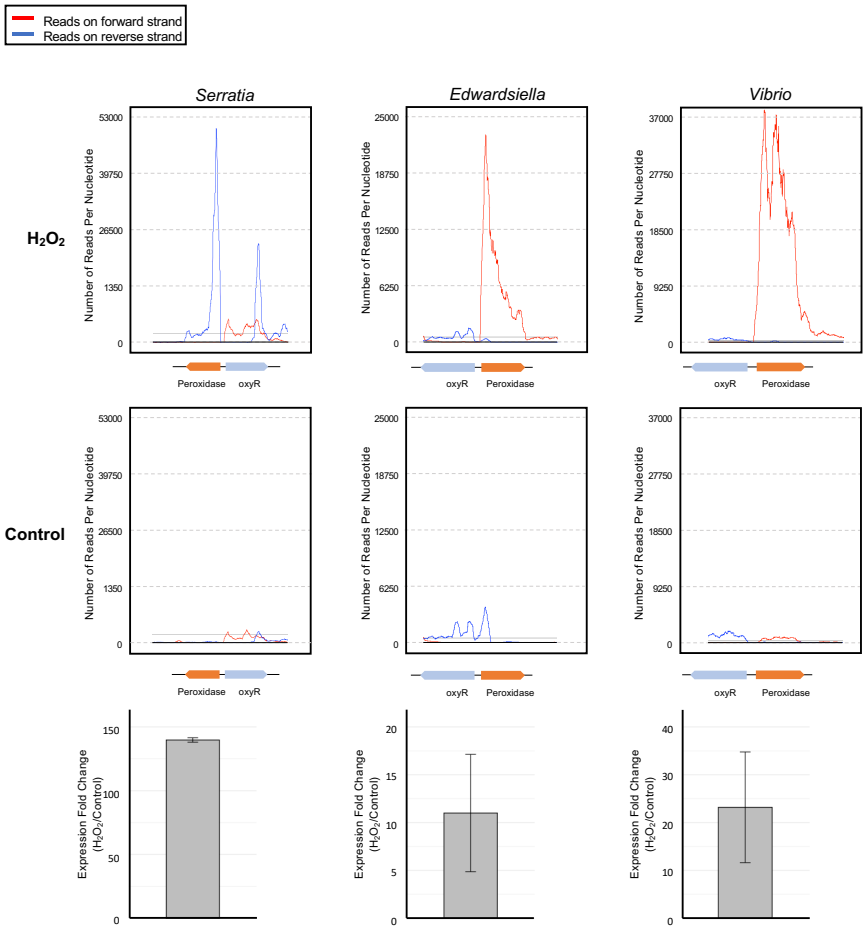

Figure 6. H_2O_2 induces peroxidase expression. Top panels: RNA-seq reads that mapped to peroxide genes (orange arrows) in *S. marcescens*, *E. hoshinae*, and *V. harveyi* when exposed to 1mM of H_2O_2 for 10 minutes. **Middle panels**: RNA-seq data from control (no H_2O_2 exposure) samples. **Bottom panels**: Confirmation of peroxidase induction by H_2O_2 via RNA-seq (*S. marcescens*) or qRT-PCR (*E. hoshinae*, and *V. harveyi*).

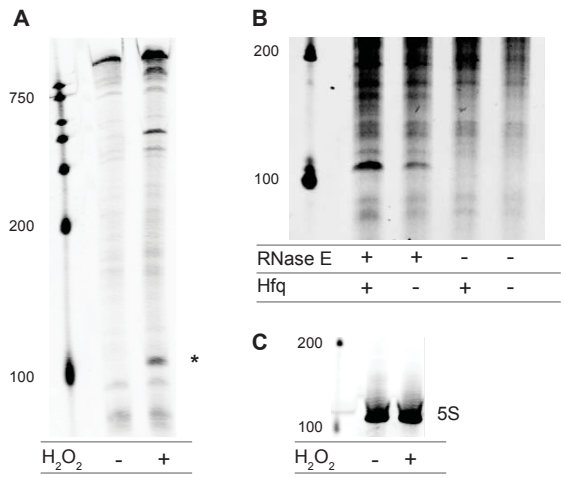

Figure 7. Peroxidase mRNA fragmentation. (A) When *S. marcescens* was exposed to H_2O_2, peroxidase mRNA was cleaved into several fragments, including a ~100 nt 3′ end fragment (marked with ∗). (**B**) *In vitro* transcribed *S. marcescens* peroxidase mRNA was cleaved into shorter fragments only in reactions that contained RNase E. Northern blotting was performed with a probe that bound to the 3′ end of peroxidase mRNA, and 5S rRNA was used as loading control (**C**).

Fragmentation of peroxidase mRNA is an ancestral trait

We investigated the prevalence of peroxidase mRNA fragmentation by testing another Enterobacterales member (*Edwardsiella hoshinae*, family Hafniaceae), and a representative from outside of Enterobacterales: *Vibrio harveyi* (order Vibrionales, family Vibrionaceae) (**Figure 3**). Upon H_2O_2 exposure, expression of peroxidase genes in both bacteria were induced several fold (**Figure 6**), and their mRNAs were cut into smaller pieces (**Figure 8**). Although the size of the cleavage products are not identical in the three bacteria, the fragmentation process itself appears to be a trait ancestral to all Enterobacterales. Cumulatively, our data show that a 3′ end fragment of an OxyR-regulated peroxidase mRNA evolved into OxyS, which is expressed as part of the OxyR regulon in Enterobacteriaceae.

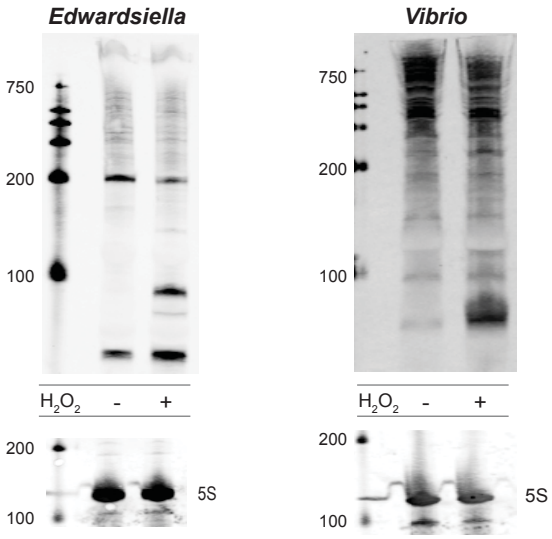

Figure 8. Peroxidase mRNA fragmentation is an ancestral trait. Short 3′ end fragments were cleaved from peroxidase mRNAs when *E. hoshinae* and *V. harveyi* were exposed to H_2O_2. Northern blotting was performed with probes that bound to 3′ ends of peroxidase mRNAs, and 5S rRNAs were used as loading controls.

DISCUSSION

Despite their importance to bacterial physiology and virulence, the evolutionary processes that produce new sRNAs are not well understood. We have previously shown that new sRNAs arise *de novo* and from degraded bacteriophage- and transposon-associated genes (23, 26, 28, 29), but it is not clear how they get integrated into regulatory networks. In this study, we report that most sRNAs in enteric bacteria arose recently and that protein-coding genes could serve as an important raw material for sRNA biogenesis. We support these conclusions by demonstrating that OxyS, a peroxide-responding sRNA in the family Enterobacteriaceae, originated from and replaced a peroxidase gene conserved across the order Enterobacterales. Collectively, our data reveal that by virtue of arising from a protein-coding gene, a new sRNA could readily be recruited into the parental protein's regulon. Thus, this work provides a new roadmap for understanding how novel sRNAs arise in bacteria and become part of preexisting regulatory networks.

One of the main reasons for the prevailing lack of clarity about sRNA origination is that unlike protein-coding genes, sRNA genes are difficult to trace across large phylogenetic distances (24). Following up on previous research that showed that enteric bacteria are at optimum distances from one another to effectively investigate sRNA prevalence (25), we traced the presence of hundreds of sRNAs across Enterobacterales and show that a majority emerged recently in a lineage-specific manner. This observation fits with earlier findings that showed that sRNAs evolve rapidly in bacteria and are typically genus- or species-specific (26, 30, 31). It should be noted however that the functions, if any, of most recently-emerged sRNAs have not been determined, and that

nearly all sRNAs with known functions belong to the "middle" and "old" categories (**Table S1**). This biased representation is probably due to the propensity of older sRNAs to be expressed at high levels, thereby making them more amenable to discovery and experimental validation (26, 32).

In addition to suggesting the recent emergence of sRNAs, our covariance-based iterative search identified protein-coding genes as a potential source from which new sRNAs could arise. Out of the 62 sRNAs that appear to have originated from protein-coding genes, in this study, we traced the evolutionary path of OxyS. This sRNA was first noticed by researchers because it is transcribed divergently from the *oxyR* gene that encodes a transcriptional regulator that orchestrates *E. coli*'s antioxidant response (10). The IGR between *oxyS* and *oxyR* genes contains OxyR-binding sites from where the oxidized form of OxyR induces the expression of OxyS in response to H_2O_2 exposure (33, 34). In this study, we show that the *oxyR-oxyS* gene arrangement is present only in the family Enterobacteriaceae, whereas a peroxidase gene occupies the genetic locus next to *oxyR* in other members of the order Enterobacterales. Furthermore, we show that OxyS originated from a 3′ end fragment of the ancestral peroxidase mRNA. Interestingly, irrespective of whether it is *oxyS* or the peroxidase gene that is situated divergently from *oxyR*, the IGRs between the genes contain putative OxyR-binding sites, and when exposed to H_2O_2, the expression of the peroxidase gene is induced several fold, similar to the induction of OxyS in *E. coli* and *S. enterica* (10, 35). Thus, our data show that although OxyS arose from and replaced an OxyR-regulated peroxidase gene, the sRNA continues to be part of the OxyR regulon.

An unresolved question in the field of sRNA biology is how new sRNAs get incorporated into regulatory networks. This study provides a possible solution: a sRNA arising from a protein-coding gene would inherently be part of the parental protein's regulon. For instance, the 3′ end fragment of the peroxidase mRNA that gave rise to OxyS would have been produced as part of the OxyR regulon even before it gained any function. The presence of the Rho-independent terminator likely increased its stability and allowed it to bind to Hfq and interact consistently with mRNAs that are also produced under oxidative stress (27, 36). Some of these interactions likely triggered mRNA decay through RNase E digestion, and because it improved bacterial fitness, the 3′ end fragment came under selection and evolved into OxyS. Later, when the rest of the peroxidase gene was eliminated via deletion, the 3′ end that encoded OxyS was retained due to its contribution to bacterial fitness, thus producing an OxyR-regulated sRNA that takes part in oxidative stress response (**Figure 9**). Several examples of 3′ UTR-derived sRNAs that are part of the same regulon as their parent genes are currently known (37–39). Based on our data, some of these 3′ UTR-derived sRNAs could evolve into IGR-located sRNAs that continue to be part of the same regulon if/when the 5′ regions of their parental genes are lost from future generations of bacteria.

New sRNAs could also arise from protein-coding genes without first functioning as 3′ end-derived sRNAs (**Figure 9**). Protein-coding genes are continually being gained and lost in bacteria (40). Genes that are critical to fitness in one environment could be rendered redundant due to alterations in bacterial ecology or lifestyle. These now-superfluous genes accumulate inactivating mutations and turn into pseudogenes;

however, many continue to produce transcripts as part of the original proteins' regulatory networks (41). For instance, it is possible that the peroxidase gene became redundant when Enterobacteriaceae diverged from the rest of the Enterobacterales, and even after being pseudogenized it continued to be transcribed as part of the OxyR regulon. Later, the 3′ end fragment could have gained functions, as outlined above, and was retained, while the rest of the gene got deleted. It is not known whether OxyS evolved via the former or the latter path (**Figure 9**), but a few other sRNA genes (e.g., *gcvB*, *sgrS*, *gadY*) in *E. coli* are also located divergently from genes that encode their transcriptional regulators (42), indicating that this process has generated new sRNAs on multiple occasions in enteric bacteria. In sum, because bacterial genomes are dominated by protein-coding genes in various stages of decay, they could function as a rich resource from which new sRNAs could arise rapidly in response to lineage-specific environmental pressures.

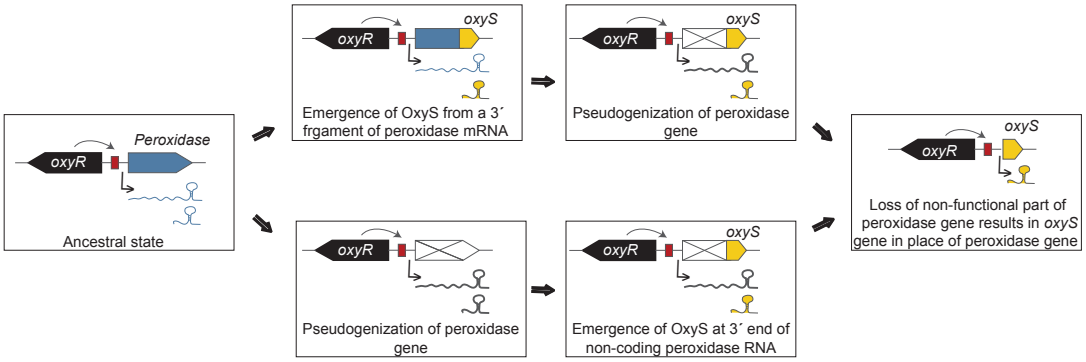

Figure 9. Two possible routes of OxyS evolution. OxyS either emerged as a 3′ end-derived sRNA prior to the pseudogenization of the peroxidase gene (top path), or it emerged at the 3′ end of an RNA transcribed from a pseudogenized peroxidase gene (bottom path). Ultimately, the non-functional part of the peroxidase gene was deleted from the genome, resulting in the formation of the *oxyS* gene in Enterobacteriaceae. Reduced and oxidized forms of OxyR binds to sites (red boxes) located in the IGR and regulate the expression of *oxyR* and peroxidase/*oxyS* genes.

MATERIALS AND METHODS

Determining sRNA presence across the order Enterobacterales

A list of candidate-sRNAs in *E. coli* K-12 MG1655 (NC_000913.3), *S. enterica* Typhimurium SL1344 (NC_016810.1), and *Yersinia pseudotuberculosis* IP32953 (NC_006155.1) were compiled from previously published studies (32, 35, 43). Several exclusion criteria were used to remove spurious and redundant sRNAs: (i) sRNAs under 60 nucleotides in length, (ii) sRNAs that overlapped each other by more than 10 nucleotides, (iii) sRNAs that were present in multiple copies, (iv) RNAs that were identified as cis-acting in the Rfam database (44), and (v) sRNAs that shared 95% or more nucleotide identity over at least 60 nucleotides of their lengths. For each sRNA of interest, a representative sequence from each hub genome was used as the query for a BLASTn (wordsize 7, maxdbsize 100Kb, dbsize normalized to 4Mb, evalue ≤1e-5) using BLAST v2.7.1 against a database of 1105 Enterobacterales genomes (**Table S4**) that met the following criteria: (i) full genome sequence was available on GenBank, and (ii) the genome was within 0.08 16S rDNA pairwise distance from the hub species. Drawing on guidance from previous studies (24, 25), non-identical but similar hits with pident >65% covering at least 95% of the length of the original query served as seed sequences from which to construct a covariance model. Candidate hits were next binned by percent identity, and a randomly selected set of sequences (one from each percent identity bin) were chosen to serve as a seed sequence for the covariance model. These sequences were aligned using ClustalW, and the Infernal suite of tools (v1.1.2) was used for subsequent covariance model construction (cmbuild), calibration (cmcalibrate), and homolog searches (cmsearch) (45). Models were constructed from the BLAST-derived seed

sequences using cmbuild, while cmscan was used to identify sRNAs already represented by existing Rfam models. These newly constructed models, plus the existing Rfam models, were then used in parallel to search the 1105-genome database for homologs. For sRNAs that were represented by an existing Rfam model, cmsearch results from this model were compared to that from the newly constructed model, and the model that yielded more hits was selected for continued iteration. Results from cmsearch with an e-value <1e-5 were used to add unrepresented sequences to the query model, which was then refined, recalibrated, and used for another round of cmsearch. This process was repeated for each sRNA until a cmsearch with its corresponding model failed to yield new unrepresented sequences. In order to ensure that any two models were not yielding the same set of hits, results from cmsearch with the finalized models were compared across sRNAs; models with redundant hits were omitted, as were any models that yielded >1e4 hits. A sRNA gene was considered present in a given organism if a hit of e-value <1e-5 was found on its chromosome and/or plasmid. All resultant hits were cross-checked by genome location to ensure that a given hit was not represented more than once in the final results. Presence/absence data for all 1105 organisms was collected, but only data for 89 Enterobacteriaceae, plus *Vibrio cholerae* El Tor str. N16961 as an outgroup, was used for downstream phylogenetic analyses.

Evolutionary reconstruction

Enterobacterales phylogenetic tree was downloaded from MicrobesOnLine, and node of origin for each sRNA was determined using the Gain and Loss Mapping Engine (GLOOME) as described previously (46, 47). For sRNAs present in a single hub genome

(*E. coli* K-12 MG1655, *S. enterica* Typhimurium SL1344 or *Y. pseudotuberculosis* IP32953), the determined gain node was the most ancestral node with a posterior probability of ≥0.6, where all nodes leading from this ancestor to the hub genome had a posterior probability ≥0.6. If a sRNA was present in more than one hub genome, and the most recent last common ancestor (LCA) of the hub genomes in which it was present had a posterior probability ≥0.6, the gain node was the most ancestral node with a posterior probability ≥0.6, where all nodes leading from this ancestor to the aforementioned LCA had a posterior probability ≥0.6. sRNAs that emerged at the root of a genus or more recently were classified as "young" (n=228), those present at the last common ancestor of all three hub genomes were deemed to be "old" (n=57), and sRNAs that emerged in between the two age groups were considered "middle" (n=73). If the last common ancestor of the hub genomes in which the sRNA was present did not have a posterior probability of ≥0.6, ages of these sRNAs were considered undetermined (n=13).

Regulatory protein tree

Using the QuickGO annotation table (48), GO:0010629 (negative regulation of gene expression) and its child terms were utilized to identify regulatory proteins in *S. enterica* Typhimurium LT2 and *Escherichia coli* K-12 MG1655. Protein sequences (UP00000104, UP000000625) were downloaded from Uniport (49). Redundancies among the regulatory proteins from the two different species were identified using a BLASTp of the regulatory protein sequence set against itself (pident >80, e-value <1e-10). Homologs of this final set of query proteins were identified using BLASTp (e-value <1e-10) against a database of protein sequences from the genomes used for phylogenetic analysis, yielding a

presence/absence matrix. Node of origin was determined using GLOOME as described above.

Finding *argH-oxyR* IGRs

To find *oxyR* and *argH* orthologs, tBLASTn searches were carried out with OxyR and ArgH sequences from *E. coli* K-12 MG1655 against Enterobacterales genomes (e-value ≤10e-10, percent positive ≥60%, percentage alignment length ≥60%). We corroborated the tBLASTn hits by confirming that they contain Pfam domains PF03466.20 and PF00126.27 (for OxyR), and PF00206.20 and PF14698.6 (for ArgH) (50). We then compiled a list of bacteria that have both *argH* and *oxyR* genes and obtained the nucleotide sequences between the two genes using the Entrez E-utilities tool.

Identifying the 5′ neighbor of *oxyR* and OxyR-binding sites

Using BioPython (51), we first determined the direction of the gene next to *oxyR*'s 5′ end. If the neighboring gene was oriented divergently, we determined the identity of the encoded protein using Pfam as described above. All Enterobacterales and Vibrionales (except Morganellaceae) included in this study contained a peroxidase gene with Glutaredoxin (PF00462) and Redoxin (PF08534) domains in this locus. To identify putative OxyR-binding sites located in the IGR between *oxyR* and its neighbor, we extracted 50 bp at the 5′ end of *oxyR* along with 100 bp of the adjoining IGR from each bacterium. Sequences were aligned with MUSCLE (52) and the 50 bp *oxyR* sequence was trimmed from each sequence. Using Multiple Em for Motif Elicitation (MEME) (53), we first detected ~37 bp palindromic sequences in IGRs, and then used the output from

MEME in Find Individual Motif Occurrences (FIMO) (54) to identify putative OxyR-binding site in each bacterium. Sequence logos were generated using WebLogo 3 (55).

Bacterial growth and gene expression

Bacteria were inoculated 1:100 from an overnight culture into fresh media. *S. marcescens* ATCC 13880 was grown at 37°C in Lysogeny broth (LB), *E. hoshinae* ATCC 35051 at 26°C in LB, and *V. harveyi* ATCC 43516 at 26°C in Marine Broth, all shaking at 200 rpm. All cultures were grown to an OD600 value of 0.4-0.5, and split into two. One half was allowed to grow under the same conditions for 10 minutes, while the other half was exposed to 1mM of H_2O_2 for 10 minutes. RNA Stop Solution (5% phenol, 95% ethanol) was added and total RNA was extracted using TRI reagent (Thermo Fisher Scientific). RNA was treated with TURBO DNase (Thermo Fischer Scientific), and was either sent to Yale Center for Genome analysis for RNA sequencing (RNA-seq), or cDNA was synthesized and quantitative PCR (qRT-PCR) was performed. PCR primers used in this study are listed in **Table 1**. RNA-seq reads were processed as described previously (56), and deposited in NCBI (PRJNA665492).

Table 1. PCR primers used in this study.

Species	Product	Application	Primer 1*	Primer 2
Serratia marcescens	5S Probe	Northern blot	TAATACGACTCACTATAGGGCATCGGCGCTACGGCGTT	CCTGGCGGCAATAGCGCG
Edwardsiella hoshinae	5S Probe	Northern blot	TAATACGACTCACTATAGGGCTGGCAGTTCCCTACTCTCG	GATAGCGCGGTGGTCCCA
Vibrio harveyi	5S Probe	Northern blot	TAATACGACTCACTATAGGGTTCTACTCTCACATGGGGAAGC	CGACCATAGCGCTTTGGACC
Serratia marcescens	Peroxidase 3'-end probe	Northern blot	TAATACGACTCACTATAGGGTAGCCCGCCAAAGTTCACG	GCGACGATCTGGAAACCTTC
Edwardsiella hoshinae	Peroxidase 3'-end probe	Northern blot	TAATACGACTCACTATAGGGGTTCTGGACAGCTAAATTGCCA	GTAGCGAAGAATTAGAAGCT
Vibrio harveyi	Peroxidase 3'-end probe	Northern blot	TAATACGACTCACTATAGGGCCAATGTGTTGGCTTTCATGA	GAGAATTTGGAGCATTAC
Serratia marcescens	Peroxidase mRNA	In vitro transcription	TAATACGACTCACTATAGGGATGTTTACCAGTCAAGAAGGCA	AAAAAAAGGGCGGGCCGTAGCCCGCCAAAGTTCACG
Edwardsiella hoshinae	Peroxidase qPCR primers	qRT-PCR	ATGTTTGGCCGTCAAGAAGG	TCATCGCTGGTGATTTCGAC
Vibrio harveyi	Peroxidase qPCR primers	qRT-PCR	AGAAGGTCAAGCTGTACCACAAG	CGTGGTTACGTTCACCCATG

*T7 promoter sequence

In vitro transcription and RNase E digestion

In vitro transcription was performed using MAXIscript T7 Transcription Kit (Thermo Fischer Scientific) and 1 µg DNA as per manufacturer's instructions. Reactions were incubated overnight at 37°C, treated with TURBO DNase and purified using a Monarch RNA Cleanup Kit (New England Biolabs). RNA was loaded onto a 10% TBE Urea gel, and bands were excised and RNA was eluted and precipitated. Purified RNase E-NTD protein provided by Ben Luisi (University of Cambridge) and Boris Görke (University of Vienna) was used according to established protocols (Bandyra et al. 2018, Durica-Mitic and Gorke 2019). Briefly, 0.5 µg of *in vitro* transcribed RNA was heated for 1 minute at 95°C, chilled on ice for 3 minutes, and reaction buffer (25 mM Tris pH 7.5, 50 mM NaCl, 50 mM KCl, 10 mM MgCl2, 1 mM DTT, 0.5 U/µl RNaseOUT) was added to a total volume of 10 µl. Purified *Serratia marcescens* Hfq (ABclonal) (0.5uM diluted in reaction buffer) was added and incubated at 30°C for 10 minutes to allow Hfq-RNA complexes to form. 0.5uM RNase E was added to this mixture and incubated for 30 minutes at 30°C and 12.8 µl of 2x RNA loading dye was added directly to the reaction and incubated at 70°C for 10 minutes. 20 µl of each sample was loaded on a 10% TBE-Urea gel (Thermo Fischer Scientific) and run at 180V for 80-minutes in 1x TBE buffer and Northern blotting was performed as described above.

Northern blot

RNA samples were loaded onto either 6% or 10% TBE-Urea Gel (Thermo Fischer Scientific) with a biotinylated RNA ladder (Kerafast). Gels were run in 1x TBE buffer at 180V for 60 minutes (6% gels) or 180V for 80 minutes (10% gels). Membranes and filter

paper were pre-soaked and RNA was transferred to a Biodyne B Nylon Membrane (Thermo Fischer Scientific) overnight. Membranes were UV-crosslinked using a Stratalinker 2400 UV Crosslinker (1200 mj) and RNA probes were hybridized overnight at 45°C with rotation. Hybridization solution was removed and membranes were washed, blocked in Licor Intercept Blocking Buffer and treated with Streptavidin-IRDye 800 CW and examined on a Licor Odyssey scanner.

ACKNOWLEDGEMENTS

We thank Samantha Fancher for assistance with *S. marcescens* experiments, and Jim Archuleta for bioinformatics support. We are grateful to Boris Görke, Svetlana Durica, Katarzyna Bandyra and Ben Luisi for providing us with RNase E. This project was supported in part by NIH grants AI133023 and DE028409 to R.R.

SUPPLEMENTAL TABLES NOT PRINTED

The following tables are available as supplemental files for download in csv format. Information about these tables can be found in the Appendix (page 193).

Table S1. Predicted age of sRNAs.

Table S2. Predicted age of regulatory proteins**.**

Table S3. sRNAs that overlap ORFs.

Table S4. Accessions for Enterobacterales genomes used to determine sRNA age and prevalence.

REFERENCES

1. J. Hör, G. Matera, J. Vogel, S. Gottesman, G. Storz, Trans-acting small RNAs and their effects on gene expression in *Escherichia coli* and *Salmonella enterica*. *EcoSal Plus* **9** (2020).

2. E. Holmqvist, E. G. H. Wagner, Impact of bacterial sRNAs in stress responses. *Biochem. Soc. Trans.* **45**, 1203–1212 (2017).

3. J. A. Imlay, Transcription factors that defend bacteria against reactive oxygen species. *Annu. Rev. Microbiol.* **69**, 93–108 (2015).

4. J. A. Imlay, Cellular defenses against superoxide and hydrogen peroxide. *Annu. Rev. Biochem.* **77**, 755–776 (2008).

5. G. Storz, J. A. Imlay, Oxidative stress. *Curr. Opin. Microbiol.* **2**, 188–194 (1999).

6. M. Zheng, F. Aslund, G. Storz, Activation of the OxyR transcription factor by reversible disulfide bond formation. *Science* **279**, 1718–1721 (1998).

7. S. Altuvia, A. Zhang, L. Argaman, A. Tiwari, G. Storz, The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. *EMBO J.* **17**, 6069–6075 (1998).

8. N. De Lay, S. Gottesman, A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. *Mol Microbiol* **86**, 524–538 (2012).

9. B. González-Flecha, B. Demple, Role for the *oxyS* gene in regulation of intracellular hydrogen peroxide in *Escherichia coli*. *J. Bacteriol.* **181**, 3833–3836 (1999).

10. S. Altuvia, D. Weinstein-Fischer, A. Zhang, L. Postow, G. Storz, A small, stable RNA induced by oxidative stress: Role as a pleiotropic regulator and antimutator. *Cell* **90**, 43–53 (1997).

11. A. Zhang, *et al.*, The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. *EMBO J.* **17**, 6061–6068 (1998).

12. L. Argaman, S. Altuvia, *fhlA* repression by OxyS RNA: Kissing complex formation at two sites results in a stable antisense-target RNA complex. *J. Mol. Biol.* **300**, 1101–1112 (2000).

13. A. Zhang, K. M. Wassarman, J. Ortega, A. C. Steven, G. Storz, The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol. Cell* **9**, 11–22 (2002).

14. C. A. Santiviago, *et al.*, Analysis of pools of targeted *Salmonella* deletion mutants identifies novel genes affecting fitness during competitive infection in mice. *PLoS Pathogens* **5**, e1000477 (2009).

15. S. Barshishat, *et al.*, OxyS small RNA induces cell cycle arrest to allow DNA damage repair. *EMBO J.* **37**, 413–426 (2018).

16. M. Adeolu, S. Alnajar, S. Naushad, R.S. Gupta, Genome-based phylogeny and taxonomy of the 'Enterobacteriales': Proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. *Int J Syst Evol Microbiol*. **66**:5575-5599 (2016).

17. M. Miyakoshi, Y. Chao, J. Vogel, Regulatory small RNAs from the 3’ regions of bacterial mRNAs. *Curr. Opin. Microbiol.* **24**, 132–139 (2015).

18. Y. Chao, *et al.*, In vivo cleavage map illuminates the central role of RNase E in coding and non-coding RNA pathways. *Mol Cell* **65**, 39–51 (2017).

19. Y. Chao, J. Vogel, A 3’ UTR-derived small RNA provides the regulatory noncoding arm of the inner membrane stress response. *Mol. Cell* **61**, 352–363 (2016).

20. I. A. Iosub, *et al.*, Hfq CLASH uncovers sRNA-target interaction networks linked to nutrient availability adaptation. *Elife* **9**, (2020).

21. Y. Chao, K. Papenfort, R. Reinhardt, C. M. Sharma, J. Vogel, An atlas of Hfq-bound transcripts reveals 3’ UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* **31**, 4005–4019 (2012).

22. M. Hoyos, M. Huber, K.U. Förstner, K. Papenfort, Gene autoregulation by 3' UTR-derived bacterial small RNAs. *Elife* **9**, (2020).

23. H. A. Dutcher, R. Raghavan, Origin, evolution, and loss of bacterial small RNAs. *Microbiol Spectr* **6** (2018).

24. L. Barquist, S. W. Burge, P. P. Gardner, Studying RNA homology and conservation with Infernal: From single sequences to RNA families. *Curr Protoc Bioinformatics* **54**, 12.13.1-12.13.25 (2016).

25. S. Lindgreen, *et al.*, Robust identification of noncoding RNA from transcriptomes requires phylogenetically-informed sampling. *PLoS Comput. Biol.* **10**, e1003907 (2014).

26. F. R. Kacharia, J. A. Millar, R. Raghavan, Emergence of new sRNAs in enteric bacteria is associated with low expression and rapid evolution. *J. Mol. Evol.* **84**, 204–213 (2017).

27. T. B. Updegrove, S. A. Shabalina, G. Storz, How do base-pairing small RNAs evolve? *FEMS Microbiol Rev* **39**, 379–391 (2015).

28. R. Raghavan, F. R. Kacharia, J. A. Millar, C. D. Sislak, H. Ochman, Genome rearrangements can make and break small RNA genes. *Genome Biol Evol* **7**, 557–566 (2015).

29. S. Wachter, R. Raghavan, J. Wachter, M. F. Minnick, Identification of novel MITEs (miniature inverted-repeat transposable elements) in *Coxiella burnetii*: implications for protein and small RNA evolution. *BMC Genomics* **19** (2018).

30. S. Gottesman, G. Storz, Bacterial small RNA regulators: Versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* **3** (2011).

31. E. Skippington, M. A. Ragan, Evolutionary dynamics of small RNAs in 27 *Escherichia coli* and *Shigella* genomes. *Genome Biol Evol* **4**, 330–345 (2012).

32. R. Raghavan, E. A. Groisman, H. Ochman, Genome-wide detection of novel regulatory RNAs in *E. coli*. *Genome Res* **21**, 1487–1497 (2011).

33. M. B. Toledano, *et al.*, Redox-dependent shift of OxyR-DNA contacts along an extended DNA-binding site: A mechanism for differential promoter selection. *Cell* **78**, 897–909 (1994).

34. M. Zheng, *et al.*, Computation-directed identification of OxyR DNA binding sites in *Escherichia coli. J Bacteriol* **183**, 4571–4579 (2001).

35. C. Kröger, *et al.*, An infection-relevant transcriptomic compendium for *Salmonella enterica* serovar Typhimurium. *Cell Host Microbe* **14**, 683–695 (2013).

36. B. R. Jose, P. P. Gardner, L. Barquist, Transcriptional noise and exaptation as sources for bacterial sRNAs. *Biochem. Soc. Trans.* **47**, 527–539 (2019).

37. M. Huber, K. S. Fröhlich, J. Radmer, K. Papenfort, Switching fatty acid metabolism by an RNA-controlled feed forward loop. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8044–8054 (2020).

38. C. Wang, Y. Chao, G. Matera, Q. Gao, J. Vogel, The conserved 3’ UTR-derived small RNA NarS mediates mRNA crossregulation during nitrate respiration. *Nucleic Acids Res.* **48**, 2126–2143 (2020).

39. H. M. Kim, J.-H. Shin, Y.-B. Cho, J.-H. Roe, Inverse regulation of Fe- and Ni-containing SOD genes by a Fur family regulator Nur through small RNA processed from 3'UTR of the *sodF* mRNA. *Nucleic Acids Res.* **42**, 2003–2014 (2014).

40. A. Mira, H. Ochman, N. A. Moran, Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **17**, 589–596 (2001).

41. D. L. Williams *et al.*, Implications of high level pseudogene transcription in *Mycobacterium leprae*. *BMC Genomics* 10, 397 (2009).

42. L. S. Waters, G. Storz, Regulatory RNAs in bacteria. *Cell* **13**, 615-628 (2009).

43. A. M. Nuss, *et al.*, Transcriptomic profiling of *Yersinia pseudotuberculosis* reveals reprogramming of the Crp regulon by temperature and uncovers Crp as a master regulator of small RNAs. *PLoS Genet.* **11**, e1005087 (2015).

44. I. Kalvari, *et al.*, Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **46**, D335–D342 (2018).

45. E. P. Nawrocki, S. R. Eddy, Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).

46. A. Peer, H. Margalit, Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. *RNA* **20**, 994–1003 (2014).

47. O. Cohen, H. Ashkenazy, F. Belinky, D. Huchon, T. Pupko, GLOOME: Gain loss mapping engine. *Bioinformatics* **26**, 2914–2915 (2010).

48. D. Binns, *et al.*, QuickGO: A web-based tool for gene ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).

49. The UniProt Consortium, UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).

50. R. D. Finn, *et al.*, Pfam: The protein families database. *Nucleic Acids Res.* **42**, D222-230 (2014).

51. P. J. A. Cock, *et al.*, Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).

52. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).

53. T. L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28–36 (1994).

54. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

55. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

56. S. Wachter, L. D. Hicks, R. Raghavan, M. F. Minnick, Novel small RNAs expressed by *Bartonella bacilliformis* under multiple conditions reveal potential mechanisms for persistence in the sand fly vector and human host. *PLoS Negl Trop Dis* **14**, e0008671 (2020).

57. K. J. Bandyra, B. F. Luisi, RNase E and the high-fidelity orchestration of RNA metabolism. *Microbiol Spectr* **6** (2018).

58. S. Durica-Mitic, B. Görke, Feedback regulation of small RNA processing by the cleavage product. *RNA Biol* **16**, 1055–1065 (2019).

CHAPTER THREE

Environmental stress elicits a sRNA-based response in *Streptococcus mutans*

Madeline C Krieger[1], Justin Merritt[2], Rahul Raghavan[1,3]

[1]Department of Biology, Portland State University, Portland, OR, USA.
[2]Department of Restorative Dentistry, School of Dentistry, Oregon Health and Science University, Portland, OR, USA.
[3]Department of Biology, The University of Texas at San Antonio, San Antonio, TX, USA.

Keywords: small RNAs, *Streptococcus mutans*, 6S RNA, sugar-phosphate stress, sRNA transcriptome, sRNA function

ABSTRACT

Streptococcus mutans is a causative agent of dental caries, and its ability to utilize numerous sugars and to effectively respond to environmental stress promotes its proliferation in oral biofilms. Because of their quick action and low energetic cost, small non-coding RNAs (sRNAs) represent an ideal mode of regulation in stress response networks; yet their role in modulating gene expression in oral pathogens has remained largely unexplored. We identified 22 novel sRNAs in *S. mutans* that respond to four stress-inducing conditions commonly encountered by this pathogen in the human mouth: low pH, high temperature, hydrogen peroxide, and sugar-phosphate stress induced by xylitol, a common component of dental care products. Transient exposure to at least one of these stressors caused differential sRNA expression, indicating that the non-coding elements likely have regulatory functions. This study also explored the function of one novel sRNA, SmsR4, and showed that SmsR4 negatively regulates the EIIA component of the sorbitol phosphotransferase (PTS) system, which transports and phosphorylates the sugar alcohol sorbitol for use in glycolysis. The fine-tuning of EIIA availability by SmsR4 promotes *S. mutans* growth in media containing the sugar-alcohols sorbitol and xylitol. Our work lays a foundation for understanding the role of sRNAs in regulating gene expression in stress response networks in *S. mutans*. Additionally, our findings highlight the importance of the underexplored phenomenon of post-transcriptional gene regulation in oral species.

INTRODUCTION

The gram-positive oral pathogen *Streptococcus mutans* plays a principal role in the formation of dental caries and is often considered to be the primary causative agent of the disease (1–3). Central to *S. mutans*'s cariogenicity is its ability to ferment a wide variety of sugars, resulting in the formation of acidic microenvironments that drive the decline of commensal species and the proliferation of aciduric bacteria (4). One factor that allows *S. mutans* to metabolize a variety of carbon sources is the presence of 14 phosphotransferase systems (PTSs) (5, 6). PTSs transport carbohydrates across the cell membrane and immediately phosphorylates them for intracellular retention and entry into glycolysis (7). Sugar transport and metabolism are linked through phosphoenolpyruvate (PEP), which is both a glycolytic end product and the initial phosphate donor in the PTS phosphorylation chain.

Sorbitol (glucitol), a sugar-alcohol found naturally in many fruits, is a popular low-calorie sweetener. Relatively low amounts of acid are produced during sorbitol fermentation by *S. mutans*, thereby decreasing the rate of caries formation (8, 9). *S. mutans* transports sorbitol through a dedicated PTS that is under the control of a catabolic responsive element (CRE), which coordinates the use of carbon sources contingent on the availability of the preferred carbohydrate glucose (7, 9-11). Unlike sorbitol, the sugar-alcohol xylitol, a common component of dental care products, is not metabolized by *S. mutans*. Even small amounts of xylitol significantly inhibit *S. mutans*'s growth due to a futile cycle of phosphorylation and expulsion, which causes sugar-phosphate stress (10–13).

The success of *S. mutans* in the oral cavity is enhanced not only by proficient sugar utilization but also by its ability to quickly respond to rapidly changing environmental conditions (14). Small RNAs (sRNAs) are non-coding transcripts that typically base-pair with their target mRNAs to regulate gene expression (15). Due to their low energetic cost, fast action, and co-degradation with mRNA targets, sRNAs are an ideal mode of regulation when a bacterium is under stress (16). However, it is unknown how pervasive sRNA-based regulation is in *S. mutans* during adaptation to common stressors of the oral cavity, and there have been no functional analyses of sRNA utility in *S. mutans* to date. In this study, we identified several putative sRNAs in *S. mutans* that respond to multiple stress conditions and describe a novel sRNA that promotes bacterial growth in the presence of sorbitol.

RESULTS

Genome-wide identification of putative sRNAs in *S. mutans*

Using a previously developed RNA-seq-based approach (**Figure 1**) (17–19), we identified 22 putative sRNAs in *S. mutans* (**Figure 2**). A search against the Rfam database identified three of the sRNAs as RNase P, 6S RNA, and tmRNA (20); in addition, 13 of the sRNAs identified in this study overlap candidates predicted by a previous genome-wide scan (21) (**Table 1**). Based on their location relative to flanking genes, the novel sRNAs could be placed into one of five categories modified from previously defined groupings (22): (I) transcripts located in IGRs on the opposite strand of the two flanking genes, (II) sRNAs found in IGRs on the same strand as their flanking genes, (III) sRNAs on the opposite strand as their 5′ flanking gene but the same strand as their 3′ flanking gene, (IV) transcripts on the same strand as their 5′ flanking gene but the opposite strand as their 3′ flanking gene, and (V) sRNAs located antisense to a previously annotated gene (**Figure 2**). For all sRNAs we were able to identify putative -10 promoter sites, and 15 had predicted intrinsic terminators (**Table 1)** (23), supporting the legitimacy of these small transcriptions.

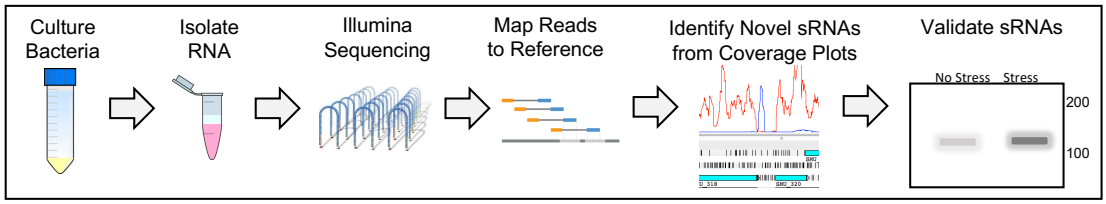

Figure 1. sRNA detection pipeline. Total RNA was isolated from bacterial cultures, then sequenced using Illumina NovaSeq following DNase treatment and rRNA depletion (not pictured). The RNA-seq reads were mapped to a reference genome and coverage plots of reads per nucleotide are visually examined to detected novel sRNA peaks. Putative sRNAs are generally identified from intergenic regions or the 5′ or 3′ ends of annotated protein coding genes. Novel sRNAs were validated using Northern blot.

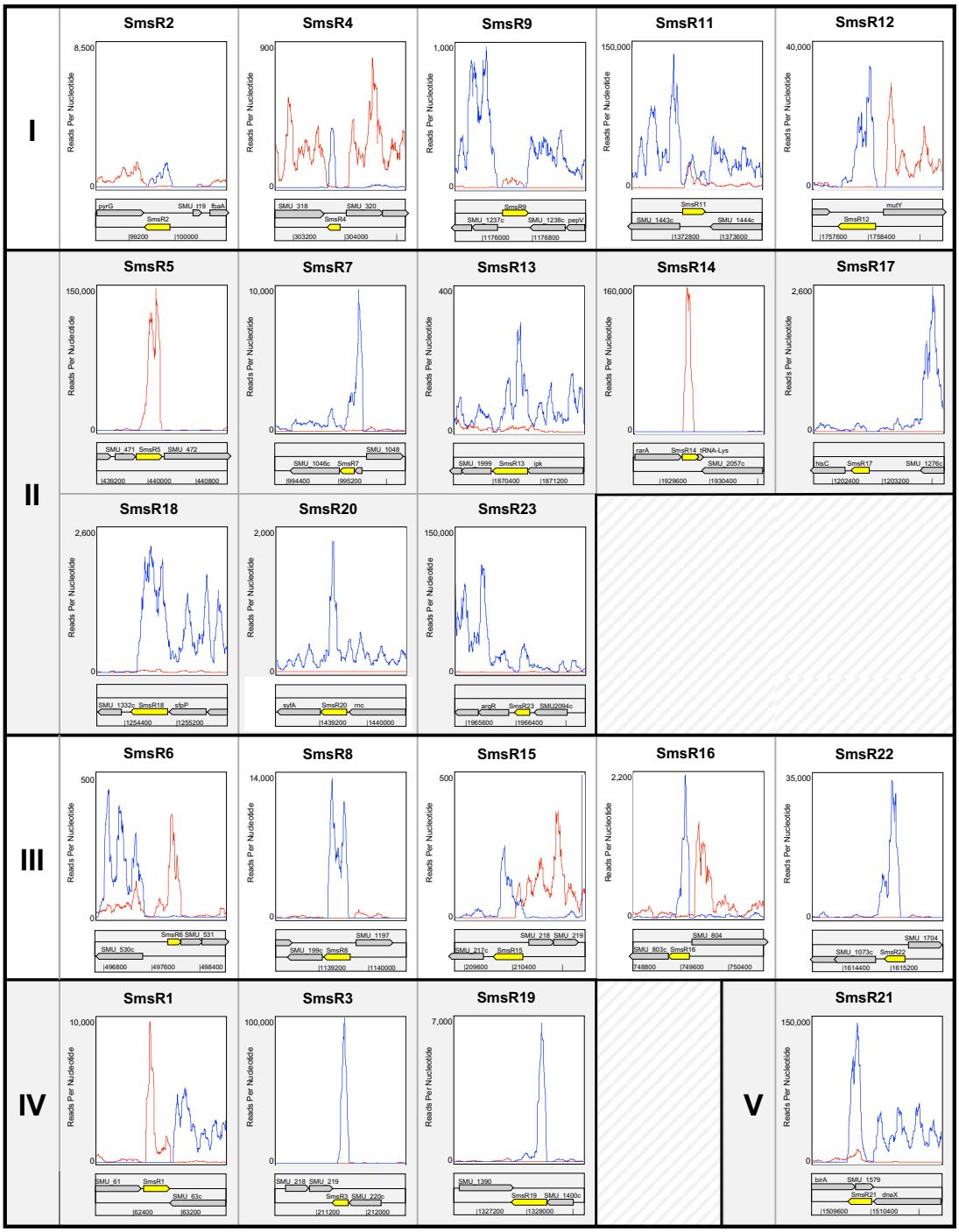

Figure 2. Expression profiles of novel sRNAs in *S. mutans*. RNA-seq reads mapped to forward (red) and reverse (blue) strands of the *S. mutans* genome are shown. The y-axes denote coverage at each nucleotide position and the locations of novel sRNA genes (yellow arrows) along with flanking genes (grey arrows) are shown below the coverage plots. sRNAs are classified based on their location relative to flanking genes: sRNAs located in (I) intergenic regions (IGRs) on the opposite strand of flanking genes, (II) IGRs on the same strand as both flanking gens, (II) sRNAs on the same strand as the 3′ flanking gene, (IV) sRNAs on the same strand as the 5′ flanking gene, and (V) those located antisense to an annotated gene.

Table 1. Novel sRNAs in *S. mutans*.

Name[a]	Rfam[b]	Previously Predicted[c]	Category[d]	Gene Left	Left Bound[e]	Gene Right	Right Bound[f]	Strand	Predicted Size (nt)	-10 Sequence[g]	Terminator Site[h]
SmsR1			4	SMU_61	62597	SMU_63c	63016	F	420	TATAAT	CTAACTAAAACAGAGACT CACATTACAATCACACGT GAATCTCTgTTTTTCTTA GCTG
SmsR2		psRNA-31	1	pyrG	99533	SMU_t19	99909	R	377	TAAACT	CAAAAAAAGAAACACCTT CTTAAATCTAGTAAATGA GATTTAAGAAGGTGTTTC TTTATAAGCA
SmsR3		psRNA-54	4	SMU_219	211522	SMU_220c	211709	R	188	TATAAT	ATAATTTTGACGGTTTAT CTTTTTTGATAAGCCaTT TTTTTATTTT
SmsR4		psRNA-62	1	SMU_318	303729	SMU_320	303848	R	120	TATACT	GTGTAAAAAAAGCCTTAG CTCTGCCAAGCTAGGGCT TTTCCGTTGCC
SmsR5	RNase P	psRNA-73	2	SMU_471	439813	SMU_472	440193	F	381	GATAAT	TTCTTAAAAAACCTTGCA GACTTAAATCTGCAAGGT TTTTTAATTCG
SmsR6		psRNA-78	3	SMU_530c	497858	SMU_531	498085	F	228	TATAAT	ATAATAGGTGAGCTAGCT TTGGCTAGCTTTTATTGT CTT
SmsR7			2	SMU_1046c	995289	SMU_1048	995586	R	298	TATAAT	AGATTTAACGCCCTCACA CAGATTTTCTGTGTGAGG TTTTTTGTTATC
SmsR8	tmRNA		3	SMU_1196c	1139336	SMU_1197	1139677	R	342	TATACT	No terminator found
SmsR9			1	SMU_1237c	1176296	SMU_1238c	1176661	F	366	TATAAT	TCTTCCAAGTAGCAGAAG CATTGATGTTTCTGCcaT TTTTAACACAG

SmsR11		psRNA-146	1	SMU_1443c	1373021	SMU1444c	1373194	F	174	TAAACT	No terminator found
SmsR12		psRNA-204 psRNA-205	1	SMU_1862	1757912	mutY	1758495	R	584	TAAAAT	TCCATAAAAATACCTCTGCCTTCTAGTTTACTAGATCTCAGCAGAGGTaTTTTTATATGTTC
SmsR13		psRNA-219 psRNA-220	2	SMU_1995c	1870476	ipk	1871026	R	551	TAAAGT	TGTGTGTCAGCGTCCTTGCAAGTTTTTTGCAAGGGCgcTTTTTTGAATAG
SmsR14	6S	psRNA-235	2	SMU_2056	1929950	SMU_2057c	1930143	F	194	TAACAT	No terminator found
SmsR15		psRNA-53	3	SMU_217c	210058	SMU_218	210532	R	475	TAAAAT	AATAAGCAAAAGACACTTGAAGCAATAATTCAAGTGTCTTTTATGGGACTT
SmsR16			3	SMU_803c	749489	SMU_804	749732	R	244	TAAACT	No terminator found
SmsR17		psRNA-123	2	hisC	1202700	SMU_1276c	1202906	R	207	TAAAAT	No terminators found
SmsR18		psRNA-132	2	SMU_1332c	1254531	sfp	1255035	R	505	TATAAT	TCCACAAAGGTCACTTGTCTATCTAGGACGAGTGaTTTTTCTTTGTAG
SmsR19			4	SMU_1398	1327789	SMU_1400c	1328337	R	549	TATAAT	TCTTTAGGCCTTCTTTCGATTTGTAAAAATTGGAGGAaTTTTTTTATGAA
SmsR20		psRNA-150	2	syfA	1439315	smc	1439585	R	271	TAAATT	AAATAATAAGAGACCCCCAACGATGAGCGTGTAGATTGTTGGGGGTCTTAATTGTATTGA
SmsR21			5	birA	1510015	dnaX	1510326	R	312	AATAAT	No terminator found
SmsR22			3	SMU_1703	1614919	SMU_1704	1615314	R	396	TATTTT	AAAAAGAAAAAGCCCCTCAGTGTGGGCTTTTTCTTTTTA
SmsR23			2	argR	1966415	SMU_2094c	1966590	R	176	TATAAT	No terminator found

[a]Small RNAs were named SmsR1-SmsR23 (Note: SmsR10 was originally annotated but removed from analysis).
[b]Existing Rfam entries for the putative sRNAs.
[c]Previously predicted sRNA transcripts that overlap with our results (21).
[d]sRNA category as defined in **Figure 2**.
[e, f]Left and right bounds were determined by RNA-seq.
[g]-10 promoter elements were determined from a manual scan of sites upstream of the RNA-seq determined transcription start site that contained at least 4/6 of the TATAAT promoter sequence, including an A at position two and a T at position six.
[h]Terminator elements were predicted by ARNold (23).

sRNAs respond to environmental stress

We evaluated the expression of the 22 novel sRNAs when *S. mutans* is exposed to four stress-inducing conditions that occur in human oral cavity: short-term exposures to low pH, high temperature, hydrogen peroxide (H_2O_2), and xylitol (**Table 2**). All sRNAs showed patterns of differential expression in at least one condition, indicating their likely utility in stress tolerance networks (**Figure 3, Figure 4**). Acid and sugar-phosphate stress induced the differential expression of 19 and 17 sRNAs, respectively, while six sRNAs were affected by oxidative stress and nine showed variance in expression under heat stress (**Figure 3**). Upregulation of sRNA expression was more common than downregulation across all conditions except for heat stress. In most cases, the size of transcripts detected via Northern blot corresponded to the estimated sRNA size from the RNA-seq analysis. However, in many cases larger transcripts were observed in addition to sRNAs, especially in transcripts with no predicted terminator (**Figure 4**). This observation suggests that these mature sRNAs were cleaved from larger mRNA, as reported in other bacteria (22, 24–27).

Table 2. Biologically relevant stress conditions tested.

Stress Type	Condition	Exposure Time
Sugar Phosphate	6% xylitol	30 minutes
Oxidative	1mM H_2O_2	15 minutes
Heat	45°C	15 minutes
Acid	pH 5	30 minutes

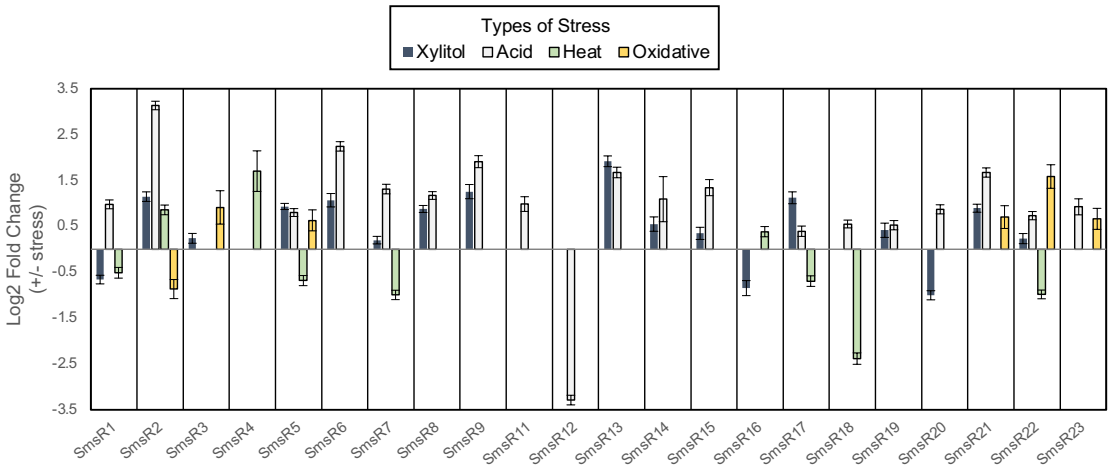

Figure 3. Differential expression of novel sRNAs in response to stress. Expression values for sRNAs were calculated from the total number of mapped RNA-seq reads. For each sRNA, Log2 fold change for three replicates as determined by DESeq2 (28) (+/- Log2 fold change standard error) are shown. Expression values that were significantly different ($p \leq 0.05$, Wald test) between stress and control conditions are shown.

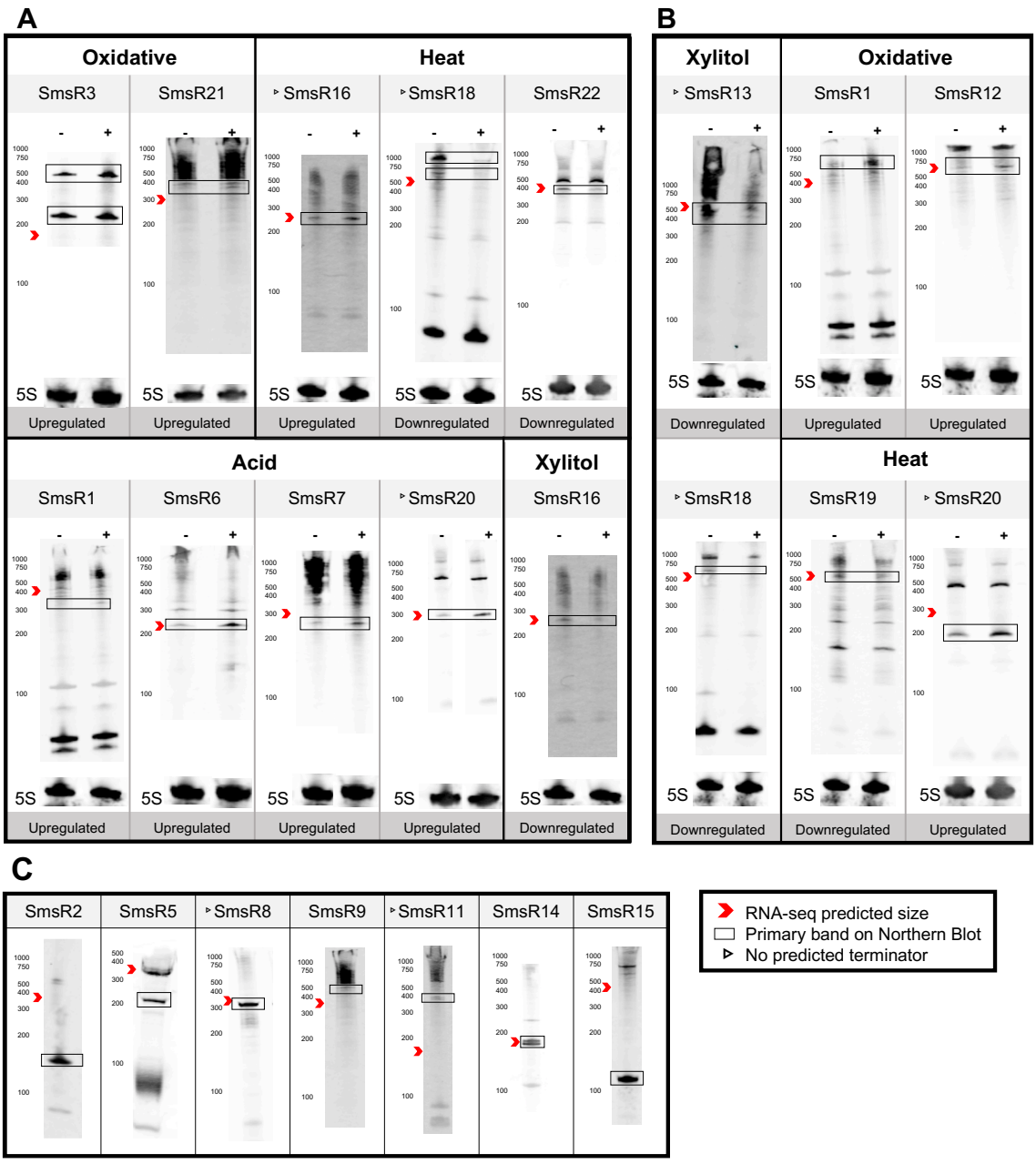

Figure 4. Validation of sRNA differential expression. **A)** sRNAs with differential expression in both Northern blot and RNA-seq data (**Figure 3**). Samples in lanes labeled "-" were not exposed to stress, while samples in "+" labeled lanes had stress induced. **B)** sRNAs that showed differential expression in Northern blot but not in RNA-seq. **C)** sRNAs that showed differential expression in RNA-seq but not in Northern blot. Transcripts for SmsR17 and SmsR23 were not visualized via Northern blot. A Northern blot of SmsR4 is shown in **Figure 10**. (Note: intensity between each blot is not comparable as exposure times differed between experiments).

6S occupies a conserved genome location in *Streptococcus*

6S RNA (SmsR14), a global regulator of transcription, was one of the most highly expressed sRNAs in our dataset (**Figure 5**) (29–31). We performed a covariance modeling (cm)-based search for 6S homologs across the genus *Streptococcus*, which showed that the sRNA is conserved in all genomes we analyzed (**Figure 6**). Our analysis also revealed that the location of 6S (between genes for RarA and tRNA-Lys) is widely conserved across all *Streptococcus* genomes. This gene arrangement is different than that in other gram-positive bacteria such as *Bacillus subtilis*, which encodes two copies of 6S, one in the intergenic region (IGR) between genes for a RecQ helicase and an azoreductase (AzoR) and another between two tRNA-associated genes (32). Although we identified only one 6S gene in all *Streptococcus* genomes, we detected two similar-sized 6S transcripts in *S. mutans* (**Figure 4**), suggesting that, as observed in *B. subtilis* (30), 6S RNA in *S. mutans* is processed by an RNase, resulting in the formation of a longer immature product in addition to the shorter, functional transcript. In contrast to gram-positive bacteria, the 6S genes in gram-negative bacteria such as *Escherichia coli* and *Salmonella enterica* are typically found upstream of the gene that encodes 5-formyl tetrahydrofolate cyclo-ligase (29, 30, 33). Interestingly, the putative sRNA, SmsR4, occupies the locus next to the 5-formyl tetrahydrofolate cyclo-ligase gene (SMU_320) in *S. mutans* (**Table 1, Figure 2**). SmsR4, however, is much shorter (110 nt) than 6S (194 nt), and their secondary structures differ considerably (30) (**Figure 7**), indicating that the two sRNAs are not homologous.

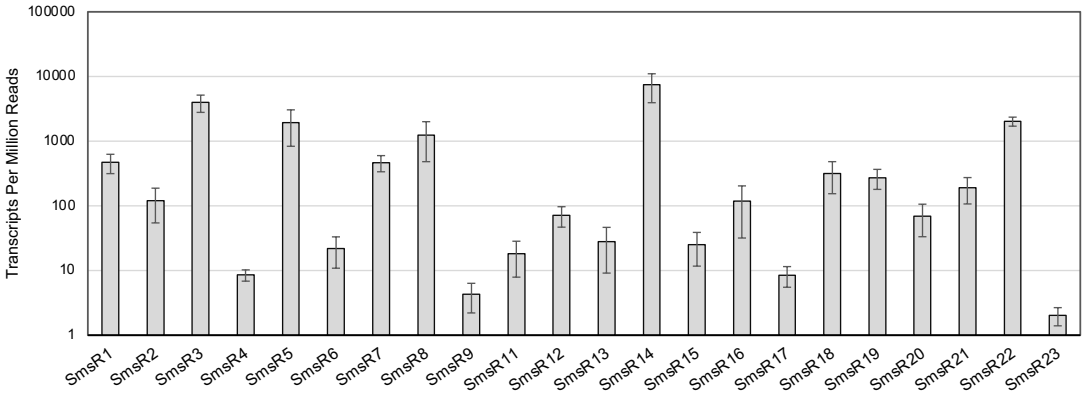

Figure 5. Expression of sRNAs during growth in Todd Hewitt (TH) broth. Out of the 22 novel sRNAs we discovered, SmsR14 (6S RNA) was the most abundant during growth in TH broth. Mean transcripts per million reads (+/- standard deviation) from three replicate experiments is shown.

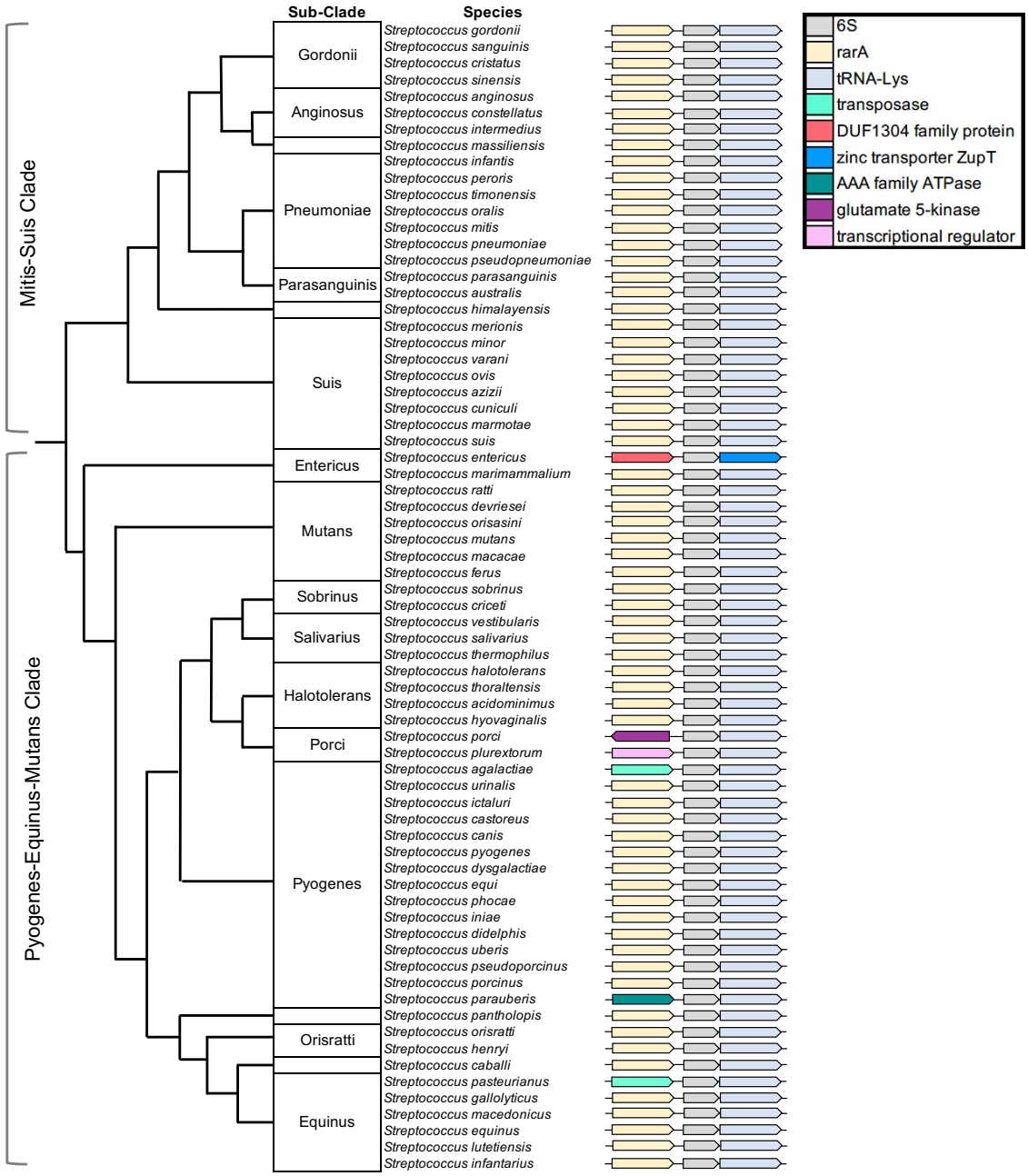

Figure 6. 6S RNA is conserved in *Streptococcus*. 6S RNA is present in all surveyed *Streptococcus* species and is located in the same genetic locus in most species. The whole genome-based phylogenetic tree is from Patel et al. 2018 (34). (Note: genes are not drawn to scale.)

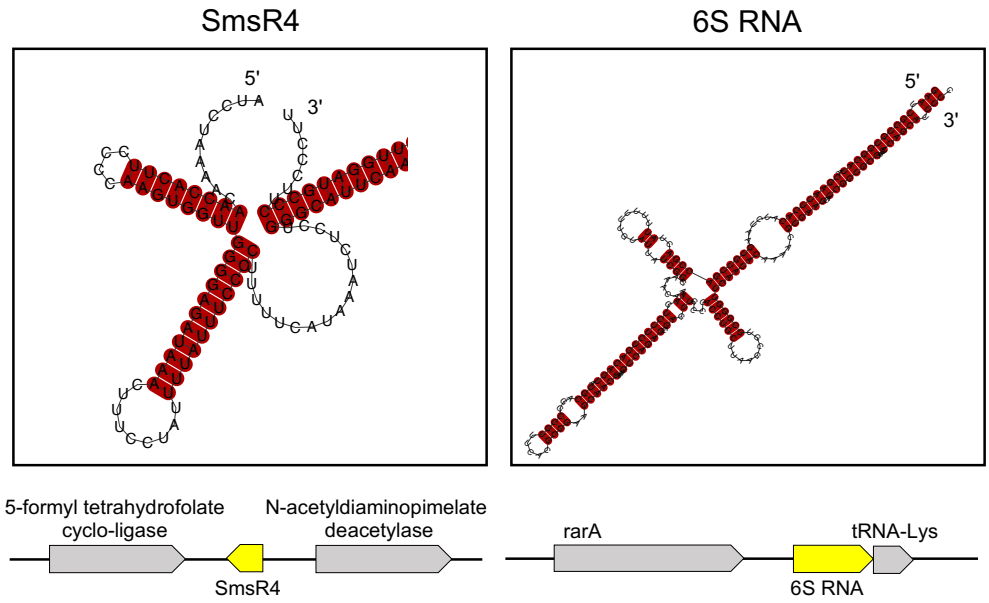

Figure 7. 6S RNA and SmsR4 have different predicted secondary structures. The consensus secondary structures of 6S and SmsR4 predicted using RNAalifold (35) are shown. Nucleotide pairs highlighted in red have perfect base-pairing. The genomic locations of both sRNAs (yellow arrows) and their flanking genes (grey arrows) are also depicted.

SmsR4 arose in the Pyogenes-Equinus-Mutans clade of *Streptococcus*

Based on whole-genome sequencing data, the genus *Streptococcus* could be classified into Mitis-Suis and Pyogenes-Equinus-Mutans clades (34). Our cm-based search identified SmsR4 homologs only in the Pyogenes-Equinus-Mutans clade, and an evolutionary reconstruction showed that SmsR4 arose at the root of this clade but was later lost in the common ancestor of Sobrinus, Salivarus, and Halotolerans and Porci subclades (**Figure 8**). While the rest of the Pyogenes subclade members contain SmsR4, it is absent in *Streptococcus equi*; similarly, in Entericus subclade*, Streptococcus marimmalium* has lost the sRNA, but it is retained by *Streptococcus entericus*. Despite these disparate cases of sRNA loss, the broad pattern of conservation of SmsR4 across a major clade of *Streptococcus* suggests that the novel sRNA has important functions.

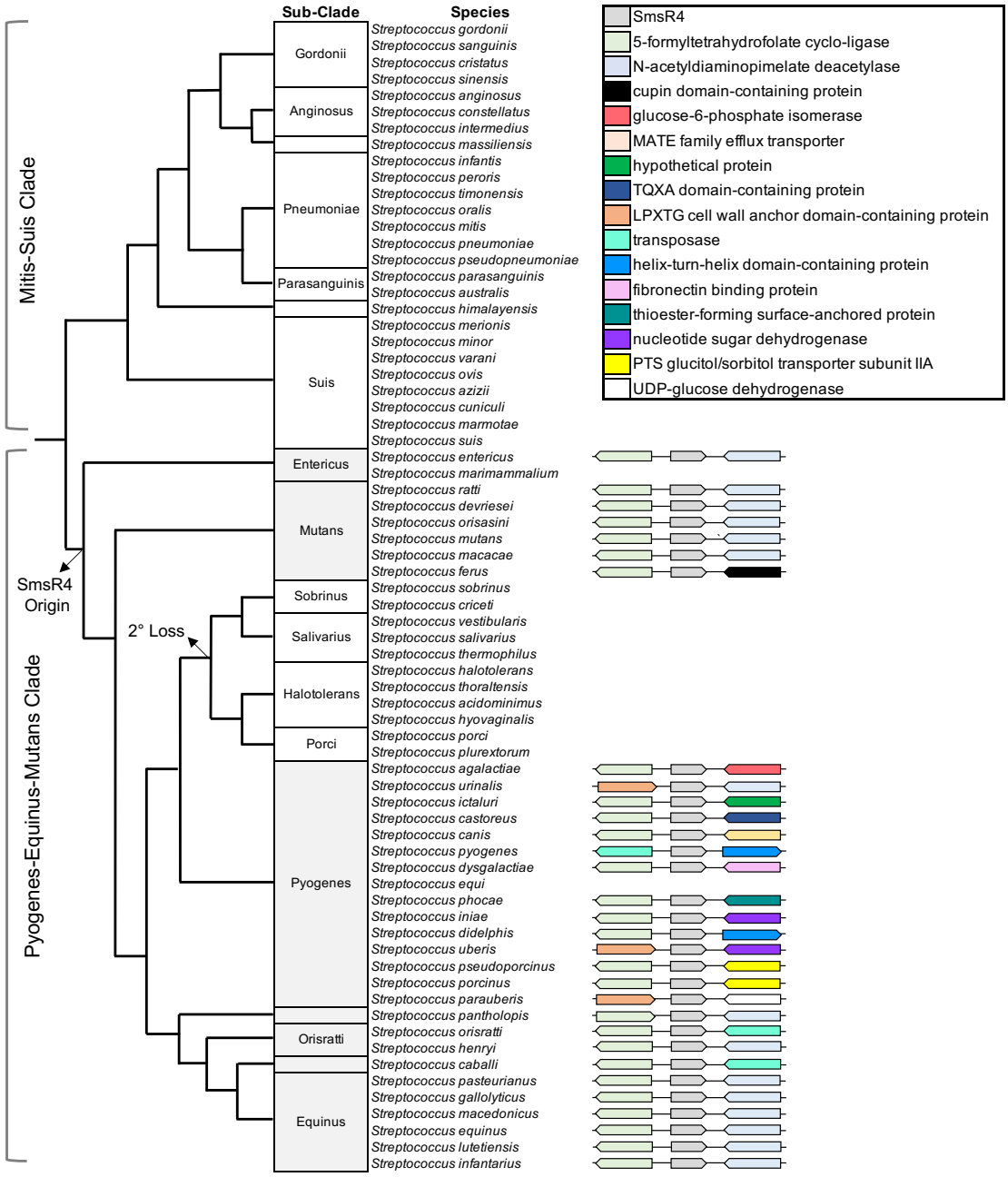

Figure 8. SmsR4 arose in the Pyogenes-Equinus-Mutans clade of *Streptococcus*. The prevalence pattern of SmsR4 indicate that the sRNA evolved at the base of Pyogenes-Equinus-Mutans clade but was later lost in the common ancestor of Sobrinus, Salivarus, Halotolerance and Porci sub-clades. In most species, SmsR4 is located in the same genetic locus. The whole genome-based phylogenetic tree is from Patel et al. 2018 (34). (Note: genes are not drawn to scale).

SmsR4 promotes *S. mutans* growth in sorbitol-containing media

To identify the function of SmsR4, we measured the growth of wild-type (WT) and SmsR4-deletion (DEL) strains of *S. mutans* using a phenotypic microarray (36). In this analysis, DEL showed reduced growth in comparison to WT in media containing sorbitol as the sole carbon source (**Figure 9**). We confirmed this phenotype by assaying bacterial growth in BTR medium that contained either 0.5% glucose (BTR-G) or 0.5% sorbitol (BTR-S). There was no meaningful difference in growth between WT and DEL in BTR-G with or without xylitol, but growth of DEL was substantially delayed in BTR-S (**Figure 10**). The growth defect of DEL was overcome by a complementation strain (COMP) in which SmsR4 was expressed from a plasmid. In fact, COMP grew better than WT, probably because multiple copies of the SmsR4-carrying plasmid are present in each cell (**Figure 10**). Similar to the improved growth of COMP in sorbitol, addition of the sugar-alcohol xylitol to BTR-G caused COMP to grow considerably better than WT or DEL. Finally, highest levels of SmsR4 expression were observed at time-points with maximal bacterial growth in all media (**Figure 10**). Taken together, our data indicate that SmsR4 is important for *S. mutans*'s growth in sorbitol, a sugar-alcohol that can be utilized in the absence of glucose, and may play a role when the bacterium is exposed xylitol, a non-metabolizable sugar-alcohol that induces sugar-phosphate stress.

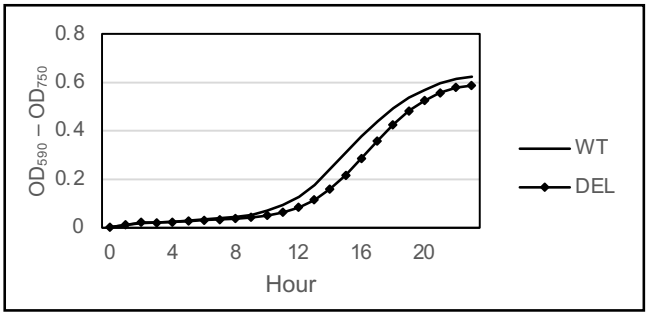

Figure 9. Phenotypic microarray. Growth of wild-type (WT) and SmsR4-deletion (DEL) strains in Biolog (36) media with sorbitol as sole carbon source is shown. Values represent the average of two independent growth experiments.

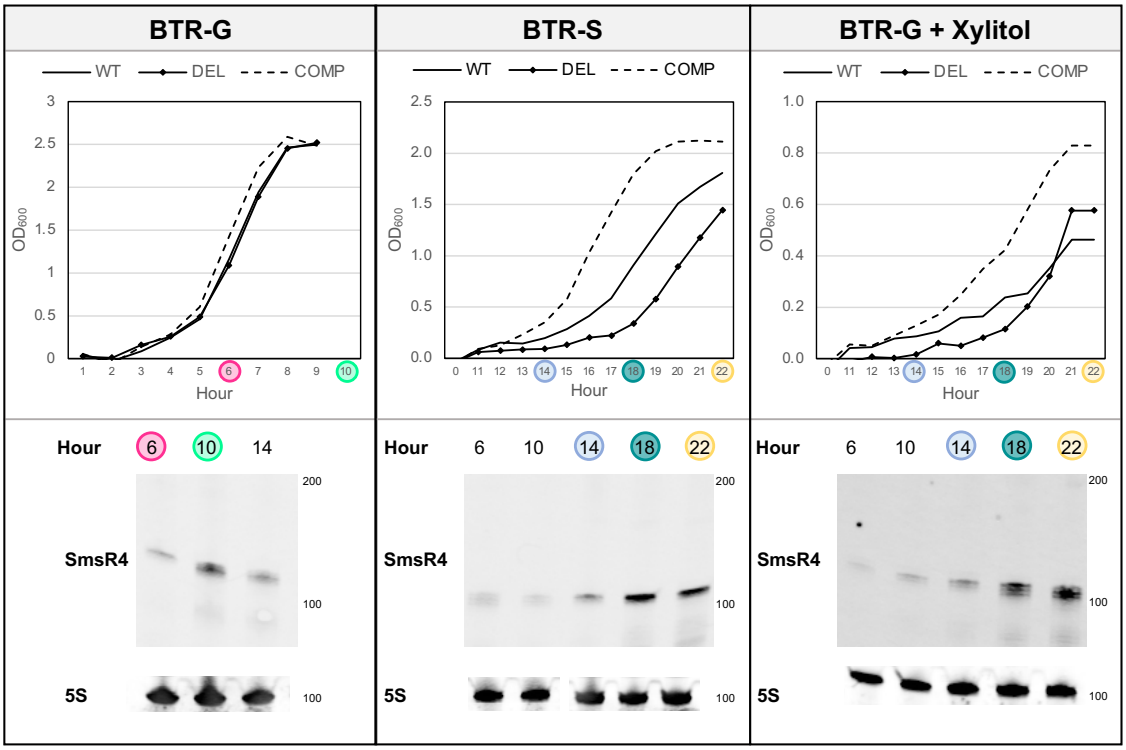

Figure 10. SmsR4 promotes *S. mutans* growth in sorbitol- and xylitol-containing media. Top panels: Growth of Wild-type (WT), SmsR4-deletion (DEL), and SmsR4 complement (COMP) strains in BTR medium with glucose (BTR-G), BTR media with sorbitol (BTR-S), or BTR-G containing xylitol. **Bottom panels**: Maximum expression of SmsR4 in WT occurred during early stationary phase in all three media. Growth time-points (hours) that are common to both panels have been highlighted for easy comparison. 5S RNA is shown as a loading control.

SmsR4 regulates sorbitol PTS

In accordance with SmsR4's potential role in regulating bacterial growth on sorbitol, *in-silico* sRNA target prediction indicated that SmsR4 could bind to SMU_313, the gene that encodes the sorbitol-specific enzyme IIA in the sorbitol PTS (**Figure 11A**). We confirmed this interaction with an electrophoretic mobility shift assay (EMSA), which showed that the binding of SmsR4 to the 5′ UTR of SMU_313 mRNA requires the putative "seed sequence" identified by the *in-silico* analyses (**Figure 12**). Further, a Crosslink-seq assay (37, 38) identified SMU_309, a transcriptional regulator located upstream of the sorbitol PTS operon as a potential target of SmsR4 (**Figure 11B, Figure 13**), and the expression of SMU_313 and other genes in the sorbitol PTS operon were significantly higher in DEL in comparison to WT in both BTR-S and BTR-G + xylitol (**Figure 14, Figure 15**). To test SmsR4's impact on sorbitol utilization, we measured levels of fructose-6-phosphate (F6P), a metabolite generated from sorbitol (**Figure 16)**, in WT and DEL grown anaerobically in BTR-S. As shown in **Figure 17**, significantly higher level of F6P was present in WT in comparison to DEL, indicating a potential role for SmsR4 in modulating sorbitol metabolism.

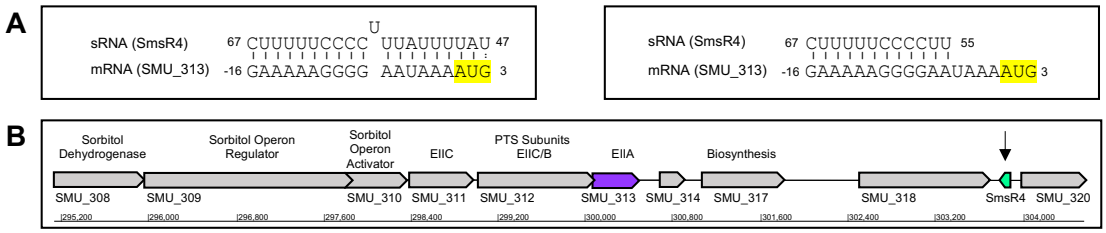

Figure 11. SmsR4 could bind to SMU_313. A) Two target prediction algorithms, IntaRNA (39) (left) and TargetRNA2 (40) (right), identified SMU_313 as a potential target of SmsR4. The predicted interaction sites on SmsR4 and 5′ untranslated region of SMU_313 are shown and the start codon (AUG) of SMU_313 has been highlighted. **B)** Genomic locations of sorbitol phosphotransferase (PTS) system operon, including SMU_313 (purple), and SmsR4 (green, arrow) in *S. mutans*.

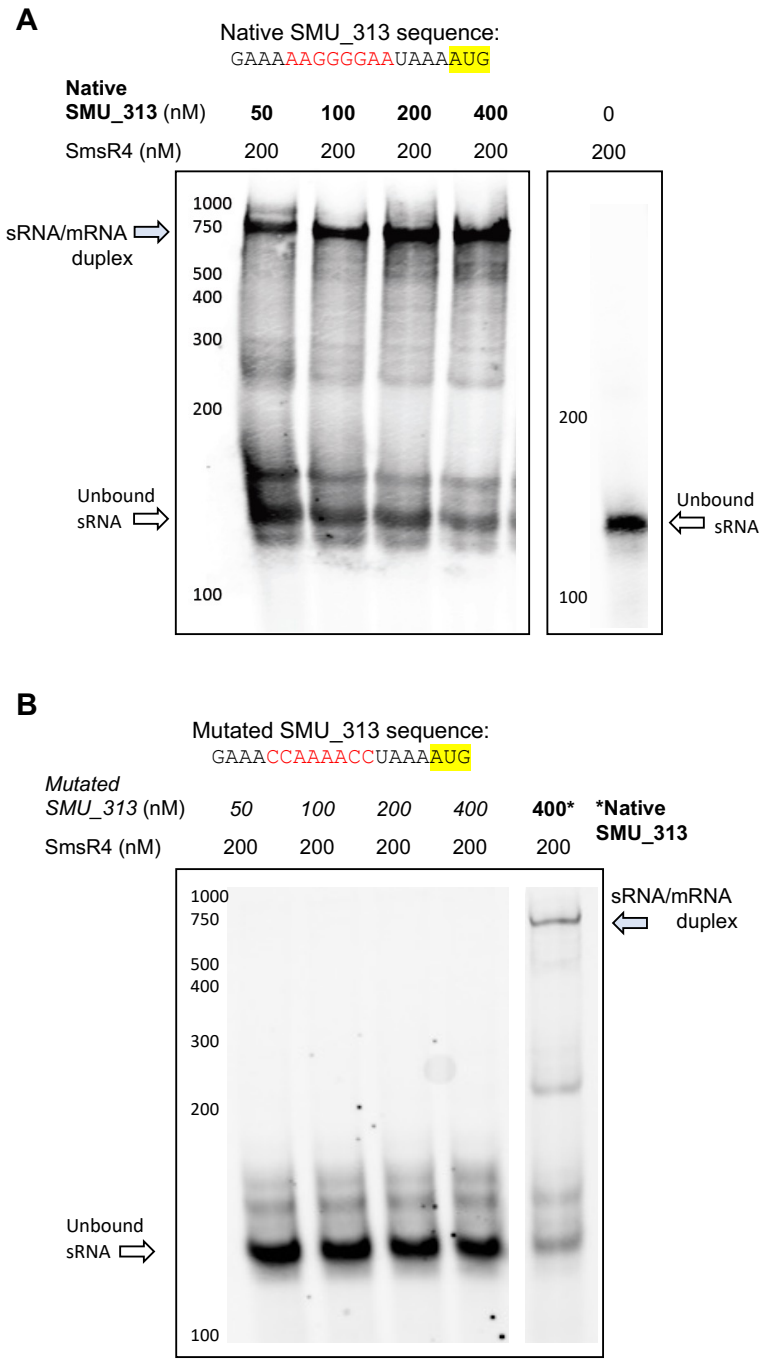

Figure 12. SmsR4 binds to SMU_313. **A)** Electrophoretic mobility shift assay (EMSA) shows that increasing concentrations of SMU_313 transcript with native 5′ untranslated region sequence bind well with SmsR4. An unbound SmsR4 transcript is shown to the right. **B)** The SMU_313 transcript with a mutated binding site does not interact with SmsR4. A control lane from the same gel with native SMU_313 transcript is shown. The native and mutated SmsR4-binding sites on SMU_313 are shown in red, and its start codon is highlighted in yellow.

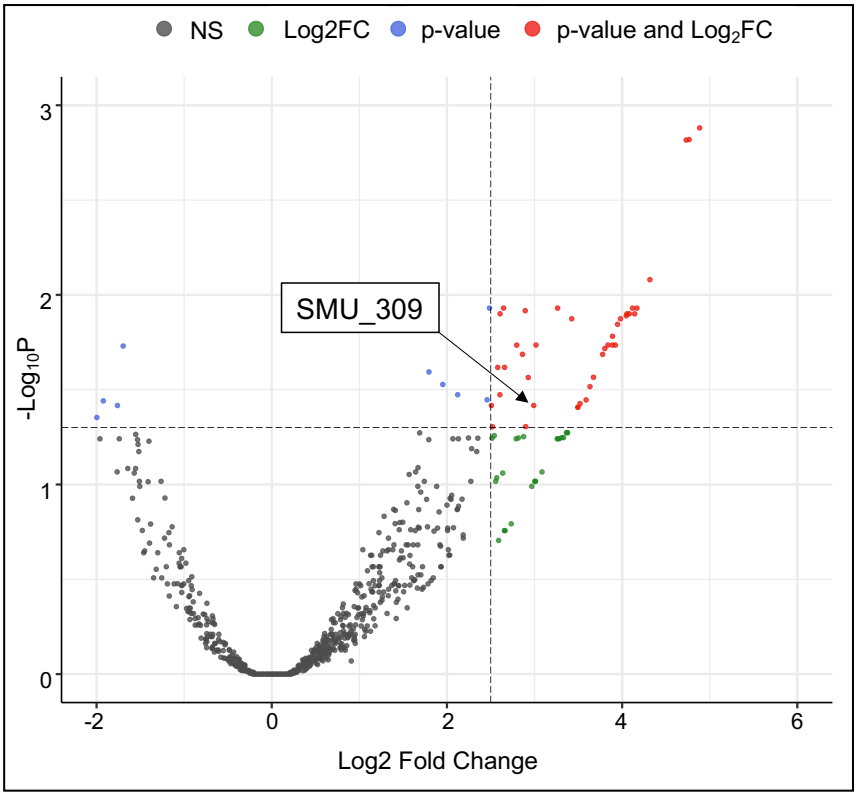

Figure 13. SMU_309 is a potential target of SmsR4. Targets were identified using Crosslink-seq, an RNA-RNA crosslinking assay that uses immunoprecipitation with sRNA-specific probes to pulldown sRNA targets which are then identified by RNA-seq. The DESeq2 package in R was used to analyze the Log2 fold change and p-value of genes enriched in the SmsR4-deletion (DEL) strain compared to the Wild-type (WT) (28). Genes identified via Crosslink-seq to be likely targeted by SmsR4 are shown in red ($p < 0.05$, Log2 fold change >2.5). Among the targets is SMU_309, a member of the sorbitol phosphotransferase system. Not significant (NS) genes are shown as grey dots, genes that only meet the criteria for Log2 fold change are in green, genes that have a p-value<0.05 (Wald test) but do not fulfil the Log2 fold change criteria are in blue, and red dots represent genes with both Log2 fold change >2.5 and $p<0.05$. The full list of potential targets is in **Table 3**.

Table 3. Complete list of crosslink-seq targets of SmsR4.

Gene[a]	Log2 Fold Change	Log2 Fold Change Standard Error	Adjusted p-value
SMU_2126c	4.8848	1.7267	0.0013
recN	4.7649	1.7330	0.0015
SMU_1521	4.7309	1.7343	0.0015
era	4.3169	1.7617	0.0083
SMU_923	4.1687	1.7904	0.0117
celR	4.1428	1.8307	0.0126
deoC	4.1186	1.7762	0.0117
tyrS	4.0818	1.7853	0.0126
SMU_702c	4.0546	1.7891	0.0126
SMU_1784c	4.0467	1.7989	0.0129
SMU_586	3.9809	1.7914	0.0134
dexB	3.9475	1.7942	0.0143
SMU_105	3.9223	1.8443	0.0184
efp	3.8901	1.7999	0.0165
trk	3.8889	1.8385	0.0184
bglP	3.8380	1.8054	0.0184
fabD	3.8026	1.8127	0.0192
psaB	3.7774	1.8197	0.0206
SMU_1623c	3.6725	1.8285	0.0272
dnaB	3.6328	1.8395	0.0305
SMU_770c	3.5884	1.8604	0.0358
SMU_718c	3.5197	1.8394	0.0374
acpP	3.4947	1.8521	0.0391
SMU_2158c	3.4933	1.8508	0.0391
SMU_1289c	3.4243	1.5300	0.0134
pgdA	3.2631	1.3715	0.0117
SMU_1303c	3.0169	1.4160	0.0184
accA	2.9905	1.5640	0.0383
SMU_309	2.9280	1.4510	0.0273
SMU_1449	2.8985	1.5711	0.0495
grpE	2.8938	1.2432	0.0121
pttB	2.8621	1.3668	0.0206
rl3	2.7969	1.3027	0.0184

trkB	2.6576	1.2868	0.0241
bccP	2.6462	1.1220	0.0117
SMU_1909c	2.6066	1.1364	0.0126
SMU_2024c	2.6046	1.3199	0.0336
SmsR7	2.5796	1.2499	0.0241
uvrA	2.5207	1.3618	0.0496
nusA	2.5077	1.3057	0.0383

[a]All results with a Log2 fold change of 2.5 of greater and $p<0.05$ (Wald test) as determined by DeSeq2 (28) are displayed.

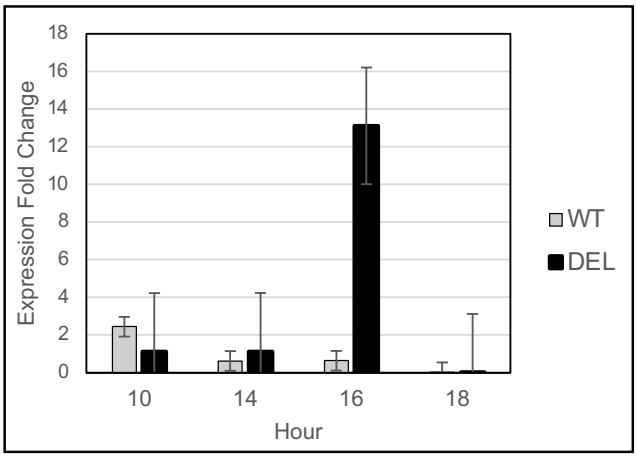

Figure 14. Effect of SmsR4 deletion on levels of SMU_313 during growth in sorbitol. Expression of SMU_313 was evaluated using qPCR. Levels of SMU_313 were elevated in SmsR4-deletion strain (DEL) in comparison to the Wild-type (WT) at 16 hours of growth in media containing sorbitol (BTR-S). Mean (+/- standard error) expression fold change from the average of two replicates are displayed.

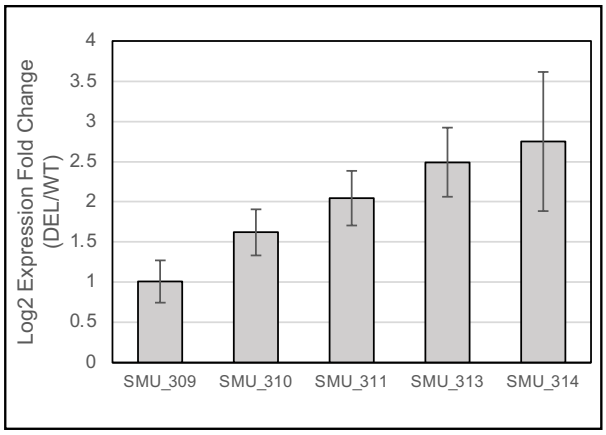

Figure 15. Effect of SmsR4 deletion on expression of sorbitol phosphotransferase (PTS) system components. Five genes of the sorbitol PTS, including SMU_313, are upregulated in the SmsR4-deletion strain (DEL) compared to the Wild-type (WT) at 19 hours of growth in glucose supplemented media (BTR-G) + xylitol. Log2 expression fold change values and mean (+/- standard error) of three RNA-seq replicates analyzed by DESeq2 (28) ($p<0.05$, Wald test) are shown.

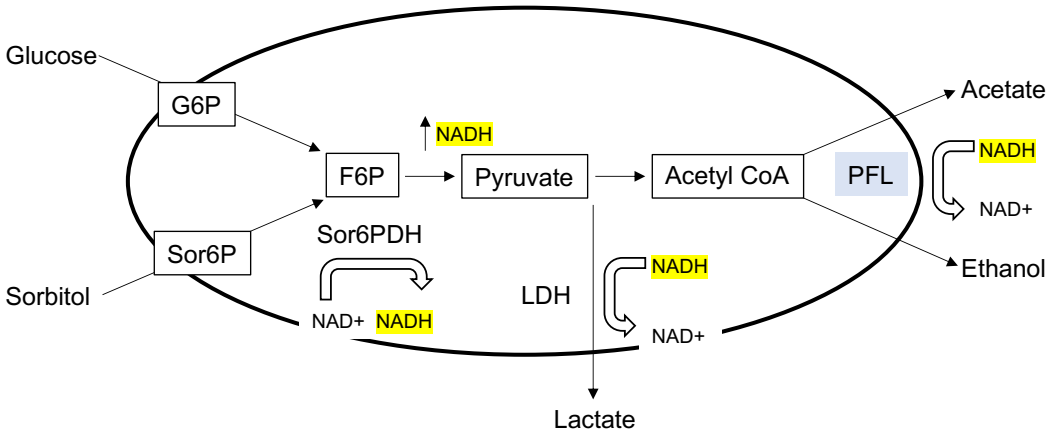

Figure 16. Comparison of sorbitol and glucose transport and metabolism during anaerobic growth. Glucose is phosphorylated to glucose-6-phosphate (G6P) before conversion to fructose-6-phosphate (F6P). Sorbitol-6-Phosphate (Sor6P) must undergo an additional processing step by sorbitol-6-phosphate dehydrogenase (Sor6PH) before entering glycolysis as F6P, generating an additional NADH. As glycolysis proceeds, the ratio of NADH/NAD+ increases. Lactate dehydrogenase (LDH) releases lactate and converts NADH to NAD+, and the oxygen-sensitive pyruvate formate lyase (PFL, blue) produces acetate and ethanol as byproducts while regenerating NAD+, thereby helping to restore the balance of NADH/NAD+.

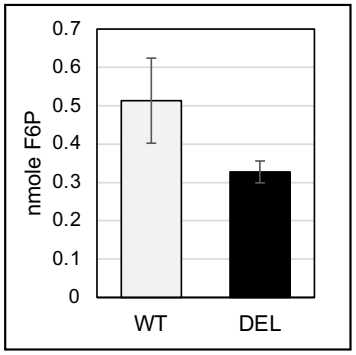

Figure 17. Fructose-6-phosphate measurement. Wild-type *S. mutans* (WT) contained higher amount of intracellular F6P than the SmsR4-deltion strain (DEL) at 16 hours of growth in media containing sorbitol (BTR-S). A standard curve was used to determined F6P concentration, and the mean (+/- standard deviation) of two experiments is displayed. Y-axis values are normalized to bacterial count.

DISCUSSION

sRNAs are critical to gene regulation in bacteria, but their roles in the dental pathogen *S. mutans* are largely unknown. In this study, we identified 22 putative sRNAs in *S. mutans* and show that most respond to environmental stressors commonly present in human oral cavity. A detailed investigation of one sRNA, SmsR4, showed that it likely regulates a sorbitol-specific PTS to promote *S. mutans*'s growth in sorbitol and during prolonged exposure to xylitol. Both sorbitol and xylitol are used in food and dental products to prevent tooth decay; hence, the discovery of SmsR4 is an important step in potentially developing new interventions that improve dental health by targeting the sRNA or other components in its regulatory network.

Prior to our study, 243 sRNA candidates were identified by a largely bioinformatics-based analysis of *S. mutans* grown with various carbon sources (21). In contrast, our more conservative method of sRNA detection that combined genome-scale approaches with experimental validation, uncovered only 22 novel sRNAs. This number matches earlier findings in bacteria with similarly sized genomes (~2 million bp) (17, 19, 22, 41, 42); however, it is highly likely that *S. mutans* contains many more sRNAs that are produced under conditions that were not tested in this study. Out of 22 predicted putative sRNAs we validated 20 of these via Northern blot and noticed that several sRNAs could be generated via RNase processing from larger transcripts as is typical of 3′ UTR-derived sRNAs (category IV in **Figure 2**) (43, 44). The genomic context of the remaining sRNAs also has implications for their functions; for instance, many of the 5′ UTR sRNAs (category III) could be riboswitch-like elements (45), and antisense sRNAs (category V) often regulate cis-encoded targets (46, 47). Because most sRNAs identified

in this study showed differential expression under stress and were validated via Northern blot, it is likely that they represent bona fide sRNAs that participate in regulatory networks. Further studies are required to elucidate the functions of these sRNAs, but preliminary *in silico* target prediction indicate that many of them are involved in processes critical to adaptation and virulence. For instance, SmsR11 likely targets a late competence protein (SMU_498), suggesting a potential function in regulating extracellular DNA uptake. Similarly, SmsR2, SmSR9, and SmR22 are all predicted to bind to PTS components, indicating that additional sRNAs may regulate sugar transport (**Table 4**), a process critical to *S. mutans*'s ability to cause dental caries.

Table 4. Targets predicted by IntaRNA for novel sRNAs.

sRNA	Target[a]	Gene	Annotation	mRNA Seed Region[b]	sRNA Seed Region[c]	P-value	FDR[d]	Hybr. Energy[e]
SmsR1	SMU_1701c			4-13	209-218	0.007	0.687	-16.33
	SMU_2049c		16S rRNA (uracil(1498)-N(3))-methyltransferase	10-19	215-224	0.010	0.687	-16.63
	SMU_1145c		histidine kinase	4-13	174-183	0.019	0.713	-16.38
	SMU_992			3-12	69-78	0.020	0.713	-12.85
	SMU_1471c			8-17	113-122	0.042	0.735	-18.49
	SMU_1311	asnS	asparaginyl-tRNA synthetase	6-15	132-141	0.049	0.735	-13.22
	SMU_1744	fabH	3-oxoacyl-ACP synthase III	11-20	265-274	0.050	0.735	-12.93
SmsR2	SMU_959c			1-10	103-112	0.001	0.448	-18.52
	SMU_154		30S ribosomal protein S15	4-13	116-125	0.005	0.743	-19.06
	SMU_2153c		peptidase	8-17	254-263	0.005	0.743	-18.6
	SMU_1700c		LrgB family protein	5-14	281-290	0.006	0.743	-16.02
	SMU_2165		SpoJ	5-14	185-194	0.013	0.775	-14.1
	SMU_102		PTS system transporter subunit IID	11-20	179-188	0.021	0.775	-14.3
	SMU_950		GTP-binding protein YsxC	11-20	187-196	0.021	0.775	-13.54
	SMU_1233	deoB	phosphopentomutase	9-18	111-120	0.021	0.775	-16.39
	SMU_1596	ptcC	cellobiose phosphotransferase system IIC component	5-14	202-211	0.022	0.775	-14.66

	SMU_827	rgpC	polysaccharide ABC transporter permease	7-16	180-189	0.022	0.775	-13.02
	SMU_1823	pncA	pyrazinamidase/nicotinamidase	10-19	188-197	0.023	0.775	-13.49
	SMU_1627	rl11	50S ribosomal protein L11	6-15	147-156	0.025	0.775	-12.98
	SMU_1178c		amino acid ABC transporter ATP-binding protein	8-17	308-317	0.026	0.775	-13.71
	SMU_1680c			5-14	185-194	0.028	0.775	-12.98
	SMU_685			6-15	161-170	0.035	0.775	-14.69
	SMU_506		type II restriction endonuclease	6-15	184-193	0.035	0.775	-13.9
	SMU_415			7-16	255-264	0.036	0.775	-14.89
	SMU_2067	csbB	stress response protein	5-14	90-99	0.037	0.775	-14.54
	SMU_1709	trkH	potassium uptake protein TrkH	5-14	299-308	0.038	0.775	-11.93
	SMU_1205c			1-10	260-269	0.042	0.775	-13.96
	SMU_104		alpha-glucosidase	2-11	189-198	0.044	0.775	-11.97
	SMU_793			9-18	189-198	0.044	0.775	-11.95
	SMU_1284c			5-14	197-206	0.045	0.775	-12.21
	SMU_863		ABC transporter ATP-binding protein	10-19	184-193	0.046	0.775	-12.76
	SMU_934		amino acid ABC transporter permease	6-15	161-170	0.047	0.775	-16.26
	SMU_902		ABC transporter ATP-binding protein	9-18	189-198	0.047	0.775	-12.57
	SMU_1073	fthS	formate--tetrahydrofolate ligase	5-14	299-308	0.049	0.775	-13.75
	SMU_880	msmG	multiple sugar-binding ABC transporter permease MsmG	1-10	216-225	0.049	0.775	-12.83

	SMU_341		deoxyribonuclease	5-14	217-226	0.049	0.775	-14.3
SmsR3	SMU_1956c			1-10	44-53	0.010	0.573	-18.94
	SMU_1450		amino acid permease	10-19	35-44	0.010	0.573	-18.96
	SMU_2108c		transcriptional regulator	8-17	28-37	0.017	0.573	-16.38
	SMU_2061			9-18	47-56	0.024	0.573	-15.77
	SMU_1768c			10-19	159-168	0.025	0.573	-17.18
	SMU_1593c		CDP-diglyceride synthetase	8-17	57-66	0.028	0.573	-15.36
	SMU_75		D-alanyl-D-alanine carboxypeptidase	10-19	46-55	0.033	0.573	-15.36
	SMU_207c		transposon protein	10-19	129-138	0.036	0.573	-15.94
SmsR5	SMU_02	dnaN	DNA polymerase III subunit beta	11-20	21-30	0.002	0.427	-19.65
	SMU_1474c		ribonuclease Z	5-14	40-49	0.010	0.692	-16.84
	SMU_292		transcriptional regulator	9-18	130-139	0.011	0.692	-19.12
	SMU_418	nusA	transcription elongation factor NusA	9-18	353-362	0.025	0.840	-14.9
	SMU_336	rnpA	ribonuclease P	2-11	348-357	0.027	0.840	-15.56
	SMU_84	truA	tRNA pseudouridine synthase A	6-15	26-35	0.032	0.840	-15.13
	SMU_1102	ascB	6-phospho-beta-glucosidase	11-20	21-30	0.045	0.840	-14.95
	SMU_346		NADH dehydrogenase NAD(P)H nitroreductase	8-17	24-33	0.047	0.840	-14.97
SmsR6	SMU_1700c		LrgB family protein	1-10	67-76	0.007	0.493	-13.84
	SMU_1460	rmlC	dTDP-4-keto-L-rhamnose reductase	11-20	198-207	0.015	0.698	-17.44

	SMU_953c		transcriptional regulator/aminotransferase	1-10	22-31	0.031	0.698	-13.92
	SMU_1286c		permease	6-15	22-31	0.033	0.698	-14.24
	SMU_344			3-12	65-74	0.035	0.698	-12.21
	SMU_671	citZ	citrate synthase	2-11	214-223	0.041	0.698	-11.97
	SMU_795		esterase	4-13	205-214	0.044	0.698	-14.45
	SMU_94c		transposase fragment	10-19	199-208	0.045	0.698	-16.13
	SMU_2085	recA	recombinase A	7-16	78-87	0.050	0.698	-12.33
SmsR7	SMU_1820c		aspartyl/glutamyl-tRNA amidotransferase subunit A	3-12	101-110	0.004	0.400	-19.98
	SMU_759		protease	6-15	65-74	0.012	0.593	-16.43
	SMU_509			11-20	93-102	0.021	0.687	-17.97
	SMU_1531	atpE	ATP synthase F0F1 subunit delta	3-12	101-110	0.030	0.718	-17.67
SmsR8	SMU_874		bifunctional homocysteine S-methyltransferase/510-methylenetetrahydrofolate reductase	10-19	308-317	0.003	0.496	-27.25
	SMU_495	gldA	glycerol dehydrogenase	5-14	19-28	0.010	0.639	-17.82
	SMU_906		ABC transporter ATP-binding protein	5-14	19-28	0.010	0.639	-16.5
	SMU_1163c		ABC transporter ATP-binding protein	2-11	189-198	0.026	0.846	-17.87
	SMU_1659c			3-12	21-30	0.026	0.846	-16.47
	SMU_1149		transporter trans-membrane domain bacteriocin immunity protein	5-14	16-25	0.028	0.846	-15.42

	SMU_758c			5-14	313-322	0.035	0.846	-18.22
	SMU_1585c		transcriptional regulator	5-14	19-28	0.036	0.846	-14.3
	SMU_1858	rs18	30S ribosomal protein S18	9-18	238-247	0.046	0.851	-11.92
SmsR9	SMU_475			6-15	337-346	0.005	0.591	-18.19
	SMU_1956c			2-11	167-176	0.009	0.591	-16.3
	SMU_55			8-17	176-185	0.013	0.591	-16.27
	SMU_1997	comX1	ComX1 transcriptional regulator of competence-specific genes	6-15	29-38	0.015	0.591	-16.88
	SMU_103		PTS system transporter subunit IIA	8-17	176-185	0.019	0.591	-15.57
	SMU_2064c		transmembrane protein	5-14	21-30	0.020	0.591	-17.02
	SMU_1114	gyrA	DNA gyrase subunit A	1-10	146-155	0.033	0.790	-16.21
	SMU_1587c			11-20	15-24	0.037	0.790	-15.32
	SMU_553	ylmE		1-10	35-44	0.045	0.820	-15.08
	SMU_471			4-13	339-348	0.047	0.820	-15.64
SmsR11	SMU_1434c		glycosyltransferase	10-19	261-270	0.001	0.359	-19.23
	SMU_885	galR	LacI family transcriptional repressor	8-17	76-85	0.003	0.470	-18.96
	SMU_247		ABC transporter ATP-binding protein	9-18	69-78	0.007	0.779	-23.68
	SMU_1768c			2-11	341-350	0.011	0.837	-16.47
	SMU_1801c		GTPase YqeH	1-10	356-365	0.020	0.837	-14.86
	SMU_415			6-15	355-364	0.022	0.837	-14.8

	SMU_104		alpha-glucosidase	8-17	257-266	0.022	0.837	-17.74
	SMU_1541	pulA	pullulanase	10-19	255-264	0.024	0.837	-16.36
	SMU_1727		OxaA-like protein precursor	3-12	19-28	0.029	0.837	-16.22
	SMU_1024c		transposase fragment	3-12	277-286	0.039	0.837	-15.21
	SMU_498	comF	late competence protein	8-17	257-266	0.041	0.837	-15.12
	SMU_641		oxidoreductase	2-11	90-99	0.042	0.837	-14.8
	SMU_2048			7-16	153-162	0.043	0.837	-14.38
	SMU_1236c			1-10	157-166	0.045	0.837	-14.57
	SMU_1965c		histidine kinase	1-10	74-83	0.046	0.837	-14.81
	SMU_730			5-14	17-26	0.048	0.837	-16.13
SmsR12	SMU_1905c		bacteriocin secretion protein	6-15	22-31	0.003	0.508	-16.43
	SMU_1602		NAD(P)H-flavin oxidoreductase	7-16	355-364	0.003	0.508	-15.28
	SMU_1623c			2-11	26-35	0.007	0.580	-16.53
	SMU_656	mutE2	protein MutE	5-14	514-523	0.008	0.580	-14.2
	SMU_1647c		transcriptional regulator	1-10	408-417	0.011	0.658	-19.96
	SMU_38c		transcriptional regulator	8-17	20-29	0.014	0.708	-14.31
	SMU_1496	lacA	galactose-6-phosphate isomerase subunit LacA	6-15	356-365	0.019	0.788	-13.63
	SMU_1818c			10-19	45-54	0.027	0.858	-11.44
	SMU_526c		transcriptional regulator	8-17	175-184	0.032	0.858	-11.54

	SMU_1454c			3-12	245-254	0.033	0.858	-12.32
	SMU_1498	lacR	lactose repressor	8-17	494-503	0.033	0.858	-12.09
	SMU_1314			2-11	237-246	0.035	0.858	-13.72
	SMU_02	dnaN	DNA polymerase III subunit beta	8-17	270-279	0.042	0.897	-12.64
	SMU_334		argininosuccinate synthase	1-10	484-493	0.048	0.897	-12.87
	SMU_1373c			1-10	310-319	0.049	0.897	-11.77
SmsR13	SMU_1588c		hexosyltransferase	1-10	368-377	0.001	0.205	-17.93
	SMU_870		sugar metabolism transcriptional regulator	11-20	123-132	0.004	0.664	-14.99
	SMU_1042			8-17	176-185	0.005	0.664	-14.76
	SMU_1810	scnE	bacteriocin operon component ScnE-like protein	7-16	226-235	0.015	0.730	-15.5
	SMU_1734	accA	acetyl-CoA carboxylase carboxyl transferase subunit alpha	1-10	225-234	0.017	0.730	-14.84
	SMU_233	ilvC	ketol-acid reductoisomerase	9-18	27-36	0.019	0.730	-14.85
	SMU_414		ABC transporter permease	3-12	531-540	0.021	0.730	-13.75
	SMU_606			7-16	156-165	0.022	0.730	-15.16
	SMU_550	ftsQ	cell division protein FtsQ	2-11	262-271	0.028	0.730	-15.34
	SMU_697		translation initiation factor IF-3	8-17	118-127	0.028	0.730	-14.23
	SMU_1506c			10-19	355-364	0.030	0.730	-15.34
	SMU_898			2-11	225-234	0.031	0.730	-14.7
	SMU_1357		transposase fragment	4-13	540-549	0.034	0.730	-12.35

	SMU_1709	trkH	potassium uptake protein TrkH	4-13	351-360	0.035	0.730	-14.17
	SMU_985	bglA	beta-glucosidase	9-18	7-16	0.035	0.730	-12.16
	SMU_1363c		transposase	10-19	524-533	0.036	0.730	-12.81
	SMU_1105c		phosphoglycerate mutase	6-15	330-339	0.037	0.730	-17.29
	SMU_863		ABC transporter ATP-binding protein	9-18	154-163	0.039	0.730	-13.59
	SMU_1515	covX		5-14	88-97	0.040	0.730	-14.05
	SMU_624		1-acylglycerol-3-phosphate O-acyltransferase	7-16	378-387	0.042	0.730	-11.81
	SMU_538	trpA	tryptophan synthase subunit alpha	4-13	89-98	0.042	0.730	-13.95
	SMU_1708	trkA	potassium transporter peripheral membrane protein	4-13	89-98	0.043	0.730	-14.22
	SMU_389			10-19	3-12	0.044	0.730	-12.13
	SMU_482	sunL	16S rRNA (cytosine(967)-C(5))-methyltransferase	5-14	415-424	0.046	0.730	-14.09
	SMU_1153c			5-14	58-67	0.046	0.730	-13.2
	SMU_704c		autolysin amidase	7-16	537-546	0.048	0.734	-13.18
SmsR14	SMU_1587c			3-12	182-191	0.013	0.742	-18.1
	SMU_981	bglB1	BglB fragment	8-17	177-186	0.027	0.742	-16.62
	SMU_244	bacA	undecaprenyl pyrophosphate phosphatase	8-17	177-186	0.032	0.742	-15.46
	SMU_925			10-19	175-184	0.033	0.742	-16.56
	SMU_457			3-12	63-72	0.035	0.742	-17.2
	SMU_813		transcriptional regulator	3-12	182-191	0.038	0.742	-14.46

	SMU_2067	csbB	stress response protein	3-12	7-16	0.038	0.742	-14.67
	SMU_493	pfl2	formate acetyltransferase	3-12	182-191	0.039	0.742	-15.16
	SMU_1712c		segregation and condensation protein B	1-10	180-189	0.039	0.742	-18.14
	SMU_1201c			10-19	31-40	0.043	0.742	-14.07
	SMU_2075c			4-13	181-190	0.043	0.742	-15.26
	SMU_1176		cation efflux transporter	7-16	3-12	0.044	0.742	-14.15
	SMU_487		response regulator	7-16	178-187	0.046	0.742	-14.33
	SMU_1196c			4-13	181-190	0.047	0.742	-13.74
	SMU_336	rnpA	ribonuclease P	4-13	152-161	0.049	0.742	-13.41
	SMU_738			1-10	179-188	0.050	0.742	-14.85
SmsR15	SMU_1003	gid	tRNA (uracil-5-)-methyltransferase Gid	3-12	309-318	0.003	0.545	-14.79
	SMU_889	pbpX	penicillin-binding protein class C fmt-like protein	5-14	423-432	0.008	0.545	-16.73
	SMU_1226c			1-10	322-331	0.009	0.545	-13.7
	SMU_1427c			2-11	266-275	0.012	0.545	-18.1
	SMU_352		ribulose-phosphate 3-epimerase	9-18	395-404	0.014	0.545	-11.95
	SMU_1065c		GntR family transcriptional regulator	9-18	267-276	0.019	0.545	-18.01
	SMU_1013c		Mg2+/citrate transporter	9-18	395-404	0.019	0.545	-13.51
	SMU_339			9-18	395-404	0.020	0.545	-12.02
	SMU_93c		transposase fragment	6-15	14-23	0.022	0.545	-12.59

	SMU_1918	dedA	membrane-associated protein DedA	10-19	456-465	0.032	0.723	-11.58
	SMU_562	clpE	ATP-dependent protease ClpE	2-11	397-406	0.041	0.743	-11.56
	SMU_1946			10-19	322-331	0.041	0.743	-11.5
	SMU_529			8-17	268-277	0.044	0.743	-18.6
	SMU_317		2345-tetrahydropyridine-26-dicarboxylate N-acetyltransferase	11-20	380-389	0.046	0.743	-13.04
SmsR16	SMU_865		30S ribosomal protein S16	10-19	185-194	0.011	0.740	-16.94
	SMU_233	ilvC	ketol-acid reductoisomerase	11-20	184-193	0.013	0.740	-16.47
	SMU_2104a		50S ribosomal protein L32	10-19	185-194	0.018	0.740	-15.51
	SMU_777	aroD	3-dehydroquinase	5-14	53-62	0.028	0.740	-16.26
	SMU_1743	acp	acyl carrier protein	10-19	185-194	0.029	0.740	-14.33
	SMU_627			5-14	202-211	0.033	0.740	-14.78
	SMU_1040c		oxidoreductase	8-17	187-196	0.039	0.740	-13.84
	SMU_1443c		tributyrin esterase	6-15	189-198	0.039	0.740	-14.52
	SMU_888	galE	UDP-galactose 4-epimerase GalE	3-12	67-76	0.043	0.740	-14.3
	SMU_1624	rrf1	ribosome recycling factor	7-16	188-197	0.046	0.740	-13.32
SmsR17	SMU_2056		recombination factor protein RarA	10-19	88-97	0.003	0.227	-17.51
	SMU_497c			5-14	93-102	0.019	0.676	-13.08
	SMU_618			11-20	87-96	0.041	0.797	-13.28
SmsR18	SMU_696		cytidylate kinase	4-13	308-317	0.001	0.201	-21.76

	SMU_1223	pyrDB	dihydroorotate dehydrogenase 1B	4-13	308-317	0.001	0.201	-21.37
	SMU_1312	aspB	aspartate aminotransferase	5-14	304-313	0.003	0.366	-20.26
	SMU_1404c			11-20	311-320	0.004	0.366	-19.92
	SMU_803c		ABC transporter ATP-binding protein	1-10	401-410	0.005	0.435	-19.88
	SMU_1498	lacR	lactose repressor	8-17	37-46	0.015	0.935	-16.08
	SMU_870		sugar metabolism transcriptional regulator	7-16	480-489	0.022	0.935	-15.43
	SMU_526c		transcriptional regulator	7-16	480-489	0.024	0.935	-15.11
	SMU_1400c			4-13	272-281	0.028	0.935	-17.48
	SMU_1054		glutamine amidotransferase	5-14	397-406	0.037	0.935	-16.17
	SMU_948			8-17	89-98	0.039	0.935	-15.88
	SMU_1037c		histidine kinase	10-19	29-38	0.041	0.935	-19.31
	SMU_2101	aspS	aspartyl-tRNA synthetase	4-13	272-281	0.045	0.935	-16.06
	SMU_1525	murA	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	3-12	333-342	0.046	0.935	-18
	SMU_399			3-12	48-57	0.048	0.935	-14.29
	SMU_1303c		dipeptidase	9-18	303-312	0.048	0.935	-15.76
SmsR19	SMU_2022	rl22	50S ribosomal protein L22	6-15	269-278	0.001	0.155	-18.5
	SMU_1676c			11-20	272-281	0.005	0.413	-17.02
	SMU_229			8-17	504-513	0.021	0.890	-17.35
	SMU_835			5-14	336-345	0.024	0.890	-13.81

	SMU_587			1-10	522-531	0.035	0.890	-13.7
	SMU_1239	pepV	dipeptidase PepV	6-15	354-363	0.036	0.890	-14.32
	SMU_1738	bccP	biotin carboxyl carrier protein of acetyl-CoA carboxylase	9-18	274-283	0.040	0.890	-14.79
	SMU_257	oppC	transmembrane permease OppC	8-17	270-279	0.041	0.890	-15.47
SmsR20	SMU_1771c			10-19	6-15	0.000	0.022	-28.29
	SMU_1529	atpC	ATP synthase F0F1 subunit gamma	6-15	202-211	0.007	0.566	-21.32
	SMU_816		transaminase	11-20	95-104	0.026	0.899	-14.04
	SMU_1382	leuC	3-isopropylmalate dehydratase large subunit	5-14	203-212	0.027	0.899	-19.46
	SMU_81	grpE	heat shock protein GrpE	9-18	110-119	0.044	0.899	-13.78
	SMU_986c			2-11	161-170	0.050	0.899	-13.83
SmsR21	SMU_1579			8-17	197-206	0.000	0.008	-32.37
	SMU_1345c		peptide synthetase	6-15	292-301	0.008	0.898	-17.37
	SMU_167			3-12	292-301	0.018	0.898	-15.53
	SMU_943c		hydroxymethylglutaryl-CoA synthase	6-15	280-289	0.021	0.898	-15.87
	SMU_1587c			6-15	292-301	0.024	0.898	-15.03
	SMU_2090c			1-10	282-291	0.036	0.898	-19.02
	SMU_672	idh	isocitrate dehydrogenase	6-15	162-171	0.036	0.898	-19.96
	SMU_916c		7-cyano-7-deazaguanosine (preQ0) biosynthesis protein QueE	6-15	223-232	0.038	0.898	-14.4
	SMU_970	folA	dihydroneopterin aldolase	1-10	274-283	0.042	0.898	-15.16

	SMU_1652	ogt	methylated DNA-protein cysteine methyltransferase	3-12	292-301	0.046	0.898	-16
	SMU_923		ABC transporter ATP-binding protein	5-14	224-233	0.049	0.898	-14.79
SmsR22	SMU_1273	hisC	histidinol-phosphate aminotransferase	5-14	315-324	0.001	0.347	-16.84
	SMU_246	rgpG	glycosyl transferase N-acetylglucosaminyltransferase RgpG	5-14	225-234	0.002	0.347	-17.54
	SMU_1048			9-18	230-239	0.006	0.699	-14.92
	SMU_312		PTS system sorbitol phosphotransferase transporter subunit IIBC	3-12	39-48	0.017	0.843	-17.95
	SMU_2127		succinate semialdehyde dehydrogenase	3-12	39-48	0.018	0.843	-18.35
	SMU_1205c			9-18	237-246	0.025	0.843	-13.58
	SMU_582		farnesyl diphosphate synthase	9-18	237-246	0.028	0.843	-14.25
	SMU_2043c		D-tyrosyl-tRNA(Tyr) deacylase	4-13	250-259	0.029	0.843	-14.97
	SMU_883	dexB	dextran glucosidase DexB	11-20	205-214	0.030	0.843	-14.98
	SMU_1822	gatA	aspartyl-tRNA synthetase	9-18	237-246	0.031	0.843	-13.38
	SMU_363	glnR	transcriptional regulator glutamine synthetase repressor	9-18	237-246	0.032	0.843	-13.43
	SMU_1301c		methyltransferase	8-17	214-223	0.034	0.843	-14.76
	SMU_573			6-15	224-233	0.034	0.843	-14.34
	SMU_2008	rl30	50S ribosomal protein L30	3-12	39-48	0.037	0.843	-15.5
	SMU_1697c		Best Blastp Hit: gb|AAC05769.1| (AF051356) YtqB [*Streptococcus* mutans]	7-16	313-322	0.044	0.843	-13.04

	SMU_1790c		transcriptional regulator	8-17	238-247	0.044	0.843	-14.06
	SMU_840c			3-12	39-48	0.045	0.843	-15.13
	SMU_1257c			4-13	295-304	0.046	0.843	-15.49
	SMU_706c			5-14	225-234	0.047	0.843	-13.12
	SMU_1299c		acetate kinase	2-11	230-239	0.049	0.843	-14.07
SmsR23	SMU_45			6-15	153-162	0.008	0.493	-13.1
	SMU_2164	htrA	serine protease HtrA	3-12	143-152	0.010	0.493	-12.3
	SMU_1508c		coenzyme PQQ synthesis protein	6-15	107-116	0.025	0.724	-17.86
	SMU_1902c			2-11	159-168	0.037	0.724	-11.13
	SMU_862		permease	4-13	162-171	0.041	0.724	-12.58
	SMU_2001	rpoA	DNA-directed RNA polymerase subunit alpha	8-17	49-58	0.042	0.724	-12.25

[a]Putative mRNA targets with sRNA binding regions around the start codon of annotated genes were obtained using IntaRNA (39). Results with a $p<0.05$ as determined by IntaRNA were retained. SmsR4 targets are shown in **Table 5**.
[b,c]Location of the sRNA and mRNA seed regions from the 5′ end.
[d]False discovery rate.
[e]Hyrbidization free ene rgy in kcal/mol of the interaction RNA seed regions.

Table 5. Complete list of SmsR4 targets predicted by IntaRNA.

Target[a]	Gene	Annotation	mRNA Position[c]	sRNA Position[d]	P-value	FDR Value[b]	Energy[e]
SMU_1485c		Endonuclease	48-73	43-68	0.000	0.000	-26.4
SMU_313		PTS system sorbitol-specific transporter subunit IIA	60-78	48-67	0.001	0.685	-16.2
SMU_607			67-79	66-78	0.003	0.950	-14.3
SMU_379			60-82	43-66	0.003	0.950	-14.2
SMU_157	cysE	serine acetyltransferase	107-131	46-70	0.004	0.950	-14.1
SMU_489			82-110	43-69	0.004	0.950	-14.0
SMU_2139c		50S ribosomal protein L9	116-132	50-66	0.005	0.950	-13.6
SMU_1442c			92-105	61-74	0.007	0.950	-13.1
SMU_1762c			59-79	47-67	0.008	0.950	-12.9
SMU_1757c			66-83	62-78	0.009	0.950	-12.8
SMU_1533	atpG	ATP synthase F0F1 subunit A	58-82	43-67	0.010	0.950	-12.6
SMU_984			116-137	42-66	0.010	0.950	-12.6
SMU_415			56-71	55-69	0.010	0.950	-12.5
SMU_1084	hemK	N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase	120-147	41-68	0.010	0.950	-12.5
SMU_129	adhC	branched-chain alpha-keto acid dehydrogenase E2 subunit	59-79	47-65	0.011	0.950	-12.4
SMU_796			59-85	44-66	0.011	0.950	-12.3
SMU_1804c			104-123	61-77	0.014	0.950	-11.9
SMU_930c		transcriptional regulator	65-80	65-78	0.016	0.950	-11.7
SMU_790			59-83	43-67	0.016	0.950	-11.7
SMU_1729c		aminodeoxychorismate lyase	121-131	59-69	0.017	0.950	-11.7
SMU_1169c		thioredoxin	65-85	60-78	0.017	0.950	-11.6
SMU_1014			112-133	44-66	0.018	0.950	-11.6
SMU_206c			58-76	62-78	0.018	0.950	-11.6
SMU_610	spaP	cell surface antigen SpaP	67-82	63-77	0.019	0.950	-11.5
SMU_1773c			85-107	41-68	0.021	0.950	-11.3
SMU_132		hippurate amidohydrolase	84-110	54-78	0.021	0.950	-11.3

SMU_1405c			65-83	62-78	0.022	0.950	-11.2
SMU_1054		glutamine amidotransferase	66-80	65-78	0.023	0.950	-11.1
SMU_862		permease	84-99	42-58	0.024	0.950	-11.1
SMU_2123			65-83	62-78	0.024	0.950	-11.1
SMU_1209c			36-60	4-25	0.026	0.950	-11.0
SMU_521			67-83	62-78	0.026	0.950	-10.9
SMU_150			54-65	58-69	0.028	0.950	-10.8
SMU_220c			54-82	43-69	0.028	0.950	-10.8
SMU_955			66-83	62-78	0.029	0.950	-10.7
SMU_85	thiD	phosphomethylpyrimidine kinase	64-79	66-78	0.030	0.950	-10.7
SMU_1108c			64-90	40-62	0.031	0.950	-10.6
SMU_981	bglB1	BglB fragment	45-66	62-78	0.031	0.950	-10.6
SMU_1683c			68-89	42-62	0.031	0.950	-10.6
SMU_1906c			54-65	58-69	0.031	0.950	-10.6
SMU_447			66-81	64-78	0.032	0.950	-10.6
SMU_804			107-125	62-78	0.032	0.950	-10.6
SMU_231	ilvB	acetolactate synthase catalytic subunit	65-83	62-78	0.033	0.950	-10.5
SMU_1306c		glmZ(sRNA)-inactivating NTPase	56-70	62-77	0.033	0.950	-10.5
SMU_781		prephenate dehydrogenase	67-83	62-77	0.034	0.950	-10.5
SMU_333			66-79	66-78	0.034	0.950	-10.4
SMU_416		tRNA (guanine-N(7)-)-methyltransferase	65-84	40-65	0.036	0.950	-10.4
SMU_107			95-109	61-76	0.038	0.950	-10.3
SMU_1595	cah	carbonic anhydrase	66-84	61-78	0.039	0.950	-10.2
SMU_366	gltB	glutamate synthase	63-79	47-62	0.040	0.950	-10.2
SMU_2023c		30S ribosomal protein S19	58-81	45-64	0.040	0.950	-10.2
SMU_1068c		ABC transporter ATP-binding protein	57-81	44-67	0.041	0.950	-10.1
SMU_809	uvrB	excinuclease ABC subunit B	56-91	42-67	0.043	0.950	-10.1
SMU_2089	hexB	DNA mismatch repair protein	130-144	62-76	0.044	0.950	-10.0
SMU_832			91-109	63-78	0.044	0.950	-10.0
SMU_494		fructose-6-phosphate aldolase	58-81	44-67	0.044	0.950	-10.0
SMU_287		ComB accessory factor for ComA	60-79	47-68	0.044	0.950	-10.0

SMU_561c		hydrolase (MutT family)	126-139	66-78	0.044	0.950	-10.0
SMU_1975c			67-82	63-78	0.045	0.950	-10.0
SMU_1449		fibronectin/fibrinogen-binding protein	61-72	55-66	0.046	0.950	-10.0
SMU_1915	comC	competence stimulating peptide	64-83	62-78	0.046	0.950	-9.9
SMU_1142c		transcriptional regulator Spx	52-72	51-70	0.046	0.950	-9.9
SMU_817		amino acid ABC transporter substrate-binding protein	53-74	42-61	0.046	0.950	-9.9
SMU_1525	murA	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	115-137	42-64	0.046	0.950	-9.9
SMU_1614	fpg	formamidopyrimidine/5-formyluracil/ 5-hydroxymethyluracil DNA glycosylase	51-68	60-78	0.047	0.950	-9.9

The SmsR4 sequence as determined by RACE was input into IntaRNA to determined mRNA binding partners using default criteria (39). All results with a p-value<0.05 as determined by IntaRNA are displayed.

[a]Putative mRNA targets with SmsR4 binding regions +/-75 nucleotides from the start codon were obtained using IntaRNA (39).

[b]Location of the mRNA seed regions from the 5′ end, starting at -75.

[c]Location of the sRNA seed regions from the 5′ end, starting at 1.

[d]False discovery rate.

[e]Energy score in kcal/mol of the interaction.

Among the 22 novel sRNAs, SmsR4 was intriguing because its genomic location is similar to that of 6S RNA in many proteobacteria (32), and as observed for 6S, SmsR4 was expressed during transition from exponential to stationary phase (31). In gammaproteobacteria, the expression profile of 6S is thought to be controlled by its linkage to the neighboring 5-formyl tetrahydrofolate cyclo-ligase gene, which responds to nutrient limitation during the transition to stationary phase (30). In a similar manner, SMU_320, the 5-formyl tetrahydrofolate cyclo-ligase gene in *S. mutans,* could influence the expression of SmsR4; however, unlike in gram-negative bacteria, SMU_320 and SmsR4 are encoded on opposite strands and hence are not co-transcribed. Instead, a native promoter likely initiates SmsR4's transcription. The 5′ IGR of SmsR4 contains a putative -10 promoter site (**Figure 18A**), and a potential CRE, which binds CcpA and coordinates the utilization of carbohydrates in a hierarchical manner (48), is located at position -234 within the flanking 5-formyl tetrahydrofolate cyclo-ligase gene (**Figure 18B**). This putative CRE site shares 56% sequence homology, including a continuous string of eight identical nucleotides, with the CRE site located upstream of the sorbitol PTS operon, suggesting that CcpA binding may regulate the expression of both sorbitol PTS and SmsR4 (**Figure 18B**). Regulation of both a sRNA and its targets by the same transcriptional factor is a feature shared by sRNAs in several other bacteria (49–54) indicating that SmsR4 has been integrated into the CcpA regulatory circuit in *S. mutans*.

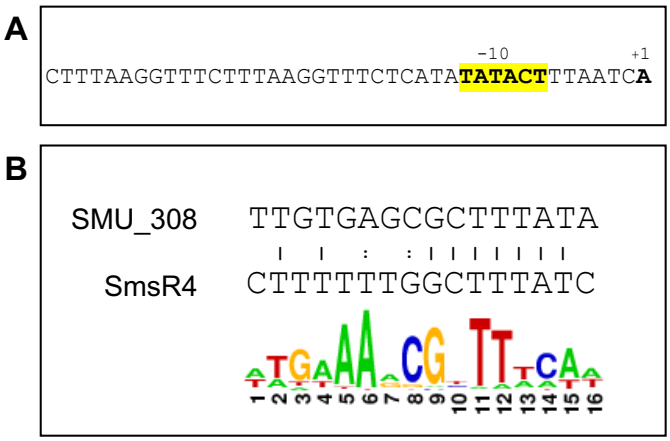

Figure 18. SmsR4 promoter structure and putative catabolite responsive element (CRE) sites for CcpA binding. A) A potential -10 promoter element is present immediately prior to the transcription start site (TSS) of SmsR4. **B)** A putative CRE site, which shares homology with the CRE site before the TSS of SMU_308, is found upstream of SmsR4 (79).

Because sugar transport is central to the proliferation and cariogenicity of *S. mutans*, it is not surprising to find a sRNA that acts to fine-tune sugar usage in the dental pathogen. Based on our data, SmsR4 regulates certain, as of yet undefined, aspects of sorbitol metabolism. While further work is required to delineate its molecular details, our results indicate that SmsR4 functions to alleviate sugar-phosphate stress in *S. mutans*. When sugars are transported through PTS systems, EIIA (e.g., SMU_313) participates in phosphorylation of the incoming sugar molecule as it enters into the cytoplasm (55). If left unchecked, phosphorylated sugars can accumulate in the cell and cause sugar-phosphate stress (56). In *E. coli* and *S. enterica,* the sRNA SgrS plays a key role in restoring glycolytic balance during sugar-phosphate stress by blocking the translation of mRNAs that encode sugar-transporters (57, 58). In a similar fashion, SmsR4 could relieve *S. mutans* of sugar-phosphate stress when the bacterium is growing in the presence of sorbitol by regulating the expression of SMU_313 ($EIIA^{sorb}$) and/or other components of the sorbitol PTS. Alternatively, modulation of SMU_313 expression by SmsR4 could prevent the accumulation of EIIA, which has been shown to negatively impact glycerol metabolism in *Klebsiella pneumoniae* (59), and to reduce *S. enterica*'s virulence (60).

In addition to SmsR4's role in regulating bacterial growth in the presence of sorbitol, we noticed that the sRNA is expressed when xylitol is added to the medium. Xylitol is a common anticariogenic component of dental-care products that inhibits *S. mutans's* growth through energy depletion during a futile cycle of entry, phosphorylation, dephosphorylation, and expulsion (11). Previous studies have shown that xylitol is transported into *S. mutans* mainly through a fructose PTS system (13); interestingly, from

our data, the sorbitol PTS also seems to have a role in bacterial growth during prolonged exposure to xylitol. Transport of xylitol through sorbitol PTS is possible because both sorbitol and xylitol are sugar-alcohols with similar structures (61). We noticed that the expression of sorbitol PTS genes were induced while *S. mutans* was growing in media containing xylitol (**Figure 13**, **Figure 15, Figure 19**), and the expression profile of SmsR4 in BTR-G + xylitol is similar to that in BTR-S (**Figure 10**). As observed in this study, concomitant effects of growth in sorbitol and xylitol have been previously reported in *S. mutans* (62), suggesting that the transport and metabolism of sorbitol and xylitol may utilize overlapping cellular components, including possibly the sorbitol PTS and its regulator, SmsR4. In conclusion, this report provides the first detailed analyses of sRNA function in *S. mutans* and has the laid the foundation for future studies aimed at understanding sRNA-based regulation and its potential application in preventing dental caries.

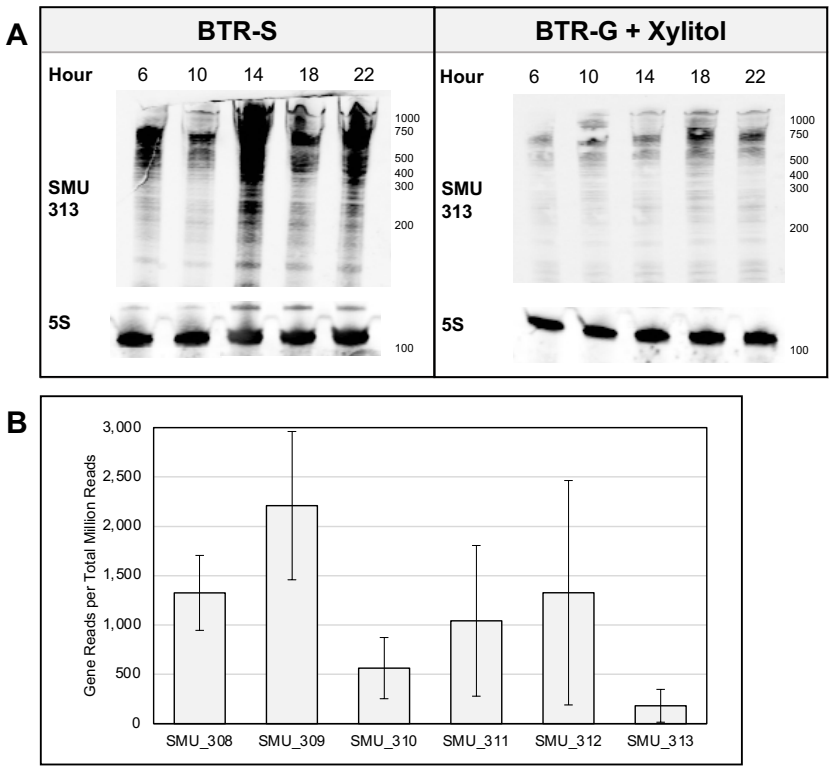

Figure 19. Transcription of sorbitol phosphotransferase (PTS) components during growth with xylitol. **A)** SMU_313 is expressed in BTR media with glucose (BTR-G), but at lower levels than seen in BTR with sorbitol (BTR-S). 5S RNA is shown as a loading control. **B)** RNA-seq analysis of WT grown for 19 hours in BTR-G + xylitol shows the expression of all sorbitol PTS components.

MATERIALS AND METHODS

Bacterial strains and growth assays

Growth curves were conducted by inoculating fresh media 1:100 from overnight growths of *S. mutans* UA159. Glucose growths were done in BTR broth base (1% Bacto-Tryptone, 0.1% Yeast, 0.61% K_2HPO_4, 0.2% KH_2PO_4) supplemented with 0.5% glucose (BTR-G) (63). To induce long-term sugar-phosphate stress, 0.6% xylitol was added to the BTR broth supplemented with 0.2% glucose. Sorbitol media was formulated using BTR broth base with 0.5% sorbitol (BTR-S). When appropriate, spectinomycin was added to the media to retain the pDL278 plasmid. For all growth curves comparing the growth of WT, DEL, and COMP strains, both the WT and DEL strains contained the empty pDL278 vector. OD600 measurements were taken using a BioRad SmartSpec 3000 spectrophotometer. All growth assays were done at 37°C in an anaerobic chamber (5% hydrogen, 5% carbon dioxide, 90% nitrogen).

Phenotypic microarray

Biolog Phenotypic microarray assays were conducted per manufacturer's recommendations (36). Briefly, overnight growths from single colonies in BHI agar were diluted in fresh BHI broth and grown to an OD600 of 0.6-0.7 at 37°C in an anaerobic chamber. 10 mL of cells were collected via centrifugation (3,000xg for 5 minutes) and washed twice with 10 mL of PBS. Cells were resuspended to an OD600 of 0.4 in IF-0a GN/GP base. Inoculating fluid was prepared and combined with cells at 81% turbidity. 100 µL of the mixture was added into each well and overlaid with 40 µL mineral oil. The temperature was maintained at 37°C while readings at 590 nm and 750 nm were collected

every 20 minutes for 24 hours using a Thermo Fischer Scientific Multiskan Spectrum plate reader. Results were obtained by subtracting the measurements at 750 nm from those at 590 nm, and an average of two replicates was used to construct a growth curve for each well.

Stress induction

S. mutans UA159 was diluted 1:100 from an overnight culture in Brain Heart Infusion Broth (BHI) and grown to an OD600 of 0.3-0.4 in BTR-G at 37°C in an anerobic chamber. Anaerobic conditions were maintained throughout stress induction. To induce heat stress, one half of the culture was incubated at 37°C for 15 minutes (control), while the other was incubated at 45°C for 15 minutes (treatment). Similarly, for oxidative stress induction, 1mM H_2O_2 was added to one half of the culture (treatment) while an equal volume of water was added to the control half and incubated for 15 minutes. For sugar-phosphate stress, 6% xylitol in BTR broth was added to one half (treatment) and grown for 15 minutes at 37°C, while an equal volume of BTR broth without xylitol was added to the control and incubated in the same conditions (control). To induce acid stress, cells were spun down and one half was resuspended in BTR-G, pH 7 (control), and the other half was resuspended in BTR-G, pH 5 (treatment), and grown for 30 minutes at 37°C. To confirm stress induction, RNA was isolated and qRT-PCR was performed as described below to measure the expression of stress marker genes previously associated with each tested stress condition (10, 64–66) (**Figure 20**).

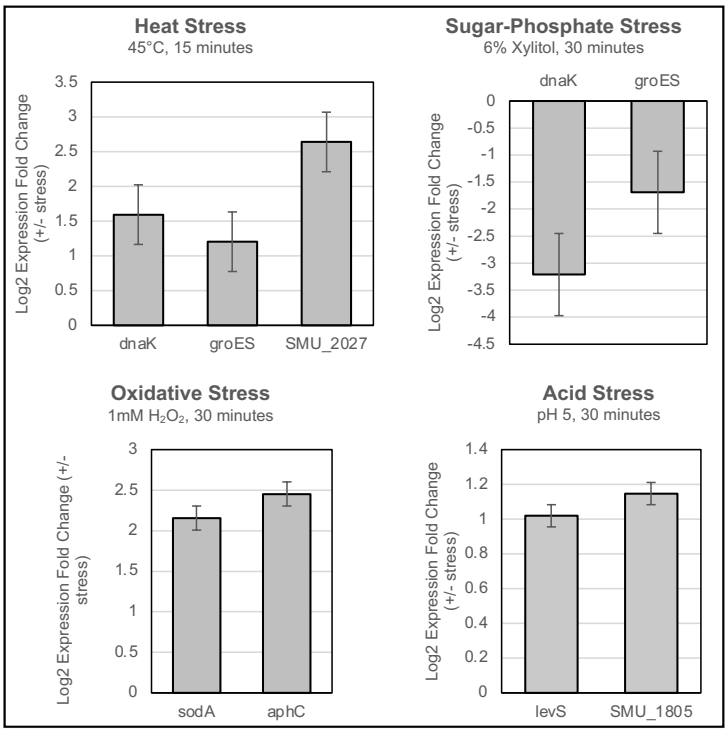

Figure 20. Confirmation of stress induction. Either upregulation or downregulation of genes known to be associated with each stress condition were used as indicators of stress induction prior to RNA-seq analysis. Values represent mean (+/- standard error) from three independent qPCR assays, except for groES and SMU_1805, which were from two replicates.

RNA isolation and qRT-PCR

RNA Stop Solution (5% Phenol, 95% Ethanol) was added to bacterial cultures (1.25 mL per 10 mL) and cells were pelleted by centrifugation at 10,000xg for 10 minutes at 4°C. Bacterial pellets were resuspended in 1 mL of TRI reagent (Thermo Fischer Scientific) and total RNA was extracted using manufacturer's protocol. RNA was resuspended in 50-100 µl nuclease-free water and concentration determined using Nanodrop (Thermo Fischer Scientific). Contaminating DNA was removed using the TURBO DNase Kit (Thermo Fischer Scientific), and cDNA was synthesized with a High-Capacity cDNA Reverse Transcription Kit (Thermo Fischer Scientific) using random primers and MultiScribe Reverse Transcriptase. qRT-PCR was performed using SYBR Green master mix (Thermo Fischer Scientific) and gene-specific primers (**Table 6**).

Table 6. Primers used in this study.

Target	Application	Sequence - Primer 1	Sequence - Primer 2
SmsR4 Complement	Cloning	GGTGGTGGATCCAAGGAAATAAGTAGCCAAAAGGTCA	GGTGGTGAATTCGGCTTCTTGAAGACCAGAACG
M13 site	Cloning	TGTAAAACGACGGCCAGT	CAGGAAACAGCTATGACC
Crosslink-seq Probe A	Crosslink-seq Probe	TAATACGACTCACTATAGGGTATGAAAAAGGGGA	ATCCTAAAACAACCACTTCCCCAAGTG
Crosslink-seq Probe B	Crosslink-seq Probe, Northern Blot Probe	TAATACGACTCACTATAGGGAAGGGAAGGGCATCC	CTTTCCTATTTTATTTCCCCTTTTTCATAAATCTCCTGG
5S probe	Northern Blot Probe	TAATACGACTCACTATAGGGCGTGCTAGGGCTTAACTGCT	TTAAGTGATGATAGCCTAGGGGAG
SmsR1 Probe	Northern Blot Probe	AATACGACTCACTATAGGGCTGTGCCGTGTTAGCAAACA	GGTCTTGACTGGTGGAGCTT
SmsR11 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGAGGCAAAACATTTTAGGGCGAC	TGACAAGGAAACAAGGTGCA
SmsR12 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGTCAATCATTAACATAAGCGCCACC	TGGTAAGATTAAATATTGCGCGGT
SmsR14 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGGGGCAACTCGCTCTGACTTT	TCACATCTAAGTGTTGACCGACT
SmsR15 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGGCTCCAACCCCCTTTGACTT	TCAGACAACTCTGCAGGCAG
SmsR16 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGAGCCCTACTAGATAAAGTAGTGGT	ACTAGGCGACAAACCACAGG
SmsR17 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGCGTGAATAACCCGTCGAGCA	TGCTCGACGGGTTATTCACG
SmsR18 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGTAACTAGCACCACGGGTTAGA	TGCAATACAGAACGCAATCATCC
SmsR19 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGGCTCTAAAACCTCGTGGACCA	TCTGGCTGAGATGAATGGCG
SmsR2 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGTCAGCACAATCTCTTTATAATGGCG	ATCTTTGGATGCTGATTGCTTTGT
SmsR20 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGCCCCGCATTCCCTGTCAAAT	TCTCGTTCCCCATAGAATCTTTT
SmsR21 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGGCGAAAGAGCAAAATCCCCC	TGCTGATAACGACTGTGACTGA
SmsR22 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGTTTCACCACATCAGCTTGCG	CGGTGACAAAGGCGATGATT
SmsR23 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGGCTTGATCACAATACGAAGCTGT	TGTCAGAGTAAAAAGCTGTTCTGA
SmsR3 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGCGCAAGCACTATCCCAACAA	GAGGGGGTGAGTTGCCAAAT
SmsR5 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGCGCAAAGCACATCCCTAGAC	TCGCGTGGTAAATATTGCAATT
SmsR6 Probe	Northern Blot Probe	TAATACGACTCACTATAGGGAGGAGTTGCACGATAAGGTGT	ACAGGCAGCAGAATTTTGGG

SmsR7 Probe	Northern Blot Probe	TAATACGACTCACTATAGGG CGGACAGCTGCAAAAAGATG AG	AGACACAAAAGGCATTGGGC
SmsR8 Probe	Northern Blot Probe	TAATACGACTCACTATAGGG AGGCTGGTTTTTAGGCAGCT	ACGGATTCGACAGGCATTATGA
SmsR9 Probe	Northern Blot Probe	TAATACGACTCACTATAGGG TTCATCTTGAGCAGAATACAG TTCT	GCACTTAGCGAACCAACAAGA
SMU_313 Probe	Northern Blot Probe	TAATACGACTCACTATAGGG GCTTCAAGCTGAACAGGGAA	ACAGGACTTGGCCATATCACT
SmsR13 Probe	Northern Blot Probe	TAATACGACTCACTATAGGG TCCCAGGCCATTTGAGATAA CA	ACTTTGAAATCTTACCTTATCGT TGCT
16S PCR, universal Primer	qPCR	AGG TGS TGC ATG GYT GTC G	GAC GGG CGG TGW GTR CA
SmsR4	qPCR	GTGGTTGGGGAGATAAACTT TCC	AAGGGAAGGGCATCCAATTTC
SMU_1093	qPCR	CCGCGACTACATGATGAAAA	TCT CAA ATI TCGGAG GTI GG
SMU_1129	qPCR	TCCCTTTTTCACGCAGTTCT	GCAGAGAGTGGCGTTTATGA
SMU_113	qPCR	AGCTGTTGGCAGTGTTGTTG	ACCTGCTTCAAGCTGAACAG
SMU_115	qPCR	GCTTTAAAATTGCGGGAGTC	TGTCTTTGGCATGAGGAATG
SMU_115	qPCR	ACGGGTGTTGAATGGCATTC	AATGTGTTGTCCCTGCTTCC
SMU_1603	qPCR	TCATCAAGCGCATCAAAAAG	ATCGGGGTCTGTGATGAAGT
SMU_1805	qPCR	TCATTGGGCTGAGAA ACTCC	GTA TTGGGC AAC CACCGTTA
SMU_1955	qPCR	TTGACTAGC ACC AGC AAG GA	TTAGGAGACCGAGTGGTCGT
SMU_1965c	qPCR	TGGCTGAGAAAATCGCTGAG	AACTGCGTCAATTCCTGCTG
SMU_2027	qPCR	ATCAGCAGGCACAGGCTATT	GATCCAAGACGCAAAATCGT
SMU_313	qPCR	AGCTGTTGGCAGTGTTGTTG	ACCTGCTTCAAGCTGAACAG
SMU_318F	qPCR	GAAGGGCGTATGCATGCTTG	AGTTGCAATTCCTCGAGCCA
SMU_320	qPCR	TTTGGCCTCAAGGAACCAAG	TTAAAAGCAAGCCCCGGAAC
SMU_494	qPCR	TTCATGCACAAGTGGTTGCG	AGATTGCCGCCAAGTTCATC
SMU_500	qPCR	TTGGATGCACGTATCAATCTG	ACGAAGAGTAACGGGAGCAA
SMU_629	qPCR	GCAACTTATGTCGCAAACGC	TCCGCTGGAATTTGTTCCAC
SMU_764	qPCR	TGGTCGTGATGCTAGTACCT TG	TTCCATTTGGCAGGGCAAAC
SMU_80	qPCR	GTGGAGCCTAAAAAGCAACG	ACCGTTACAGGACTGGCTTC
SMU_82	qPCR	ATCGAAACAATGGGTGGTGT	AACGGCTGGTTGGTTATCTG
3' RACE GSP	RACE	ATCCTAAAACAACCACTTCCC CA	
5' RACE GSP 1	RACE	AAGGGAAGGGCATCCA	
5' RACE GSP 2	RACE	CATTGAATGCCCAGGAGAT	

Abridged Universal Amplification Primer	RACE	GGCCACGCGTCGACTAGTAC	
IDFC2 Cassette	SmsR4 Deletion	CCGAGCAACAATAACACTCAT	GAAGCTGTCAGTAGTATACCTA
SmsR4 Deletion 1	SmsR4 Deletion	TGTTGAGCAGAGGACTTGGCGT	GTACAGGAGCAGAAATCGCCAC
SmsR4 Deletion 2	SmsR4 Deletion	CTCATATATACTTTAATCATCCTAAA GACTTAAATCTGCAAGGTTTTTTA	TAAAAAACCTTGCAGATTTAAGTC TTTAGGATGATTAAAGTATATATGAG
Native SMU_313 with 5' UTR	EMSA	TAATACGACTCACTATAGGGAGCATGGATTGCAAGTATTGGAC	AGTGTCTAACGTTAATTGCATCGT
SMU_313 transcript with mutated 5' UTR, for Q5 mutagenesis	EMSA	CCTAAAATGACTAAAATTTTTGAAGCAAAA	TTTTGGTTTCTAATAAACAATTAAGTTTACACATTTTATTG
SmsR4 Transcript	EMSA	TAATACGACTCACTATAGGGATCCTAAAACAACCACTTCCCCA	AAGGGAAGGGCATCCAATTTCA

RNA-seq, sRNA detection and gene expression

RNA sequencing was performed at the Yale Center for Genomic Analysis using Illumina NovaSeq. Raw reads were deposited in NCBI Sequence Read Archive (PRJNA72673) (67). RNA-seq reads were processed using Trimmomatic to remove low-quality reads and adapters (68). CLC Genomics workbench was used to map the reads to the *S. mutans* UA159 genome (NC_004350.2) and to determine the total read count for each gene. Coverage plots were generated by calculating reads mapped per nucleotide across the entire genome using an in-house Perl script, as described previously (17, 69). The Artemis genome browser (70) was used for visual inspection of transcriptomics data. Novel sRNAs were identified as discrete transcriptional peaks originating from un-annotated regions of the genome. The DESeq2 package in R was used to determine differential expression of genes and sRNA transcripts (28).

***In vitro* transcription**

Amplification of gDNA for *in vitro* transcription was performed using PCR primers designed to incorporate a T7 promoter (**Table 6**). PCR products to be used as DNA templates were gel purified from a 1% agarose gel using a NucleoSpin Gel and PCR Clean-up kit (Takarabio). *In vitro* transcription was performed using MAXIscript T7 Transcription Kit per manufacturer protocol (Thermo Fischer Scientific) with a maximum of 1 μg of DNA used as template. Following TURBO DNase treatment, RNA was purified using a Monarch RNA Cleanup Kit (New England Biolabs).

Northern blot

Equal amounts of RNA from each sample were adjusted to 10 µl with nuclease-free water, and 10 µl of 2x RNA loading dye (Thermo Fischer Scientific) was added and incubated for 10 minutes at 70°C followed by 3 minutes on ice. Samples were loaded onto either 6% or 10% TBE-Urea Gel (Thermo Fischer Scientific) with a biotinylated sRNA ladder (Kerafast). Gels were run in 1x TBE buffer at 180V for 60 minutes (6% gels) or 180V for 80 minutes (10% gels). RNA was transferred to a Biodyne B Nylon Membrane (Thermo Fischer Scientific) using the BioRad Mini-Trans Blot at 12V overnight, 4°C in 0.5x TBE buffer. Membranes were UV-crosslinked using a Staralinker 2400 UV Crosslinker (1200 mJ) and were moved to glass hybridization chambers and prehybridized using 10 mL of ULTRAhyb-Oligo Buffer (Thermo Fischer Scientific) at 45°C for 2 hours with rotation. RNA probes were heated at 95°C for 5 minutes and cooled on ice for 3 minutes, then added to fresh hybridization buffer. Membranes were incubated overnight at 45°C with rotation. After washing, membranes were incubated for 2 hours with shaking in Licor Intercept Blocking Buffer with 1% SDS at room temperature. Blocking buffer was removed and membranes were incubated in Streptavidin-IRDye 800 CW diluted 1:20,000 in Licor Intercept Blocking Buffer with 1% SDS for 30 minutes. Blots were washed and viewed on a Licor Odyssey scanner.

EMSA and mutagenesis

Electromobility shift assay (EMSA) was performed as described previously (38). Briefly, RNA was transcribed using the MAXI T7 Transcription Kit that incorporated biotinylated Uracil into the sRNA transcript and purified with a Monarch RNA Cleanup

Kit. Purified RNA was concentration and resuspended in TE buffer. sRNA and mRNA transcripts were combined and heated for 5 minutes at 85°C, then plunged in ice for 30 seconds. The reaction volume was adjusted to 10 μL with 5x TMN buffer and incubated at 37°C for 30 minutes. Samples were run on an 8% TBE gel (Thermo Fischer Scientific) for 90 minutes at 100 V in 1x TBE buffer. The gel was transferred to Biodyne B Membrane overnight at 12V, 4°C in 0.5x TBE. Membranes were crosslinked, blocked, and probed with Strepatvidin-IRDye 800CW as described above. Images were viewed on a Licor Odyssey scanner. mRNA transcripts with binding site mutations were constructed using the Q5 Mutagenesis Kit (New England Biolabs) using the manufacturer's primer design tool and protocol recommendations. Mutations were confirmed through Sanger sequencing.

Gene deletion and complementation

Mutant strains were constructed using the markerless-mutagenic system and an IDFC2 selection and counter-selection cassette as described previously by Xie et al. 2011 (71). Complementation was preformed using the pDL278 plasmid designed for expression in both *E. coli* and *S. mutans* (72, 73). The plasmid was purified from *E. coli* using a Plasmid MiniPrep (Thermo Fischer Scientific). pDL278 was linearized using BamHI and EcoRI, and PCR products (**Table 6**) were ligated into the linearized plasmid using T4 ligase (Thermo Fischer Scientific). Plasmids were then transformed into competent DH5-alpha cells following manufacturers protocol (New England Biolabs). Sanger sequencing was used to confirm plasmid construction. Purified plasmids were transformed into *S.*

mutans using the same methods described for mutant strain construction by Xie et al. 2011 (71).

Crosslink-seq

Crosslink-seq was performed as described previously (37, 38). WT and DEL strains were grown in BTR-G with 0.6% xylitol and grown for 21 hours at 37°C in an anaerobic chamber. Cells were washed, resuspended in PBS, and RNA crosslinking was performed by the addition of 0.2 mg/mL 4′-Aminomethyltroxsalen (AMT) and subsequent exposure to 365 UV light (UVP, 115V, 60Hz). TRI Reagent was used to isolate RNA, which was then treated with TURBO DNase. Two biotinylated RNA probes that are complementary to the first and last 70 nucleotides of SmsR4 were constructed using T7 transcription and purified with a Monarch RNA Cleanup Kit (New England Biolabs). 10.9 μg of DNased RNA was resuspended in hybridization buffer, heated to 85°C for 5 minutes and then cooled on ice for 5 minutes. 200 U RNAse OUT was added to each reaction, and 1 μg of either 5′ or 3′ end probe was added to each reaction. The probe/RNA solution was incubated overnight at 4°C with rotation. NeutraAvidin beads were washed 3x in WB100 and blocked for 2 hours with blocking buffer at room temperature with rotation. Beads were combined with the probe/RNA solution, and the mixture was incubated at 4°C for 4 hours with rotation, then washed 5x in WB400. Hyrbidized RNA was isolated using TRIreagent and exposed to UV 254nm for 15 minutes to uncrosslink RNA. RNA was sequenced and mapped to the *S. mutans* UA159 genome as described above. DESeq2 was used to determined genes enriched in the DEL strain compared to the WT as previously described (28, 37, 38).

RACE assay

Rapid amplification of cDNA ends (RACE) was preformed using the RACE System for Rapid Amplification of cDNA Ends (Thermo Fisher Scientific) per the manufacturer's recommendations with slight modifications to account for small transcript size. 5′ RACE assays were conducted using a gene specific primer (GSP) complementary to the 3′ end of the sRNA sequence. RNA degradation, cDNA synthesis and TdT tailing was accomplished using kit components and manufacturer protocols. A second nested GSP was used to amplify the tailed cDNA, and PCR products from this reaction were cloned into competent DH5-alpha cells (New England Biolabs) following manufacturer's protocols. Inserts in successfully transformed colonies were screened using M13 PCR primers and Sanger sequencing.

For the 3′ RACE assay, Poly-A polymerase (New England Biolabs) was used to poly-A tail 1 μg of total RNA. Precipitated poly-A tailed RNA was reverse transcribed using SuperScript II reverse transcriptase provided with the kit, and RNA was subsequently degraded using RNase H. A GSP designed to bind to the 5′ end of the sRNA and an Abridged Universal Amplification Primer (**Table 6**) were used to amplify the cDNA with 2x PCR Mastermix (Thermo Fischer Scientific), and PCR products were transformed and screened as described above.

Covariance Modeling

Covariance models of sRNAs were constructed as previously described (74). Briefly, the SmsR4 and 6S sequences from *S. mutans* UA159 were used as the query in a blast search (75) against *Streptococcus* genomes available in RefSeq (76). Hits with >65% identity

and >70% coverage were retained. A list of seed sequences was compiled from these results by selecting four to six sequences from unique *Streptococcus* species in addition to the *S. mutans* sRNA sequence. The seed sequences were aligned using default conditions on the WAR webserver (77). Using the consensus sequence produced from this alignment, the Infernal suite of tools (78) was used to build (cmbuild) and calibrate (cmcalibrate) a covariance model. This first model was then used to search for hits within a database constructed from the 37 full representative *Streptococcus* genomes available on RefSeq (using cmsearch with the flags: -Z 4 -E 0.00001) (76). The first model was re-calibrated to incorporate new hits from this search, and the search and calibration process continued until no new hits were generated from the database. Using the same search parameters, the finalized model was then used to search a database created from full and partial genomes of *Streptococcus* species with at least one member from each subclade found in **Figures 4 and 6** included. The presence and absence of sRNAs from each *Streptococcus* species were mapped onto a previously published phylogenetic tree (34) and nodes of sRNA origin and secondary loss were determined through maximum parsimony.

In-silico Target Prediction

For target prediction of SmsR4 using IntaRNA (39) (**Table 5**), the transcript sequence as determined by RACE assay was used as input and searched against the *S. mutans* UA159 genome using default parameters (75 nt upstream and downstream from the translation start site, one interaction per pair, 7 nt hybridization seed). Target RNA2 (40) was run

using the same query sequence and default parameters (80 nt upstream and 20 nt downstream from the translation start site, 7 nt hybridization seed).

IntaRNA (39) was used to predict targets for the remaining 21 novel sRNAs (**Table 4**) using stricter parameters to screen for interactions adjacent to the TSS of mRNA targets. sRNA sequence was defined within the left and right bounds determined by RNA-seq (**Table 1**). For all *in silico* target predictions, only significant results ($p < 0.05$, as determined by IntaRNA or Target RNA2) were retained.

Fructose-6-phosphate Assay

Overnight cultures of WT and DEL grown in BHI were diluted 1:100 into BTR-S in duplicate and cultured for 16 hours in anerobic conditions at 37°C. Cells were pelleted and immediately frozen at -80°C. Once thawed, cell pellets were washed twice in cold PBS and resuspended to an OD600 of 0.55. 1 mL of resuspended cells were broken using bead beating and filtered through a 10kD MW spin column (Millipore Sigma). 50 µL of flow-through was immediately used to quantify fructose-6-phosphate using a fluorometric fructose-6-phosphate assay kit (Millipore Sigma). Results were obtained using a BioTek Synergy HT plate reader and averaged over duplicate samples.

ACKNOWLEDGEMENTS

We thank Zhengzhong Zou and Samantha Fancher for their assistance with experiments.

This work was supported in part by National Institute of Dental and Craniofacial Research grants DE028409 to RR and DE028252 to JM.

REFERENCES

1. S. D. Forssten, M. Björklund, A. C. Ouwehand, *Streptococcus mutans*, caries and simulation models. *Nutrients* **2**, 290–298 (2010).

2. S. Hamada, T. Koga, T. Ooshima, Virulence factors of *Streptococcus mutans* and dental caries prevention. *J Dent Res* **63**, 407–411 (1984).

3. W. J. Loesche, Role of *Streptococcus mutans* in human dental decay. *Microbiol Rev* **50**, 353–380 (1986).

4. W. H. Bowen, R. A. Burne, H. Wu, H. Koo, Oral biofilms: Pathogens, matrix, and polymicrobial interactions in microenvironments. *Trends in Microbiology* **26**, 229–242 (2018).

5. D. Ajdić, *et al.*, Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. *PNAS* **99**, 14434–14439 (2002).

6. J. Lemos, *et al.*, The biology of *Streptococcus mutans*. *Microbiol Spectr* **7** (2019).

7. J. Deutscher, C. Francke, P. W. Postma, How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev* **70**, 939–1031 (2006).

8. K. K. Mäkinen, Sugar alcohols, caries incidence, and remineralization of caries lesions: A literature review. *Int J Dent* **2010** (2010).

9. S. Takahashi-Abbe, K. Abbe, N. Takahashi, Y. Tamazawa, T. Yamada, Inhibitory effect of sorbitol on sugar metabolism of *Streptococcus mutans* in vitro and on acid production in dental plaque in vivo. *Oral Microbiology and Immunology* **16**, 94–99 (2001).

10. E.-M. Decker, C. Klein, D. Schwindt, C. von Ohle, Metabolic activity of *Streptococcus mutans* biofilms and gene expression during exposure to xylitol and sucrose. *Int J Oral Sci* **6**, 195–204 (2014).

11. A. Pihlanto-Leppälä, E. Söderling, K. K. Mäkinen, Expulsion mechanism of xylitol 5-phosphate in *Streptococcus mutans*. *Eur J Oral Sci* **98**, 112–119 (1990).

12. E. M. Söderling, T. C. Ekman, T. J. Taipale, Growth inhibition of *Streptococcus mutans* with low xylitol concentrations. *Curr. Microbiol.* **56**, 382–385 (2008).

13. J. M. Tanzer, A. Thompson, Z. T. Wen, R. A. Burne, *Streptococcus mutans*: Fructose transport, xylitol resistance, and virulence. *J Dent Res* **85**, 369–373 (2006).

14. J. A. Lemos, R. A. Burne, A model of efficiency: Stress tolerance by *Streptococcus mutans*. *Microbiology* **154**, 3247–3255 (2008).

15. G. Storz, J. Vogel, K. M. Wassarman, Regulation by small RNAs in bacteria: Expanding frontiers. *Mol Cell* **43**, 880–891 (2011).

16. E. Holmqvist, E. G. H. Wagner, Impact of bacterial sRNAs in stress responses. *Biochem. Soc. Trans.* **45**, 1203–1212 (2017).

17. R. Raghavan, E. A. Groisman, H. Ochman, Genome-wide detection of novel regulatory RNAs in *Escherichia coli*. *Genome Res* **21**, 1487–1497 (2011).

18. S. Wachter, L. D. Hicks, R. Raghavan, M. F. Minnick, Novel small RNAs expressed *by Bartonella bacilliformis* under multiple conditions reveal potential mechanisms for persistence in the sand fly vector and human host. *PLoS Negl Trop Dis* **14**, e0008671 (2020).

19. I. Warrier, L. D. Hicks, J. M. Battisti, R. Raghavan, M. F. Minnick, Identification of novel small RNAs and characterization of the 6S RNA of *Coxiella burnetii*. *PLoS One* **9** (2014).

20. I. Kalvari, *et al.*, Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **46**, D335–D342 (2018).

21. L. Zeng, *et al.*, Gene regulation by CcpA and catabolite repression explored by RNA-Seq in *Streptococcus mutans*. *PLoS One* **8** (2013).

22. D. Sinha, *et al.*, Redefining the small regulatory RNA transcriptome in *Streptococcus pneumoniae* serotype 2 strain D39. *Journal of Bacteriology* **201** (2019).

23. M. Naville, A. Ghuillot-Gaudeffroy, A. Marchais, D. Gautheret, ARNold: A web tool for the prediction of Rho-independent transcription terminators. *RNA Biol* **8**, 11–13 (2011).

24. B. C. Jester, P. Romby, E. Lioliou, When ribonucleases come into play in pathogens: A survey of gram-positive bacteria. *International Journal of Microbiology* **2012**, e592196 (2012).

25. B. K. Mohanty, S. R. Kushner, Enzymes involved in post-transcriptional RNA metabolism in gram-negative bacteria. *Microbiol Spectr* **6** (2018).

26. J. Merritt, Z. Chen, N. Liu, J. Kreth, Posttranscriptional regulation of oral bacterial adaptive responses. *Curr Oral Health Rep* **1**, 50–58 (2014).

27. R. A. T. Mars, P. Nicolas, E. L. Denham, J. M. van Dijl, Regulatory RNAs in *Bacillus subtilis*: A gram-positive perspective on bacterial RNA-mediated regulation of gene expression. *Microbiol. Mol. Biol. Rev.* **80**, 1029–1057 (2016).

28. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

29. A. E. Trotochaud, K. M. Wassarman, A highly conserved 6S RNA structure is required for regulation of transcription. *Nature Structural & Molecular Biology* **12**, 313–319 (2005).

30. J. E. Barrick, N. Sudarsan, Z. Weinberg, W. L. Ruzzo, R. R. Breaker, 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA* **11**, 774–784 (2005).

31. A. T. Cavanagh, K. M. Wassarman, 6S RNA, a global regulator of transcription in *Escherichia coli, Bacillus subtilis*, and beyond. *Annual Review of Microbiology* **68**, 45–60 (2014).

32. E. Belda, *et al.*, An updated metabolic view of the *Bacillus subtilis* 168 genome. *Microbiology (Reading)* **159**, 757–770 (2013).

33. H. Chae, *et al.*, Rho-dependent termination of *ssrS* (6S RNA) transcription in *Escherichia coli*. *J Biol Chem* **286**, 114–122 (2011).

34. S. Patel, R. S. Gupta, Robust demarcation of fourteen different species groups within the genus Streptococcus based on genome-based phylogenies and molecular signatures. *Infection, Genetics and Evolution* **66**, 130–151 (2018).

35. S. H. Bernhart, I. L. Hofacker, S. Will, A. R. Gruber, P. F. Stadler, RNAalifold: Improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* **9**, 474 (2008).

36. B. R. Bochner, P. Gadzinski, E. Panomitros, Phenotype MicroArrays for high-throughput phenotypic testing and assay of gene function. *Genome Res* **11**, 1246–1255 (2001).

37. F. R. Kacharia, J. A. Millar, R. Raghavan, Emergence of new sRNAs in enteric bacteria is associated with low expression and rapid evolution. *J. Mol. Evol.* **84**, 204–213 (2017).

38. S. Wachter, *et al.*, A CsrA-binding, trans-acting sRNA of *Coxiella burnetii* is necessary for optimal intracellular growth and vacuole formation during early infection of host cells. *J. Bacteriol.* **201**, e00524-19 (2019).

39. M. Mann, P. R. Wright, R. Backofen, IntaRNA 2.0: Enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Res* **45**, W435–W439 (2017).

40. M. B. Kery, M. Feldman, J. Livny, B. Tjaden, TargetRNA2: Identifying targets of small regulatory RNAs in bacteria. *Nucleic Acids Res* **42**, W124–W129 (2014).

41. R. A. Tesorero, *et al.*, Novel regulatory small RNAs in *Streptococcus pyogenes*. *PLoS One* **8**, e64021 (2013).

42. K. M. Wassarman, F. Repoila, C. Rosenow, G. Storz, S. Gottesman, Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.* **15**, 1637–1651 (2001).

43. M. Miyakoshi, Y. Chao, J. Vogel, Regulatory small RNAs from the 3′ regions of bacterial mRNAs. *Current Opinion in Microbiology* **24**, 132–139 (2015).

44. Y. Chao, K. Papenfort, R. Reinhardt, C. M. Sharma, J. Vogel, An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* **31**, 4005–4019 (2012).

45. A. D. Garst, A. L. Edwards, R. T. Batey, Riboswitches: Structures and mechanisms. *Cold Spring Harb Perspect Biol* **3** (2011).

46. B. R. Jose, P. P. Gardner, L. Barquist, Transcriptional noise and exaptation as sources for bacterial sRNAs. *Biochem. Soc. Trans.* **47**, 527–539 (2019).

47. M. Lybecker, I. Bilusic, R. Raghavan, Pervasive transcription: Detecting functional RNAs in bacteria. *Transcription* **5** (2014).

48. J. Abranches, *et al.*, CcpA regulates central metabolism and virulence gene expression in *Streptococcus mutans*. *J Bacteriol* **190**, 2340–2349 (2008).

49. G. Desnoyers, A. Morissette, K. Prévost, E. Massé, Small RNA-induced differential degradation of the polycistronic mRNA iscRSUA. *The EMBO Journal* **28**, 1551–1561 (2009).

50. A. Gaballa, *et al.*, *The Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *PNAS* **105**, 11927–11932 (2008).

51. C. Bækkedal, P. Haugen, The Spot 42 RNA: A regulatory small RNA with roles in the central metabolism. *RNA Biol* **12**, 1071–1077 (2015).

52. G. Porcheron, C. M. Dozois, Interplay between iron homeostasis and virulence: Fur and RyhB as major regulators of bacterial pathogenicity. *Vet Microbiol* **179**, 2–14 (2015).

53. M. I. Camacho, *et al.*, Effects of the global regulator CsrA on the BarA/UvrY two-component signaling system. *Journal of Bacteriology* **197**, 983–991 (2015).

54. A. Brosse, A. Korobeinikova, S. Gottesman, M. Guillier, Unexpected properties of sRNA promoters allow feedback control via regulation of a two-component system. *Nucleic Acids Res* **44**, 9650–9666 (2016).

55. J. Deutscher, *et al.*, The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system: Regulation by protein phosphorylation and phosphorylation-dependent protein-protein interactions. *Microbiol. Mol. Biol. Rev.* **78**, 231–256 (2014).

56. G. R. Richards, M. V. Patel, C. R. Lloyd, C. K. Vanderpool, Depletion of glycolytic intermediates plays a key role in glucose-phosphate stress in *Escherichia coli*. *Journal of Bacteriology* **195**, 4816–4825 (2013).

57. M. Bobrovskyy, C. K. Vanderpool, The small RNA SgrS: Roles in metabolism and pathogenesis of enteric bacteria. *Front Cell Infect Microbiol* **4** (2014).

58. J. B. Rice, C. K. Vanderpool, The small RNA SgrS controls sugar–phosphate accumulation by regulating multiple PTS genes. *Nucleic Acids Res* **39**, 3806–3819 (2011).

59. W.-Y. Jeng, *et al.*, The negative effects of KPN00353 on glycerol kinase and microaerobic 1,3-propanediol production in *Klebsiella pneumoniae*. *Front. Microbiol.* **8** (2017).

60. J. Choi, *et al.*, *Salmonella* pathogenicity island 2 expression negatively controlled by EIIANtr-SsrB interaction is required for *Salmonella* virulence. *Proceedings of the National Academy of Sciences* **107**, 20506–20511 (2010).

61. M. Grembecka, Sugar alcohols—their role in the modern world of sweeteners: A review. *Eur Food Res Technol* **241**, 1–14 (2015).

62. S. Assev, G. Rølla, Sorbitol increases the growth inhibition of xylitol on *Streptococcus mutans* OMZ 176. *Acta Pathol Microbiol Immunol Scand B* **94**, 231–237 (1986).

63. Y. Sato, Y. Yamamoto, R. Suzuki, Harutosi Kizaki, H. K. Kuramitsu, Construction of scrA::lacZ gene fusions to investigate regulation of the sucrose PTS of *Streptococcus mutans*. *FEMS Microbiology Letters* **79**, 339–346 (1991).

64. J. K. Kajfasz, T. Ganguly, E. L. Hardin, J. Abranches, J. A. Lemos, Transcriptome responses of *Streptococcus mutans* to peroxide stress: Identification of novel antioxidant pathways regulated by Spx. *Scientific Reports* **7**, 16018 (2017).

65. Y. Gong, *et al.*, Global transcriptional analysis of acid-inducible genes in *Streptococcus mutans*: Multiple two-component systems involved in acid adaptation. *Microbiology* **155**, 3322–3332 (2009).

66. C. Liu, *et al.*, *Streptococcus mutans* copes with heat stress by multiple transcriptional regulons modulating virulence and energy metabolism. *Sci Rep* **5** (2015).

67. R. Leinonen, H. Sugawara, M. Shumway, The sequence read archive. *Nucleic Acids Res* **39**, D19–D21 (2011).

68. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

69. R. Raghavan, D. B. Sloan, H. Ochman, Antisense transcription is pervasive but rarely conserved in enteric bacteria. *mBio* **3** (2012).

70. K. Rutherford, *et al.*, Artemis: Sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).

71. Z. Xie, T. Okinaga, F. Qi, Z. Zhang, J. Merritt, Cloning-independent and counterselectable markerless mutagenesis system in *Streptococcus mutans*. *Appl Environ Microbiol* **77**, 8025–8033 (2011).

72. L. Zhou, D. A. Manias, G. M. Dunny, Regulation of intron function: Efficient splicing in vivo of a bacterial group II intron requires a functional promoter within the intron. *Molecular Microbiology* **37**, 639–651 (2000).

73. D. J. LeBlanc, L. N. Lee, A. Abu-Al-Jaibat, Molecular, genetic, and functional analysis of the basic replicon of pVA380-1, a plasmid of oral streptococcal origin. *Plasmid* **28**, 130–145 (1992).

74. L. Barquist, S. W. Burge, P. P. Gardner, Studying RNA homology and conservation with Infernal: From single sequences to RNA families. *Curr Protoc Bioinformatics* **54**, 12.13.1-12.13.25 (2016).

75. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).

76. N. A. O'Leary, *et al.*, Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733-745 (2016).

77. E. Torarinsson, S. Lindgreen, WAR: Webserver for aligning structural RNAs. *Nucleic Acids Res* **36**, W79–W84 (2008).

78. E. P. Nawrocki, S. R. Eddy, Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).

79. P. S. Novichkov, *et al.*, RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* **14**, 745 (2013).

CHAPTER FOUR

Discussion and future considerations

I. CONCLUSIONS

A. OxyS evolved from the 3′ end of a protein coding gene

Although the *de novo* emergence of sRNAs from degraded bacteriophage and transposon associated genes has been well documented (1–4), the mechanisms through which sRNAs evolve are still not clearly understood. Our evolutionary reconstruction found that 61% of 371 previously annotated sRNAs in enteric bacteria arose recently (**Chapter 2, Figure 1, Table S1**). The high frequency of recently emerged sRNAs suggests the existence of other sources for new sRNAs besides the previously identified transposable elements and *de novo* formation events. sRNA scans have traditionally examined intergenic regions (IGRs) for novel transcripts and considered the 5′ untranslated region (UTR) of protein coding genes for potential riboswitches. Recently, 3′ UTRs have been shown as an additional source of sRNAs that are processed from the longer mRNA strands by RNases (5). However, the possibility of protein coding genes themselves providing the framework for sRNA development has not been investigated. **My first aim was to explore the potential for sRNAs to originate from within protein coding genes.** As an initial survey, we performed a covariance modeling (cm) based iterative search to identify sRNAs that are located in IGRs in core genomes (*Escherichia coli* K-12 MG1655, *Salmonella enterica* Typhimurium SL1344, or *Yersinia pseudotuberculosis* IP32953) but oriented within coding regions of annotated genes in other Enterobacterales

members (**Chapter 2, Table S3**). We identified 62 such sRNAs with sRNA-ORF overlaps, suggesting that the sRNAs currently located in IGRs in the core genomes likely originated from a now-decayed protein coding gene.

One of these sRNAs that overlapped an annotated protein-coding gene was OxyS. Although the role of OxyS in oxidative stress has been detailed in *E. coli* (6–10), the evolutionary origins of OxyS within the family Enterobacteriaceae remain unknown. Interestingly, a peroxidase gene is present at the same genetic locus as OxyS in the families Erwiniaceae, Pectobacteriaceae, Yersiniaceae, Hafniaceae, Budviciaceae, Pasteurellales and the order Vibrionales (**Chapter 2, Figure 3**). Since OxyS is only present in Enterobacteriaceae, which lack the peroxidase gene, OxyS likely evolved from this peroxidase. Thus, as the ancestral gene lost its function during oxidative stress response and decayed, the sRNA was retained and integrated into the same functional network.

To test this hypothesis, we used cm-based search to detect OxyS-like sequences within the peroxidase gene in Enterobacteriaceae species. An OxyS-like sequence (proto-OxyS) was identified at the peroxidase 3′ UTR in Pectobacteriaceae and Yersiniaceae. In order to better understand the origins of OxyS, my **second aim was to investigate the proto-OxyS sequence found in *Serratia marcescens*** (family Yersiniaceae). This proto-OxyS sequence overlaps with 56 bp of the peroxidase coding sequence (CDS) and extends downstream into the 3′ UTR, ending in an intrinsic terminator (**Chapter 2, Figure 4**). Similar to OxyS in *E. coli*, in *S. marcescens* the peroxidase gene is located divergently from *oxyR*, a master regulator of oxidative stress (11). The IGR between *oxyR* and the peroxidase contains a putative OxyR-binding site (**Chapter 2, Figure 3, 4,**

5). Additionally, H_2O_2 exposure induces the expression of the peroxidase gene in *S. marcescens*, as observed with OxyS in *E. coli* (**Chapter 2, Figure 6**), suggesting it is also regulated by OxyR. Exposure to H_2O_2 induces 3′ fragmentation of the peroxidase mRNA into a ~100 nt transcript that contains the proto-OxyS segment (**Chapter 2, Figure 7A**). No promoters exist within the peroxidase gene, but *in vitro* RNase E digestion of the peroxidase transcript produced a small 3′ end similar to the transcript observed *in vivo*, suggesting that RNase E is responsible for the cleavage event (**Chapter 2, Figure 7B**). Within this assay, the presence or absence of Hfq, an important sRNA binding protein that often functions in concert with RNase E (12), did not impact cleavage. However, an immunoprecipitation assay demonstrated that proto-OxyS associates with Hfq in *S. marcescens* (**Figure 1**), indicating that Hfq is likely required for OxyS function.

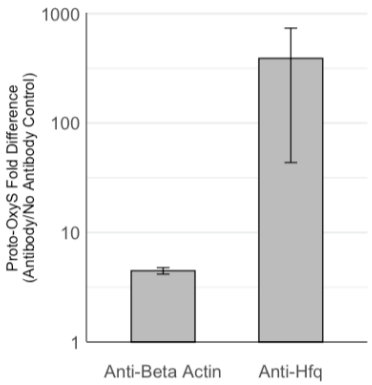

Figure 1. Proto-OxyS binds to Hfq. Proto-OxyS transcripts pulled down by anti-Hfq antibody or by anti-beta actin antibody (control) were quantified using qPCR. Data is shown as fold-difference in comparison to no-antibody control. Mean (+/- standard deviation) from two replicates are shown.

In order to further investigate the evolution of OxyS, my **third aim was to investigate the ancestral peroxidase gene in species without a proto-OxyS sequence**. To this end, we selected two species without a proto-OxyS (*Vibrio harveyi*, order Vibrionales, family Vibrionaceae, and *Edwardsiella hoshinae*, family Hafniaceae) (**Chapter 2, Figure 3**). Expression of the peroxidase gene in both species was significantly induced by H_2O_2 exposure, indicating that these genes retain function during oxidative stress and are controlled by OxyR (**Chapter 2, Figure 6**). Additionally, the H_2O_2 induced fragmentation observed in *S. marcescens* was still present in these two species (**Chapter 2, Figure 8**). Despite variations in fragment size, it appears that peroxidase fragmentation is an ancestral trait in Enterobacterales.

Our investigation into peroxidase cleavage suggests that the RNase-induced fragmentation of the mRNA occurs prior to the maturation of the sRNA and its integration into regulatory networks. Even in species without proto-OxyS, like *V. harveyi* and *E. hoshinae*, the functional peroxidase gene is cleaved at its 3′ end, suggesting that the small, 3′ UTR-derived transcript is freely available within the cell during states of oxidative stress. Since sRNAs only require a small seed region to bind to mRNA partners, it is not hard to imagine that this proto-OxyS transcript might easily acquire mRNA binding partners through chance encounters with already complementary mRNAs, or that slight mutations in sequence could produce a functional seed region. Additionally, the association of the proto-OxyS sequence with Hfq in *S. marcescens* highlights that this pre-sRNA transcript is already associating with an important sRNA binding protein, increasing the likelihood that it may gain functionality (**Figure 1**). We do not yet know if *S. marcescens* proto-OxyS has utility within regulatory networks;

however, preliminary experiments using the RIL-seq technique (13) to detect functional partners of sRNAs *in vivo* found that proto-OxyS was not associated with any genes aside from its parental peroxidase (**Table 1**). It is possible that proto-OxyS may regulate its peroxidase gene as previously described (14); however, future research is needed to confirm its function.

Little is known about the rise of sRNAs aside from *de novo* emergence from bacteriophage- and transposon- associated genes, and the process by which sRNAs are incorporated into regulatory networks remains largely a mystery (1–4). Tracing the evolution of sRNAs can be challenging due to their fast rate of evolution and significant sequence divergence between functional homologs (15), and our cm-based search confirmed early observations that sRNAs evolve rapidly and that most have emerged recently within bacterial lineages (2, 16, 17). We also found that many of these sRNAs likely arose from protein-coding genes, uncovering coding sequences as an unexplored reservoir for sRNA emergence. This finding shifts the traditional paradigm of sRNA emergence primarily occurring within non-coding or transposable genetic elements and opens a whole new avenue for investigating protein-coding genes as a source of sRNA sequences. Specifically, this study provided evidence to show that the well-studied sRNA OxyS arose from a peroxidase gene that is conserved across Enterobacterales. These results also indicate that the oxidative stress responding OxyS became integrated into the same regulatory network as its ancestral peroxidase gene, providing unique insight into how sRNAs become ingrained in specific regulatory pathways. Additionally, we uncovered a likely method of OxyS evolution from this coding sequence, as the 3′ end fragmentation of the peroxidase gene was shown to be an ancestral trait present even

before the development of OxyS homologs within that region. Taken together, our results present evidence for an understudied form of sRNA emergence from protein-coding genes. Because traditional sRNA discovery scans focus largely on finding new transcripts from IGRs, this study offers important novel insight into how sRNAs may evolve from previously overlooked regions of the genome and become integrated into functional networks.

Table 1. Chimeric reads that mapped to Proto-OxyS.

Locus 1	Locus 2	Interactions[a]	Other interactions of RNA 1[b]	Other interactions of RNA 2[c]	Total other interactions[d]	Odds ratio	Fisher's exact test p-value
Proto-OxyS	Peroxidase	32	96	397890	4044274	3.4	0.0060
Peroxidase	Proto-OxyS	9	449	6262	4435572	14.2	0.0007

[a]Total number of interactions between locus 1 and 2.
[b, c]Other interactions of RNAs 1 and 2, respectively.
[d]Total other interactions excluding [b] and [c].

B. Small RNAs are expressed in response to stress in *S. mutans*

It is likely that sRNAs participate in most, if not all, of the regulatory pathways within the bacterial cell, but most functional studies of sRNAs have discovered their roles within stress-tolerance networks (18). Despite the majority of research being focused on gram-negative bacteria, there have been some investigations of sRNA utility in streptococci. The presence of sRNAs in *Streptococcus pneumoniae* and *Streptococcus pyogenes* has been validated, and some evidence exists for their role in virulence (19, 20). Although several studies of *S. mutans* sRNA expression have focused on identifying extremely small microRNAs (21, 22), these studies have failed to move past qPCR validation and *in silico* target prediction to truly confirm the presence of these transcripts and to define their functional roles. One whole genome scan investigating the impact of CcpA on carbohydrate metabolism identified 243 putative sRNA transcripts, 114 of which were differentially expressed in response to the presence/absence of CcpA (23). Aside from these few studies, little work has been done to define sRNA functions in *S. mutans*.

The **fourth aim of my work was to identify sRNAs produced in *S. mutans*.** To this end, we used our established RNA-seq based sRNA discovery pipeline (**Chapter 3, Figure 1**) to identify novel sRNAs from *S. mutans* UA159 isolated at mid-log growth in nutrient rich Todd Hewitt (TH) broth. Despite the lack of stress during this growth condition, we discovered 22 novel sRNAs (**Chapter 3, Figure 2**). Many of these putative sRNAs overlapped with candidates predicted by previous genome-wide scans, and three were annotated in the Rfam database as RNAse P (SmsR5), tmRNA (SmsR8), and 6S RNA (SmsR14) (**Chapter 3, Table 1**). The sRNAs we identified fell into five categories, which were modified from previous designations (19), based on their location relative to

flanking genes: sRNAs located in (I) IGRs on the opposite strand of flanking genes, (II) IGRs on the same strand as both flanking gens, (III) sRNAs on the same strand as their 3′ flanking gene, (IV) sRNAs on the same strand as their 5′ flanking gene, and (V) those located antisense to an annotated gene. The genomic locations of sRNAs are an important consideration because it can provide clues about how sRNAs are produced and provide potential insights into their functionality. sRNAs found in 5′ UTRs of protein-coding genes (III) frequently act as riboswitches or RNA thermometers, controlling the expression of the adjacent mRNA based on environmental conditions or the presence of certain metabolites. Trans-acting sRNAs at 3′ UTRs of other genes (IV) are often processed from larger mRNAs, specifically the 3′ UTRs that follow coding sequence (5). sRNAs located in an IGR on the same strand as the flanking genes (II) can also be subject to RNAse processing, commonly of polycistronic transcripts encoding multiple coding RNAs (27). The most easily distinguishable category of sRNAs are those located in IGRs found on the opposite strand from the flanking genes (I). These discrete sRNAs are easily visible on coverage plot scans and often are under the control of their own native promoter; however, this promoter may be located inside the coding sequence of the 5′ flanking gene (12, 28), likely allowing for some degree of transcriptional coordination with neighboring genes. Additionally, the genes flanking sRNAs can provide important clues to their function: sRNAs that are closely tied to flanking genes, such as those in 5′ or 3′ UTRs or those processed from polycistronic mRNAs, often participate in the same regulatory network as their mRNA neighbors or function during the same period of growth (29). The temporal link between expression of adjacent mRNA and sRNAs is

clearly demonstrated with 6S RNA, which is transcribed at the same phase of growth as its flanking genes *ygfA* in gammaproteobacteria (30).

Because sRNAs typically participate in stress-response networks, our **fifth aim was to determine the differential expression of our novel sRNAs during four short-term stress conditions** that we identified as biologically relevant to the oral cavity (heat, acid, oxidative, and xylitol stress) (**Chapter 3, Table 2**). Remarkably, each one of our putative sRNAs were differentially expressed in at least one of the stress conditions tested, suggesting that *S. mutans* utilizes sRNAs during stress adaptation (**Chapter 3, Figure 3, 4**). The down- or upregulation of sRNAs in response to stress exposure implies that transcriptional control of the sRNA is linked to some aspect of the stress response, suggesting that these transcripts participate in regulatory networks relevant to the stressor. Since sRNAs typically act as negative regulators of mRNAs, their transcriptional patterns associated with stress may be slightly counterintuitive: a sRNA that represses expression of an mRNA involved with a stress response will be downregulated once the stress condition is induced and the mRNA is needed to regulate the response. An example of this pattern of expression is SmsR12, which is predicted to bind with an enzyme involved in arginine synthesis (SMU_334) around the ribosomal binding site, blocking translation. Arginine metabolism is linked to acid stress adaptation (31), and since SmsR12 was highly downregulated in response to acid stress, it is likely that SmsR12 may be suppressing arginine metabolism until needed (i.e. during acid stress).

Differential expression data is incomplete without considering transcript abundance. Although a sRNA may be differentially expressed in response to a stress

condition, if the difference is between zero in the control and less than 10 transcripts in the stressed samples, it is hard to conclude that the sRNA is making any meaningful contribution to the stress response, despite a statistically significant difference in expression. For example, SmsR4 appears to be upregulated in response to heat stress (**Chapter 3, Figure 3**), but the extremely low abundance of the transcript suggests it does not play a role in temperature stress networks. Conversely, sRNAs with extremely high expression likely play an essential role in multiple pathways at many stages of growth, as was seen with RNAse P (SmsR5), tmRNA (SmsR8), and 6S RNA (SmsR14).

To verify the presence of our putative sRNAs and to validate their differential expression during stress, we performed Northern blotting on all of the novel sRNAs and verified the presence of 20 out of 22 novel sRNA transcripts (**Chapter 3, Figure 4**). Through this method we observed the differential expression of nine sRNAs that confirmed our results from the RNA-seq based determination of differential expression (**Chapter 3, Figure 4A**). All but two sRNAs (SmsR17 and SmsR22) showed either differential expression via Northern blot analysis that were not found to be significantly different through RNA-seq analysis (**Chapter 3, Figure 4B**), or were observed as discrete transcripts without differential expression (**Chapter 3, Figure 4C**). Additionally, Northern blots of the sRNAs gave us a picture of the probable transcriptional and cleavage events leading to their formation. Because of their organization in relation to flanking genes many of our novel sRNAs appeared to be transcribed as part of polycistronic mRNAs, and the majority of these sRNAs did not have predicted intrinsic terminators (**Chapter 3, Table 1**), a feature which suggests RNase processing is essential to the maturation of the sRNA transcript. Many of our sRNA blots confirmed this

observation and displayed multiple bands; in almost all instances one band corresponded to the RNA-seq-predicted size of the sRNA, but the presence of multiple bands provides evidence of RNase processing.

Whole-genome sRNA scans, like the one presented in this study, lay a critical foundation for future sRNA research. Prior to identifying the function of individual sRNAs in regulatory networks, researchers must first understand the abundance of novel sRNA transcripts within a genome. Our study has uncovered almost two dozen sRNAs that are differentially expressed in response to one or more short-term stress conditions. These transcripts have been visualized with Northern blots (**Chapter 3, Figure 4**), further confirming their validity as sRNAs which likely play important functional roles in stress response networks.

Previous work had suggested the presence of several hundred putative sRNAs in *S. mutans* (23), but no studies to date have verified the existence of any sRNA transcripts in the species. Because multiple studies of sRNA utility in other *Streptococcus* species have confirmed the presence of multiple sRNAs (19, 32) we were confident that these transcripts were also present in *S. mutans*. This study identified almost two dozen sRNAs that are differentially expressed in response to common stressors of the oral cavity, providing valuable framework for understanding the role of sRNAs in multiple stress responses of *S. mutans*. We were able to validate the presence of these transcripts through Northern blotting of all but two candidate sRNAs, and combined with differential expression analysis, this data further strengthens our confidence in the existence and importance of our putative sRNAs.

Future work should investigate the function of these novel sRNAs and describe any RNase mediated processing necessary for the formation of mature sRNA transcripts. *S. mutans* contains multiple RNases, including RNase J1 and J2, which likely have significant impacts on the cleavage of sRNAs from larger transcripts (33). Of particular interest is 6S RNA (SmsR14), which has been studied extensively in *E. coli* and *Salmonella enterica* but not functionally classified in *S. mutans* (34). Our Northern blot of 6S revealed two similarly sized transcripts (**Chapter 3, Figure 4**), suggesting that RNase processing is necessary to produce the mature 6S transcript, as documented in other species. Additionally, the expression pattern of 6S has not been studied in *Streptococcus* species, and the possible conservation or divergence of 6S expression and function in *S. mutans* requires future exploration.

C. SmsR4 regulates components of the sorbitol PTS

Our **sixth aim was to define the function of a novel sRNA in *S. mutans***. Out of the 22 putative transcripts identified in our whole-genome scan, SmsR4 was of particular interest due to its genome location: in gammaproteobacteria, 6S RNA is linked to a formate biosynthesis gene, *ygfA*, whose homolog in the *S. mutans* genome is SMU_320. SmsR4 resides in the IGR between SMU_318 and SMU_320, while 6S is located at a different, conserved locus in streptococci (**Chapter 3, Figure 6, 7**). Additionally, SmsR4 was upregulated in response to long-term xylitol stress (**Figure 2**), indicating that it may have a functional role in sugar phosphate stress or sugar metabolism.

A cm-based scan revealed that SmR4 is conserved in other streptococci species, and it appears that SmsR4 originates at the Pyogenes-Equinus-Mutans clade (**Chapter 3,**

Figure 8). A secondary loss event is evident with SmsR4 disappearing from the Sobrinus, Salivarus, and Halotolerans subclades. Since IGRs tend to mutate much faster than protein coding genes (1), conservation of sRNAs in these regions suggests that the transcript is functionally important in *Streptococcus*. Interestingly, the location of SmsR4 was highly conserved, with a formate biosynthesis gene located at the 5′ end of 28 out of 31 species with SmsR4 homologs. In two closely related species, *S. pseudoporcinus* and *S. porcinus*, SmsR4 was located adjacent to a gene for the sorbitol PTS component EIIA, a homolog of its mRNA binding partner in *S. mutans*, SMU_313. The interaction of SmsR4 and SMU_313 in *S. mutans* is described in the following section.

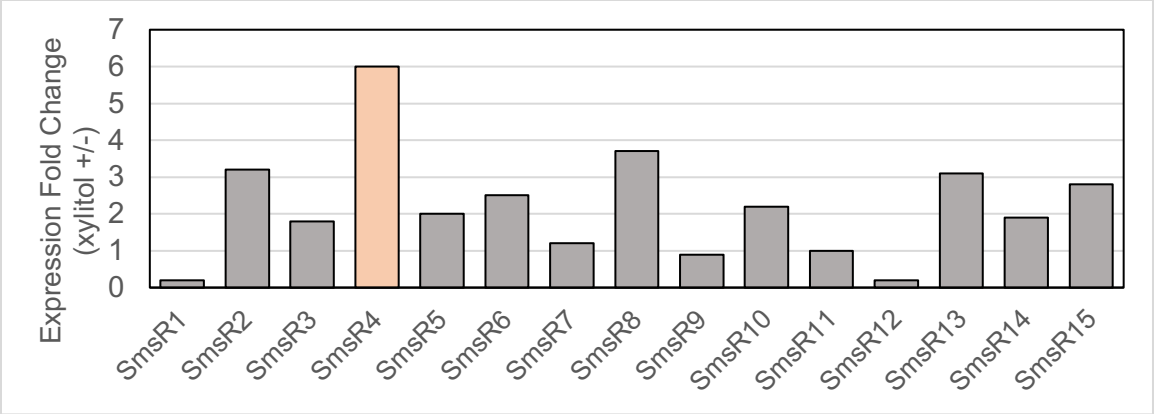

Figure 2. Preliminary data indicates that SmsR4 is highly upregulated during long-term xylitol stress. The upregulation of the novel sRNAs SmsR1-SmsR15 in response to growth in media containing xylitol compared to cells unexposed to xylitol stress is shown. Out of the 15 sRNAs surveyed, SmsR4 showed the highest level of upregulation in response to this stress condition.

Although the overall abundance of the SmsR4 transcript is low compared to other sRNAs such as 6S and tmRNA (**Chapter 3, Figure 5**), we were able to define the 5′ and 3′ bounds of the SmsR4 transcript through rapid amplification of cDNA ends (RACE) assays. The definitive bounds of the SmsR4 transcript were used to inform probe synthesis for subsequent *in vivo* experiments. The first of these was Crosslink-seq (2, 35), an assay that combines RNA crosslinking with immunoprecipitation of sRNA-mRNA duplexes from Wild-type (WT) and SmsR4-deletion (DEL) strains grown for 21 hours in media containing xylitol. This approach identified multiple mRNAs that were potential targets for SmsR4, including SMU_309, the sorbitol PTS operon regulator (**Chapter 3, Figure 13**). We also performed RNA-seq on DEL and WT cells grown for 19 hours in media containing xylitol and found that multiple components of the sorbitol PTS system were upregulated in the DEL strain, a pattern characteristic of negative sRNA regulation (**Chapter 3, Figure 15**). Since xylitol is only known to be transported by the fructose phosphotransferase system (PTS) (36), it was surprising for us to discover components of the sorbitol PTS upregulated in multiple experiments involving xylitol stress.

In silico target prediction programs provide information about potential mRNA binding partners of novel sRNAs. We performed a predictive analysis using two separate target prediction algorithms, IntaRNA and Target RNA2 (37, 38). Both tools suggested that SMU_313, a gene for the sorbitol PTS component EIIA, was a likely target for SmsR4 (**Chapter 3, Figure 11A**), confirming our RNA-seq data that showed components of the sorbitol PTS to be upregulated in the SmsR4-deletion strain compared to WT (**Chapter 3, Figure 15**).

To confirm that SmsR4 can bind to SMU_313 in the region predicted by the algorithms, we performed an EMSA using SmsR4 and the native transcript of SMU_313. The two transcripts associated tightly and increasing levels of mRNA produced progressively more sRNA-mRNA duplexes (**Chapter 3, Figure 12**). However, a mutated version of the SMU_313 with a disrupted binding site failed to bind significantly to SmsR4 (**Chapter 3, Figure 12**).

After gathering strong experimental evidence implicating SMU_313 as a target of SmsR4, we moved on to investigating the role of SmsR4 in metabolism. As a preliminary survey, we used Biolog Phenotypic Microarray plates to compare the growth of the WT and DEL strain on various carbon sources. Among the growth conditions displaying differential growth, the growth of the DEL strain was weaker in media containing sorbitol as the sole carbon source (**Chapter 3, Figure 9**). To confirm the results of the Biolog assay we performed a series of growth curves with the WT, DEL, and a complement strain containing the SmsR4 IGR on the plasmid pDL278 (COMP) in anaerobic conditions in media containing glucose (BTR-G), glucose and xylitol (BTR-G + xylitol), or sorbitol (BTR-S) (**Chapter 3, Figure 10**). While the glucose-grown cells reached maximum density around 10 hours of growth, it took 18+ hours for sorbitol- and xylitol-exposed cells to reach the same phase of growth due to the lag time in sorbitol PTS induction and growth inhibition caused by xylitol. Although there was no difference between the three strains during growth in glucose, the growth of the COMP strain was clearly the most robust in the sorbitol media, with the DEL strain growing worse than the WT. A similar, although less dramatic, pattern was seen in media containing glucose and xylitol. Furthermore, there was a growth advantage for the COMP strain, which supplied

additional copies of SmsR4 on the plasmid vector, in both BTR-S and BTR-G with xylitol. These data clearly illustrate the importance of SmsR4 during sugar-alcohol metabolism. Next, we determined the expression profile of SmsR4 during growth in these three media types by analyzing SmsR4 abundance (**Chapter 3, Figure 10**). From these data it appears that SmsR4 is expressed during the transition to stationary phase: in the glucose grown cells SmsR4 was only moderately expressed around hour 10 of growth when the cells begin to enter stationary phase, while in the sorbitol and xylitol grown samples SmsR4 appeared around 18 hours of growth, corresponding to the same stage of growth. The expression of SmsR4 in media containing the sugar-alcohols was significant and appeared to increase over time, and interestingly the blots of SmsR4 from xylitol and sorbitol grown cells looked very similar (**Chapter 3, Figure 10**).

In order to gain a better understanding of how SmsR4 action may be influencing downstream metabolism, we performed an assay to test the intracellular concentration of fructose-6-phosphate (F6P) in WT and DEL cells grown for 16 hours in sorbitol (**Chapter 3, Figure 17**). Despite having observed an increase of SMU_313 transcript in the DEL strain at this same timepoint (**Chapter 3, Figure 14, 15**), we found more F6P in the WT strain. The increased level of this important glycolytic substrate in the WT suggests that SmsR4 contributes to the efficiency of glycolysis and energy metabolism.

Further analysis is required to fully elucidate the mechanism by which the binding of SmsR4 to SMU_313 regulates sugar-alcohol metabolism; however, previous research informs two plausible theories of action. The first is that by regulating the availability of SMU_313, which participates in the phosphorylation chain that generates intracellular phosphosugars, SmsR4 is able to prevent or respond to sugar-phosphate stress. It has

been well documented that an accumulation of phosphorylated sugars can be damaging to a cell, and the sRNA SgrS responds to high levels of glucose-6-phosphate (G6P) in some enteric pathogens (39, 40). SgrS represses the translation of glucose PTS components in order to prevent the buildup of additional G6P, and it is likely that SmsR4 plays a similar role in alleviating sugar phosphate stress during sorbitol metabolism. Secondly, it is possible that by downregulating the expression of SMU_313, SmsR4 prevents the accumulation of EIIA, which can be deleterious to bacteria at high concentrations. For example, work in *Klebsiella pneumoniae* has shown that overexpression of EIIA during glycerol metabolism negatively impacts the formation of useful metabolic end products (41), and EIIA overexpression in *Salmonella enterica* decreased its virulence (42). In a similar manner, it is possible that EIIA overexpression has deleterious effects on *S. mutans*, and that negative regulation of SMU_313 by SmsR4 helps maintain low intracellular levels of $EIIA^{sorb}$.

Defining the functional utility of a sRNA can be one of the most challenging tasks in the field of sRNA biology, and to date no sRNAs have been functionally described in *S. mutans*. Our work defining the regulatory role of SmsR4 provides valuable context for future work investigating the function of sRNAs within the species. Because we have shown that SmsR4 binds to SMU_313, a component of the sorbitol PTS (**Chapter 3, Figures 11, 12, 13**), and that the absence of SmsR4 from the genome negatively impacts *in vivo* growth in the sugar-alcohol sorbitol and, to a lesser extent, xylitol (**Chapter 3, Figure 10)**, it is clear that SmsR4 plays a valuable role in regulating some aspect of sugar metabolism. Combined with data describing the role of SgrS in controlling sugar-phosphate stress in gram-negative bacteria (39), the discovery of SmsR4 in a gram-

positive species suggests that the regulation of sugar metabolism through sRNAs is abundant throughout most, if not all, bacterial species. Although the exact molecular mechanism by which SmsR4 facilitates growth is currently only speculative, with future research SmsR4 may join the well-studied sRNA SgrS on the list of important sRNAs that regulate metabolic networks (40).

II. FUTURE CONSIDERATIONS FOR sRNA RESEARCH

An increasing abundance of new research illuminates the importance of sRNAs in regulating a variety of networks within the prokaryotic cell. Despite the recent advances, a lack of standardization for sRNA discovery and functional analyses brings challenges to the field. The following section explores considerations for further research into sRNA biology as advised by my doctoral work on documenting the evolution, prevalence and function of sRNAs in both gram-positive and gram-negative bacteria.

A. Improving whole-genome scans

A multitude of bioinformatical pipelines exist for discovering novel sRNAs, creating a spectrum of approaches that vary in how liberally or conservatively they define a putative sRNA transcript. Some purely computational methods bin transcripts based on size and abundance, detecting sometimes thousands of tiny transcripts (<50 nt) that are then automatically classified as sRNAs. Our sRNA detection pipeline falls on the other end of the spectrum: we only define new sRNA transcripts that we are confident have expression profiles similar to previously reported sRNAs, such as those in IGRs or originating from 5′ or 3′ UTRs. However, because of the visual nature of our sRNA detection method, it has the potential to miss novel sRNA transcripts due to human error in scanning a genome several million nucleotides in length. Additionally, the nature of coverage plot mapping can cause significantly expressed sRNAs to become buried under highly expressed neighboring transcripts. sRNA scans must take both sensitivity and specificity into account in order to avoid classifying spurious transcripts as novel sRNAs while still retaining transcripts of interest for analysis. A future recommendation for

whole-genome sRNA scans would be to combine a bioinformatics approach to identify potential transcripts of interested based on criteria about expression level and genome location (i.e. location in an IGR), followed by a visual scan of transcriptomics data to further validate the expression profile of these putative transcripts and confirm small transcripts of interest.

The most basic challenge of sRNA biology is to identify the tiny sRNA transcripts of interest among a sea of transcriptomics data. Traditionally, IGRs have been the target of most scans for novel sRNAs as they consistently provide a location for easily identifiable small transcripts (43). As demonstrated by my work examining the function of the intergenically encoded sRNA SmR4 in *S. mutans*, these IGRs are indeed a rich source for novel sRNAs; however, focusing solely on IGRs overlooks the majority genome. Our sRNA detection pipeline does not always have the sensitivity required to detect putative sRNAs arising from the RNAase cleavage of 3′ end of protein coding genes. However, our pipeline could be expanded to incorporate known RNase cleavage sites as well as predicted stem-loop features that facilitate the cleavage of mature sRNAs from mRNA transcripts. RNase processing of sRNA transcripts has been well documented in gram-negative bacteria (5, 44, 45), but gram-positive species like *S. mutans* also encode multiple active RNases (33, 46), so this approach to increase sRNA detection sensitivity could be implemented across all prokaryotic species. Additionally, comparing traditional RNA-seq data with a dataset of reads processed with terminator exonuclease (TEX), which digests 5′ monophosphates, can help enrich for primary transcripts and provide data on which transcripts may arise from RNase processing (28). Even more in-depth studies of transcriptional start and stop sites from experimental

approaches like term-seq, which identifies the 3′ end of all transcripts (47), can provide an additional layer of transcriptional mapping to sRNA detection pipelines. The more layers of detail investigators are able to add to transcriptomics data the more complete and useful sRNA scans become in describing novel sRNA expression.

Some pipelines include an enrichment step of sRNA targets before sequencing to remove larger mRNA transcripts, thus increasing the prevalence of small transcripts. This approach comes with the obvious potential benefit of increasing the number of sRNAs found within a transcriptomics dataset, but also presents drawbacks. Primarily, enriching for small transcripts disrupts the natural balance of transcriptomics data generated in favor of these tiny RNAs. This enrichment eliminates the usefulness of RNA-seq data for measuring gene expression of known genes so that the data may only be used for sRNA detection. Additionally, the process of enriching for small transcripts may also lead to the inadvertent selection for transcriptional noise over useful sRNAs. Being able to view transcriptomics data on genes located close to sRNAs can often yield useful clues about the expression of sRNAs relative to flanking genes, such as with sRNAs cleaved from protein-coding genes. Finally, the hallmark of sRNA regulation is the decreasing abundance of its target mRNA. Enriching for small fragments distorts the RNA-seq data on these potential targets, eliminating the ability to utilize the same dataset to observe mRNA expression.

B. Limitations to sRNA research

An obvious limitation to any whole-genome sRNA scan is that it can only detect transcripts that are actually being produced at the moment of RNA isolation, and

furthermore survive the isolation and RNA-seq library preparation process. Since sRNAs are often expressed only in very specific conditions during a small window in time, picking the correct growth conditions to survey is essential. Discovering novel sRNAs is like finding a needle in a haystack – only first, you must pick the right haystack to search. This problem was illustrated by our hunt for SmsR4 in short-term stress conditions. SmsR4 was first identified in our preliminary scan in nutrient-rich media, leading us to assume that SmsR4 was a fairly abundantly expressed sRNA that would play a role during short term stress conditions, since stress experiments were conducted at a timepoint similar to the mid-log TH growth initially surveyed. However, the transcript was almost completely absent from our short-term stress data. This result led us to reconsider the expression profile of SmsR4 and search for the sRNA in different growth conditions, focusing instead on sugar-alcohol metabolism after an initial phenotypic screen using Biolog plates suggested differential growth in sorbitol and *in silico* models predicted a sorbitol PTS component as an SmsR4 target. Choosing the correct haystack in which to search was critical to finding and eventually defining the function of this sRNA.

Once the specific growth conditions and time point have been chosen and RNA-seq libraries constructed, finding putative sRNA transcripts is not difficult. Depending on the detection pipeline, dozens to thousands of novel short RNA transcripts can be pulled form RNA-seq datasets; however, the usefulness of these detected transcripts then comes into question. Several common molecular biology techniques can be used to confirm the presence of putative sRNA transcripts, with Northern blot being considered the gold standard. The usefulness of Northern blotting above simply quantitative measurements, like qPCR, is the ability to detected multiple transcripts produced from RNase processing

of larger transcripts. In species that contain known RNA binding proteins such as Hfq, protein/RNA co-immunoprecipitation assays allow for the isolation of sRNAs that can then be confirmed through qPCR or Northern blot. We were able to employ this technique to isolate proto-OxyS sequences bound to Hfq in *S. marcescens*, confirming the presence of the sRNA transcript and its prevalence during oxidative stress. Because of the generally brief nature of sRNA expression and their inherently unstable form, sRNAs can be difficult to detect; however, I advise that a clear pattern of expression for a putative sRNA transcript be first established via Northern blot before proceeding to functional analysis. As I learned while describing a functional role for SmsR4, it can be nearly impossible to discern a phenotype for a sRNA without first having a clear understanding of the time and conditions in which it is expressed.

Due to the challenge of determining sRNA function, *in silico* target prediction can seem like an easy way to gain vast amounts of information about sRNA targets, and it is tempting to rely solely on such predictive methods for defining functional partners of sRNAs. Many whole-genome screens provide a predictive list of potential sRNA targets along with the putative transcripts they have identified. But the usefulness of these predictive lists is extremely limited: since sRNA-mRNA interactions are generally predicted based on base-pairing in small, single stranded seed-regions that are available for binding based on secondary structure, it is easy for computational methods to identify false targets for sRNAs. Equally as important is identifying the expression pattern of the sRNA and considering the putative mRNA targets that may exist during the same time period that the sRNA is transcribed. Additionally, a sRNA may have multiple targets during different stages of growth or during different conditions. For example, the

potential role for SmsR4 to regulate other PTS components with homology to SMU_313 is an intriguing avenue for future research.

The most helpful and productive functional analyses of sRNAs combine a bioinformatical and wet-lab approach to functional classification. Although sRNAs were first identified decades ago (48), the advent of RNA-seq technology has illuminated their central role in prokaryotic biology. Deep sequencing techniques are necessity for studying sRNA transcripts, which often have no easily observable phenotypic impact on the cell. The most canonical function of sRNAs is negative regulation of mRNA transcripts, but researchers must consider other potential modes of sRNA regulation aside from merely looking for downregulation of targets in response to the presence of the sRNA, as that is not always a reliable measure of sRNA regulation (49). The most helpful technique for sRNA functional analysis is not reliance on one assay, but a holistic approach that utilizes many different experimental techniques. Because sRNA regulation is often incredibly nuanced, results from one experiment should be verified by a supplemental technique. However, contradictory or negative results should not always be considered definitive, but instead advise further exploration.

C. Moving forward

How does understanding sRNA utility in a bacterial species contribute to overall scientific knowledge? Some sRNAs provide only the most nuanced fine-tuning to regulatory networks, so much so that a deletion of the sRNA often does not provide a noticeable growth phenotype. Other sRNAs, like SgrS, contribute significantly to the survival and fitness during growth in the conditions they regulate. It may be tempting for

those outside the field to want to ignore examples of sRNAs like the former as minor players in already complex regulatory networks. More tempting still may be the impulse to direct overall efforts on understanding transcriptional profiles of prokaryotes away from sRNAs and to instead focus on the effect of transcripts for protein coding genes, which arguably play a much more significant role in stress responses and important cellular pathways.

It is not only useful to study all regulatory mechanisms used by a prokaryotic cell in order to thoroughly understand the biology of the organism, but it is essential to do so to fully describe changes the cell goes through in response to stress and evolving environmental conditions. Ignoring the role of sRNAs in prokaryotic regulatory networks is akin to studying traffic flow at an intersection while ignoring the role of yellow lights — they are not the definitive start and stop signal but have significant bearings on the flow of cars on roadways. The fact that the impact of sRNAs may be small does not discount their importance in helping the organism grow efficiently by tightly regulating mRNA expression.

By understanding the role of sRNAs in regulating cellular networks, we can better understand the ebb and flow of the entire pathway. During its quest for survival and replication a cell does not want to waste energy on useless transcription, and yet pervasive transcription of noncoding RNA remains a ubiquitous phenomenon across the prokaryotic domain (50). These genetic elements serve no function as mRNAs, but the retention of this phenomenon suggests its utility for bacteria in general. One likely explanation for pervasive transcription is the potential benefit to a cell when these small transcripts can become integrated into regulatory networks. Despite the energetic cost of

spurious transcription, the potential reward for acquiring new sRNAs seems to outweigh the initial energy investment. Thus, it seems that fine-tuning transcription is an essential component of prokaryotic biology.

My findings on SmR4 expression and utility illustrate the importance of investigating sRNA biology in the context of overall stress response. Despite being experimentally well-documented, sugar-phosphate stress is still a relatively poorly understood phenomenon; however, the role of the sRNA SgrS in regulating sugar-phosphate stress in some enteric bacteria has been clearly defined (42). Finding another sRNA in the gram-positive bacterium *S. mutans* that likely participates in the same type of regulation highlights the importance of tightly regulating sugar-phosphate stress in all bacteria. Discovering SmsR4 regulation of the EIIA component specific for sorbitol phosphorylation during growth in xylitol further demonstrates the significance of investigating sRNA utility in order to understand overall metabolic processes. Because the sorbitol PTS is thought to be sugar-specific, the discovery of sorbitol PTS genes as a target of SmsR4 under conditions of xylitol stress was initially surprising. However, this finding corroborated decades old research exploring the effects of concomitant growth in xylitol or sorbitol (51).

Some functional sRNA studies cite investigating sRNAs as drug targets as a purpose for study and targeting bacterial sRNAs has been proposed as a promising avenue for combating infections by multiple drug resistant bacteria (52, 53). When considering the potential of sRNAs as therapeutic targets, it is most helpful to think about the entire regulatory network that the sRNA participates in instead of focusing narrowly on the sRNA transcript. Our discoveries of the evolutionary path of OxyS indicates that

the ability to respond to oxidative stress may be conserved in Enterobacteriaceae, so therapeutic targets for such bacteria may focus on abating their response to oxidative stress. The utility of SmR4 in *S. mutans* and its conservation in other streptococci suggests the importance of tight regulation of metabolic networks to bacterial fitness, and thus targets may be developed to enhance the effects of metabolic disorders, such as sugar-phosphate stress, in these bacteria. When a sRNA has become ingrained into a regulatory network it can emphasize the importance of that particular stress response to the bacteria, leading to future discoveries of novel therapeutics that can take advantage of these specific regulatory networks.

The central dogma of molecular biology - that DNA is transcribed into RNA which is translated into proteins – provides a straightforward paradigm to explain gene regulation; however, my work has shown that this process is vastly more complicated, even in relatively simple prokaryotic cells. Although no scientific knowledge is gained in a truly straightforward path my research experience uncovering the evolution, prevalence, and function of sRNAs has been anything but linear, and over the past four years I have found my determination, dedication, knowledge, and resilience continually tested. However, I have been left with a deep and enduring respect and fascination for the complexity of the bacterial cell. By studying sRNA biology, I gained an intimate look into the inner workings of prokaryotic transcriptional control and my appreciation for the nuisance and intricacy of transcriptional regulation within all cells grew immensely. I believe that this perspective is not unique to my experience, and that micro- and molecular- biologists from all disciplines could benefit from gaining an understanding of

how post-transcriptional regulation by sRNAs impacts their area of study as science continues to reveal the staggering complexity of life.

MATERIALS AND METHODS

All methods not detailed in chapters two and three are found below.

Hfq co-immunoprecipitation

Serratia marcescens ATCC 13880 was cultured and exposed to oxidative stress as described in chapter two. Cells were pelleted by centrifugation at 10,000xg for 10 minutes at 4°C and washed with sterile 4°C PBS. The Hfq-co-IP protocol was performed as described previously (54) with several modifications. Cells were mechanically disrupted in cycles of 1-minute bead-beating followed by a 2-minute chill on ice repeated four times using 0.1 mm zirconium beads. Cleared lysate was removed and combined with either 10 µg of purified *Serratia marcescens* Hfq antibody diluted to 200 µl in PBST (ABclonal), or anti-β-actin monoclonal antibody diluted to 200 µl in PBST (Thermo Fischer Scientific), or 200 µl of PBST without any antibodies. The mixtures were incubated with rotation at 4°C for 1 hour and were combined with 50 µl (1.5 mg) of Dynabeads Protein A (Thermo Fischer Scientific) and incubated with rotation at 4°C for 1 hour. After supernatant removal and washing, the Dynabeads were resuspended in 200 µl of nuclease-free water followed by a 5-minute incubation at room temperature. RNA was isolated from the beads using 1 mL of TRI reagent as described in chapters two and three.

RIL-seq

Serratia marcescens ATCC 13880 was cultured and exposed to oxidative stress as described in chapter two. Cell lysate was prepared as previously described (58) using a

Stratalinker UV Crosslinker 2400 for RNA-protein crosslinking and 0.1 mm zirconium beads for mechanical disruption. For immunoprecipitation of Hfq-RNA complexes, the cleared lysate was removed and combined with 10 µg of purified *Serratia marcescens* Hfq (ABclonal) antibody diluted in 200 µl of PBST. Cell lysate-antibody mixtures were incubated with rotation at 4°C for 1 hour, then combined with 50 µl (1.5 mg) of Dynabeads Protein A (Thermo Fischer Scientific) and incubated with rotation at 4°C for 1 hour. Trimming, ligation, and Hfq release were performed as previously described (58) using 1 µl of RNase A/T1 mix (Thermo Fischer Scientific) for trimming and 100 units of T4 RNA Ligase (New England Biolabs) for ligation. RNA was extracted with 1 mL TRI reagent as described in chapters two and three. Paired-end RNA-seq reads were generated at the Yale Center for Genome Analysis. Adapters were trimmed and reads were sorted based on quality with Trimmomatic (55) and rRNA was removed with SortMeRNA (56). The remaining paired reads were processed using the RILseq package as described previously using the *Serratia marcescens* ATCC 13880 chromosome (CP041233.1) and plasmid (CP041234.1) genomes, with a read length of 30 and allowed mismatches set at 3 for the single fragment mapping (64). All chimeric reads containing rRNA were discarded from the final analysis.

REFERENCES

1. H. A. Dutcher, R. Raghavan, Origin, evolution, and loss of bacterial small RNAs. *Microbiol Spectr* **6** (2018).

2. F. R. Kacharia, J. A. Millar, R. Raghavan, Emergence of new sRNAs in enteric bacteria is associated with low expression and rapid evolution. *J. Mol. Evol.* **84**, 204–213 (2017).

3. R. Raghavan, F. R. Kacharia, J. A. Millar, C. D. Sislak, H. Ochman, Genome rearrangements can make and break small RNA genes. *Genome Biol Evol* **7**, 557–566 (2015).

4. S. Wachter, R. Raghavan, J. Wachter, M. F. Minnick, Identification of novel MITEs (miniature inverted-repeat transposable elements) in *Coxiella burnetii*: implications for protein and small RNA evolution. *BMC Genomics* **19** (2018).

5. Y. Chao, *et al.*, In vivo cleavage map illuminates the central role of RNase E in coding and non-coding RNA pathways. *Mol Cell* **65**, 39–51 (2017).

6. S. Altuvia, A. Zhang, L. Argaman, A. Tiwari, G. Storz, The *Escherichia coli* OxyS regulatory RNA represses fhlA translation by blocking ribosome binding. *EMBO J.* **17**, 6069–6075 (1998).

7. L. Argaman, S. Altuvia, fhlA repression by OxyS RNA: Kissing complex formation at two sites results in a stable antisense-target RNA complex. *J. Mol. Biol.* **300**, 1101–1112 (2000).

8. S. Barshishat, *et al.*, OxyS small RNA induces cell cycle arrest to allow DNA damage repair. *EMBO J.* **37**, 413–426 (2018).

9. A. Zhang, *et al.*, The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein. *EMBO J.* **17**, 6061–6068 (1998).

10. A. Zhang, K. M. Wassarman, J. Ortega, A. C. Steven, G. Storz, The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol. Cell* **9**, 11–22 (2002).

11. S. O. Kim, *et al.*, OxyR: A molecular code for redox-related signaling. *Cell* **109**, 383–396 (2002).

12. Y. Chao, K. Papenfort, R. Reinhardt, C. M. Sharma, J. Vogel, An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* **31**, 4005–4019 (2012).

13. S. Melamed, *et al.*, Mapping the small RNA interactome in bacteria using RIL-seq. *Nature Protocols* **13**, 1–33 (2018).

14. S. Durica-Mitic, B. Görke, Feedback regulation of small RNA processing by the cleavage product. *RNA Biol* **16**, 1055–1065 (2019).

15. L. Barquist, S. W. Burge, P. P. Gardner, Studying RNA homology and conservation with Infernal: From single sequences to RNA families. *Curr Protoc Bioinformatics* **54**, 12.13.1-12.13.25 (2016).

16. S. Gottesman, G. Storz, Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harb Perspect Biol* **3** (2011).

17. E. Skippington, M. A. Ragan, Evolutionary dynamics of small RNAs in 27 *Escherichia coli* and *Shigella* genomes. *Genome Biol Evol* **4**, 330–345 (2012).

18. E. Holmqvist, E. G. H. Wagner, Impact of bacterial sRNAs in stress responses. *Biochem. Soc. Trans.* **45**, 1203–1212 (2017).

19. D. Sinha, *et al.*, Redefining the small regulatory RNA transcriptome in *Streptococcus pneumoniae* serotype 2 strain D39. *Journal of Bacteriology* **201** (2019).

20. R. A. Tesorero, *et al.*, Novel regulatory small RNAs in *Streptococcus pyogenes*. *PLoS One* **8**, e64021 (2013).

21. S. S. Liu, *et al.*, Analysis of sucrose-induced small RNAs in *Streptococcus mutans* in the presence of different sucrose concentrations. *Appl. Microbiol. Biotechnol.* **101**, 5739–5748 (2017).

22. S. Liu, *et al.*, Analysis of small RNAs *in Streptococcus mutans* under acid stress—a new insight for caries research. *International Journal of Molecular Sciences* **17**, 1529 (2016).

23. L. Zeng, *et al.*, Gene regulation by CcpA and catabolite repression explored by RNA-Seq in *Streptococcus mutans*. *PLoS One* **8** (2013).

24. A. D. Garst, A. L. Edwards, R. T. Batey, Riboswitches: Structures and mechanisms. *Cold Spring Harb Perspect Biol* **3** (2011).

25. E. Loh, F. Righetti, H. Eichner, C. Twittenhoff, F. Narberhaus, RNA thermometers in bacterial pathogens. *Microbiology Spectrum* **6** (2018).

26. T. S. Lotz, B. Suess, Small-molecule-binding riboswitches. *Microbiol Spectr* **6** (2018).

27. A. DeLoughery, J.-B. Lalanne, R. Losick, G.-W. Li, Maturation of polycistronic mRNAs by the endoribonuclease RNase Y and its associated Y-complex in *Bacillus subtilis*. *PNAS* **115**, E5585–E5594 (2018).

28. D. Ryan, L. Jenniches, S. Reichardt, L. Barquist, A. J. Westermann, A high-resolution transcriptome map identifies small RNA regulation of metabolism in the gut microbe *Bacteroides thetaiotaomicron*. *Nat Commun* **11**, 3557 (2020).

29. S. Chareyre, P. Mandin, Bacterial iron homeostasis regulation by sRNAs. *Microbiol Spectr* **6** (2018).

30. J. E. Barrick, N. Sudarsan, Z. Weinberg, W. L. Ruzzo, R. R. Breaker, 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA* **11**, 774–784 (2005).

31. J. A. Lemos, R. A. Burne, A model of efficiency: Stress tolerance by *Streptococcus mutans*. *Microbiology* **154**, 3247–3255 (2008).

32. E. W. Miller, T. N. Cao, K. J. Pflughoeft, P. Sumby, RNA-mediated regulation in gram-positive pathogens: An overview punctuated with examples from the group A Streptococcus. *Mol Microbiol* **94**, 9–20 (2014).

33. J. Merritt, Z. Chen, N. Liu, J. Kreth, Posttranscriptional regulation of oral bacterial adaptive responses. *Curr Oral Health Rep* **1**, 50–58 (2014).

34. A. T. Cavanagh, K. M. Wassarman, 6S RNA, a global regulator of transcription in *Escherichia coli*, *Bacillus subtilis*, and beyond. *Annual Review of Microbiology* **68**, 45–60 (2014).

35. S. Wachter, *et al.*, A CsrA-binding, trans-acting sRNA of *Coxiella burnetii* is necessary for optimal intracellular growth and vacuole formation during early infection of host cells. *J. Bacteriol.* **201**, e00524-19 (2019).

36. J. M. Tanzer, A. Thompson, Z. T. Wen, R. A. Burne, *Streptococcus mutans*: fructose transport, xylitol resistance, and virulence. *J Dent Res* **85**, 369–373 (2006).

37. M. B. Kery, M. Feldman, J. Livny, B. Tjaden, TargetRNA2: Identifying targets of small regulatory RNAs in bacteria. *Nucleic Acids Res* **42**, W124–W129 (2014).

38. M. Mann, P. R. Wright, R. Backofen, IntaRNA 2.0: Enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Res* **45**, W435–W439 (2017).

39. J. B. Rice, C. K. Vanderpool, The small RNA SgrS controls sugar–phosphate accumulation by regulating multiple PTS genes. *Nucleic Acids Res* **39**, 3806–3819 (2011).

40. M. Bobrovskyy, C. K. Vanderpool, The small RNA SgrS: Roles in metabolism and pathogenesis of enteric bacteria. *Front Cell Infect Microbiol* **4** (2014).

41. W.-Y. Jeng, *et al.*, The negative effects of KPN00353 on glycerol kinase and microaerobic 1,3-propanediol production in *Klebsiella pneumoniae*. *Front. Microbiol.* **8** (2017).

42. J. Choi, *et al.*, *Salmonella* pathogenicity island 2 expression negatively controlled by EIIANtr-SsrB interaction is required for *Salmonella* virulence. *Proceedings of the National Academy of Sciences* **107**, 20506–20511 (2010).

43. C.-H. Tsai, R. Liao, B. Chou, M. Palumbo, L. M. Contreras, Genome-wide analyses in bacteria show small-RNA enrichment for long and conserved intergenic regions. *Journal of Bacteriology* **197**, 40–50 (2015).

44. N. De Lay, S. Gottesman, RNase E finds some sRNAs stimulating. *Mol Cell* **47**, 825–826 (2012).

45. K. J. Bandyra, B. F. Luisi, RNase E and the high-fidelity orchestration of RNA metabolism. *Microbiol Spectr* **6** (2018).

46. X. Chen, N. Liu, S. Khajotia, F. Qi, J. Merritt, RNases J1 and J2 are critical pleiotropic regulators in *Streptococcus mutans*. *Microbiology (Reading)* **161**, 797–806 (2015).

47. D. Dar, *et al.*, Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science* **352**, aad9822 (2016).

48. K. M. Wassarman, Small RNAs in bacteria: Diverse regulators of gene expression in response to environmental changes. *Cell* **109**, 141–144 (2002).

49. R. Faigenbaum-Romm, *et al.*, Hierarchy in Hfq chaperon occupancy of small RNA targets plays a major role in their regulation. *Cell Rep* **30**, 3127-3138.e6 (2020).

50. M. Lybecker, I. Bilusic, R. Raghavan, Pervasive transcription: Detecting functional RNAs in bacteria. *Transcription* **5** (2014).

51. S. Assev, G. Rølla, Sorbitol increases the growth inhibition of xylitol on *Streptococcus mutans* OMZ 176. *Acta Pathol Microbiol Immunol Scand B* **94**, 231–237 (1986).

52. G. Parmeciano Di Noto, M. C. Molina, C. Quiroga, Insights into non-coding RNAs as novel antimicrobial drugs. *Front. Genet.* **10** (2019).

53. P. Dersch, M. A. Khan, S. Mühlen, B. Görke, Roles of regulatory RNAs for antibiotic resistance in bacteria and their potential value as novel drug targets. *Front Microbiol* **8** (2017).

54. V. Pfeiffer, *et al.*, A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the Salmonella core genome. *Mol. Microbiol.* **66**, 1174–1191 (2007).

55. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

56. E. Kopylova, L. Noé, H. Touzet, SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).

APPENDIX

Chapter Two Supplemental Tables

Table S1. Predicted age of sRNAs
File name: TableS1.csv
File size: 7 kb
Required software: None

Table S2. Predicted age of regulatory proteins
File name: TableS2.csv
File size: 2 kb
Required software: None

Table S3. sRNAs that overlap ORFs
File name: TableS3.csv
File size: 8 kb
Required software: None

Table S4. Accessions for Enterobacterales genomes used to determine sRNA age and prevalence
File name: TableS4.csv
File size: 175 kb
Required software: None