Genome-wide identification of novel sRNAs in *Streptococcus mutans*

Madeline C Krieger[1,2], 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 Molecular Microbiology and Immunology, The University of Texas at San Antonio, San Antonio, TX, USA.

Author for correspondence:

Rahul Raghavan | Email: rahul.raghavan@utsa.edu

Key words: sRNA, *Streptococcus mutans*, 6S RNA, sugar-phosphate stress, sorbitol, transcriptome

ABSTRACT

Streptococcus mutans is a major pathobiont involved in the development of dental caries. Its ability to utilize numerous sugars and to effectively respond to environmental stress promotes *S. mutans* proliferation in oral biofilms. Because of their quick action and low energetic cost, non-coding small RNAs (sRNAs) represent an ideal mode of gene regulation in stress response networks, yet their roles in oral pathogens have remained largely unexplored. We identified 15 novel sRNAs in *S. mutans* and show that they respond to four stress-inducing conditions commonly encountered by the pathogen in human mouth: sugar-phosphate stress, hydrogen peroxide exposure, high temperature, and low pH. To better understand the role of sRNAs in *S. mutans*, we further explored the function of the novel sRNA, SmsR4. Our data demonstrate that SmsR4 regulates the EIIA component of the sorbitol phosphotransferase system, which transports and phosphorylates the sugar alcohol sorbitol. The fine-tuning of EIIA availability by SmsR4 likely promotes *S. mutans* growth while using sorbitol as the main carbon source. Our work lays a foundation for understanding the role of sRNAs in regulating gene expression in stress response networks in *S. mutans* and highlights the importance of the underexplored phenomenon of posttranscriptional gene regulation in oral bacteria.

IMPORTANCE

Small RNAs (sRNAs) are important gene regulators in bacteria, but the identities and functions of sRNAs in *Streptococcus mutans*, the principal bacterium involved in the formation of dental caries, are unknown. In this study, we identified 15 putative sRNAs in *S. mutans* and show that they respond to four common stress-inducing conditions present in human mouth: sugar-phosphate stress, hydrogen peroxide exposure, high temperature, and low pH. We further show that the novel sRNA SmsR4 likely modulates sorbitol transport into the cell by regulating SMU_313 mRNA, which encodes the EIIA subunit of the sorbitol phosphotransferase system. Gaining a better understanding of sRNA-based gene regulation may provide new opportunities to develop specific inhibitors of *S. mutans* growth, thereby improving oral health.

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* 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 act by transporting carbohydrates across the cell membrane and immediately phosphorylating them for intracellular retention and entry into glycolysis (7). While glucose is the preferred carbon source for *S. mutans,* other PTSs, including the sorbitol (glucitol) PTS, are inducible in its absence. Sorbitol is a sugar-alcohol found naturally in many fruits and is a popular low-calorie anti-cariogenic sweetener (7, 8). The anticariogenic action of sorbitol is due to the relatively low amounts of acid produced during sorbitol fermentation by *S. mutans* compared to that of glucose or sucrose metabolism (8, 9).

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 and multiple stressors (10). Small RNAs (sRNAs) are non-coding transcripts that typically bind to their target mRNAs to regulate gene expression (11). Due to their low energetic cost, fast action, and capacity for co-degradation with mRNA targets, sRNAs are an ideal mode of posttranscriptional control when a bacterium is experiencing environmental stress (12). Surprisingly little is known about sRNA-based regulation in *S. mutans* and there have been no functional analyses of sRNA utility in *S. mutans* to date. In this study, we identified 15 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

Identification and validation of novel sRNAs

We used an RNA-seq-based approach that we developed previously (13–15) to identify 15 novel sRNAs — named SmsR for "*S. mutans* small RNA" — expressed in *S. mutans* during exponential (OD_{600} 0.5-0.6) growth in Brain Heart Infusion (BHI) broth (**Figure 1**). Nine of the novel sRNAs overlapped with candidates predicted by a previous genome-wide scan and a search against the Rfam database identified three of the sRNAs as RNase P, 6S RNA, and tmRNA (16, 17) (**Table 1**). All candidate sRNAs have putative -10 promoter sites, and 10 have predicted rho-independent terminators at their 3′ ends (**Table 2**), indicating these as authentic sRNAs. We further verified all 15 putative sRNA transcripts using Northern blot (**Figure 1**). Interestingly, in a few cases, higher molecular weight bands in addition to the sRNAs were observed, presumably because the sRNAs were cleaved from longer transcripts, as observed in other bacteria (18–23).

sRNAs respond to environmental stress

We evaluated the expression of the 15 novel sRNAs following *S. mutans* exposure to four stress-inducing conditions relevant to the human oral cavity: sugar-phosphate stress, hydrogen peroxide (H_2O_2), high temperature, and low pH (**Table 3**). All sRNAs showed patterns of differential expression in at least one of these conditions, suggesting likely roles in stress tolerance networks (**Figure 2**). Acid and sugar-phosphate stress induced differential expression of 12 sRNAs, while

97 eight sRNAs were affected by heat stress and four responded to oxidative stress (**Figure 2**).
98 Upregulation of sRNA expression was more common than downregulation across all conditions
99 except for heat stress.
100
101 **SmsR14 and SmsR4 are unrelated sRNAs**
102 6S RNA, a global regulator of transcription, is widely conserved among bacteria (24–26). Two
103 copies of 6S RNA are present in *Bacillus subtilis*, a model gram-positive bacterium: one in the
104 intergenic region between genes for azoreductase and RecQ helicase, and the other between two
105 tRNA-associated genes (**Figure 3**) (27). The prevalence of 6S RNA in *Streptococcus* has not
106 been examined previously. SmsR14, one of the newly-discovered sRNAs, is a 6S homolog and is
107 located between *rarA* and tRNA-Lys genes in *S. mutans* (**Figure 3**). A covariance modeling
108 (cm)-based search for 6S homologs in other *Streptococcus* species revealed only one copy of the
109 sRNA in all genomes we analyzed (**Figure 4**). Interestingly, another novel sRNA (SmsR4)
110 occupies the genomic location at which 6S RNA is present in *Escherichia coli*, i.e., upstream of
111 the gene encoding 5-formyltetrahydrofolate cyclo-ligase (5-FTC) (**Figure 3**) (24, 26, 28).
112 SmsR4, however, is much shorter than 6S (110 nt vs 194 nt), has a predicted secondary structure
113 that is distinct from that of 6S, and is encoded on the opposite strand as 5-FTC gene (**Figure 3**).
114 These differences suggest that SmsR4 is unrelated to 6S RNA and could have distinct functions
115 in *S. mutans*.
116
117 **SmsR4 arose in the Pyogenes-Equinus-Mutans clade of *Streptococcus***
118 The genus *Streptococcus* could broadly be classified into two clades: Mitis-Suis and Pyogenes-
119 Equinus-Mutans (29). Our cm-based search identified SmsR4 homologs only in the Pyogenes-

Equinus-Mutans clade (**Figure 4**). Further, an evolutionary reconstruction showed that SmsR4 arose at the root of this clade but was later lost in the common ancestor of Sobrinus, Salivarus, Halotolerans and Porci subclades. 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.

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

To identify the functions of SmsR4 in *S. mutans*, we first confirmed the 5′ and 3′ boundaries of SmsR4 and generated an SmsR4-deletion (DEL) strain. We measured the growth of wild-type (WT) and DEL strains using a phenotypic microarray (30). In this analysis, DEL had reduced growth in comparison to WT in media containing sorbitol as the sole carbon source (**Figure S1**). We verified this phenotype by measuring bacterial growth in BTR medium that contained either 0.5% glucose (BTR-G) or 0.5% sorbitol (BTR-S). In this assay DEL had weaker growth than WT in BTR-S despite displaying a growth pattern identical to that of WT in BTR-G (**Figure 5**). The growth defect of DEL in BTR-S was overcome by a complementation strain (COMP) in which SmsR4 was expressed on the shuttle vector pDL278 (31, 32), indicating that the loss of SmsR4 caused DEL's reduced growth. Probably because multiple copies of the SmsR4-carrying plasmid were present in each cell, COMP grew considerably better than WT in BTR-S, again denoting a role for the sRNA in promoting robust bacterial growth when using sorbitol as the carbon source.

SmsR4 regulates the EIIA component of the 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 encoding enzyme IIA of the sorbitol PTS (**Figure 6**; **Table S1)**. An SMU_313-deletion strain failed to achieve meaningful growth in BTR-S but grew at comparable levels to WT in BTR-G (**Figure S2**), demonstrating that SMU_313 is essential for growth on sorbitol but not glucose. We confirmed the interaction between SmsR4 and SMU_313 mRNA with an RNA-RNA electrophoretic mobility shift assay (EMSA), which showed that SmsR4 binds well to the 5′ UTR of SMU_313 mRNA (**Figure 7A**). The putative-SmsR4-binding site identified by *in silico* analyses (**Figure 6**) is likely required for this interaction, as a mutation of the predicted SmsR4-binding site in SMU_313 inhibited its *in vitro* interaction with SmsR4 (**Figure 7B**). Taken together, our data indicate that SmsR4 modulates sorbitol import into the cell, likely by antagonizing translation of SMU_313 mRNA, which encodes the EIIA subunit of the sorbitol PTS.

DISCUSSION

sRNAs are critical for posttranscriptional gene regulation in bacteria, but their roles in the dental pathogen *S. mutans* have remained largely unknown. Prior to our study, a mostly bioinformatics-based analysis of gene expression in *S. mutans* grown with various carbon sources identified 243 sRNA candidates (17). In contrast, our approach combined genome-wide analysis with experimental validation and uncovered 15 sRNAs expressed under a single growth condition (BHI, OD_{600} 0.5-0.6) (**Figure 1**); hence, it is certainly possible that *S. mutans* transcribes additional undetected sRNAs in growth conditions different from those examined here. We

166 noticed several sRNAs that appeared to be processed from larger parent transcripts, e.g., SmsR3
167 and SmsR20 (**Figure 1**). The generation of sRNAs from processed mRNAs has been well
168 documented; for example, OppZ and CarZ are produced by RNase E cleavage of the *oppABCDF*
169 and *carAB* mRNAs in *Vibrio cholera*, ArgX is produced from *argR* mRNA in *Lactococcus*
170 *lactis*, and RsaC is formed via RNase III digestion of *mntACB* mRNA in *Staphylococcus aureus*
171 (20, 33–37). Alternatively, some of these sRNAs may be regulatory elements with multiple
172 products formed from premature transcriptional termination and transcriptional read-through
173 (38).

174 The genomic contexts of sRNAs likely have implications for their functions (**Figure 1**).
175 For instance, some of the sRNAs located in the 5′ UTRs of downstream genes (e.g., SmsR5,
176 SmsR6, SmsSR8, SmsR16, SmsR20, SmsR22) could function as riboswitch-like elements (39),
177 or sRNAs that are transcribed divergently from downstream protein-coding genes (e.g., SmsR1,
178 SmsR3, SmsR19) could function as antisense sRNAs that regulate cis-encoded targets (40, 41).
179 A few sRNA genes in other bacteria have been shown to contain small open reading frames
180 (ORFs) that encode proteins (42). We searched the 15 sRNA sequences for ORFs and found that
181 SmsR12 potentially encodes a small protein (33 amino acids). Although we could not detect a
182 ribosome binding site upstream of the ORF and a BlastP search did not produce any hits in other
183 bacteria, further studies would be required to determine whether the ORF is indeed functional.
184 Because sRNAs identified in this study exhibited differential expression under stress and their
185 existence as discrete transcripts were validated via Northern blot, it is likely that they represent
186 bona fide sRNAs participating in regulatory networks. Aside from SmsR4, the functions of these
187 novel sRNAs are currently unknown, but preliminary *in silico* target predictions suggest that
188 many of them are involved in processes critical to adaptation and virulence. For instance, SmsR2

and SmR22 are predicted to bind to PTS components, indicating that additional sRNAs may regulate sugar transport (**Table S1**), a process critical for *S. mutans* cariogenicity.

Among the 15 novel sRNAs, SmsR4 was intriguing because its genomic location is similar to that of 6S RNA in many proteobacteria (24), and as observed for 6S, SmsR4 was maximally expressed during the transition from exponential to stationary phase (**Figure S3**) (25). In gammaproteobacteria, expression of 6S is thought to be controlled by its linkage to the neighboring 5-FTC gene, which responds to nutrient limitation during the transition to stationary phase (24). In a similar manner, SMU_320, the 5-FTC homolog in *S. mutans,* could influence the expression of SmsR4; however, unlike with 5-FTC and 6S in gram-negative bacteria, SMU_320 and SmsR4 are encoded on opposite DNA strands and hence are not co-transcribed. Instead, a unique promoter likely initiates SmsR4 transcription (**Table 2**).

Because sugar transport is central to the proliferation and cariogenicity of *S. mutans*, it is unsurprising to find an sRNA regulator of its sorbitol usage. While further work is required to delineate the molecular details of its action, our results indicate that SmsR4 likely functions by modulating translation of the EIIA component of the sorbitol PTS. 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 (43). If left unchecked, phosphorylated sugars can accumulate in the cell and trigger sugar-phosphate stress (44). In *E. coli* and *Salmonella enterica,* the sRNA SgrS plays a key role in restoring glycolytic balance during sugar-phosphate stress by blocking the translation of mRNAs encoding the corresponding sugar-transporters (45, 46). In a similar fashion, SmsR4 could be modulating sorbitol intake by regulating the expression of SMU_313 ($EIIA^{sorb}$) to relieve *S. mutans* sugar-phosphate stress during the transition from exponential to stationary phase growth in sorbitol-containing growth medium. Alternatively,

212 modulation of SMU_313 expression by SmsR4 could impact alternative functions of EIIA. For
213 example, this protein has been shown to negatively impact glycerol metabolism in *Klebsiella*
214 *pneumoniae* and to reduce *S. enterica* virulence (47, 48). In conclusion, our *in silico* and
215 biochemical data both support a role for sRNAs as posttranscriptional regulators of sugar
216 transport through PTS systems. Further insights into these regulatory mechanisms may provide
217 new opportunities to develop specific inhibitors of *S. mutans* growth in the oral cavity.
218

219 **MATERIALS AND METHODS**

220 **Bacterial strains and growth assays**

221 Growth experiments were conducted by diluting overnight cultures of *S. mutans* UA159 in fresh
222 media. BTR broth base (1% Bacto-Tryptone, 0.1% yeast extract, 0.61% K_2HPO_4, 0.2% KH_2PO_4)
223 supplemented with either 0.5% glucose (BTR-G) or 0.5% sorbitol (BTR-S) was utilized for
224 growth assays. For experiments comparing the growth of WT, DEL, and COMP strains, 1 mg/ml
225 spectinomycin was added to the growth media to retain the pDL278 plasmid. All growth assays
226 were done at 37°C in an anaerobic chamber (5% hydrogen, 5% carbon dioxide, 90% nitrogen).
227

228 **Phenotypic microarray**

229 Biolog Phenotypic microarray assays were conducted per manufacturer's recommendations (36).
230 Briefly, overnight cultures from single bacterial colonies were diluted in fresh Brain Heart
231 Infusion (BHI) broth and grown to an OD_{600} of 0.6-0.7 at 37°C in an anaerobic chamber. Cells
232 were collected via centrifugation (3,000xg, 5 min) and washed twice with PBS and resuspended
233 to an OD_{600} of 0.4 in IF-0a GN/GP base. Inoculating fluid was prepared and combined with cells
234 at 81% turbidity and 100 µl of the mixture was added to each well and overlaid with 40 µl

mineral oil. Temperature was maintained at 37°C while absorbance values at 590 nm and 750 nm were collected every 20 min for 24 h using a Multiskan Spectrum plate reader (Thermo Fischer Scientific). Results were obtained by subtracting the measurements at 750 nm from those at 590 nm, and the average of two replicates was used to construct a growth curve for each well.

sRNA discovery

Bacterial cultures were grown in BHI broth to an OD_{600} of 0.5-0.6. RNA stop solution (5% Phenol, 95% Ethanol) was added to bacterial cultures (1.25 mL stop solution per 10 mL culture) and cells were pelleted by centrifugation at 10,000xg for 10 min at 4°C. Bacterial pellets were resuspended in 1 ml of TRI reagent (Thermo Fischer Scientific) and total RNA was extracted using the manufacturer's protocol. RNA was resuspended in nuclease-free water and DNA was removed by TURBO DNase (Thermo Fischer Scientific) treatment. RNA sequencing was performed at the Yale Center for Genome Analysis using Illumina NovaSeq (paired-end, 150 bp). Raw reads have been deposited in the NCBI Sequence Read Archive under the accession number PRJNA726731. RNA-seq reads were processed using Trimmomatic to remove low-quality reads and adapters (49). 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 (13). The Artemis genome browser (50) was used for visual inspection of transcriptomics data and gggenes package in R (version 0.4.1) was used to draw Figure 1. The putative 5′ end of each sRNA was estimated from sites on transcriptional coverage plots with sharp increases in reads mapped per nucleotide. RNAalifold was used to predict secondary structure of SmsR4 and 6S RNA (51), and ARNold

was used to predict 3′ Rho-independent terminator structures (52). For sRNAs without predicted terminators, 3′ ends were defined as sites on transcriptional coverage plots with a sharp decrease in reads mapped per nucleotide. Potential open reading frames (ORFs) within sRNA genes were detected using ORF Finder (53).

Stress induction

Overnight cultures of *S. mutans* UA159 were diluted 1:100 in BTR-G and grown to an OD_{600} of 0.3-0.4 at 37°C in an anaerobic chamber. To induce 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 under the same conditions. 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 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). To induce acid stress, cultures were centrifuged and one half was resuspended in BTR-G, pH 7 (control), while the other half was resuspended in BTR-G, pH 5 (treatment), and incubated for 30 minutes at 37°C. Anaerobic conditions were maintained throughout the stress induction assays. RNA was extracted as described above. To confirm stress induction, qRT-PCR was used to measure the expression of stress marker genes previously associated with each tested stress condition (54–57) (**Figure S4**). Contaminating DNA was removed by TURBO DNase (Thermo Fischer Scientific) treatment and cDNA was synthesized with a High-Capacity cDNA Reverse Transcription Kit (Thermo Fischer Scientific) using random primers. qRT-PCR was performed using SYBR Green master mix (Thermo Fischer Scientific) and gene-specific primers (**Table S2**). RNA was

sequenced and RNA-seq reads were processed as described above. The DESeq2 package in R was used to determine differential gene expression of sRNAs under the four stress-inducing conditions compared to controls (58). Experiments were performed in triplicate for all conditions except for acid stress, which was performed in duplicate.

In vitro transcription

Amplification of gDNA for *in vitro* transcription was performed using PCR primers designed to incorporate a T7 promoter (**Table S2**). PCR products to be used as DNA templates were cleaned using a NucleoSpin Gel and PCR Clean-up kit (Takara Bio). *In vitro* transcription was performed using MAXIscript T7 Transcription Kit (Thermo Fischer Scientific) per manufacturer's protocol 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

RNA was isolated from *S. mutans* cells under sugar-phosphate stress, oxidative, heat, or acid stress as described above. Equal amounts of RNA 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 min at 70°C followed by 3 min on ice. Samples were loaded onto either 6% or 10% TBE-Urea Gel (Thermo Fischer Scientific) along with a biotinylated sRNA ladder (Kerafast). Gels were run in 1x TBE buffer at 180V for 60 min (6% gels) or 180V for 80 min (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 h with rotation. RNA probes produced from *in vitro* transcription as described above were heated at 95°C for 5 min and cooled on ice for 3 min, then added to fresh hybridization buffer. Membranes were incubated overnight at 45°C with rotation. After washing, membranes were incubated for 2 h 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 min. Blots were washed and viewed on a Licor Odyssey scanner.

EMSA and mutagenesis

Electromobility shift assay (EMSA) was performed as described previously (59). Briefly, a DNA template was amplified from *S. mutans* UA159 gDNA using primers with a T7 tag (**Table S2**). RNA was transcribed from this template using the MAXI T7 Transcription Kit that incorporated biotinylated uracil into the sRNA transcript. RNA was purified with a Monarch RNA Cleanup Kit and resuspended in TE buffer. SmsR4 and SMU_313 transcripts were combined at ratios shown in **Figure 7** and heated for 5 min at 85°C, then immersed in ice for 30 sec. The reaction volume was adjusted to 10 µl with 5x TMN buffer and incubated at 37°C for 30 min. Samples were run on an 8% TBE gel (Thermo Fischer Scientific) for 90 min at 100 V in 1x TBE buffer. The gel was transferred to a 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 for Northern blot assays, and images were examined on a Licor Odyssey scanner. SMU_313 with mutated SmsR4-binding site was constructed using the Q5 Mutagenesis Kit (New England Biolabs) and mutations were confirmed through Sanger sequencing.

Gene deletion and complementation

SmsR4- and SMU_313-deletion strains were constructed using the markerless-mutagenic system and an IDFC2 selection and counter-selection cassette, as described previously (60). Complementation was preformed using the pDL278 plasmid designed for expression in both *E. coli* and *S. mutans* (31, 32). The plasmid was purified from *E. coli* using a Plasmid MiniPrep kit (Thermo Fischer Scientific). pDL278 was linearized using BamHI and EcoRI, and PCR products (**Table S2**) were ligated into the linearized plasmid using T4 ligase (Thermo Fischer Scientific). Plasmids were then transformed into competent *E. coli* DH5-alpha cells following manufacturer's protocol (New England Biolabs). Sanger sequencing was used to confirm plasmid construction. Purified plasmids were transformed into *S. mutans* using competence stimulating peptide as described previously (60).

RACE assay

Rapid Amplification of cDNA Ends (RACE) was performed using a RACE kit (Thermo Fisher Scientific) per the manufacturer's recommendations. 5′ RACE assay was conducted using a gene specific primer (GSP) complementary to the 3′ end of SmsR4 (**Table S2**). RNA degradation, cDNA synthesis, and TdT tailing were accomplished using kit components and manufacturer protocols. A second nested GSP was used to amplify the tailed cDNA, and the PCR product from this reaction was cloned into pGEM T-Easy vector (Promega), transformed into competent DH5-alpha *E. coli* and the 5′ end of SmsR4 was determined using Sanger sequencing. For the 3′ RACE assay, Poly-A polymerase (New England Biolabs) was used to add poly-A tails to all transcripts. Precipitated poly-A tailed RNA was reverse transcribed using an oligo-dT adapter

primer and SuperScript II reverse transcriptase. RNA was subsequently degraded using RNase H. A GSP designed to bind to the 5′ end of SmsR4 and an Abridged Universal Amplification Primer (**Table S2**) were used to amplify the cDNA, and PCR products were cloned into pGEM T-Easy vector, transformed into competent *E. coli* DH5-alpha cells and the 3′ end of SmsR4 was identified using Sanger sequencing.

Covariance modeling

Covariance models of sRNAs were constructed as previously described (61, 62). Briefly, SmsR4 and 6S sequences from *S. mutans* UA159 were used as queries in BlastN searches against all *Streptococcus* genomes available in RefSeq (63, 64). Hits with >65% identity and >70% coverage were retained and five and six sequences were randomly selected to serve as seed sequences for constructing 6S and SmsR4 covariance models, respectively (**Table S3**). The WAR webserver was used to align the seed sequences and the Infernal suite of tools (v1.1.2) was used to construct (cmbuild) and calibrate (cmcalibrate) an initial covariance model for each sRNA (65, 66). This model was used to search (cmsearch) a database constructed from 37 representative *Streptococcus* full genomes available on RefSeq (**Table S4**). 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 cmsearch failed to yield new unrepresented sequences. The final models were used to determine the prevalence of 6S and SmsR4 in 62 full and partial *Streptococcus* genomes (**Table S5**). The presence or absence of each sRNAs was mapped on a previously published phylogenetic tree (29) and nodes of sRNA origin and secondary loss were determined through maximum parsimony.

373

374 ***In silico* sRNA target prediction**

375 For SmsR4 target prediction using IntaRNA (67), SmsR4 sequence as determined by RACE
376 assay was used as input and searched against the *S. mutans* UA159 genome using default
377 parameters (75 nt upstream and downstream from the translation start site, one interaction per
378 pair, 7 nt hybridization seed). Target RNA2 (68) was run using the same query sequence and
379 default parameters (80 nt upstream and 20 nt downstream from the translation start site, 7 nt
380 hybridization seed). IntaRNA was also used to predict targets for 11 other sRNAs using stricter
381 parameters to identify potential interactions adjacent to translation start sites of mRNA targets.
382 For all *in silico* target predictions, only significant results ($p < 0.05$, as determined by IntaRNA
383 or Target RNA2) were retained.

384

385 **ACKNOWLEDGEMENTS**

386 We thank Zhengzhong Zou and Samantha Fancher for their assistance with experiments, and
387 Shaun Wachter for technical advice. This work was supported in part by National Institute of
388 Dental and Craniofacial Research grants DE028409 to RR and DE028252 to JM.

389

390

REFERENCES

1. Forssten SD, Björklund M, Ouwehand AC. 2010. *Streptococcus mutans*, caries and simulation models. Nutrients 2:290–298.

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

3. Loesche WJ. 1986. Role of *Streptococcus mutans* in human dental decay. Microbiol Rev 50:353–380.

4. Bowen WH, Burne RA, Wu H, Koo H. 2018. Oral biofilms: pathogens, matrix, and polymicrobial interactions in microenvironments. Trends Microbiol 26:229–242.

5. Ajdić D, McShan WM, McLaughlin RE, Savić G, Chang J, Carson MB, Primeaux C, Tian R, Kenton S, Jia H, Lin S, Qian Y, Li S, Zhu H, Najar F, Lai H, White J, Roe BA, Ferretti JJ. 2002. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc Natl Acad Sci U S A 99:14434–14439.

6. Lemos JA, Palmer SR, Zeng L, Wen ZT, Kajfasz JK, Freires IA, Abranches J, Brady LJ. 2019. The biology of *Streptococcus mutans*. Microbiol Spectr 7.

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

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

9. Mäkinen KK. 2010. Sugar alcohols, caries incidence, and remineralization of caries lesions: a literature review. Int J Dent 2010:981072.

10. Lemos JA, Burne RA. 2008. A model of efficiency: stress tolerance by *Streptococcus mutans*. Microbiology (Reading) 154:3247–3255.

11. Storz G, Vogel J, Wassarman KM. 2011. Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell 43:880–891.

12. Holmqvist E, Wagner EGH. 2017. Impact of bacterial sRNAs in stress responses. Biochem Soc Trans 45:1203–1212.

13. Raghavan R, Groisman EA, Ochman H. 2011. Genome-wide detection of novel regulatory RNAs in *Escherichia coli*. Genome Res 21:1487–1497.

14. Warrier I, Hicks LD, Battisti JM, Raghavan R, Minnick MF. 2014. Identification of novel small RNAs and characterization of the 6S RNA of *Coxiella burnetii*. PLoS One 9:e100147.

15. Wachter S, Hicks LD, Raghavan R, Minnick MF. 2020. 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.

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

17. Zeng L, Choi SC, Danko CG, Siepel A, Stanhope MJ, Burne RA. 2013. Gene regulation by CcpA and catabolite repression explored by RNA-Seq in *Streptococcus mutans*. PLoS One 8:e60465.

18. Jester BC, Romby P, Lioliou E. 2012. When ribonucleases come into play in pathogens: a survey of gram-positive bacteria. International Journal of Microbiology 2012:e592196.

19. Mars RAT, Nicolas P, Denham EL, van Dijl JM. 2016. Regulatory RNAs in *Bacillus subtilis*: a gram-positive perspective on bacterial RNA-mediated regulation of gene expression. Microbiol Mol Biol Rev 80:1029–1057.

20. Menendez-Gil P, Toledo-Arana A. 2021. Bacterial 3′ UTRs: a useful resource in post-transcriptional regulation. Front Mol Biosci 7:617633.

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

22. Mohanty BK, Kushner SR. 2018. Enzymes involved in post-transcriptional RNA metabolism in gram-negative bacteria. Microbiol Spectr 6:10.1128/microbiolspec.RWR-0011–2017.

23. Sinha D, Zimmer K, Cameron TA, Rusch DB, Winkler ME, De Lay NR. 2019. Redefining the small regulatory RNA transcriptome in *Streptococcus pneumoniae* serotype 2 strain D39. J Bacteriol 201:e00764-18.

24. Barrick JE, Sudarsan N, Weinburg Z, Ruzzo WL, Breaker RR. 2005. 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. RNA 11:774–784.

25. Cavanagh AT, Wassarman KM. 2014. 6S RNA, a global regulator of transcription in *Escherichia coli*, *Bacillus subtilis*, and beyond. Annu Rev Microbiol 68:45–60.

26. Trotochaud AE, Wassarman KM. 2005. A highly conserved 6S RNA structure is required for regulation of transcription. Nat Struct Mol Biol 12:313–319.

27. Belda E, Sekowska A, Le Fèvre F, Morgat A, Mornico D, Ouzounis C, Vallenet D, Médigue C, Danchin A. 2013. An updated metabolic view of the *Bacillus subtilis* 168 genome. Microbiology (Reading) 159:757–770.

28. Chae H, Han K, Kim K, Park H, Lee J, Lee Y. 2011. Rho-dependent termination of ssrS (6S RNA) transcription in *Escherichia coli*. J Biol Chem 286:114–122.

29. Patel S, Gupta RS. 2018. 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.

30. Bochner BR, Gadzinski P, Panomitros E. 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res 11:1246–1255.

466 31. LeBlanc DJ, Lee LN, Abu-Al-Jaibat A. 1992. Molecular, genetic, and functional analysis of
467 the basic replicon of pVA380-1, a plasmid of oral streptococcal origin. Plasmid 28:130–
468 145.

469 32. Zhou L, Manias DA, Dunny GM. 2000. Regulation of intron function: efficient splicing in
470 vivo of a bacterial group II intron requires a functional promoter within the intron. Mol
471 Microbiol 37:639–651.

472 33. Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J. 2012. An atlas of Hfq-bound
473 transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. EMBO J
474 31:4005–4019.

475 34. Hoyos M, Huber M, Förstner KU, Papenfort K. 2020. Gene autoregulation by 3′ UTR-
476 derived bacterial small RNAs. Elife 9:e58836.

477 35. Lalaouna D, Baude J, Wu Z, Tomasini A, Chicher J, Marzi S, Vandenesch F, Romby P,
478 Caldelari I, Moreau K. 2019. RsaC sRNA modulates the oxidative stress response of
479 *Staphylococcus aureus* during manganese starvation. Nucleic Acids Res 47:9871–9887.

480 36. Miyakoshi M, Chao Y, Vogel J. 2015. Regulatory small RNAs from the 3′ regions of
481 bacterial mRNAs. Curr Opin Microbiol 24:132–139.

482 37. van der Meulen SB, Hesseling-Meinders A, de Jong A, Kok J. 2019. The protein regulator
483 ArgR and the sRNA derived from the 3′-UTR region of its gene, ArgX, both regulate the
484 arginine deiminase pathway in *Lactococcus lactis*. PLoS One 14:e0218508.

485 38. Adams PP, Baniulyte G, Esnault C, Chegireddy K, Singh N, Monge M, Dale RK, Storz G,
486 Wade JT. 2021. Regulatory roles of *Escherichia coli* 5′ UTR and ORF-internal RNAs
487 detected by 3′ end mapping. eLife 10:e62438.

488 39. Garst AD, Edwards AL, Batey RT. 2011. Riboswitches: structures and mechanisms. Cold
489 Spring Harb Perspect Biol 3:a003533.

490 40. Jose BR, Gardner PP, Barquist L. 2019. Transcriptional noise and exaptation as sources for
491 bacterial sRNAs. Biochem Soc Trans 47:527–539.

492 41. Lybecker M, Bilusic I, Raghavan R. 2014. Pervasive transcription: detecting functional
493 RNAs in bacteria. Transcription 5:e944039.

494 42. Raina M, King A, Bianco C, Vanderpool CK. 2018. Dual-Function RNAs. Microbiology
495 Spectrum 6:6.5.06.

496 43. Deutscher J, Aké FMD, Derkaoui M, Zébré AC, Cao TN, Bouraoui H, Kentache T,
497 Mokhtari A, Milohanic E, Joyet P. 2014. The bacterial phosphoenolpyruvate:carbohydrate
498 phosphotransferase system: regulation by protein phosphorylation and phosphorylation-
499 dependent protein-protein interactions. Microbiology and Molecular Biology Reviews
500 78:231–256.

501 44. Richards GR, Patel MV, Lloyd CR, Vanderpool CK. 2013. Depletion of glycolytic
502 intermediates plays a key role in glucose-phosphate stress in *Escherichia coli*. J Bacteriol
503 195:4816–4825.

504 45. Bobrovskyy M, Vanderpool CK. 2014. The small RNA SgrS: roles in metabolism and
505 pathogenesis of enteric bacteria. Front Cell Infect Microbiol 4:61.

506 46. Rice JB, Vanderpool CK. 2011. The small RNA SgrS controls sugar–phosphate
507 accumulation by regulating multiple PTS genes. Nucleic Acids Res 39:3806–3819.

508 47. Jeng W-Y, Panjaitan NSD, Horng Y-T, Chung W-T, Chien C-C, Soo P-C. 2017. The
509 negative effects of KPN00353 on glycerol kinase and microaerobic 1,3-propanediol
510 production in *Klebsiella pneumoniae*. Front Microbiol 8:2441.

511 48. Choi J, Shin D, Yoon H, Kim J, Lee C-R, Kim M, Seok Y-J, Ryu S. 2010. *Salmonella*
512 pathogenicity island 2 expression negatively controlled by EIIANtr-SsrB interaction is
513 required for *Salmonella* virulence. Proc Natl Acad Sci U S A 107:20506–20511.

514 49. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
515 sequence data. Bioinformatics 30:2114–2120.

516 50. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B. 2000.
517 Artemis: sequence visualization and annotation. Bioinformatics 16:944–945.

518 51. Hofacker IL. 2007. RNA consensus structure prediction with RNAalifold. Methods Mol
519 Biol 395:527–544.

520 52. Naville M, Ghuillot-Gaudeffroy A, Marchais A, Gautheret D. 2011. ARNold: a web tool
521 for the prediction of Rho-independent transcription terminators. RNA Biol 8:11–13.

522 53. Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and
523 formatting protein and DNA sequences. Biotechniques 28:1102, 1104.

524 54. Decker E-M, Klein C, Schwindt D, von Ohle C. 2014. Metabolic activity of *Streptococcus*
525 *mutans* biofilms and gene expression during exposure to xylitol and sucrose. Int J Oral Sci
526 6:195–204.

527 55. Gong Y, Tian X-L, Sutherland T, Sisson G, Mai J, Ling J, Li Y-H. 2009. Global
528 transcriptional analysis of acid-inducible genes in *Streptococcus mutans*: multiple two-
529 component systems involved in acid adaptation. Microbiology (Reading) 155:3322–3332.

530 56. Kajfasz JK, Ganguly T, Hardin EL, Abranches J, Lemos JA. 2017. Transcriptome
531 responses of *Streptococcus mutans* to peroxide stress: identification of novel antioxidant
532 pathways regulated by Spx. Sci Rep 7:16018.

533 57. Liu C, Niu Y, Zhou X, Zheng X, Wang S, Guo Q, Li Y, Li M, Li J, Yang Y, Ding Y,
534 Lamont RJ, Xu X. 2015. *Streptococcus mutans* copes with heat stress by multiple
535 transcriptional regulons modulating virulence and energy metabolism. Sci Rep 5:12929.

536 58. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion
537 for RNA-seq data with DESeq2. Genome Biology 15:550.

538 59. Wachter S, Bonazzi M, Shifflett K, Moses AS, Raghavan R, Minnick MF. A CsrA-binding,
539 trans-acting sRNA of *Coxiella burnetii* is necessary for optimal intracellular growth and
540 vacuole formation during early infection of host cells. Journal of Bacteriology 201:e00524-
541 19.

542 60. Xie Z, Okinaga T, Qi F, Zhang Z, Merritt J. 2011. Cloning-independent and
543 counterselectable markerless mutagenesis system in *Streptococcus mutans*. Appl Environ
544 Microbiol 77:8025–8033.

545 61. Barquist L, Burge SW, Gardner PP. 2016. Studying RNA homology and conservation with
546 Infernal: from single sequences to RNA families. Curr Protoc Bioinformatics 54:12.13.1-
547 12.13.25.

548 62. Krieger MC, Dutcher HA, Ashford AJ, Raghavan R. 2021. A peroxide-responding sRNA
549 evolved from a peroxidase mRNA. bioRxiv doi: https://doi.org/101101/20211018464853
550 2021.10.18.464853.

551 63. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search
552 tool. J Mol Biol 215:403–410.

553 64. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B,
554 Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O,
555 Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T,
556 Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P,
557 McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD,
558 Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR,
559 Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P,
560 Murphy TD, Pruitt KD. 2016. Reference sequence (RefSeq) database at NCBI: current
561 status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733-745.

562 65. Torarinsson E, Lindgreen S. 2008. WAR: Webserver for aligning structural RNAs. Nucleic
563 Acids Res 36:W79–W84.

564 66. Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches.
565 Bioinformatics 29:2933–2935.

566 67. Mann M, Wright PR, Backofen R. 2017. IntaRNA 2.0: enhanced and customizable
567 prediction of RNA-RNA interactions. Nucleic Acids Res 45:W435–W439.

568 68. Kery MB, Feldman M, Livny J, Tjaden B. 2014. TargetRNA2: identifying targets of small
569 regulatory RNAs in bacteria. Nucleic Acids Res 42:W124–W129.

570

571

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

sRNA[a]	Rfam annot-ation[b]	Previous prediction[c]	Bordering gene (left)	sRNA left bound[d]	sRNA right bound[d]	Bordering gene (right)	Strand	Predicted size (nt)
SmsR1			SMU_61	62597	62953	SMU_63c	F	357
SmsR2		psRNA-31	SMU_97	99488	99909	SMU_t19	R	422
SmsR3		psRNA-54	SMU_219	211498	211709	SMU_220c	R	212
SmsR4		psRNA-62	SMU_318	303748	303857	SMU_320	R	110
SmsR5	RNase P	psRNA-73	SMU_471	439833	440221	SMU_472	F	389
SmsR6		psRNA-78	SMU_530c	497858	498119	SMU_531	F	262
SmsR7			SMU_1046c	995311	995586	SMU_1048	R	276
SmsR8	tmRNA		SMU_1196c	1139336	1139677	SMU_1197	R	342
SmsR12		psRNA-204 psRNA-205	SMU_1862	1757885	1758495	SMU_1865	R	611
SmsR14	6S	psRNA-235	SMU_2056	1929960	1930160	SMU_2057c	F	200
SmsR16			SMU_803c	749489	749732	SMU_804	R	244
SmsR18		psRNA-132	SMU_1332c	1254431	1255035	SMU_1334	R	605
SmsR19			SMU_1398	1327742	1328337	SMU_1400c	R	596
SmsR20		psRNA-150	SMU_1512	1439375	1439585	SMU_1513	R	211
SmsR22			SMU_1703	1614938	1615314	SMU_1704	R	377

573 [a]Small RNAs were named SmsR1-SmsR22 (Note: some of the original candidate sRNAs were

574 later excluded from analysis).

575 [b]Existing Rfam entries for the putative sRNAs.

576 [c]Previously predicted sRNAs that overlap with our results (17).

577 [d]The putative 5′ ends of sRNAs were estimated using visual scans of RNA-seq coverage plots for

578 sites with sharp increases in reads mapped per nucleotide. 3′ ends were estimated based on the

579 locations of predicted terminators (**Table 2**) or for sRNAs without predicted terminators visual

580 scans of the RNA-seq coverage plots were used to identify sites with sharp drops in reads per

581 nucleotide. SmsR4 boundaries were confirmed using 5′ and 3′ RACE assays.

582

583 **Table 2. Predicted promoter and terminator elements of novel sRNAs.**

sRNA	-10 Promoter[a]	Terminator[b]
SmsR1	TGTCCTGTTCTTTTTTTGAAGGATCAT T**TATAAT**GAATGATATCAAAAAG**A**AA	CTAACTAAAAC**AGAGACTC**ACATTACAATCAC ACGTGAATCTCTgTTTTTCTTAGCTG
SmsR2	AC**TAAACT**ATACGTGTATTCGTTTTGT TGGCAGCCAATTTCTTTAACTATAAC	CAAAAAAAGAAACACCTTCTTAAATCTAGTAA ATGAGATTTAAGAAGGTGTTTCTTTATAAGCA
SmsR3	ACAGTTGTTTTATCGTTTGTGGAGAAT ATGA**TATAAT**ACTAACGGCACAA**A**CT	GTGTAAAAAAAGCCTTAGCTCTGCCAAGCTAG GGCTTTTCCGTTGCC
SmsR4	ACTTTAAGGTTTCTTTAAGGTTTCTCA TA**TATACT**TTAATCATCCTAAAA**C**AA	TTCTTAAAAAACCTTGCAGACTTAAATCTGCA AGGTTTTTTAATTCG
SmsR5	TGGTAAACAAATTACCGAATAGATTAA GAAAACGATGCAATTTTTG**GATAAT**C	ATAATAGGTGAGCTAGCTTTGGCTAGCTTTTA TTGTCTT
SmsR6	GCAAATATGCTTGCAATTCTTTTTTAG AAAGTG**TATAAT**CGTAAGAAATA**A**AT	AGATTTAACGCCCTCACACAGATTTTCTGTGT GAGGTTTTTTGTTATC
SmsR7	CTTTGTTAAGCTTATTTATTATGA**TAT AAT**GAAGTATTCAATTGAAGAAA**A**AG	TCTTCCAAGTAGCAGAAGCATTGATGTTTCTG CcaTTTTTAACACAG
SmsR8	TTTTTCCTTTATTTTGC**TATACT**ATTT TCACACAATATGTACTGGGGTCG**T**TA	No terminators found
SmsR12	TCTTGACAAATGTAAGCGGTTACGA**TA AAAT**AAATTTAGAAAAAAGATAA**G**AA	AAATAATAAGAGACCCCCAACGATGAGCGTGT AGATTGTTGGGGGTCTTAATTGTATTGA
SmsR14	TTAACTTGAATTTTTT**TAACAT**TATGG TATCATATTTAAAGAAGAATTGC**T**GT	No terminators found
SmsR16	TGAAAGCGTTTTGT**TAAACT**GACTTTA GCAAATTATTTGGAGGTAAGGTG**C**AT	No terminators found
SmsR18	GATAGATAATTTTCATAGTTATTTGTT AAAAGTGA**TATAAT**AAGAATTAT**C**CG	No terminators found
SmsR19	GCTAGAAAGATTGATTTCTAGCGATTT TTTAGG**TATAAT**ATAAGCAATCA**A**AA	AATAAGCAAAAGACACTTGAAGCAATAATTCA AGTGTCTTTTATGGGACTT
SmsR20	CAAGGAAACAAGGCGGGAACGACAAAA TCATTTCTG**TAAATT**GCAATGGA**A**TC	No terminators found
SmsR22	AAGAAAACTTGTTATTAACTT**TATTTT** TTGATATGCTATAAGTGTCTTCA**G**GG	TCTTTAGGCCTTCTTTCGATTTGTAAAAATTG GAGGAaTTTTTTTATGAA

584 [a] -10 promoter elements were determined from a manual scan of sites upstream of transcription
585 start sites (in green) estimated from RNA-seq data. Sites that contained at least 4/6 of the
586 TATAAT consensus promoter sequence, including an A at position two and a T at position six
587 were identified and are in bold.

588 [b]Intrinsic terminators were predicted by ARNold (52). The predicted stem-loop structure is
589 shown in blue (stem) and red (loop).

590

591 **Table 3. Stress conditions tested in this study.**

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

592

593

FIGURE LEGENDS

Figures 1A and 1B. Expression profiles of novel sRNAs in *S. mutans*. RNA-seq reads mapped to forward (red) and reverse (blue) strands of the *S. mutans* UA159 genome are shown. Y-axes denote the number of RNA-seq reads mapped to each nucleotide. Genomic locations of novel sRNA genes (black arrows) and flanking genes (grey arrows) along with their nucleotide positions are shown below the coverage plots. Northern blot performed for each sRNA is shown to the right. White arrows denote estimated sizes of sRNAs (**Table 1**). (Note: intensity between each blot is not comparable as exposure times differed between experiments.)

Figure 2. 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 between treatment (stress) and control (no stress) samples (+/- standard error) from three replicate experiments are shown, except for acid stress, which used two replicates. Only expression values that were significantly different ($p \leq 0.05$, Wald test) between stress and control conditions are displayed.

Figure 3. 6S RNA and SmsR4 are distinct sRNAs. The genome locations and predicted secondary structures of 6S RNAs in *Bacillus subtilis*, *Streptococcus mutans,* and *Escherichia coli* are shown in the top three panels. The genome location and secondary structure of SmsR4 in *S. mutans* is displayed in the bottom panel. Genes that encode N-acetyldiaminopimelate deacetylase and 5-formyltetrahydrofolate cyclo-ligase are abbreviated as N-Ad and 5-FTC, respectively. (Note: genes are not drawn to scale.)

Figure 4. Prevalence of 6S RNA and SmsR4 in *Streptococcus*. Presence or absence of 6S and SmsR4 as determined by a covariance modeling-based search is shown. In most species, 6S is present between genes for RarA and tRNA-Lys, whereas SmsR4 is located between genes that encode 5-formyltetrahydrofolate cyclo-ligase (5-FTC) and N-acetyldiaminopimelate deacetylase (N-Ad). Flanking genes other than *rarA*, tRNA-Lys, 5-FTC, and N-Ad are shown as white arrows. The cladogram is from Patel and Gupta 2018 (29). (Note: genes are not drawn to scale.)

Figure 5. SmsR4 promotes *S. mutans* growth in sorbitol-containing medium. Growth of wild-type (WT), SmsR4-deletion (DEL), and SmsR4 complementation (COMP) strains in BTR medium with glucose (BTR-G, left) or sorbitol (BTR-S, right).

Figure 6. SmsR4 is predicted to bind to SMU_313. A) Two target prediction algorithms, IntaRNA (67) (left) and TargetRNA2 (68) (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. The nucleotides highlighted in red are required for efficient binding of SmsR4 to SMU_313 (see **Figure 7**). **B)** Genomic locations of sorbitol phosphotransferase (PTS) system operon, including SMU_313 (blue), and SmsR4 (teal) in *S. mutans*.

Figure 7. SmsR4 binds to SMU_313. **A)** An RNA-RNA electrophoretic mobility shift assay (EMSA) shows that increasing concentrations of SMU_313 with native 5′ untranslated sequence binds well to SmsR4. A control lane with unbound SmsR4 is shown to the right. **B)** SMU_313 transcript with a mutated binding site does not interact with SmsR4. A control lane with native

SMU_313 transcript ran on the same gel is shown on the right. The native and mutated SmsR4-binding sites on SMU_313 are shown in red, and start codon is highlighted in yellow.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Phenotypic microarray. Growth of wild-type (WT) and SmsR4-deletion (DEL) strains in Biolog medium with sorbitol as sole carbon source (30). Values represent the average of two independent growth experiments.

Figure S2. Effect of SMU_313 deletion on growth in sorbitol. Growth of SMU_313-deletion (SMU_313 DEL) and wild-type (WT) strains of *S. mutans* in BTR medium with glucose (BTR-G, left), or sorbitol (BTR-S, right).

Figure S3. SmsR4 expression over time. Northern blot for SmsR4 in wild-type *S. mutans* grown in BTR-S was performed at time-points shown. 5S RNA was used as a loading control.

Figure S4. Confirmation of stress induction. Either upregulation or downregulation of genes known to be associated with each stress condition were used to confirm stress induction (54–57). Values represent means (+/- standard error) from three independent qPCR assays, except for *groES* and SMU_1805, which were from two replicates.

SUPPLEMENTAL TABLES

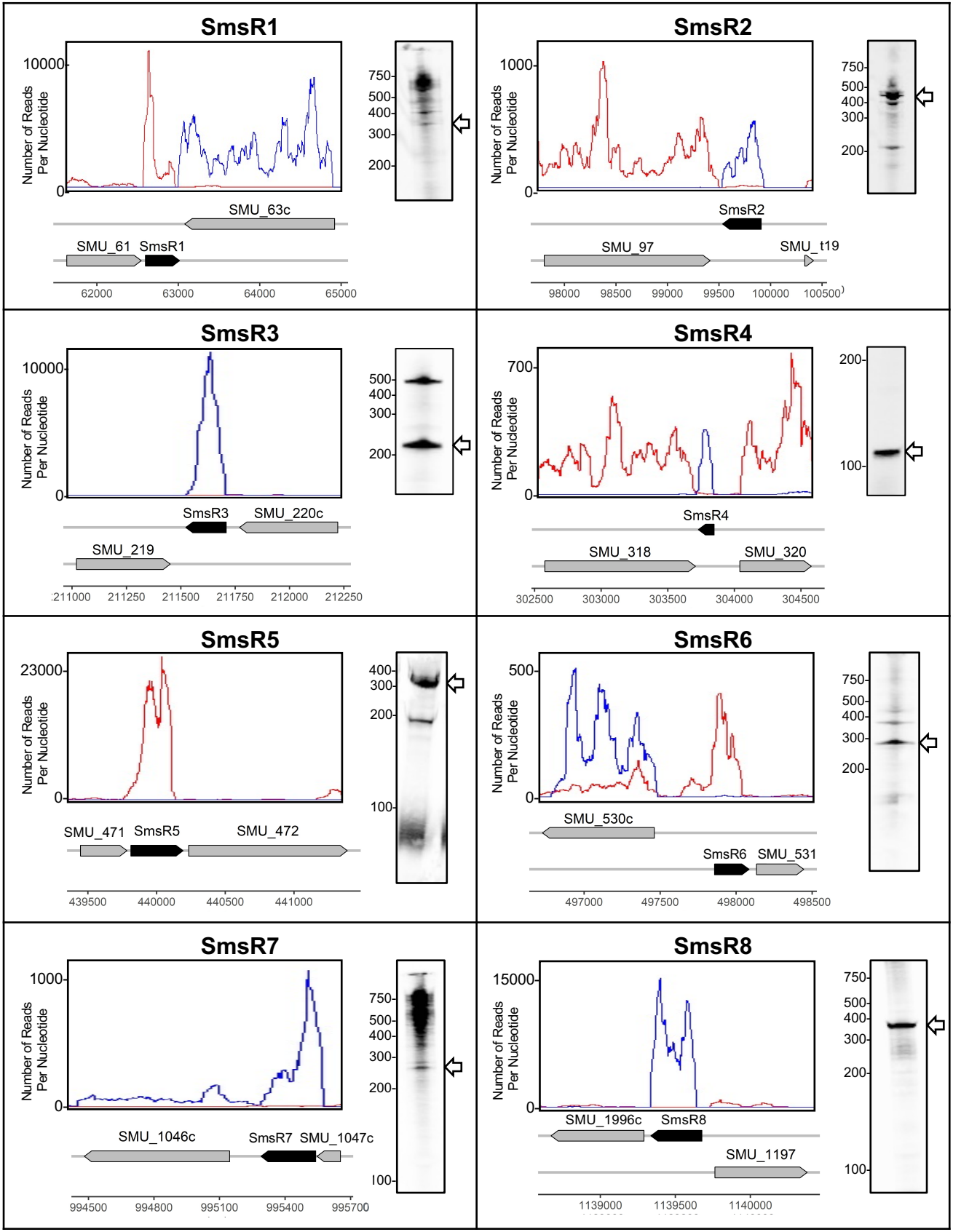
Table S1. Targets predicted by IntaRNA for novel sRNAs.

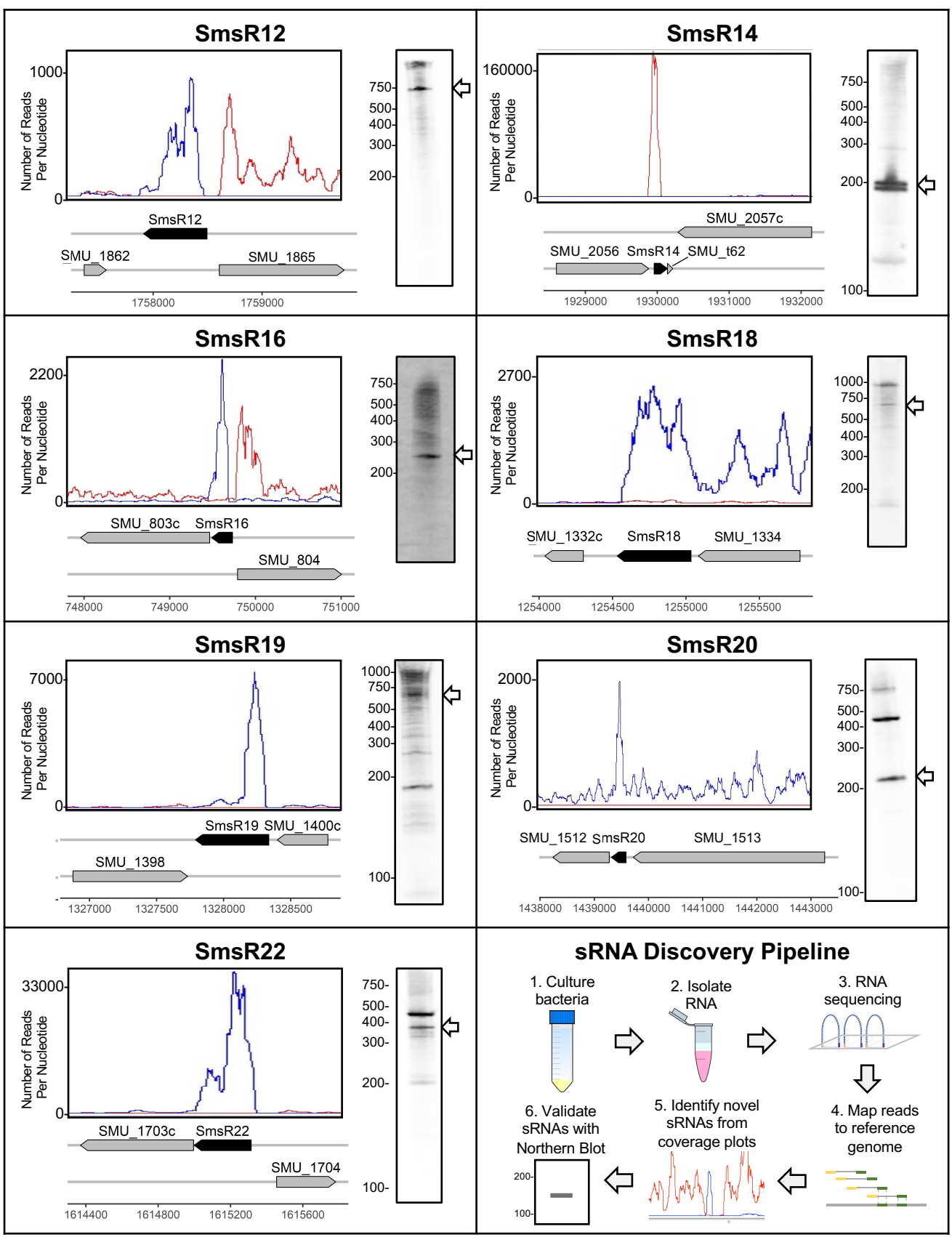
Table S2. Primers used in this study.

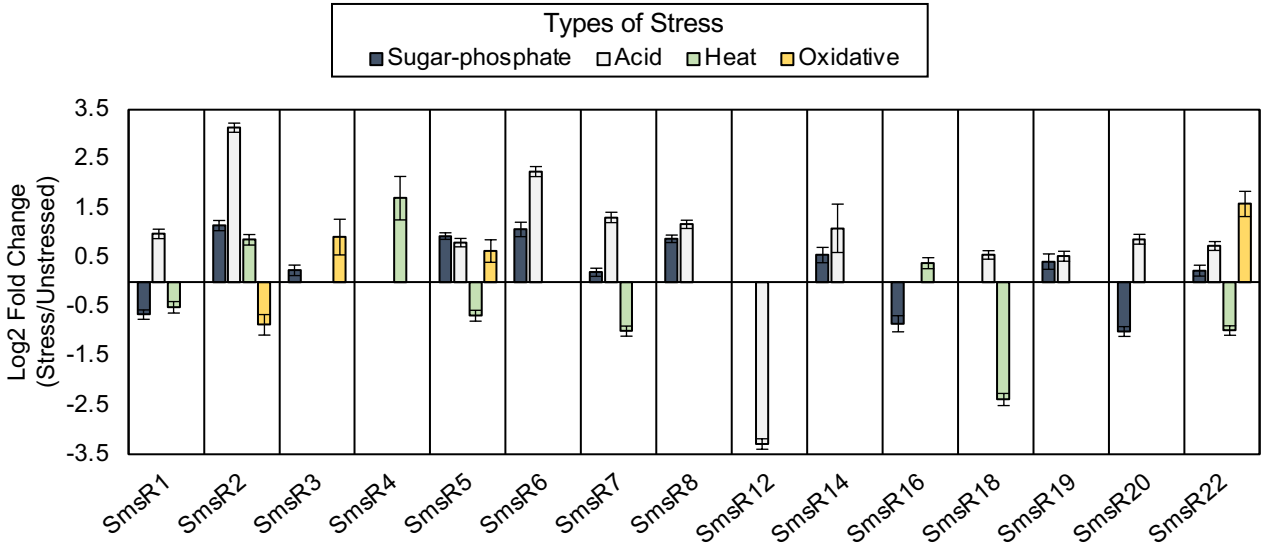
Table S3. Genomes used for initial SmsR4 and 6S covariance model construction.

Table S4. Genomes used for covariance model calibration.

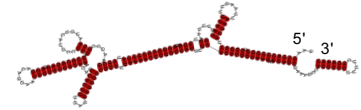
Table S5. Genomes used to determine the prevalence of 6S and SmsR4 in *Streptococcus*.

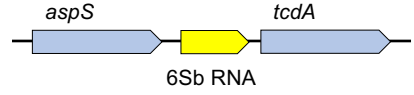

SmsR1
Number of Reads Per Nucleotide
10000
0
SMU_63c
SMU_61
SmsR1
62000
63000
64000
65000
750
500
400
300
200
SmsR2
Number of Reads Per Nucleotide
1000
0
SmsR2
SMU_97
SMU_t19
98000
98500
99000
99500
100000
100500
750
500
400
300
200
SmsR3
Number of Reads Per Nucleotide
10000
0
SmsR3
SMU_220c
SMU_219
211000
211250
211500
211750
212000
212250
500
400
300
200
SmsR4
Number of Reads Per Nucleotide
700
0
SmsR4
SMU_318
SMU_320
302500
303000
303500
304000
304500
200
100
SmsR5
Number of Reads Per Nucleotide
23000
0
SMU_471
SmsR5
SMU_472
439500
440000
440500
441000
400
300
200
100
SmsR6
Number of Reads Per Nucleotide
500
0
SMU_530c
SmsR6
SMU_531
497000
497500
498000
498500
750
500
400
300
200
SmsR7
Number of Reads Per Nucleotide
1000
0
SMU_1046c
SmsR7
SMU_1047c
994500
994800
995100
995400
995700
750
500
400
300
200
100
SmsR8
Number of Reads Per Nucleotide
15000
0
SMU_1996c
SmsR8
SMU_1197
1139000
1139500
1140000
750
500
400
300
200
100


SmsR12
Number of Reads Per Nucleotide
1000
0
SmsR12
SMU_1862
SMU_1865
1758000
1759000
750
500
400
300
200
SmsR14
Number of Reads Per Nucleotide
160000
0
SMU_2057c
SMU_2056
SmsR14
SMU_t62
1929000
1930000
1931000
1932000
750
500
400
300
200
100
SmsR16
Number of Reads Per Nucleotide
2200
0
SMU_803c
SmsR16
SMU_804
748000
749000
750000
751000
750
500
400
300
200
SmsR18
Number of Reads Per Nucleotide
2700
0
SMU_1332c
SmsR18
SMU_1334
1254000
1254500
1255000
1255500
1000
750
500
400
300
200
SmsR19
Number of Reads Per Nucleotide
7000
0
SmsR19
SMU_1400c
SMU_1398
1327000
1327500
1328000
1328500
1000
750
500
400
300
200
100
SmsR20
Number of Reads Per Nucleotide
2000
0
SMU_1512
SmsR20
SMU_1513
1438000
1439000
1440000
1441000
1442000
1443000
750
500
400
300
200
100
SmsR22
Number of Reads Per Nucleotide
33000
0
SMU_1703c
SmsR22
SMU_1704
1614400
1614800
1615200
1615600
750
500
400
300
200
100
sRNA Discovery Pipeline
1. Culture bacteria
2. Isolate RNA
3. RNA sequencing
4. Map reads to reference genome
5. Identify novel sRNAs from coverage plots
6. Validate sRNAs with Northern Blot
200
100

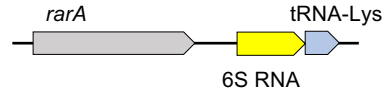

Types of Stress
Sugar-phosphate
Acid
Heat
Oxidative
Log2 Fold Change (Stress/Unstressed)
3.5
2.5
1.5
0.5
-0.5
-1.5
-2.5
-3.5
SmsR1
SmsR2
SmsR3
SmsR4
SmsR5
SmsR6
SmsR7
SmsR8
SmsR12
SmsR14
SmsR16
SmsR18
SmsR19
SmsR20
SmsR22

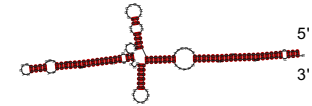
Bacillus subtilis



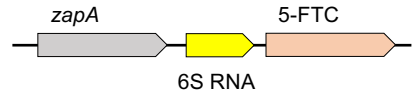


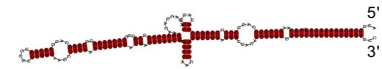
Streptococcus mutans



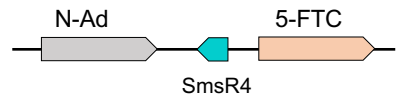


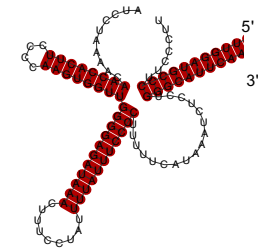
Escherichia coli

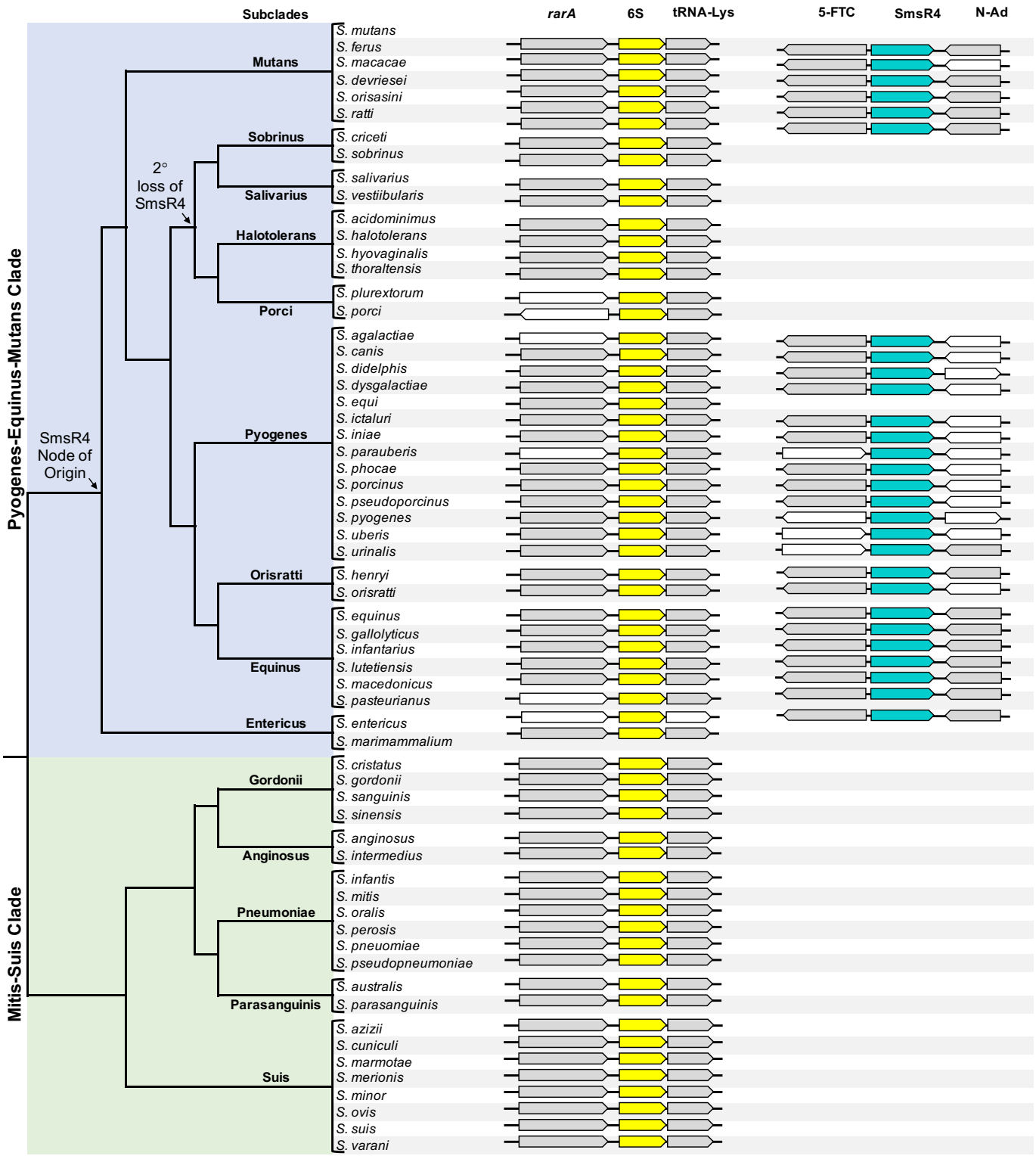


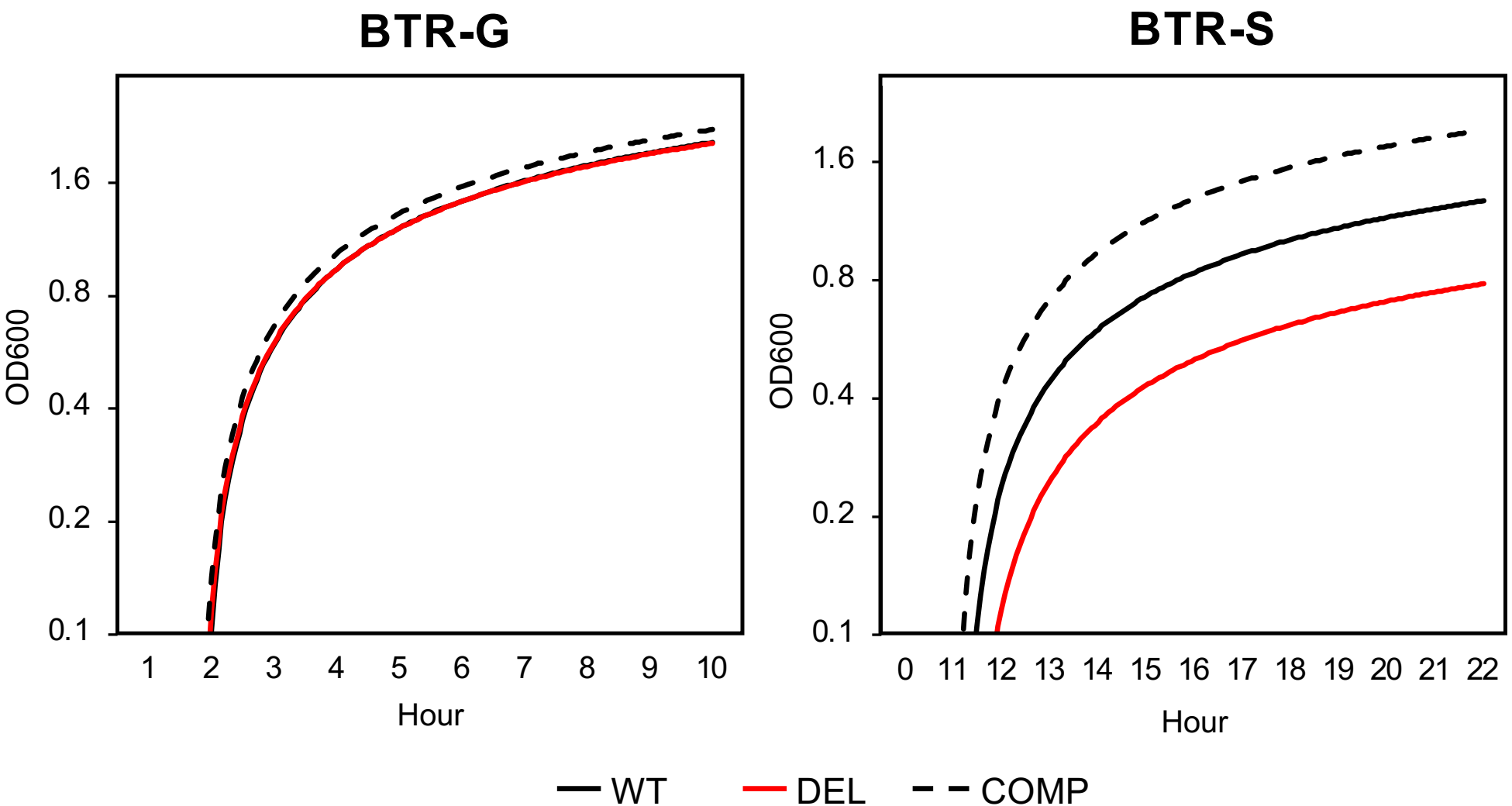


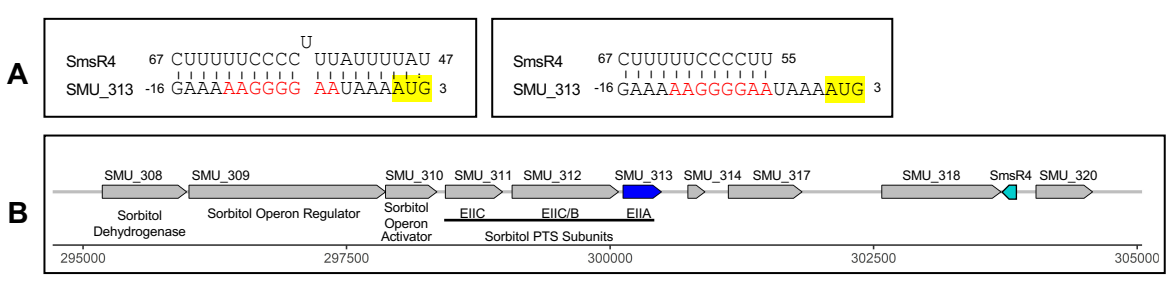
Streptococcus mutans

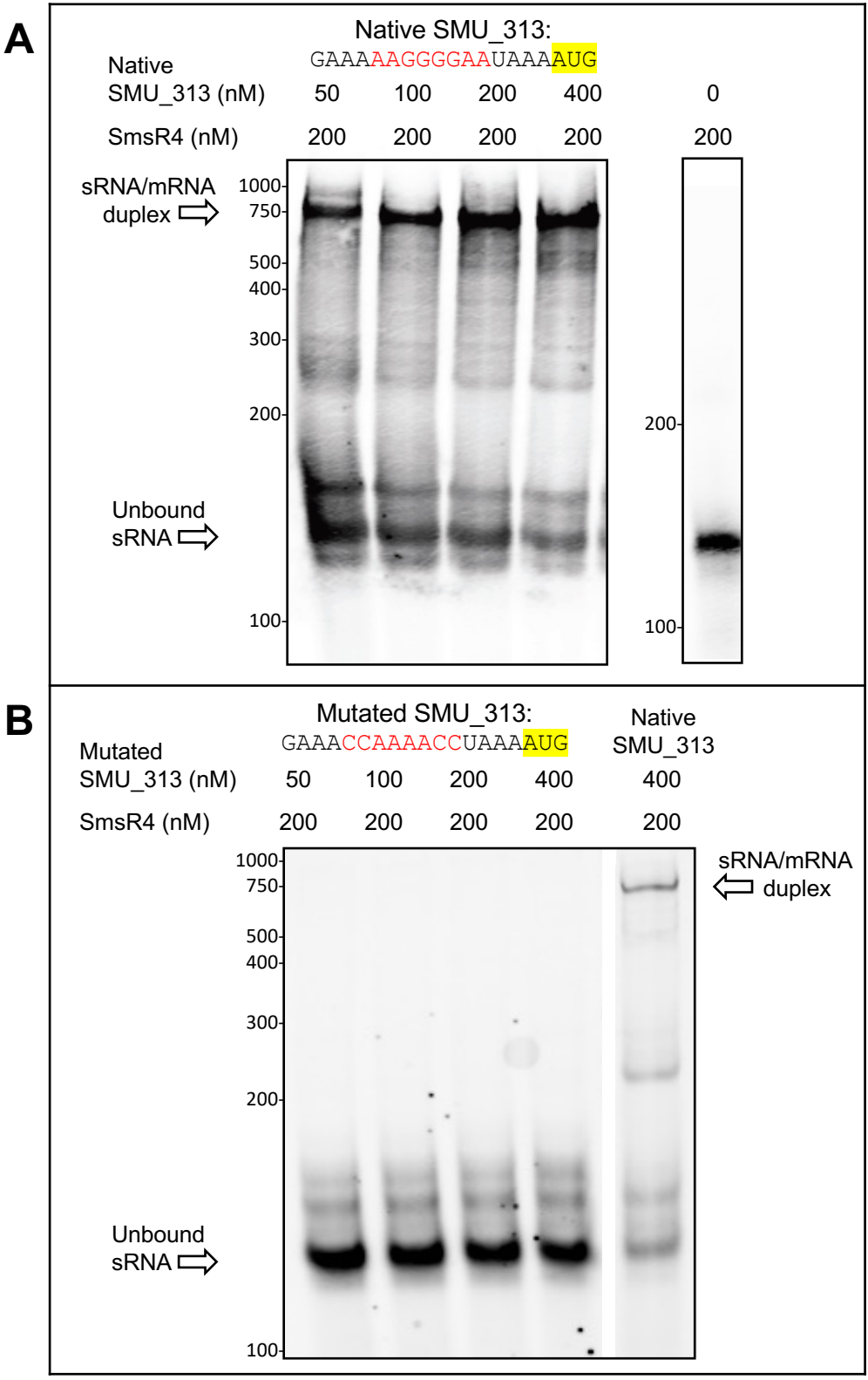







BTR-G
BTR-S
OD600
1.6
0.8
0.4
0.2
0.1
1 2 3 4 5 6 7 8 9 10
0 11 12 13 14 15 16 17 18 19 20 21 22
Hour
WT
DEL
COMP


A
SmsR4 67 CUUUUUCCCC U UUAUUUUAU 47
SMU_313 -16 GAAAAAGGGG AAUAAAAUG 3
SmsR4 67 CUUUUUCCCCUU 55
SMU_313 -16 GAAAAAGGGGAAUAAAAUG 3
B
SMU_308
SMU_309
SMU_310
SMU_311
SMU_312
SMU_313
SMU_314
SMU_317
SMU_318
SmsR4
SMU_320
Sorbitol Dehydrogenase
Sorbitol Operon Regulator
Sorbitol Operon Activator
EIIC
EIIC/B
EIIA
Sorbitol PTS Subunits
295000
297500
300000
302500
305000


A
Native SMU_313:
GAAAAAGGGGAAUAAAAUG
Native SMU_313 (nM)
50
100
200
400
0
SmsR4 (nM)
200
200
200
200
200
sRNA/mRNA duplex
1000
750
500
400
300
200
100
Unbound sRNA
200
100
B
Mutated SMU_313:
GAAACCAAAACCUAAAAUG
Native SMU_313
Mutated SMU_313 (nM)
50
100
200
400
400
SmsR4 (nM)
200
200
200
200
200
1000
750
500
400
300
200
100
sRNA/mRNA duplex
Unbound sRNA