THE UNIVERSITY OF CHICAGO

INVESTIGATION OF PROKARYOTIC NON-CODING RNA AND THEIR MODIFICATIONS

A DISSERTATION SUBMITTED TO

THE FACULTY OF DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON MICROBIOLOGY

BY

MARCUS CHUN HOE FOO

CHICAGO, ILLINOIS

AUGUST 2024

Dedicated to my beloved parents, who have been truly understanding, supportive and patient throughout my journey in science and time abroad

To my closest partner, Eos Trinidad, who has always been by my side and provided the much needed support throughout our journey through graduate school

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

aa-tRNA	aminoacyl-tRNA
aMG	α-methyl glucoside-6-phosphate
asRNA	anti-sense RNAs
bp	base pair
CMC	carbodiimide
D	dihydrouridine
DIP	dipyridyl
DM	demethylase
ISP2	International Streptomyces Project yeast malt extract agar medium
LC/MS	liquid chromatography/ mass spectromtery
lncRNA	long non-coding RNA
KO	knockout
m^1A	N1-methyladenosine
m^1G	N1-methylguanosine
m^1I	N1-methylinosine
$m^2{}_2G$	N2,2-dimethylguanosine
m^3C	N3-methylcytidine
m^5C	N5-methylcytosine
m^6A	N6-methyladenosine
m^7G	N7-methylguanosine
mRNA	messenger RNA
MSR-seq	Multiplex Small RNA sequencing
MTase	methyltransferase
ncRNA	non-coding RNA

Nm	2’-O-methylation
nt	nucleotide
OSR	oxidative stress response
Q	queuosine
rRNA	ribosomal RNA
RT	reverse transcription
RMD	tRNA-m^{1}A-demethylase
SAM	S-adenosyl-L-methionine
SV	*Streptomyces venezuelae*
sncRNA	small non-coding RNA
sRNA	small RNA
SRP	signal recognition particle
TGIRT	thermostable group II intron reverse transcriptase
tmRNA	transfer-messenger RNA
tRF	tRNA-derived RNA fragments
tRNA	transfer RNA
W	wybutosine
Ψ	pseudouridine

ACKNOWLEDGEMENT

A huge thanks to my mentor, Prof. Tao Pan who has been a great inspiration throughout my time at the University of Chicago. I have learned so much from him about everything related to science and academia, and truly cannot express how much his support has mean to me in completing my studies and research. The later years in my program has been tumultuous at best, yet Tao remained a steady anchor for continuing through these hard times and has been very understanding of the struggles I have faced. When I first started interacting with Tao, it was a difficult difference in authority barrier to overcome, given my cultural background, but I am very thankful for all the interactions and opportunities I had to see him in action as a professor, researcher, and leading expert in the field of tRNA biology. I am very grateful for having been mentored by such a great person, one who encourages his students and mentees to not be afraid and pursue the things they are passionate about in life.

Many thanks also go to all the members of the Pan lab, past and present, who have been as supportive, inspiring, and fun to hang out with as Tao. In particular, I am grateful to have learned from alumnus Dr. Chris Katanski, alumnus Dr. Chris Watkins and Dr. Wen Zhang for being the senior members of the lab at the time I joined; they provided great support for my growth as a researcher and were just great people to spend time with. Shoutout to all the other members as well, who have kept the lab lively through the fun social events they kept going as a lab tradition.

I thank my committee members, Prof. A. Murat Eren, Prof. Jingyi Fei, and Prof. Mark Mimee for their patience and understanding for my research progress.

I thank the other collaborators from the other labs at the University of Chicago and University of Illinois Urbana-Champaign for providing the much needed technical support in assisting me with completion of my experiments.

I thank my close circle graduate school friends, as well as those outside of the university, for their continuous support and friendship. I am also grateful for all the positive interactions I've had with other students and associated staff and faculty at the university

Finally, I want to thank my family, especially my parents, brother and my partner, for always believing in me and providing unconditional support. Thanks to all of my extended family back in Malaysia who have never failed to provide a warm welcome whenever I get the chance to return home.

ABSTRACT

Non-coding RNAs (ncRNA) are transcripts produced by the cell that are not translated into proteins yet play a fundamental role in maintaining proper regulation and function of cellular activities. The term encompasses a myriad of different types of RNAs, due to the different properties these molecules can be categorized by, such as size, structure or function. In general, ncRNAs can be split into two major groups: housekeeping and regulatory. Housekeeping ncRNAs encompass RNAs that are necessary to maintain fundamental cellular functions such as translation, hence includes ribosomal RNA (rRNA) and transfer RNA (tRNA). For example, tRNAs range from 70 nucleotides (nt) to 95 nt and are constitutively expressed at high levels in all cell types. Regulatory ncRNAs are typically categorized based on their size, those shorter than 200 nt classified as small non-coding RNAs (sncRNA) and long non-coding RNAs (lncRNAs) for those larger than 200 nt. The mechanism in which these molecules exert their regulatory function in maintaining cellular functions varies wildly, but many are typically studied within the context of affecting transcriptional or translational events.

The advancement in high throughput sequencing technology has allowed more encompassing and intricate studies of ncRNAs, such as through the detection and discovery of more elusive types, in depth quantification and characterization of previously characterized species, and even the profiling of RNA specific properties such as modifications. The types and properties of ncRNAs can vary wildly in between different types of bacteria as adaptations to stressors and environments exerts selective pressure in maintaining specific profile of these ncRNAs. Using different bacteria, I leverage the power of high throughput sequencing technology to investigate two facets of ncRNAs in

prokaryotes: profiling of tRNA and tRNA modifications in response to stress conditions, and unbiased discovery and characterization of small RNAs.

In Chapters 2 and 3, I investigate the N1-methyladenosine (m^1A) tRNA modification profile in *Streptomyces venezuelae* using advanced tRNA sequencing technology. This study documents the dynamic nature of the m^1A tRNA signature in a mesophilic bacterium and how it responds to changes in the environment. Moreover, this investigation leads to the discovery and characterization of the first prokaryotic RNA modification eraser enzyme, one of two necessary maintenance proteins required for proper regulation of tRNA methylation levels.

In Chapters 4 and 5, I continue the investigation of ncRNAs in *Bacteroides fragilis*, a commensal human gut bacteria capable of causing opportunistic infection. Using high throughput RNA sequencing technology, we document changes in the tRNA profiles for this bacterium in response to oxygen stress. Additionally, I utilize the same sequencing data as a case study for an unbiased approach in the identification and profiling of ncRNAs that can be applied to other organisms which represents a heuristic approach to screen potential ncRNAs of interest.

Chapter 1 Introduction to Prokaryotic Non-coding RNA Biology

Sections 1.2 and 1.4 in this chapter is derived from a Review article, "tRNA modification dynamics from individual organisms to metaepitranscriptomics of microbiomes", published in Molecular Cell on March 3rd, 2022[1]. The authors are Wen Zhang, ***Marcus Foo****, A. Murat Eren and Tao Pan.*

1.1 Overview of Non-coding RNA

Non-coding RNAs (ncRNAs) are transcripts that do not encode a protein product and instead, function in many different integrated networks in essential and in regulatory roles. Compared to their protein-coding counterparts, ncRNAs play a much wider role in maintaining cell function due to the high cellular abundance of RNA species present in this group. This is understandable given that messenger RNAs (mRNA) only constitute a very small composition of RNA content in the cell, compared to the other major categories of ncRNAs such as transfer RNA (tRNA) and ribosomal RNA (rRNA). ncRNA analysis has become increasingly popular given the advancement in key experimental techniques that have become indispensable in uncovering the vast, interconnected complex networks governed by ncRNAs.

Housekeeping ncRNAs, such as tRNAs and rRNAs, have been extensively studies given their key roles in maintaining global cellular functions, in relation to translation and protein synthesis. Housekeeping ncRNAs, in general, range from the smaller end of ~50 nt to 500 nt or more, are constitutively expressed in the cell and are essential for cell viability. Shorter RNA fragments can be derived from tRNAs, known as

tRNA derived RNA fragments (tRFs), adding an additional axis of regulatory function by tRNAs, on top of their canonical role in codon decoding.

The other major group of ncRNAs are regulatory ncRNAs, that can be further classified based on their average size: small non-coding RNAs (sncRNAs) consist of about 200 nt while long non-coding RNAs (lncRNAs) are those larger than 200 nt. SncRNAs in bacteria are typically associated in direct regulation of gene expression, typically achieved through non-consecutive base pairing mechanisms targeting mRNAs. These consist of small RNAs (sRNAs) that either act as trans- or cis-regulatory elements which respond to environmental cues to control gene expression. This can happen at the translational or transcriptional level, depending on the type of sRNA. In contrast, lncRNAs contain more complex structures to establish a variety of interactions needed to maintain their regulatory functions in relation to maintenance of the bacterial genome[2]. These types of RNAs are less extensively studied as compared to their smaller counterparts, but some notable research includes the role of lncRNAs as potential immunostimulatory molecules in eukaryotic hosts they co-habit[3].

Bacterial ncRNAs remain of interest for investigation due to their roles in diseases, and as a result, a target for development of therapeutic assets. A notable example of this is RNAIII in *Staphylococcus aureus*, a gram-positive bacterium that is implicated in a wide variety of clinical diseases. RNAIII is a key molecule of the *agr* quorum-sending system that is responsible for controlling various virulence factors[4].

1.2 tRNA and tRNA Modifications in Bacteria

In all domains of cellular life, tRNAs are the most extensively modified RNA family. On average, a bacterial tRNA contains 8 modifications per molecule, while eukaryotic tRNA average about 13 modifications, corresponding to ~1 in 10 to ~1 in 5 residues being modified, each having critical implications on cellular physiology. From a general perspective, tRNA modifications found in the anticodon stem loop are integral in fine-tuning mRNA decoding, while modifications found outside of this region are typically involved with tRNA stability, folding, localization and overall quality control for proper maturation of the transcript (**Figure 1.1.**)

From a chemical basis, tRNA modifications span a wide range of complexity with the simplest tRNA modifications being nucleobase or 2'OH methylations, to hypermodified bases requiring a multi-step process for installation. Simpler modifications like methylations also tend to act as primary modification markers for mRNA (e.g., N6-methyladenosine, m^6A), chromosomal DNA (e.g., 5-methylcytosine, m^5C) and proteins (e.g., N-methyl-lysine). Some examples of modification on the other extreme end include the wybutosine (W) modification requiring 7 enzymatic steps of de novo synthesis and queuosine (Q) that can be further glutamatylated in bacteria or glycosylated with mannose and galactose in mammals. The wide range of chemical complexity thus implicates the importance of tRNA modifications as a part of the wider cellular networks of metabolism in a cell.

Dynamic changes to the tRNA modification fractions are expected to support proper regulation of cellular activities and changes in cell physiology. Some examples of mechanisms underlying this process includes affecting translational efficiency due to

altered decoding[5], changes in translation initiation[6], or as biomarkers for altering global cellular activity[7].

In bacteria, tRNA modifications can even play an important role in modulating their activity in relation to their immediate environment, particularly for those found in the human gut microbiome. The mammalian gut microbial communities live in multiple, different microenvironments, depending on the type of diet and metabolite intake by the host. Bacteria within the gut respond to these constant changes, either to utilize the influx of new nutrients, resist changes to the existing environment or respond in different manners to ensure survival.

One such example includes the regulation of tRNA 2'-O-methylation (Nm) in *E. coli* tRNAs upon the onset of mild antibiotic stress and limited nutrient supply. The regulation of Nm modification does not significantly influence translation, but rather, the immune-stimulatory properties of the bacteria[8]. tRNA 2'-O-methlyations, specifically at position G18 in the D loop, effectively suppress Toll-like receptors (TLR7) of the innate immune system in mammalian hosts[9]. Under typical laboratory growth conditions, the lack of Gm18 minimally affects translation efficiency. However, Gm18 was upregulated upon exposure to mild antibiotic conditions. *E. coli* tRNA isolated from mild antibiotic stress conditions caused lower activation of human plasmacytoid dendritic cells than those from conditions without antibiotic exposure. This tRNA modification dependent bacterial adaptation in response to antibiotics improves its survival within a human host.

Food intake also greatly influences gut bacterial physiology[10]. The gut bacterium *Lactobacillus rhanosus* GG was found to increase in abundance in the mouse gut in response to lemon-derived exosome-like nanoparticles, likely through increased

resistance to bile[11]. The underlying mechanism was proposed to be regulated by specific tRNA decay, in which downregulation of proteins Msp1 and Msp3 lead to decreases in cell wall hydrolysis, reducing bile accessibility into cells. The minor change of Msp1 and Msp3 mRNA levels suggests a potential mechanism at the translational or post-translational level. Indeed, the tRNA processing enzyme RNase P seems to be involved in mediating $tRNA^{Ser}$ decay which decreased the expression of Msp1/3 proteins.

Pathogenic bacteria also alter their tRNA dynamics to ensure survival during infection. On a global tRNA modification scale, high levels of tRNA methylations in *E. coli* was necessary in ensuring a general resistance against antibiotics[12]. In line with the use of modifications to control translation efficiency, the N1-methylguanosine (m^1G) at position 37 modification status in the anticodon loop significantly affects the translation of several membrane proteins through proline decoding in their open reading frames. The lack of the m^1G37 writer impaired membrane structure of *E. coli* and *Salmonella enterica*, sensitizing them to various antibiotic classes.

Upon onset of infection within a human host, reactive oxygen species production is a common general response in mitigating growth and spread of pathogenic bacteria. As a result, these pathogens have adapted specific methods to deal with this environmental stressor. Mycobacteria has been observed to modify their tRNA patterns upon exposure to hypoxia, a commonly induced environment due to the inflammatory response by the immune system [13]. Each stage of hypoxia-induced changes was associated with distinct patterns in tRNA modification. Specifically, the Doc regulon genes, which control the hypoxia response, had altered translation efficiency depending on their codon composition bias, particularly through decoding by $tRNA^{Thr}$(UGU).

Pseudomonas aeruginosa was found to modulate their expression of detoxifying genes such as KatA and KatB, at the translational level upon exposure to hydrogen peroxide stress[14]. Here, the writer of the N7-methylguanosine (m^7G) modification at position 46, TrmB was necessary in effectively translating Phe- and Asp-enriched mRNAs including the catalase enzymes KatA and KatB. Exposure to hydrogen peroxide stress was associated with increases in m^7G levels, while the lack of TrmB led to decreases in KatA and KatB expression, as well as an overall reduced resistance to hydrogen peroxide stress. These studies highlight specific mechanisms in which tRNA modification dynamics can be a main driver in mediating bacterial survival responses to human hosts defenses.

1.3 sRNAs in Bacteria

Regulatory sRNAs play a large role in regulatory networks and responses to stress in many bacterial species. This group encompasses a wide variety of mechanisms in which these ncRNA transcripts exert their function, be it by affecting mRNA stability or influencing ribosome binding. Early studies of regulatory sRNAs lead to characterization of housekeeping RNAs, such as transfer-messenger RNAs (tmRNAs) that rescues stalled ribosomes[15] and 4.5S RNA that function as the scaffold for the signal recognition particle (SRP) that directs proteins to the membrane. Examples of regulatory sRNAs that act on transcription include the conserved 6S RNA that sequesters RNA polymerases to control cellular activity depending on the growth density and RteR in *Bacteroides thetaiotamicron* that inhibits the transfer of CtnDOT, a conjugative transposon[16].

sRNAs can be split into general categories based on their specific functions or general mechanism of actions: sRNAs with specialized functions, antisense RNAs

(asRNA) and sRNAs with limited base pairing. sRNAs with specialized functions include housekeeping transcripts that are vital in maintaining key cellular processes and tend to be more widely conserved. Examples include the tmRNA, 4.5S RNA and 6S RNA mentioned above, but also RNase P RNA and Y-family RNAs. The 4.5S RNA is essential for the SRP protein targeting machine and is bound to the SRP54 protein (Ffh). While the Ffh protein handles the cargo recognition through binding of signal sequences, 4.5S RNA has an elongated hairpin structure that mediates assembly and disassembly of the SRP complex with its cognate receptor protein[17]. The 6S RNA structure mimics that of DNA and traps RNA polymerases bound with the house-keeping sigma factor to help assist cell transcription into stationary phase. This process is reversible upon entry into nutrient rich environments, where the bound polymerase transcribes part of the 6S RNA, resulting in the release of the former[18].

Antisense RNAs constitute a notable group of regulatory sRNAs due to their intimacy with target genes that can include small proteins that can be toxic to the cells, forming a toxin-antitoxin regulatory mechanism. For example, the IstR-tisAB pair is induced by DNA damage. Both transcripts are transcribed in opposite directions from different strands with the IstR sRNA base pairing with the tisAB toxin mRNA[19]. These types of sRNAs are dependent on proper base pairing to their specific target, controlling regulation at the translation level by affecting mRNA stability or at the transcription level by targeting termination events. A key feature of sRNAs is the seed region, typically constituting a conserved stretch of ~8 nucleotides[20] that begins the initial RNA duplex formation.

sRNAs with limited base pairings has been mostly characterized within the context of binding to a RNA chaperone partner, such as the Hfq protein[21]. Unlike the specific base pairing mRNA targets by asRNAs, a much wide range of sRNAs can bind their chaperone partners. Hfq-bound sRNAs tend to require a factor-independent intrinsic terminator, i.e., a stem-loop followed by a stretch of U residues. This binding event allows for stabilization of the sRNA once base paired with the target mRNA, leading to one of many regulatory events, such as RNA degradation, modulation of translation due to occlusion of ribosome entry or changes to mRNA stability.

Despite the relevance of RNA chaperones in sRNA function, some organisms like Bacteroides species do not contain equivalent chaperone homologs that have been characterized in other model bacteria[22]. Investigations into these types of bacteria requires alternative approaches to sRNA characterizations that were previously dependent of RNA chaperone enrichment and analysis of bound RNAs. Such studies typically focus on the common, conserved characteristics of sRNAs as examined through transcriptomic wide approaches, followed by biochemical validation of target sRNAs of interest[23].

1.4 High throughput sequencing of tRNAs

While liquid chromatography-mass spectrometry (LC/MS) continues to be a powerful approach to study tRNA modifications, recent advances in sequencing strategies have enabled comprehensive access to the modification dynamics with unprecedented throughput (**Table 1.1**). Variations in the library preparation steps, that target the different chemical properties or response to enzyme treatments, provide much

flexibility in either detecting a broad group of modifications or targeting a specific common modification. Leveraging specific properties of tRNAs and their modifications is a common theme in recently developed sequencing methods while others aim to alleviate common issues during library preparation steps through clever use of enzymes and specific adapters. Several methods also examine tRNA dynamics on a broader, encompassing scale, such as within tissues or in naturally occurring, complex microbiomes. However, the existing sequencing methods target only a subset of tRNA modifications that are open to detection and quantitative assessment by reverse transcription. There is much room for future development for more comprehensive and more precise measurements of tRNA modifications.

Watson-Crick face methylations are abundant in tRNAs; they introduce a major challenge in the reverse transcription step during cDNA synthesis. In eukaryotic tRNAs these include m^1A, m^1G, 3-methylcytidine (m^3C), and N2,2-dimethylguanosine ($m^2{}_2G$), which all cause some degree of misincorporation or truncated cDNA products during reverse transcription. However, several sequencing techniques take advantage of these observations of RT signatures derived from modifications. Ryvkin et al. developed the HAMR pipeline that identifies modifications through detection of nucleotides with significant sequencing error rates in Illumina sequencing results[24]. Detected modifications could be grouped into large modification families by similarities in their incorporation patterns.

The introduction of an enzymatic demethylation step in the sequencing library preparation allows for comparative analysis of treated and untreated samples to further pinpoint these modifications. ARM-seq[25] and DM-tRNA-seq[26] both take a similar

approach in using the *E. coli* AlkB enzymes to demethylate RNA samples, with the latter also utilizing a highly processive thermostable group II intron reverse transcriptase (TGIRT) to increase read-through of these modifications in cDNA synthesis. These techniques solidify the characteristic properties of these methylations in producing specific RT signatures. Additional analysis of the DM-tRNA-seq results[27] introduces the modification index (MI) which combines the fraction of mutation and stop at each modification site to describe the qualitative nature of measuring modification fractions.

More recently, mim-tRNAseq[28] builds upon the analysis of modification-induced nucleotide misincorporation, with additional approaches to address cDNA synthesis issues. Mim-tRNAseq achieves a substantial increase in full length cDNA reads by using the TGIRT with DNA adapters at the 3' end of tRNA and longer reaction time. Obtaining large portion of full-length tRNA reads is important in the precise mapping of mammalian tRNAs which are derived from many tRNA isodecoder genes with diverse sequences within the same anticodon family[29].

Another fundamental theme in detecting and quantifying specific modifications by sequencing uses chemical treatments that cause targeted changes to select modification groups. Pseudouridine (Ψ) is a prevalent modification in all RNA families and is also the most abundant modification in eukaryotic tRNAs[30]. A common Ψ detection method uses the carbodiimide (CMC) reaction that specifically forms a chemical adduct for Ψ that can be detected by RT stops[31]. Ψ-seq[32] and Pseudo-seq[33] use this CMC reaction approach to sequence Ψ in mRNAs and in tRNAs. DM-Ψ-seq[34] combines demethylase treatment and CMC reaction which significantly enhanced the data quality for Ψ analysis in tRNAs. Another Ψ detection method (RBS-seq) uses a bisulfite reaction

that removes the Ψ base, thus producing a deletion signature in the RT reaction[35]. More recently published methods iterate and improve on the bisulfite reaction based approach with BID-seq[36] and PRAISE[37] further optimizing reaction and RT conditions to improve transcriptome-wide mapping and quantification of Ψ sites

m^5C is another common modification in a variety of RNA families and tRNAs. The most commonly used m^5C detection method is the bisulfite treatment which converts unmodified C into U, but m^5C remains as C in sequencing results. Several studies[38-40] applied this technique to study this modification in bacterial, archaeal and eukaryotic tRNAs.

Alkaline hydrolysis is not limited to specific modification changes, as shown by hydro-tRNAseq[41]. Here, a limited alkaline hydrolysis step generates shorter, less structured RNA fragments containing fewer modifications, making them more amenable to cDNA library preparations. AlkAniline-Seq[42] combines the alkaline hydrolysis step to generate abasic sites, followed by extensive 5'- and 3'-dephosphorylation and subsequent aniline cleavage to profile susceptible modifications: m^7G, m^3C and dihydrouridine (D).

Specific chemical reactions that target m^7G (TRAC-seq) or m^3C (HAC-seq) modifications result in highly efficient RT stops for their detection and quantitation[43, 44].

Several other approaches have been documented to overcome cDNA synthesis related issues to enrich for tRNA molecules, reduce truncated cDNA reads or increase full length tRNA reads, all resulting in clearer, higher quality sequencing results of tRNAs. A common solution to these problems includes the construction of adapter molecules that select for tRNAs in the input sample, thus increasing the portion of sequencing data for

tRNAs in biological samples. Pang et al.[45] document the effects of stressors on tRNA levels in *S. cerevisiae* using a two separate ligation steps to adapters: the first defined-sequence linker ligates selectively to the 3' end of purified tRNAs, while the second ligation to the 3' end of the cDNA products assists with PCR amplification. Taken together, these two adapter ligation steps minimize modification induced falloffs during reverse transcription, while increasing read coverage. YAMAT-seq[46] uses a Y-shaped adapter that specifically ligates to mature tRNAs using T4 RNA Ligase 2. This high selectivity for mature tRNAs proves advantages in estimating expression levels of mature tRNAs in human cells while only requiring a low level of total RNA input. LOTTE-seq[47] addresses limitations of adapters selective for mature tRNAs through construction of an adapter capable of targeting both full length and prematurely terminated tRNA transcripts. This adapter is specifically ligated to the tRNA 3'-CCA end. Similarly, QuantM-tRNAseq[48] utilizes a split ligation strategy with complementary double-stranded adapters to monitor tRNA abundance and sequence variants secondary to RNA modifications. AQRNA-seq[49] further addresses potential biases and errors introduced in the library preparation steps, namely ligation biases towards specific RNA molecules. Using a two-step linker ligation with an optional AlkB treatment, Hu et al. were able to compare their findings with commercially available small RNA sequencing kits, establishing an optimized pipeline that presents a lack of significant length bias, as well as a potential direct correlation between sequencing read counts and true RNA copy numbers.

MSR-seq[1] goes beyond capturing tRNA specific information, instead providing a versatile platform capable of multiplexing options to better increase throughput analysis of small RNAs. Key features of this approach include the use of a biotinylated capture

hairpin oligonucleotide that supports easy processing of key library preparation steps and modularity of enzymatic or chemical treatment of RNA on-bead. In turn, this allows for parallel investigations into multiple RNA families to better characterize cell state changes and improved detection of modifications based on treatment steps easily introduced into the workflow. Given the vast interaction and regulatory network between small RNAs and other RNA families, approaches like MSR-seq set a strong example in highly modular and flexible techniques that support more in depth investigations in RNA biology.

These methods highlight the importance in examining and quantifying the potential biases and errors typical tRNA library preparation steps can introduce into the sequencing data, in which future methods and analysis pipelines should aim to further address, or at the very least, recognize.

The emergence of nanopore sequencing technologies provides another framework in sequencing tRNAs including the identification and quantification of modifications. One major benefit with nanopore sequencing is to bypass the issues associated with cDNA synthesis, as tRNAs can be sequenced directly. Nanopore also sequences single RNA molecules so that the coordination of multiple tRNA modifications can be studied simultaneously. However, the stable tRNA secondary structure can be an issue when translocating tRNA through the pores. Smith et al.[50] proposed possible methods to deal with linearizing tRNAs using specific adapters and polymerases to enhance processivity, thus avoiding clogged pores. Recently, Nano-tRNAseq[51] has shown great promise by utilizing RNA adapters that extend mature tRNA lengths combined with customization of the base processing software, improving base calling and mappability of tRNAs.

With only a limited selection of complete tRNA modification profiles of mostly model organisms[52], comparative analysis using these well described tRNA profiles is a useful approach for investigating the unknown territory tRNA profiles of other organisms. Kimura et al.[53] leveraged the known information on *E. coli* tRNA modifications to fully map and also identify novel *V. cholerae* specific tRNA modifications through combining tRNA sequencing and tRNA mass spectrometry.

1.5. Brief summary of thesis

The two projects covered in this thesis work constitutes major applications of high throughput sequencing technology in investigations related to non-coding RNA and their modifications. TRNA modifications constitute a major quantifiable property of tRNA profiles in cells and the m^1A tRNA modification is a remarkably well conserved example, yet not much is known about its role in bacterial systems. Using *Streptomyces venezuelae* as a representative bacterium, I utilize advanced tRNA sequencing approaches to document the dynamic nature of m^1A tRNA modifications and how it can serve as a relevant biomarker upon the onset of stress. I follow up this initial analysis by identifying and characterizing the related maintenance enzymes necessary to convey the reversible nature of this modification. This sets a foundation in further analysis of bacterial m^1A tRNA maintenance enzymes, particularly for RNA modifications erasers that are scarcely described in current literature.

Using the same high throughput RNA sequencing pipeline, I then show how such a versatile technology can be applied to simultaneous investigations of ncRNA classes[1], with a later focus give to the identification of unannotated small RNAs. Using *Bacteroides*

fragilis as an exemplar bacterium, I first analyze the changes in the tRNA profiles of the organism when stressed under oxygen conditions. This commensal bacterium displays a robust oxidative stress response, supporting its ability to escape the gut microenvironment, causing potential intra-abdominal abscesses infections. I identify changes to tRNA thio-modifications, whose installations are tied to iron-sulfur cluster enzymes. Similar to the m^1A modifications before, these thio-modifications are of interest as a potential biomarker that is specific to oxidative stress. Using the same sequencing library, I also initiate an investigation into previously uncharacterized small RNAs of interest that may be expressed by *B. fragilis* under stress conditions. By utilizing a more unbiased approach in filtering down reads of interest, I identify differences in the presence of a subset of these RNAs when comparing control conditions to the stressed samples. This workflow serves as a broad approach to ncRNA characterization that can be applied to other organisms.

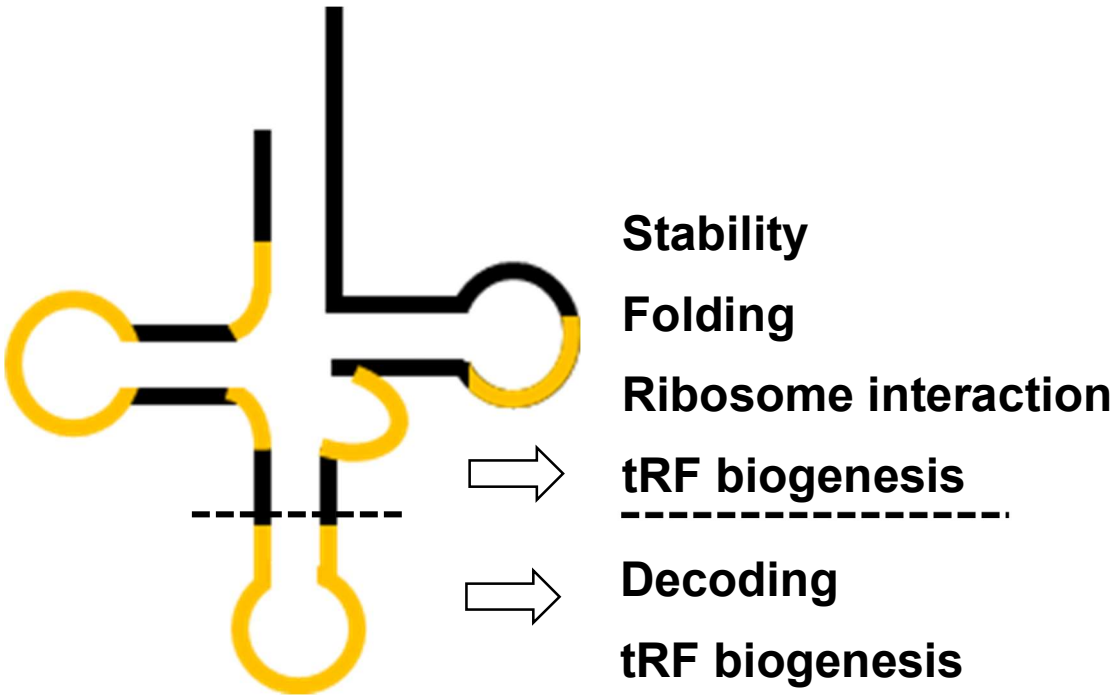

Figure 1.1: Broad categories of tRNA modifications. In the cloverleaf secondary structure of tRNA, the locations of the modified nucleotides are highlighted in orange. Dashed line separates the anticodon stem-loop from the rest of tRNA. Broad categorization of tRNA modification function is on the right.

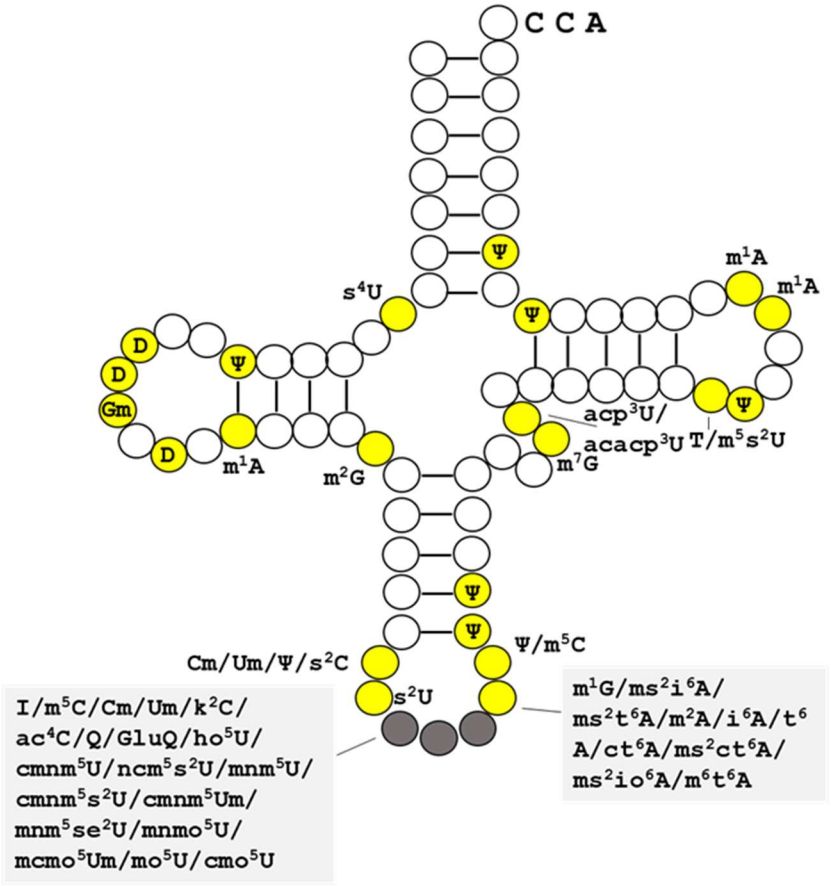

Figure 1.2: Microbial tRNA modifications. Modified positions are shown in yellow with the corresponding modification abbreviations. Result obtained from Modomics[52].

Category	Method	Year Published	Description	Capabilities	Limitations	Ref.
RT misincorporation	DM-tRNA-seq	2015	• Uses *E. coli* AlkB demethylase mixture and the highly processive TGIRT[a] in library preparation • This combination overcomes modification and tRNA structure issues during cDNA synthesis	• Identification of Watson-Crick face methylations and several other modifications; can work with total RNA • Generates higher proportion of full length tRNA reads; useful in tRNA isodecoder studies	• Additional steps required such as demethylase treatment and gel purification of cDNA	Zheng et al., 2015[26]
	ARM-seq	2015	• Uses *E. coli* AlkB demethylase in library preparation • Enables identification of methylations	• Identification of Watson-Crick face methylations • Improves tRNA fragment detection and quantitation	• Uses conventional RT[b] that stops frequently due to tRNA structure • Additional steps needed such as AlkB treatment	Cozen et al., 2015 [25]
	mim-tRNA-seq	2021	• Uses TGIRT under new conditions that read through tRNA modifications extensively • Computation pipeline addresses mapping issues of incorporating modification-induced mutation results	• Substantially increases tRNA full-length reads • Computation pipeline improves tRNA isodecoder analysis	• May need tRNA purification from total RNA	Behrens et al., 2021 [28]

Table 1.1: Summary of next-gen tRNA sequencing methods.

Category	Method	Year Published	Description	Capabilities	Limitations	Ref.
Chemical treatment to study specific modifications	Ψ-seq; Pseudo-seq	2014	• CMC reaction introduces RT stops at Ψ sites • Ψ sites in mRNA and tRNA mapped transcriptome wide across multiple RNA species	• Specifically targets Ψ sites	• Uses RT stop signature for detection; mostly short reads for tRNA	Schwartz et al., 2014 [32]; Carlile et al., 2014 [33]
	RNA bisulfite sequencing	2009, 2012, 2013	• Bisulfite reaction changes C to U, while m^5C remains as C in sequencing data • m^5C sites in mRNA and tRNA mapped transcriptome wide across multiple RNA species	• Specifically targets m^5C sites	• Uses RT stop signatures for detection; mostly short reads for tRNA	Schaefer et al., 2009 [39]; Squires et al., 2012 [40]; Edelheit et al., 2013 [38]
	DM- Ψ -seq	2020	• Uses AlkB demethylase treatment and CMC reaction to identify Ψ sites in tRNA at high efficiency	• Significantly improves tRNA coverage by using demethylase	• Ψ specific	Song et al., 2020 [34]

Table 1.1: Summary of next-gen tRNA sequencing methods. (Continued)

Category	Method	Year Published	Description	Capabilities	Limitations	Ref.
Chemical treatment to study specific modifications	AlkAniline-Seq	2018	• Alkaline hydrolysis coupled with aniline cleavage identifies m^7G and m^3C sites in tRNA	• Specifically targets m^7G and m^3C sites	• Ability to analyze modification levels unclear	Marchand et al., 2018 [42]
	RBS-seq	2019	• Bisulfite treatment induces mutation and deletion signatures for m^1A, Ψ and m^5C	• Simultaneous identification of multiple modifications	• Ability to analyze modification levels unclear	Khoddami et al., 2019 [35]
	TRAC-seq	2019	• Borohydride reduction coupled with aniline treatment identifies m^7G in tRNA	• Specifically targets m^7G sites	•	Lin et al., 2019 [44]
	HAC-seq	2021	• Hydrazine treatment coupled with aniline treatment identifies m^3C in tRNA	• Specifically targets m^3C sites	•	Cui et al., 2021 [43]
	BID-seq	2023	• New bisulfite treatment conditions generates unique deletion signature for Ψ	• Improved detection of Ψ sites	•	Qing et al., 2023 [36]
	PRAISE	2023	• New bisulfite treatment conditions and pipeline detects deletion signature for Ψ	• Improved detection of Ψ sites	•	Zhang et al., 2023 [37]

Table 1.1: Summary of next-gen tRNA sequencing methods. (Continued)

Category	Method	Year Published	Description	Capabilities	Limitations	Ref.
Library construction strategies	YAMAT-seq	2017	• Uses ligation of a Y-shaped adapter to the 5' and 3' ends of tRNA, followed by PCR of cDNA from full-length tRNA	• Specifically targets mature, full-length tRNAs	• Misses tRNA modification or structure derived short cDNA that are the majority of cDNA products	Shigematsu et al., 2017[46]
	Hydro-tRNAseq	2017	• Uses standard small RNA sequencing method with tRNAs after partial alkaline hydrolysis • Shorter tRNA hydrolysis products have less structure and fewer modifications	• Obtained short tRNA reads will be difficult for use in tRNA isodecoder analysis	•	Gogakos et al., 2017 [41]
	LOTTE-seq	2020	• Ligation a hairpin adapter to 3' CCA allows for selective library construction for tRNA	• Sequencing of full-length and prematurely terminated tRNA transcripts	• Short 3' tRNA reads due to modification or structure induced stops	Erber et al., 2020 [47]
	QuantM-tRNAseq	2021	• Utilization of split ligation strategy allows for quantification of tRNA abundance and sequence variants • Precise tRNA abundance measurements correlated with Northern blots	•	•	Pinkard et al., 2020[48]

Table 1.1: Summary of next-gen tRNA sequencing methods. (Continued)

Category	Method	Year Published	Description	Capabilities	Limitations	Ref.
Library construction strategies	AQRNA-seq	2021	• Uses exonucleases to remove excess adapters, allows for rapid and uninterrupted library construction • Direct correlation between read counts and copy number of miRNAs	• Measures tRNA abundance and modification changes	• Short 3' tRNA reads due to modification/structure induced stops without demethylase treatment • May need additional steps such as demethylase treatment	Hu et al., 2021[49]
	MSR-seq	2022	• Capture hairpin oligonucleotide supports modular treatment steps • Multiplexing of samples during library construction	• Parallel analysis of small RNA families	•	Watkins et al., 2022[1]

Table 1.1: Summary of next-gen tRNA sequencing methods. (Continued)

Chapter 2 m^1A tRNA modification profile in *Streptomyces venezuelae*

This chapter is derived from a manuscript submitted for publication on June 24th, 2024[54]. The authors are ***Marcus Foo****, Luke Frietze, Behnam Enghiad, Yujie Yuan, Christopher D. Katanski, Huimin Zhao, and Tao Pan. M. F., C. D. K., and T. P. conceived the project. M. F. performed S. venezuelae characterization, QQQ-LC-MS, and built MSR-seq libraries. L.F. analyzed MSR-seq data. B.E., H.Z. generated the knockout strains. Y. Y. performed antibiotic production measurements. M.F. and T.P. wrote the paper.*

2.1 Abstract

tRNA modifications help maintain tRNA structure, facilitate translation and stress response. Found in all three kingdoms of life, the m^1A tRNA modification occurs in the T loop of many tRNAs and stabilizes tertiary tRNA structure and impacts translation. M^1A in the T loop is reversible through the presence of demethylase enzymes, which bypasses the need of turning over the tRNA molecule to adjust their methylation levels in cells. Using *Streptomyces venezuelae*, we confirmed the presence and quantitative m^1A tRNA signatures using mass spectrometry and high throughput tRNA sequencing. Induction of stress conditions reveals global changes in tRNA methylation levels, highlighting the responsiveness of this modification in relation to environmental adaptation. Our characterization of the *S. venezuelae* m^1A tRNA profile documents the dynamic nature of this modification within a mesophilic bacterium outside the hyperthermic context in which

prokaryotic m^1A has been previously described, hinting at the extended role of the modification outside of structure maintenance.

2.2 Introduction

M^1A methylation in RNA has been a growing interest in RNA epigenetics and epitranscriptomics, given its wide distribution across the major classes of RNAs. First discovered as the base N1-methyladenine in 1961[55], this modification constitutes the simple addition of a methyl group to the N1 position of adenosine. Aided by the increasing number of methods for detecting and quantifying this modification, m^1A sites have been found present in mRNA[56-58], lncRNA[57, 58], mitochondrial transcripts[57, 59, 60], tRNA[6, 61, 62], and rRNA[59], with the highest quantity found in tRNA and rRNA.

In tRNAs, the m^1A modification is conserved in all three kingdoms of life (**Figure 2.1a**). M^1A has been mapped to positions 9, 22 and 58/59 of tRNAs (standard tRNA nomenclature is used here and below), but only m^1A58 in the T loop is conserved across the three kingdoms[62]. A key feature of m^1A is that its level can respond dynamically to cellular conditions, for example in the mouse cecum microbiome in response to diets[63]. Studies in human cells have implicated the role of m^1A58 as an additional regulator of translation initiation[6] and T-cell expansion[61], and of m^1A9 in maintaining proper mitochondrial tRNA folding[64], while in yeast, m^1A58 is required for initiator $tRNA^{Met}$ stability[65]. In thermophilic bacteria, m^1A58 is known to enhance tRNA stability at high temperatures[66]. Previous epitranscriptomic profiling studies have documented the presence of m^1A58 in other types of mesophilic bacteria[63], implying potential function beyond ensuring tRNA thermostability.

Detection of m^1A modification can be approached with a variety of methods, including antibody based assays[67], mass spectrometry methods[68] and most recently, sequencing based approaches[63]. High throughput sequencing methods allow for better resolution in mapping modifications[60] and can help with enrichment of samples through additional processing steps prior to library building[58]. These technologies take advantage of the chemical properties of this modification, either through the modification inhibiting reverse transcription by blocking Watson-Crick base pairing[56, 58] or the Dimroth rearrangement of m^1A into m^6A[56]. Incorporation of additional library preparation steps, namely addition of demethylase enzymes[25-27], have also yielded higher quality sequencing data, allowing for better mapping of methyl modifications, especially when combined with evolved reverse transcriptase, such as the highly processive thermostable group II intron reverse transcriptase (TGIRT)[26] to increase read-through.

Here, our work maps the m^1A tRNA modification landscape in *Streptomyces venezuelae*, which belongs to a bacterial genus known to contain tRNA $m^1A58/59$ modifications, initially mapped in *Streptomyces griseus* by mass spectrometry[69]. Using high throughput tRNA sequencing and mass spectrometry, we identify $m^1A58/59$ modifications in *S. venezuelae* in individual tRNAs. We document changes to this m^1A tRNA profile upon induction of stress, underscoring a potential biosensor property of the modification. Further investigation is necessary in understanding how these changes correspond to the bacterium's stress response regulation.

2.3 Results

Streptomyces venezuelae has a strong m^{1}A58/59 tRNA modification signature.

Our previous mouse cecum microbiome study documented the presence of m^{1}A in tRNAs among four bacterial classes: Actinobacteria, Bacilli, Clostridia and Mollecutes[63]. The Actinobacteria m^{1}A58/59 modification levels in the microbiome are significantly different for mice fed with a high fat or low fat diet. We hypothesized that the responsiveness of the modification levels is conserved in other related bacteria and can be observed for other types of stress conditions. From a mechanistic standpoint, this is achieved through the existence of m^{1}A eraser enzyme in these bacteria, along with a dedicated installation enzyme for the modification (**Figure 2.1a**); the maintenance aspect of the m^{1}A modification will be explored in the following chapter. Examining the model organisms in the RNA modifications base, MODOMICS[52], we noted that a representative Actinobacteria, *Streptomyces griseus*, contained m^{1}A58/59 across its tRNA population as determined by LC/MS[69]. Reasoning that m^{1}A could be a conserved feature in Streptomyces species, we selected *Streptomyces venezuelae* as our model system. *S. venezuelae* is known for its ability to produce chloramphenicol and can be grown easily under laboratory conditions, along with being amenable to genetic manipulation through a variety of CRISPR based methods[70].

To confirm that *S. venezuelae* indeed contains m^{1}A in their tRNAs, we extracted the total RNA from laboratory cultures and identified m1A as a highly abundant modified nucleotide that is primarily present in their tRNAs (**Figure 2.1b**). Given the 58/59 position specific location in which m^{1}A is found in tRNAs, we applied multiplex small RNA sequencing (MSR-seq)[1] to identify and locate the m^{1}A positions in individual tRNAs (**Figure 2.1c-g**). MSR-seq relies on Superscript IV reverse transcriptase (RT)

readthrough of the m^1A which induces a “mutation” signature compared to the reference tRNA sequence in high throughput sequencing data analysis[1]. To confirm that the mutation signature at putative m^1A positions is indeed derived from m^1A modification, we also constructed another sequencing library of the same sample in parallel, treated with the purified, recombinant *E. coli* AlkB demethylase which selectively removes m^1A[71]. Indeed, *S. venezuelae* tRNAs show moderate and high mutation rates at position 58 or 59 in almost all tRNAs in our sequencing data (**Figure 2.1c**).

The presence of the m^1A59 modification was an unexpected find given previous reports only noting m^1A58 as the conserved modification, along with m^1A57 as the only other adjacent modification found in archaea[62]. The m^1A59-modified tRNAs were GluCUC-1, GluCUC-2, GluUUC, SerCGA, SerGGA, SerUGA, and TyrGUA (**Figure 2.1e**). These 7 tRNAs share a common sequencing motif of CCN in the β-region of the D loop that is absent in all other tRNA with m^1A58 modification (**Figure 2.1h**). This motif likely serves as a potential recognition element for the writer of these 7 tRNAs. Regardless of the position, the mutation rate reversed to background levels upon demethylase treatment of the RNA, validating them as *bona fide* m^1A modifications (**Figure 2.1d, f**).

Streptomyces venezuelae tRNA m^1A levels change under stress.

To examine how tRNA m1A modification responds to stress, we subjected the *S. venezuelae* liquid cultures to commonly used heat stress (from 30°C to 42°C) or ethanol shock (addition to 6% v/v)[72, 73]. We first measured the global m1A levels using QQQ LC-MS comparing unstressed and stress conditions for the WT strain (**Figure 2.2a**). Overall, no significant changes were observed for either stress condition compared to the control.

To further investigate changes in m^1A levels among individual tRNAs, we performed MSR-seq on the same samples, given the lack of resolution with mass spectrometry approaches (**Figure 2.2b**).

Comparing the m^1A58-modified tRNAs in the wild-type strain, the heat stress induced a significant increase in m^1A levels (**Figure 2.2b, d**), indicating that tRNA m^1A levels can indeed change upon stress. Certain subsets of tRNAs showed much more dramatic changes, highlighting them as potential tRNAs of interest for future investigations as they serve as a modification profile that responds to heat stress as to support cellular activity changes. As for the ethanol stress, more subtle changes were observed, happening in both directions, with no clear indication of tRNAs that are most responsive to this type of stress.

For the m^1A59 modified tRNAs, essentially no changes were observed for any conditions in the wild-type strain, suggesting that m^1A59 is resistant to demethylation, at least under the conditions tested here.

No significant changes in tRNA abundance or charging were seen under these stress conditions for the wild-type strain (**Figure S3.3a**), suggesting that adjusting tRNA m^1A levels constitutes the major tRNA response to stress conditions.

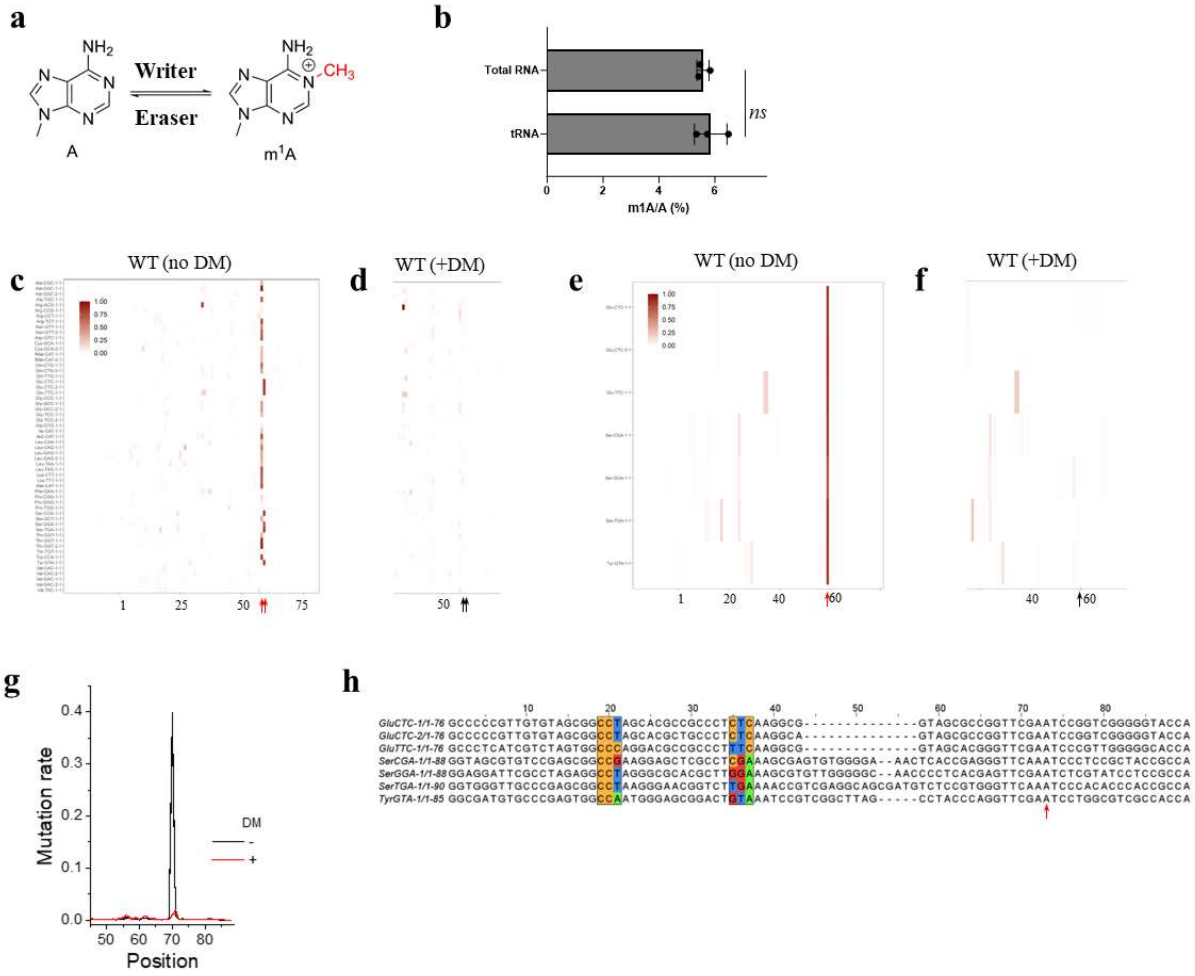

Figure 2.1: *Streptomyces venezuelae* contains high levels of tRNA m^{1}A modifications.

(a) Chemical structure of m^{1}A and schematic of m^{1}A writer and eraser.
(b) m^{1}A levels in total RNA or total tRNA isolated from SV cultures by QQQ LC-MS.
(c) MSR-seq positional mutation rate for SV tRNAs. Heatmap scale is from 0-1. Arrows indicate the position 58 and 59 (standard tRNA nomenclature).
(d) Same sample as (c) but treated with *E. coli* AlkB treatment before cDNA synthesis in library construction. Same heatmap scale as (c).
(e) MSR-seq positional mutation rate for SV tRNAs with positions m^{1}A59 only. Heatmap scale is from 0-1. Arrows indicate the position 59 (standard tRNA nomenclature).
(f) Same sample as (e) but treated with *E. coli* AlkB treatment before cDNA synthesis in library construction. Same heatmap scale as (e).
(g) Mutation rate of a tRNA exemplar, tRNALeu(GAG)-1. Sequencing replicates for both no-demethylase and DM-treated libraries are shown.

(h) Alignment of tRNAs that retain m^1A modification levels in the Trm1L-KO strain. Highlighted are the 3-nt β-region in the D loop that share C20a/C20b. Anticodon nucleotides are also highlighted. Position 59 (standard tRNA nomenclature) is indicated by the red arrow, accounting for differences in variable loop length.

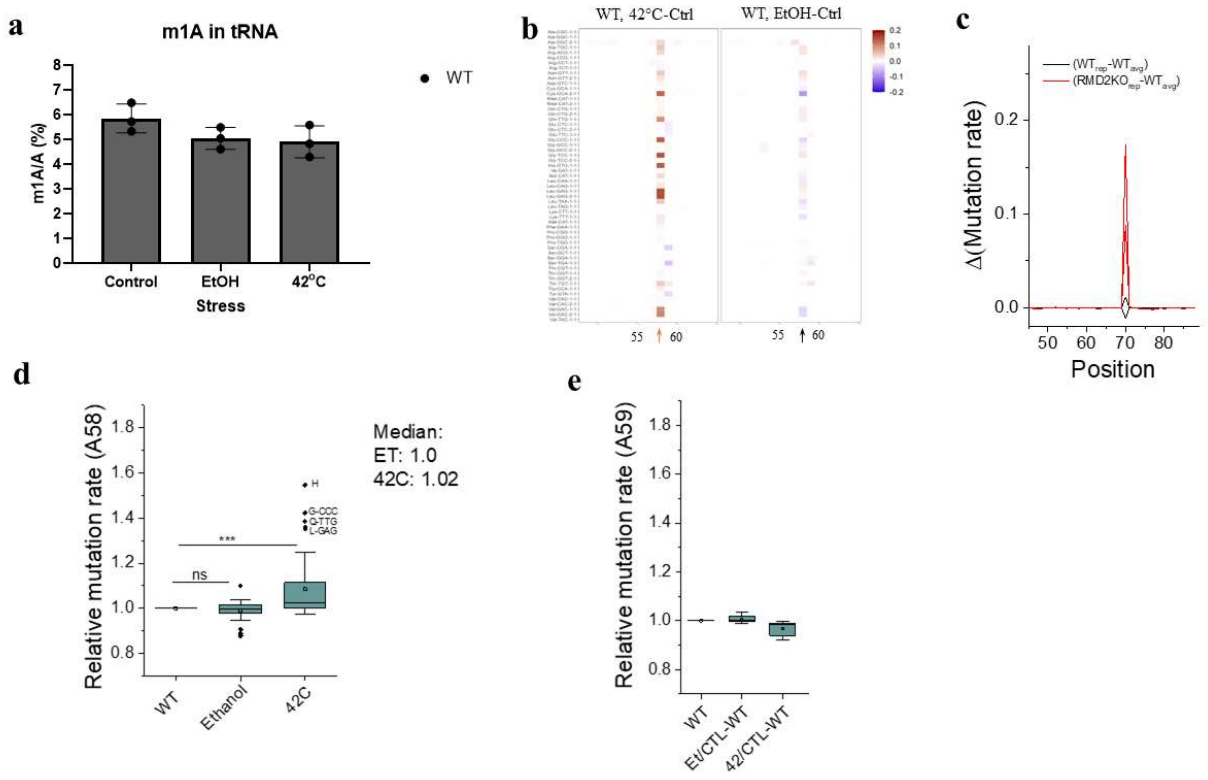

Figure 2.2: tRNA m[1]A modification response to stress.

(a) m[1]A levels in WT unstressed and stressed by QQQ LC-MS. Individual data points are shown.
(b) MSR-seq results of WT unstressed and stressed sample. Heatmap scale is from -0.2 to +0.2, mean of n=3 biological replicates for each condition. Blue: decreased mutation rate in stress compared to unstressed, red: increased mutation rate. Arrows indicate the position 58.
(c) Δmutation rate of WT (Stress-unst*) of a tRNA exemplar, tRNALeu(GAG)-1. Unst* = average mutation rate of no stress replicates.
(d) Mutation rate ratios for tRNAs with methylation at position 58 for WT strain, stress conditions normalized relative to unstressed growth. Outliers are denoted by their amino acid and anticodon.
(e) Mutation rate ratios for the 7 tRNAs with methylation at position 59 for WT strains, under stress conditions. Within each strain, stress conditions are normalized relative to unstressed growth.

2.4 Discussion

The advancement of high throughput sequencing technologies has afforded more in depth studies of cellular responses to stress conditions from the perspective of ncRNA and their corresponding properties like modifications and abundance. Currently, only a handful of bacteria has been surveyed for their tRNA modification profiles with some placing an emphasis on unique modifications to the specific species[53]. These in depth profiles allow for single nucleotide resolution characterization of modifications, opening an avenue to study the potential roles of these modifications within context of cellular physiology and regulation. The m^1A modification, while better studied in eukaryotic organisms, is mostly enigmatic with respect to bacteria, other than in providing thermostability in thermophilic bacteria[66]. Given its wide conservation within bacteria, as well as eukaryotes and archaea, we sought to investigate the potential role of this modification in mesophilic bacteria. Using MSR-seq, we show that *Streptomyces venezuelae* has a profound m^1A tRNA signature at positions 58 and 59 of tRNAs. These methylation levels were responsive to environmental stressors, with heat shock causing a subset of tRNAs to respond with dramatic increases in modification.

2.5 Materials and Methods

Bacterial strains and culture conditions

The wild-type *Streptomyces venezuelae* (ATCC10712) were grown at 30^{o}C using ISP2[74] (International Streptomyces Project yeast malt extract agar medium) for general growth and maintenance. Unless stated otherwise, liquid cultures were grown in 250 ml baffled flasks containing 50 ml of medium inoculated with starter *S. venezuelae* cultures

to a starting optical density at 600 nm (OD600) of ~ 0.05 with shaking (225 rpm). Starter cultures were prepared starting with a streak from frozen *Streptomyces* stocks onto ISP2 agar, grown up to 2 days at 30°C, then isolated colonies were inoculated into an overnight culture of ISP2 medium.

RNA modification mass spectrometry analysis

RNA samples were digested into single nucleosides using nuclease P1 (Sigma-Aldrich) incubation at 42 °C for 3 hours followed by incubation with Fast AP (Invitrogen) at 37 °C for another 3 hours. The reaction mixture was then filtered through 0.22 μm Millex-GV polyvinylidenedifluoride filters (Millipore). Samples were injected at 5 μl volumes and analyzed in triplicate using a ZORBAX SB-Aq 4.6 x 50 mm column (Agilent) on UHPLC (Agilent) coupled to a SCIEX 6500+ Triple Quadrople Mass Spectrometer in positive electrospray ionization mode. Nucleosides were quantified based on the nucleoside to base transitions and compared with calibration curves.

Extraction of S. venezuelae total RNA and tRNA

Total RNA was extracted from cell pellets using TRIzol (Thermo Fisher) by adapting manufacturer's protocol. Briefly, cell pellets were resuspended in TRIzol and ~100 mg 0.1 mm glass beads (Biospec) then homogenized in two 1-minute intervals. An equal amount of chloroform was added to the homogenate and the solution was mixed vigorously before centrifugation at maximum speed for 15 minutes. The clear aqueous layer was transferred to a fresh tube and 100% isopropanol was added at 1.5x volume excess. Nucleic acid was precipitated by incubation at -80°C for 1 hour. The sample was

then centrifuged at maximum speed for 30 minutes, forming a nucleic acid pellet that was washed with 75% EtOH, allowed to dry and resuspended in resuspended in nuclease-free H_2O. Total RNA concentrations were measured using a NanoDrop 2000 Spectrophotometer (Thermo Scientific). tRNA samples were isolated from total RNA samples from a urea denaturing gel, followed by extraction of RNA from the excised bands ranging from ~70 to ~100 bp. Briefly, total RNA samples were mix with 2x RNA loading buffer (9 M Urea, 100 mM EDTA, pH 8, 0.2% Bromophenol blue, 0.2% Xylene cyanol) loaded onto a pre-run 10% TBE-Urea Polyacrylamide gel (Criterion). The gel was stopped when the Bromophenol blue band reached the bottom of the gel. The gel was stained with SYBR gold (Thermo) and scanned using a Chemi-Doc imaging system (Bio-Rad). Excised gels were mixed with elution buffer (200 mM KCl, 50 mM KOAc, pH7) at estimated 2x volume and the gel was crushed up with a clean pipette tip. Samples were incubated overnight while spinning. Tubes were then centrifuged with the supernatant was transferred to a new tube and the nucleic acid pellet was recovered and wash, per the final recovery steps as listed above.

MSR-seq library preparation

Streptomyces venezuelae total RNA MSR-seq libraries including AlkB treatment for library construction was prepared as described by Watkins et al[1]. Major steps of the protocol were followed as stated in the general protocol for MSR-seq, in sequential order: One-pot diacylation and β-elimination for tRNA charging; first barcode ligation; binding to dynabeads; dephosphorylation; reverse transcription; RNase H digestion; periodate oxidation; second ligation; PCR; TBE-PAGE gel extraction. 1 µg of total RNA were used

as the starting input material. Each replicate was assigned a different first barcode with the strains being assigned as the second ligation index. Following the gel run of PCR products, lanes were cut at the expected tRNA band sizes of ~150 bp to 250 bp for extraction and recovery of the final library preparation samples.

Sequencing data analysis

The data analysis followed the MSR-seq data processing pipeline[1]. Libraries were sequenced from Illumina NovaSeq 6000 platforms. The resulting paired-end reads were then demultiplexed by the identification of barcode sequences using Je demultiplex[75] with the following parameters: BPOS=BOTH BM=READ_2 LEN=6:4 FORCE=true C=false. These options are optimized for samples where the barcode sequence is likely present on read 2. Barcode sequences are as previously described18. Following demultiplexing, data was aligned using bowtie233 (version 2.3.3.1) with the following parameters: -q -p 10 --local --no-unal. These reads were aligned to a curated Streptomyces Venezuelae tRNA reference which was curated from the Streptomyces venezuelae ATCC 10712 reference from the GtRNAdb: Genomic tRNA Databse[76]. This reference contained sequences of tRNA that was curated for non-redundancy, to add 5S rRNA, to remove low scoring tRNA so that all tRNA-scan SE score > 46, removing intron sequences, and appending 3' "CCA". The Bowtie2 output sam files were converted to bam files, which were then sorted by samtools' sort function[77]. IGVtools count was used to collapse reads into 1nt windows using the following parameters: -z 5 -w 1 -e 250 —bases. The resulting IGV output wig files were reformatted using a custom Python script to obtain mutation rate and read coverage compatible with R for data visualization and analysis.

To obtain a more accurate position for mutation rates, alignment was done from the 3’ CCA end. R was then used for visualization of the data and Welch’s unequal variances t-test was used for the abundance volcano plots and the global charging box and whiskers plots.

Chapter 3 Characterization of key *Streptomyces venezuelae* maintenance enzymes

This chapter is derived from a manuscript submitted for publication on June 24th, 2024[54]. The authors are ***Marcus Foo****, Luke Frietze, Behnam Enghiad, Yujie Yuan, Christopher D. Katanski, Huimin Zhao, and Tao Pan. M. F., C. D. K., and T. P. conceived the project. M. F. performed S. venezuelae characterization, QQQ-LC-MS, and built MSR-seq libraries. L.F. analyzed MSR-seq data. B.E., H.Z. generated the knockout strains. Y. Y. performed antibiotic production measurements. M.F. and T.P. wrote the paper.*

3.1 Abstract

The m^1A modification is found in in the T loop region of tRNAs and is very well conserved between organisms from all three kingdoms of life. It is normally associated with maintenance of tertiary tRNA structure and can impact translation as well as tRNA maturation. M^1A in T loop is reversible by three mammalian demethylase enzymes, which bypasses the need of turning over the tRNA molecule to adjust their m^1A levels in cells. However, no prokaryotic tRNA demethylase enzyme has been identified that acts on endogenous RNA modifications. Using *Streptomyces venezuelae* as a target organism, we identified two RNA methylases that can remove m^1A in tRNA and validated the activity of a previously annotated tRNA m^1A writer. Using single gene knockouts of this set of maintenance enzymes, we found dynamic changes in m^1A levels of many tRNAs under stress conditions. Phenotypic characterization highlighted changes in their growth and

altered antibiotic production. Our identification of the first prokaryotic tRNA demethylase enzyme paves the way for investigating new mechanisms in translational regulation in bacteria.

3.2 Introduction

tRNA methylations are installed through a family of RNA methyltransferases (MTases) that consist of more than 60 members in humans. MTases responsible for the m^1A modification can be categorized into two superfamilies: RFM or SPOUT, both of which require S-adenosyl-L-methionine (SAM) as a methyl donor. The SPOUT superfamily of proteins allows for methylation of tRNAs and rRNAs[78-80] with the SPOUT domain itself acting as the catalytically active subunit and binds the SAM cofactor. The RFM family contains RNA MTases and has a Rossmann-fold acting as the structural motif for binding cofactors. Enzymes from this family are most associated with MTases responsible for methylation in tRNAs at position 58[81]. They are be found in monomeric states, but have also been reported to exist in di-, tri-, and tetrameric oligomeric states depending on the organism in which they are characterized. For bacteria, the TrmI proteins are the dedicated m^1A58 MTases, typically existing as a homotetrameric structure[82-84]. Recognition of the target methylation side is dictated through recognition of the tRNA structure, with emphasis placed on the aminoacyl and T-stem loops[85].

As of today, m^1A is the only tRNA modification known to be reversible by demethylase enzymes or erasers. In humans, three RNA demethylases, AlkBH1, AlkBH3, and FTO have been described to reverse m^1A in both cytosolic and mitochondrial tRNAs[6, 86, 87]. The presence of m^1A erasers allows cells to control the m^1A methylation status more

rapidly than waiting for tRNA turnover, which in turn can enable faster response and adaptation to environmental changes. All three human m^1A erasers belong to the alpha-ketoglutarate-dependent dioxygenase AlkB-like superfamily. AlkB was first discovered in *E. coli* as a DNA/RNA aberrant m^1A and m^3C methylation repair enzyme[71]. *E. coli* does not have natural m^1A modification in their tRNA or rRNA, hence AlkB's natural RNA substrates are rRNA or tRNA methylations that arise from chemical damage. $M^1A58/59$ and m^1A22 are present in many bacterial taxa[83], however, a prokaryotic m^1A demethylase acting on natural m^1A modifications was unknown.

This work identifies and characterizes two m^1A demethylases in *Streptomyces venezuelae*, as well as confirms the role of an annotated m^1A methyltransferase as the major m^1A58 tRNA writer. We find two candidate genes as m^1A erasers and another as an m^1A writer using protein homology searches. We generate single gene knockout strains of these three genes and subject them to different stress conditions to monitor changes in individual tRNA m^1A levels using high throughput tRNA sequencing and mass spectrometry. We also test these strains for phenotypic differences to reveal a potential role of tRNA m^1A in this bacterium.

3.3 Results

Streptomyces venezuelae has two putative tRNA demethylase proteins targeting tRNA m^1A.

Having established that *Streptomyces venezuelae* has a strong $m^1A58/59$ tRNA signature that is responsive to environmental changes, we hypothesize that mechanistically, this is achieved through the presence of m^1A eraser that is capable of

acting on the modification installed by a dedicated m^1A writer, thus forming a concise set of m^1A maintenance enzymes. As before, we reasoned that m^1A could be a conserved feature in *Streptomyces* species, so we performed a homology search based on the *E. coli* AlkB protein for the *Streptomyces* genes. We found that this superfamily is indeed well conserved in the *Streptomyces* genus, especially for the key enzymatic domain of AlkB (alpha-ketoglutarate-dependent hydroxylase) that is responsible for catalyzing the demethylase reaction utilizing its non-heme iron active site (**Figure 3.1 a**).

To experimentally test the potential m^1A demethylase enzymes among them, we continued with *Streptomyces venezuelae* as our model system. Blastp results led to the identification of one putative, unannotated gene in *S. venezuelae*, which we designated as putative t**R**NA-**m**^{1}A-**d**emethylase 1 (RMD1) (**Figure 3.1a, d**). Unexpectedly, additional genome gazing led to another gene that contained the AlkB domain and was annotated as DEJ43_06135 in the *S. venezuelae* CP029197 genome in NCBI, which we designated as tRNA-m^1A-demethylase 2 (RMD2) (**Figure 3.1b, e**). Sequence alignments of the SV RMD proteins with the *E. coli* AlkB protein revealed major conservation in the catalytic domain, with SV RMD1 maintaining a higher overall sequence conservation as compared to SV RMD2.

We also noted a putative *S. venezuelae* methyltransferase gene for installing the m^1A modification by using the *Thermus thermophilus* TrmI as the previously known prokaryotic m^1A writer protein. This *S. venezuelae* gene has already been annotated as tRNA (adenine-N1)-methyltransferase in the CP029197 genome (**Figure 3.1c**), although there was no experimental evidence to indicate that it actually methylates tRNA.

Using the two putative SV demethylase gene sequences, we cloned, expressed, and purified the recombinant proteins using the *E. coli* T7 expression system. We verified these recombinant proteins for potential demethylase activity in vitro by QQQ-LC-MS. Total yeast tRNA was used at the biochemical substrate for this assay due to its commercial availability and the presence of high levels of m^1A58 modification. Both SV RMD proteins showed appreciable m^1A demethylase activity, although the activities were weaker relative to the recombinant, purified *E. coli* AlkB protein control (**Figure 3.1 f**).

M^1A erasers and a writer affect individual tRNAs differently.

To determine whether RMD1 or RMD2 can indeed erase tRNA m^1A modification *in vivo*, we generated single *S. venezuelae* gene knockout (KO) strains using the CRISPomyces system, a CRIPSR-Cas9 system specifically designed for Streptomyces species[88]. We also generated a KO strain of this TrmI-like gene (named as Trm1L gene) to compare with our eraser KO strains. We successfully generated all 3 single gene KO strains as shown by Sanger sequencing (**Figure S3.2 a-c**).

To determine the effect of these three single gene KOs on m^1A modification globally, we performed QQQ-LC-MS and dot blots using a specific m^1A antibody on total RNA isolated from *S. venezuelae* cultures grown in a rich medium (**Figure 3.2a, S3.2d**). Knocking out the Trm1L gene reduced the global m^1A level to very low levels, confirming that this *S. venezuelae* protein is indeed a tRNA m^1A writer. However, knocking out either the RMD1 or RMD2 gene showed little change in global m^1A levels. M^1A modification quantification using mass spectrometry measures the global m^1A levels that may not reflect m^1A changes in individual tRNAs. Therefore, we performed MSR-seq on the wild-

type and these three KO strains to determine the KO effect on the m^1A level in individual tRNAs (**Figure 3.2 b-f**).

Comparing RMD2 KO with the WT strains, we observed an overall increase in mutation rates, suggesting that RMD2 is indeed a tRNA m^1A eraser *in vivo* (**Figure 3.2b-d**). Furthermore, Trm1L knockout eliminated all m^1A58 modifications (**Figure 3.2d, e**), confirming Trm1L's role as the tRNA m^1A58 methyltransferase. On the other hand, comparing RMD1 KO with the WT stains only showed both up or down m^1A level changes (**Figure 3.2c**), with an overall decrease in m^1A58 (**Figure 3.2d**). One potential explanation for this unexpected result is that RMD1 and Trm1L form a complex to coordinate the m^1A58 modification levels. RMD1 KO would therefore impair TRM1L's ability to methylate A58 in some tRNAs. On the other hand, RMD1 likely plays a role as a stress-inducible tRNA m^1A eraser (see below).

As for the m^1A59-modified tRNAs, the m^1A levels remained unaffected in both RMD KO strains and slightly increased in the Trm1L KO (**Figure 3.2e, f**), supporting the possibility that *S. venezuelae* has a second, unknown tRNA m^1A writer that installs m^1A at position 59. This second m^1A writer likely uses the above mentioned D-loop motif as a recognition element (**Figure 2.1h**).

We also compared the changes in tRNA abundance and charging level upon writer or eraser KO, as these are part of the MSR-seq data. Very little difference in tRNA charging was observed (**Figure S3.2e**). However, many significant changes in tRNA abundance were observed for all three KO strains (**Figure 3.2g**). Venn diagrams of specific tRNAs that are up- or down-regulated in abundance upon eraser or writer KOs (**Figure S3.2f**) show counteraction of the writer Trm1L or the eraser RMD1 only on a small

number of these tRNAs, while most tRNAs only respond alone to either writer or eraser KO. This result suggests that RMD1, RMD2, and Trm1L all contribute to tRNA abundance regulation, but without much coordination among these tRNA erasers and writer.

S. venezuelae has an 8.2 Mb genome at a GC content of 72.5% and contains over 7,400 annotated genes. This very large genome and very high GC content may imply the existence of an unusual mechanism of tRNA abundance regulation. The high GC content highly biases its synonymous codons to be G-ending over A-ending (**Figure 3.2h**). We found that the tRNA abundance also accommodates this codon bias: tRNA isoacceptors with C34 wobble nucleotide are expressed at much higher levels than the corresponding isoacceptors with U34 wobble nucleotide (**Figure 3.2i**).

tRNA m^1A levels change under stress.

As before with the wild-type strain, we subjected all the generated *S. venezuelae* KO strains in liquid cultures to commonly used heat stress (from 30°C to 42°C) or ethanol shock (addition to 6% v/v). Globally, with a single exception, no significant changes were observed for all stress conditions compared to unstressed control. As described above, we performed MSR-seq to look for changes in m^1A levels among individual tRNAs (**Figure 3.3b-g**).

Comparing RMD1-KO to WT, m^1A58 levels increased under both stress conditions (**Figure 3.3d-f**). The same comparison between RMD2-KO and WT strains showed smaller differences for either stress condition (**Figure 3g, h**). These results are consistent with RMD1 being the dominant tRNA m^1A eraser under stress conditions. Combined with our results above showing RMD2 as the dominant eraser under steady-

state growth, the two tRNA m^1A58 erasers of *S. venezuelae* seems to actively demethylate under distinct cellular conditions.

For the m^1A59 modified tRNAs, no changes were again observed for any conditions in any strain, highlighting its resistance to demethylation, for the conditions tested here.

Like the wild-type strain, we did not observe any major changes in tRNA abundance under these stress conditions for the WT (**Figure S3.3A**), Trm1L-KO (**Figure S3.3b**), RMD1-KO (**Figure S3.3c**), and RMD2-KO (**Figure S3.3d**), giving emphasis to changes in tRNA methylation levels as the prominent tRNA response to stress conditions.

Eraser KO strains affect metabolic processing and antibiotic production.

Finally, we determined whether tRNA m^1A eraser or writer knockouts conferred phenotypic changes of *S. venezuelae*. Only the Trm1L-KO strain showed a moderate decrease of growth in the rich medium compared to the wild-type (**Figure S3.4a**). To examine metabolic processing and growth differences, we used the high throughput screening system of Biolog plates[89], which allowed for examining up to 1,000 different growth conditions. Briefly, each strain was grown up to early log growth conditions then transferred to the various 96 well plates. These plates were grown up over a period of 48 hours, where their growth could be tracked through a colorimetric output (**Figure 3.4a, S3.4b, c**). Identical conditions were maintained between runs of the different strains, allowing for us to compare the growth curves of each strain in each condition. These growth curves were fit to a typical growth curve formula using the Growthcurver R package, allowing comparisons of various growth attributes. In most cases, we did not

see any major differences between the 4 strains, but noted several notable differences related to use of different carbon sources (**Figure 3.4b, S3.4d, e**). For example, D-Ribose was not used effectively by the RMD2 KO strain, while RMD1 KO strain had a better growth response to this carbon source, as compared to the wildtype.

We also examined the gene KO effects on antibiotic production under steady-state growth conditions (**Figure 3.4c**). We observed a significant decrease in chloramphenicol and in Jadomycin B production for the KO strains compared to the wild-type strain. Surprisingly, both eraser and writer KO decreased antibiotic production, suggesting that a consistent tRNA m^1A modification level contributes to optimal antibiotic production.

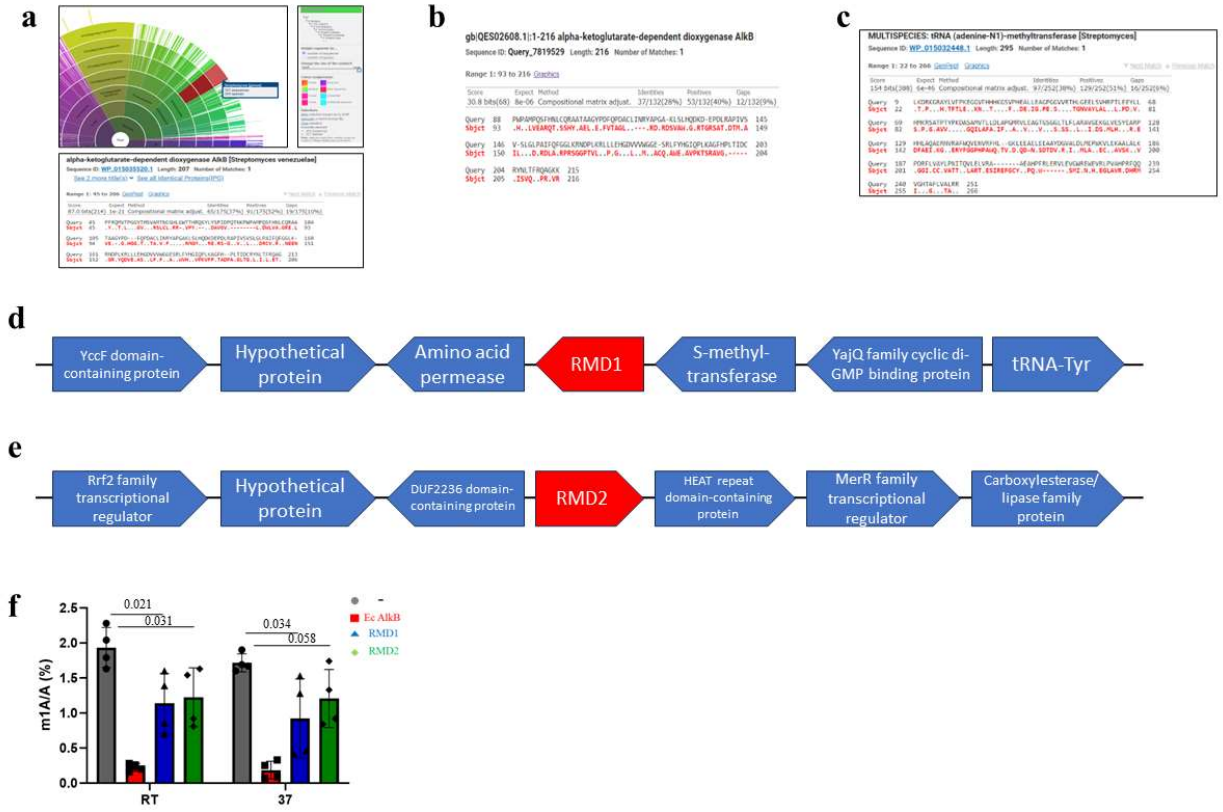

Figure 3.1: *Streptomyces venezuelae* contains one m^{1}A writer protein and two putative m^{1}A demethylase proteins.

(a) Pfam search of *E. coli* AlkB homologs in *Streptomyces* species. Protein sequence alignment of *S. venezuelae* (SV) RMD1 gene to *E. coli* AlkB.
(b) Protein sequence alignment of *S. venezuelae* RMD2 gene to *E. coli* AlkB.
(c) Protein sequence alignment of SV TrmI-like gene to *Thermus thermophilus* TrmI.
(d) Genome position map of SV RMD1 gene.
(e) Genome position map of SV RMD2 gene.
(f) In vitro demethylase assay using recombinant, purified SV RMD1 and RMD2 protein by QQQ LC-MS.

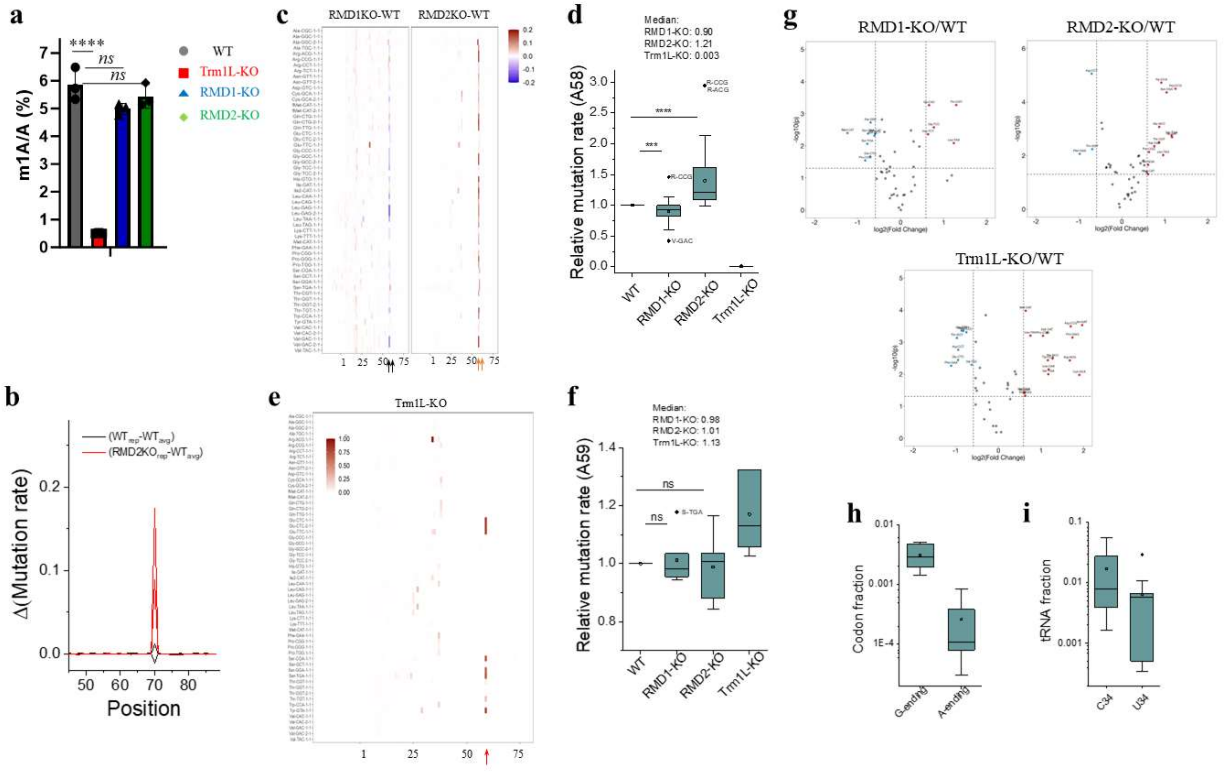

Figure 3.2: Erasers and writer knockouts affect tRNA m^{1}A modification and tRNA abundance.

ns: not significant, ***: $p<10^{-3}$, ****: $p<10^{-4}$.

(a) m^{1}A levels in total RNA isolated from WT and 3 KO strains grown in rich medium by QQQ LC-MS. Individual data points are shown.

(b) Δmutation rate (RMD2KO-WT*) of a tRNA exemplar, tRNALeu(GAG)-1. WT* = average mutation rate of WT replicates.

(c) MSR-seq results of WT and RMD1 and RMD2-KO strains grown in rich medium. Heatmap shows differential mutation rate between KO and WT strain (KO-WT), mean of n=3 biological replicates for each strain. Heatmap scale is from -0.2 to +0.2. Blue: decreased mutation rate in KO compared to WT, red: increased mutation rate. Arrows indicate the position 58 and 59.

(d) Mutation rate ratios for tRNAs with methylation at position 58 for KO strains relative to WT. Values correspond to the positional mutation rate indicated by the appropriate arrows in (c). Outliers are denoted by their amino acid and anticodon.

(e) MSR-seq positional mutation rate for Trm1L-KO tRNAs. Heatmap scale is from 0-1. Arrows indicate the position 59.

(f) Mutation rate ratios for the seven tRNAs with methylation at position 59 for KO strains relative to WT. Values correspond to the positional mutation rate indicated by the appropriate arrows in (d). Outliers are denoted by their amino acid and anticodon.

(g) Volcano plot of tRNA abundance differences of WT versus each KO strain, n=3 biological replicates for each strain. Vertical dotted lines show 1.5-fold changes from WT samples, horizontal dotted lines show p = 0.05. Red points are tRNA in KO strains

with significantly increased abundance, blue point decreased abundance compared to WT.

(h) Synonymous codons of amino acids Ala, Gly, Pro, Arg, Val, Ser, Leu, Lys, Glu, Gln in the *S. venezuelae* genome that end with G or A in the third codon position.

(i) tRNA abundance grouped for those with C34 wobble nucleotide versus their corresponding isoacceptors with U34 wobble nucleotide. These tRNAs read the codons presented in (f).

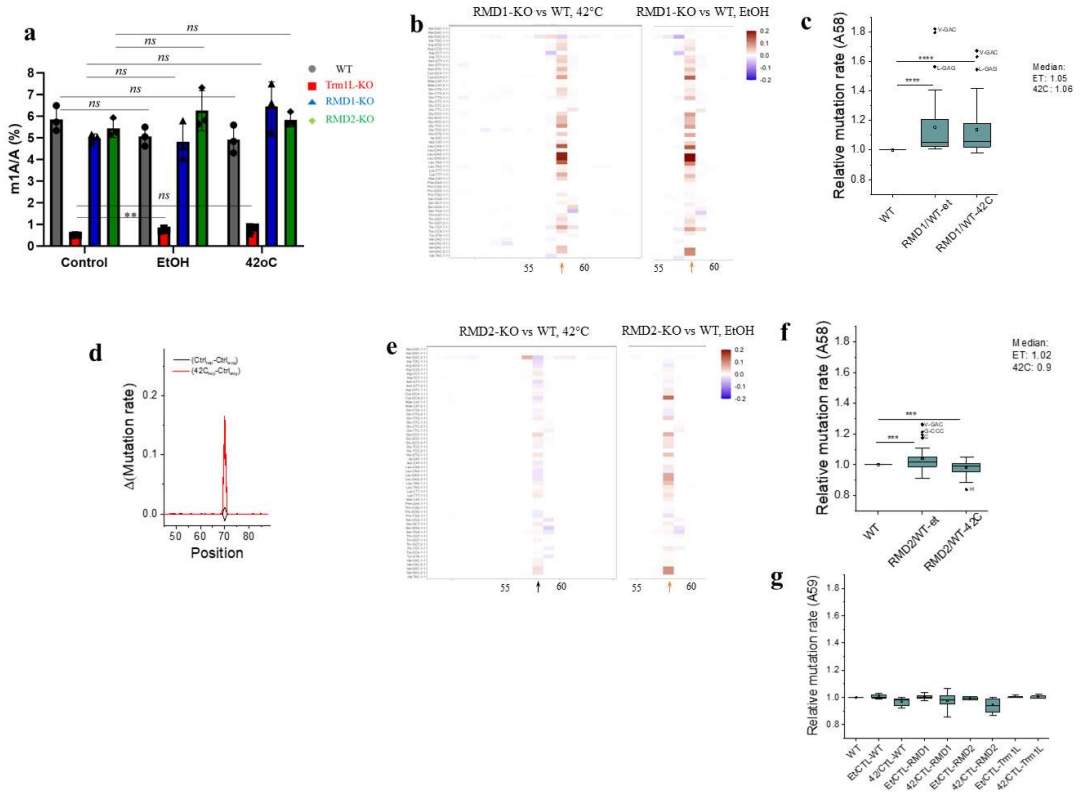

Figure 3.3: Differences in tRNA m^{1}A modification response to stress.

ns: not significant, **: p<0.01, ***: p<10^{-3}, ****: p<10^{-4}.

(a) m^{1}A levels in WT unstressed and stressed by QQQ LC-MS. Individual data points are shown.

(b) MSR-seq results of RMD1 versus WT stains under stress: mutation rate differences = [RMD1(stress)-(control)]-[WT(stress)-(control)]. Heatmap scale is from -0.2 to +0.2, mean of n=3 biological replicates for each condition. Blue: decreased mutation rate in stress compared to unstressed, red: increased mutation rate. Arrows indicate the position 58.

(c) Mutation rate ratios for tRNAs with methylation at position 58 for RMD1 KO strain in each stress condition normalized relative to that of the WT strain. Outliers are denoted by their amino acid and anticodon.

(d) Δmutation rate of WT (Stress-unst*) of a tRNA exemplar, tRNALeu(GAG)-1. Unst* = average mutation rate of no stress replicates.

(e) MSR-seq results of RMD2 versus WT stains under stress: mutation rate differences = [RMD2(stress)-(control)]-[WT(stress)-(control)]. Other descriptions same as (b).

(f) Mutation rate ratios for tRNAs with methylation at position 58 for RMD2 KO strain in each stress condition normalized relative to that of the WT strain. Outliers are denoted by their amino acid and anticodon.

(g) Mutation rate ratios for the 7 tRNAs with methylation at position 59 for WT and KO strains, under stress conditions. Within each strain, stress conditions are normalized relative.

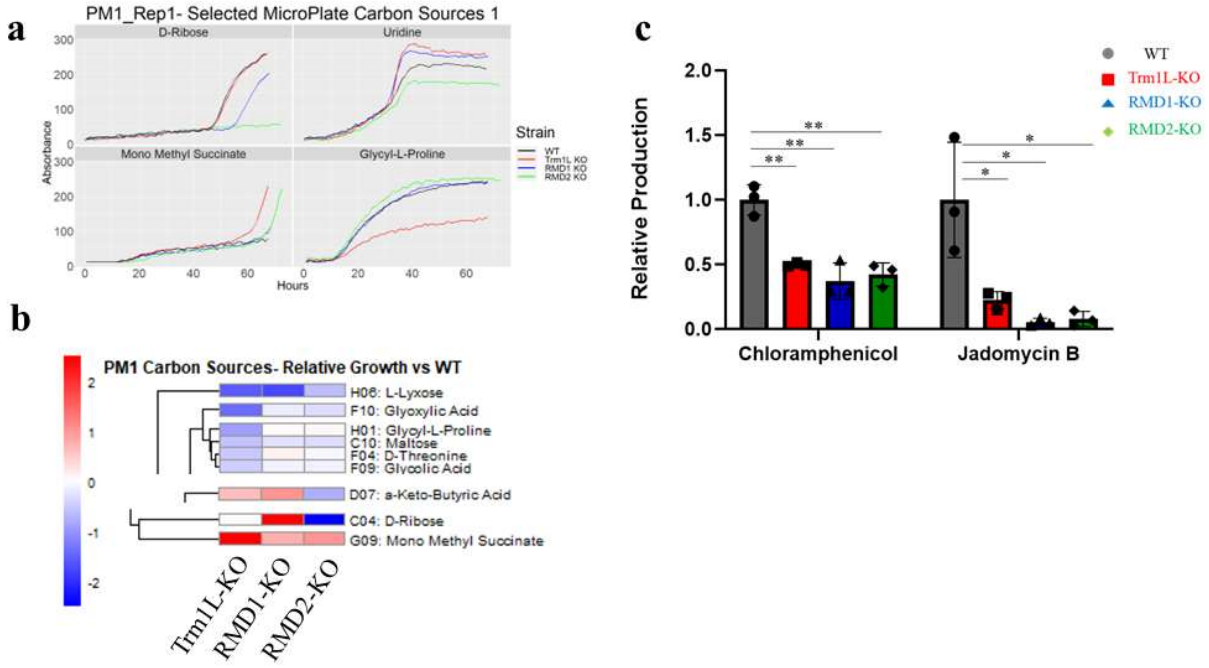

Figure 3.4: Phenotyping of *Streptomyces venezuelae* knockout strains.

(a) Representative growth curves from each strain obtained from Biolog PM1 (Carbon Source) plate runs. D-Ribose, Uridine, Mono methyl succinate and Glycyl-L-proline represent the sole carbon source in their respective wells, out of a total of 96 different carbon sources.

(b) Relative growth of each KO strain versus WT. Red indicates increases while blue indicates decreases.

(c) Mass spectrometry analysis of antibiotic extracts from *Streptomyces venezuelae* cultures of WT and KO strains. n=3 biological replicates. *: $p<0.05$, **: $p<0.01$.

3.4 Discussion

Our study provides a foray into understanding the role of tRNA modification eraser enzymes in modulating the methylation levels of m^1A in bacteria. We identify the key enzymes that enable m^1A to be a dynamic, reversible tRNA modification, including the first prokaryotic tRNA erasers that act on endogenous m^1A modifications. We show that knocking out of either eraser or the major tRNA m^1A writer significantly alters the abundance of many non-overlapping tRNAs, suggesting that the individual tRNA abundance in *S. venezuelae* is fine-tuned through the action of these enzymes. We cannot rule out, however, that these erasers may control tRNA abundance in ways other than acting on m^1A modification. How widespread this result is among prokaryotes is unclear. Streptomyces genomes are among the largest in bacteria, encode the largest number of genes, and have very high GC contents. Its codons are highly biased toward C/G-ending over A/U-ending, so that the tRNA population's need to adapt to these unique genomic features may require both writers and erasers in unique ways.

One unexpected finding was the set of 7 tRNAs with m^1A59 mutations that remained, despite the disruption of Trm1L, the tRNA m^1A58 writer. As these tRNAs share an CCN motif in the β-region of the D-loop not found in m^1A58-modified tRNAs, this motif likely acts as a recognition element for the unidentified methyltransferase, given that many tRNAs in *S. venezuelae* have adjacent A58 and A59. Currently only a handful of publications cover the tRNA modification landscape of Actinobacteria[69, 90, 91], and a single tRNA sequencing paper from us has reported m^1A59 modification in the mouse cecum microbiome[63]. The tRNA m^1A sites across all three kingdoms include the location in both D- and T-loops at 9, 14, 22, and 58, and the only other 58-adjacent site at position 57

found in archaea. A previous report has documented the *in vitro* activity of an archaeal methyltransferase on both positions 58 and 57 of bacterial tRNAs[92]. Here, m^1A57 serves as an obligatory intermediate for the installation of N1-methylinosine (m^1I). Since *S. venezuelae* does not have m^1I and its Trm1L cannot modify both m^1A58 and m^1A59, this archaeal methyltransferase could prove useful as starting point in identifying the m^1A59 methyltransferase gene in the Streptomyces genome. So far, cross-referencing homology searches of the archaeal protein in the *Streptomyces venezuelae* ATCC10712 genome with that of the first documented TrmI gene in *Thermus thermophilus* led to low level matching of an uncharacterized, putative methyltransferase (Genbank accession CCA60552). It remains to be determined whether this gene indeed codes for the tRNA m^1A59 methyltransferase.

Finally, even though some differences were evident, we did not find very large differences in phenotypes for the eraser and writer KO strains among the conditions we tested. *Streptomyces* is a common soil bacterium, and naturally lives in microbial communities, so the conditions that these erasers and writer manifest themselves more dominantly may not be easily identifiable in the laboratory monocultures. Nevertheless, our results hint a role of maintaining tRNA m^1A modification levels in facilitating bacterial adaptation to growth and to stresses.

3.5 Materials and Methods

Bacterial strains and culture conditions

All bacterial strains used in this study are listed in Table S3.1. The wild-type *Streptomyces venezuelae* (ATCC10712) and their gene deletion derivatives were grown

at 30°C using ISP2[74] (International Streptomyces Project yeast malt extract agar medium) for general growth and maintenance and in Glucose-L-isoleucine medium[72, 93] for antibiotic production. Unless stated otherwise, liquid cultures were grown in 250 ml baffled flasks containing 50 ml of medium inoculated with starter *S. venezuelae* cultures to a starting optical density at 600 nm (OD_{600}) of ~ 0.05 with shaking (225 rpm). Starter cultures were prepared starting with a streak from frozen *Streptomyces* stocks onto ISP2 agar, grown up to 2 days at 30 °C, then isolated colonies were inoculated into an overnight culture of ISP2 medium. *Escherichia coli* strains were grown at 37 °C (or different temperatures as indicated) using solid or liquid Terrific Broth (TB) medium supplemented with antibiotic where necessary.

DNA manipulation and plasmid/strain construction

The plasmids used in this study are listed in Table 1. Protein expression plasmids were designed and ordered from Genscript. DNA manipulations in expression *E. coli* vectors were carried out according to manufacturer's protocols. *Streptomyces venezuelae* mutant strains were generously provided by the Zhao lab using their developed CRISPomyces system[88].

Recombinant protein expression and purification

The RMD1 and RMD2 proteins of *S. venezuelae* were designed as N-terminally His-tagged constructs with the addition of the DesG tag to improve solubility and a TEV cleavage site (6xHis-DesG-GSS-TEV) and inserted into a pET-28a(+) plasmid backbone using Genscript services. The plasmids were then transformed into *E. coli* BL21(DE3)

(NEB) according to manufacturer's protocols. Two-liter TB media cultures were used to grow each *E. coli* expression strain supplemented with 50 µg/ml Kanamycin and 5 mM $FeSO_4$. Overnight cultures were grown up and used as starters for the expression, starting at an OD_{600} of ~0.05. Cultures were incubated at 37 °C & 225 rpm until reaching an OD600 of 0.3. Expression was induced with the addition of IPTG to a final concentration of 0.1 mM and the culture was expressed for 3 hours at 30 °C before collection by centrifugation.

For purification, each protein was isolated from a 1 liter culture pellets using affinity chromatography performed through a gravity-flow using Ni-NTA Agarose (Qiagen) resin bed. The column was washed twice using low Imidazole wash buffers containing 10 mM and 25 mM, respectively, followed by a single step elution in 250 mM Imidazole elution buffer (10 mM Tris, pH7.4, 1 M NaCl, 5% glycerol, 2 mM $CaCl_2$, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 250 mM Imidazole). The collected protein fractions were concentrated and exchanged into a storage buffer (10 mM Tris, pH7.4, 100 mM NaCl, 5% glycerol, 2 mM $CaCl_2$, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 10 mM Imidazole) using Amicon Ultra Centrifugal Filter Units and stored at -20 °C.

Demethylase assay

The activities of SV RMD proteins were assessed using m^1A demethylase assays *in vitro*. Briefly, 1 µg of total yeast tRNA (Roche) were incubated with SV RMD proteins or *E. coli* AlkB protein (purified as described by Zheng, G. et al[26]), at a substrate: protein ratio of 1:10 or 1:2, respectively. Each reaction proceeded for 2 hours, either at room temperature or 37 °C, in reaction buffer (50 mM MES pH 6.0, 2 mM L-ascorbic acid, 1

mM 2-ketoglutarate, 0.3 mM $(NH_4)_2Fe_2SO_4.6H_2O$, 0.1 M KCl). Following incubation, samples were extracted for RNA through addition of equal amount of phenol-chloroform followed by a spin down at maximum speed for 5 minutes. RNA was precipitated from the supernatant layer after transferring to a clean tube and the addition of triple the volume of 75% EtOH and 1 µl 3 M NaCl. Samples were incubated at -80 ^oC for 1 hour followed by precipitation through configuration at maximum speed for 45 minutes. Excess liquid was decanted from the sample and the remaining pellet was washed with 70% EtOH then resuspended in nuclease-free H_2O. All centrifugation steps were conducted at 4^oC.

RNA modification mass spectrometry analysis

RNA samples were digested into single nucleosides using nuclease P1 (Sigma-Aldrich) incubation at 42 ^oC for 3 hours followed by incubation with Fast AP (Invitrogen) at 37 ^oC for another 3 hours. The reaction mixture was then filtered through 0.22 µm Millex-GV polyvinylidenedifluoride filters (Millipore). Samples were injected at 5 µl volumes and analyzed in triplicate using a ZORBAX SB-Aq 4.6 x 50 mm column (Agilent) on UHPLC (Agilent) coupled to a SCIEX 6500+ Triple Quadrople Mass Spectrometer in positive electrospray ionization mode. Nucleosides were quantified based on the nucleoside to base transitions and compared with calibration curves.

Extraction of *S. venezuelae* total RNA and tRNA

Streptomyces culture strains were prepared as described by Zhu et al[94]. Briefly, *Streptomyces* strains were cultured to mid log growth phase (OD_{600} ~0.6) in ISP2 medium before being transferred to Glucose-L-isoleucine medium, setting the initial OD_{600} of 0.3.

Once the cultures reach an OD_{600} of 0.6, the cultures were split into equal volumes, then subjected to either heat shock (42^{o}C), ethanol shock (6%, v/v) or no stress for 2 hours. After 2 hours, cells were collected through centrifugation and stored at -80 ^{o}C until extraction. Total RNA was extracted from cell pellets using TRIzol (Thermo Fisher) by adapting manufacturer's protocol. Briefly, cell pellets were resuspended in TRIzol and ~100 mg 0.1 mm glass beads (Biospec) then homogenized in two 1-minute intervals. An equal amount of chloroform was added to the homogenate and the solution was mixed vigorously before centrifugation at maximum speed for 15 minutes. The clear aqueous layer was transferred to a fresh tube and 100% isopropanol was added at 1.5x volume excess. Nucleic acid was precipitated by incubation at -80 ^{o}C for 1 hour. The sample was then centrifuged at maximum speed for 30 minutes, forming a nucleic acid pellet that was washed with 75% EtOH, allowed to dry and resuspended in resuspended in nuclease-free H_2O. Total RNA concentrations were measured using a NanoDrop 2000 Spectrophotometer (Thermo Scientific). tRNA samples were isolated from total RNA samples from a urea denaturing gel, followed by extraction of RNA from the excised bands ranging from ~70 to ~100 bp. Briefly, total RNA samples were mix with 2x RNA loading buffer (9 M Urea, 100 mM EDTA, pH 8, 0.2% Bromophenol blue, 0.2% Xylene cyanol) loaded onto a pre-run 10% TBE-Urea Polyacrylamide gel (Criterion). The gel was stopped when the Bromophenol blue band reached the bottom of the gel. The gel was stained with SYBR gold (Thermo) and scanned using a Chemi-Doc imaging system (Bio-Rad). Excised gels were mixed with elution buffer (200 mM KCl, 50 mM KOAc, pH7) at estimated 2x volume and the gel was crushed up with a clean pipette tip. Samples were incubated overnight while spinning. Tubes were then centrifuged with the supernatant

was transferred to a new tube and the nucleic acid pellet was recovered and wash, per the final recovery steps as listed above.

MSR-seq library preparation

Streptomyces venezuelae total RNA MSR-seq libraries including AlkB treatment for library construction was prepared as described by Watkins et al[1]. Major steps of the protocol were followed as stated in the general protocol for MSR-seq, in sequential order: One-pot diacylation and β-elimination for tRNA charging; first barcode ligation; binding to dynabeads; dephosphorylation; reverse transcription; RNase H digestion; periodate oxidation; second ligation; PCR; TBE-PAGE gel extraction. 1 µg of total RNA were used as the starting input material. Each replicate was assigned a different first barcode with the strains being assigned as the second ligation index. Following the gel run of PCR products, lanes were cut at the expected tRNA band sizes of ~150 bp to 250 bp for extraction and recovery of the final library preparation samples.

Detection of antibiotics produced by *S. venezuelae*

Streptomyces culture strains were prepared as described above, with the change that all cultures were returned to maintenance growth conditions and cultured for 72 hours. Antibiotic samples were extracted from the cultures by mixing with acetonitrile at a 1:2 ratio, vortexed, incubated at room temperature then centrifuged to remove cell debris. The supernatant was filtered using a 0.22 µm syringe filters. Samples were analyzed using Q Exactive Orbitrap (Thermo Fisher Scientific, USA) equipped with a Vanquish HPLC system (Thermo Fisher Scientific, USA) loaded on a Thermo hypersil

GOLD aQ column (3 µm, 2.1 × 150 mm; PN 25303-152130) using the following mobile phase: solvent A (100% water, 0.1 % formic acid) and solvent B (100% acetonitrile, 0.1 % formic acid). A linear gradient from 5% to 95% buffer B over 18 min followed by an isocratic elution with 95% buffer B held 5 min was then applied at a flow rate of 0.2 mL/min. MS spectra were acquired on positive ionization modes using Q Exactive Orbitrap with the scan range *m/z* 50-750, resolution 70,000 FWHM, AGC target 3×10^6 ions capacity, Maximum IT 100 ms. The antibiotic yield was evaluated based on the standard curve using Jadomycin B and Chloramphenicol.

Biolog Phenotyping

Streptomyces venezuelae strains were prepared for use with Biolog Phenotype MicroArrays according to manufacturer recommended procedures. PM1 – 12 microplates were used for phenotype screening, with each strain assessed using three biological replicates. Briefly, strains were grown to early log phase and diluted into the corresponding inoculation fluid and 100 µl of the suspension was inoculated into the PM plate wells. All PM plates were incubated at 28°C for 72 hours in a Biolog Omni Log incubator. Cellular metabolic activity was measured every 15 minutes using the automated incubator reader. Measurements were analyzed using the Growthcurver[95] R package, summarizing the growth curve data into key metrics. The growth rates of each strain per condition were then compared to identify any major differences.

Sequencing data analysis

The data analysis followed the MSR-seq data processing pipeline[1]. Libraries were sequenced from Illumina NovaSeq 6000 platforms. The resulting paired-end reads were then demultiplexed by the identification of barcode sequences using Je demultiplex[75] with the following parameters: BPOS=BOTH BM=READ_2 LEN=6:4 FORCE=true C=false. These options are optimized for samples where the barcode sequence is likely present on read 2. Barcode sequences are as previously described18. Following demultiplexing, data was aligned using bowtie233 (version 2.3.3.1) with the following parameters: -q -p 10 --local --no-unal. These reads were aligned to a curated *Streptomyces Venezuelae* tRNA reference which was curated from the *Streptomyces venezuelae* ATCC 10712 reference from the GtRNAdb: Genomic tRNA Databse[76]. This reference contained sequences of tRNA that was curated for non-redundancy, to add 5S rRNA, to remove low scoring tRNA so that all tRNA-scan SE score > 46, removing intron sequences, and appending 3' "CCA". The Bowtie2 output sam files were converted to bam files, which were then sorted by samtools' sort function[77]. IGVtools count was used to collapse reads into 1nt windows using the following parameters: -z 5 -w 1 -e 250 —bases. The resulting IGV output wig files were reformatted using a custom Python script to obtain mutation rate and read coverage compatible with R for data visualization and analysis.

To obtain a more accurate position for mutation rates, alignment was done from the 3’ CCA end. R was then used for visualization of the data and Welch’s unequal variances t-test was used for the abundance volcano plots and the global charging box and whiskers plots.

Chapter 4 tRNA profiling analysis of *Bacteroides fragilis* under oxygen stress

Chapter 4 is a collaboration between the Pan lab and the Eren lab. Marcus Foo, A. Murat Eren and Tao Pan conceived the project. M. F. built B. fragilis MSR-seq libraries and performed initial data analysis. Luke Frietze did final MSR-seq data analysis for tRNA. Karen Lolans grew B. fragilis cultures and extracted total RNA.

4.1 Abstract

Bacteroides fragilis is common commensal bacterium in the human gut microbiome that is capable of becoming an opportunistic pathogen. A key feature in achieving this is through a robust oxidative stress response, allowing it to survive outside its typical anaerobic microenvironment. I hypothesize that one way it supports proper regulation of this response is through tRNA modification modulation through tRNA modification tunable translation of mRNA transcripts. Using high throughput sequencing, we first document changes to its tRNA profile under oxygen stress, along with a widespread plasmid found in *B. fragilis* among human gut microbiomes. We note that 2-thio modifications in the anticodon loop of several tRNAs, as installed by oxygen sensitive iron-sulphur cluster writer enzymes, fluctuate upon exposure of *B. fragilis* cultures to oxygen stress. This work documents the initial changes in tRNA modifications that may contribute to a proper oxidative stress response in *B. fragilis*.

4.2 Introduction

Bacteroides fragilis is an anaerobic, Gram-negative commensal bacterium closely associated with the mucosal layer of the human gut. Like other Bacteroides species, the presence of multiple polysaccharide utilization loci (PULs) and complex polysaccharide capsule supports its colonization and survival within the gut mucosal niche. Another prominent pathogenic factor in *B. fragilis* is its robust ability to survive in high oxygen levels, hence the reports of this bacterium escaping the gut, causing infections and forming intra-abscesses[96]. While unable to multiply in the presence of air, it remains highly aerotolerant, capable of surviving for extended periods of time in aerobic environment. This is achieved in part by induction of more than 28 proteins in a global oxidative stress response (OSR)[97]. Further transcriptomic analysis of the oxidative stress response, both to oxygen and hydrogen peroxide, implicate the upregulation of detoxification enzymes such as catalase (katB), superoxide dismutase (sod), alkyl hydroperoxide reductase (ahpCF) and thioredoxin peroxidase (tpx) to support its survival[98]. All these genes are a part of the LysR-family regulator (OxyR) operon that positively regulates their expression in response to oxidative stress[96].

While the OSR has been primarily characterized through the context of transcriptional and proteome controls and regulation, some bacteria also show regulatory mechanism for post-transcriptional regulation. One such avenue is through the use of tRNA modifications that can influence translational efficiency by altering of decoding bias. Romsang et. al.[99] showed such an example, in which *Pseudomonas aeruginosa* upregulates the catalase, KatA, and derepresses TtcA, the writer enzyme responsible for the 2-thiocytidine (s^2C) tRNA modification at position 32, increasing overall catalytic

activity by the bacterium. Disruption of the ttcA enzyme resulted in overall decreased catalase activity, despite the increase in katA transcripts. This suggests a disruption in translational control. Additionally, induction of oxidative stress leads to increases in TtcA expression. Given the immediate function of TtcA as a thio-modification writer, this result hints towards an underlying regulatory mechanism in which certain catalase enzymes require additional translational control achieved by thio-containing tRNAs. The TtcA enzyme is an example among many other RNA-modifying enzymes that contain iron-sulfur clusters. These types of proteins have been shown to be sensitive to reactive oxygen stress[100], providing a throughline with the reactive oxygen stress response.

In this work, we document the overall tRNA modification profile changes of *Bacteroides fragilis* under oxygen stress using high throughput tRNA sequencing technology. Additionally, we examine the potential impact of the presence of a cryptic, widespread *B. fragilis* plasmid, pBI143[101], on tRNA 2-thio modification in response to oxidative stress.

4.3 Results

Bacteroides fragilis tRNA thio-modifications respond to oxidative stress.

Given the robustness of *B. fragilis* in withstanding oxidative stress responses in relation to becoming an opportunistic pathogen, we were interested in examining the potential changes to its tRNA profile upon encountering such a scenario. Previous preliminary experiments showed some changes to the s^2C32 tRNA modification upon exposure to oxygen from its typical anaerobic state, so we hypothesize that this and other thio-related modifications could act as a responsive biomarker, in conjunction to supporting proper stress regulatory activities.

Using two widely used laboratory strains of *B. fragilis* (638R and 9343), we cultured each strain in a control and oxygen stress condition to compare potential changes in their tRNA profiles. In addition to the two strains used, an additional strain of each type containing the pBI143 was also included in the sample set. The pBI143 plasmid is a minimalistic plasmid of ~2 kb containing only two annotated genes: a mobilization protein (mobA) and a replication protein (repA). It is derived from *B. fragilis* in human gut microbiomes with the plasmid found to be widespread in such niches, particularly for developed nations[101]. Further investigation of this plasmid by Fogarty et al.[101], with respect to its bacterial host, showed a small but negative effect on *B. fragilis* surviving in mouse gut microbiomes. The plasmid itself, however, responds to oxidative stress by increasing its copy number. We were interested to examine the potential role of this plasmid on impacting the tRNA responses to oxidative stress and included the additional plasmid bearing strains in our analysis (**Figure 4.1a**).

We collected the total RNA samples from these cultures to build and analyze the sequencing libraries using the MSR-seq technology, allowing us to capture the major tRNA properties simultaneously[1]: modifications, charging states and abundances. Of these, changes in their modifications have been associated with adaptations to stress conditions, so we first examined the overall profiles for these samples (**Figure 4.1b, 4.3a**). The MSR-seq method works well for detecting and quantifying the difference of 2-thio modifications as revealed in a previous publication[102]. Mutation signatures in our RNA sequencing data typically correspond to chemical modifications that affect the reverse transcription step in library building, resulting in deviations from the expected tRNA sequence which can then be quantified. Across all the *B. fragilis* 9343 tRNA profiles

examined, we observe high mutation signatures centered around between positions 25 – 45, corresponding to modifications that are typically found in the anticodon stem loop region (**Figure 4.1b**). Modifications found in this region are typically associated with regulatory roles due to their direct reading of the mRNA codons during translation. The *B. fragilis* 638R samples show a similar tRNA profile to that of 9343 (**Figure 4.3a**).

Upon exposure to oxygen stress, the *B. fragilis* 9343 (-plasmid) shows an overall decrease in modifications across the tRNA species (**Figure 4.1c**). We note a similar change in modification profile when looking at the impact from the presence of pBI143. As for the changes between *B. fragilis* 9343 (+plasmid) under control and oxygen stress conditions, changes in the modification profile are more muted as comparing changes from the *B. fragilis* 9343 (-plasmid). Similar changes in the tRNA modification profile suggests that the presence of the plasmid imposes the same type of response as to oxygen stress. The influence of this tRNA profile change is a point of interest in understanding how the presence of the plasmid affects any stress response mechanisms or regulations. While *B. fragilis* 638R samples showed similar modification profiles overall (**Figure 4.3a**), changes to these modifications' levels appeared mostly consistent across all conditions, when compared to *B. fragilis* 638R (- plasmid) in control growth (**Figure 4.3b**).

As thio-related tRNA modifications are commonly associated with installation enzymes with iron-sulphur clusters, making them sensitive to oxidative conditions, we looked closer at changes in the 2-thio modifications across the different conditions. These consisted of 7 different modifications found on specific *B. fragilis* tRNAs, all of which were either s^2C or s^2U found at position 32 or 34. The most striking result is the increased levels

for *B. fragilis* 9343 (+plasmid) under oxygen stress, as compared to its respective strain under control conditions (**Figure 4.1d**). The presence of the plasmid alone under regular growth conditions cause a slight decrease thio modifications for both strain types, similar to how the presence of oxygen causes the same direction of change in the absence of the plasmid. Given the previously implicated role of s^2C32 modification in modulating the Arg-codon decoding by $tRNA^{Arg}$, further investigation into codon bias and decoding for *B. fragilis* may highlight the role for these modifications.

Finally, MSR-seq also is capable of capturing the changes in tRNA abundances and charging states for these samples. For both *B. fragilis* 638R (**Figure 4.3c-f**) and 9343 (**Figure 4.2a-d**), we do not observe any major differences in abundances when comparing different conditions to one another. No notable differences were also observed for the overall charging states in *B. fragilis* 638R (**Figure 4.3g, h**), but we do see a notable decrease in tRNA charging for *B. fragilis* 9343 (-plasmid) under oxygen stress (**Figure 4.2e, f**).

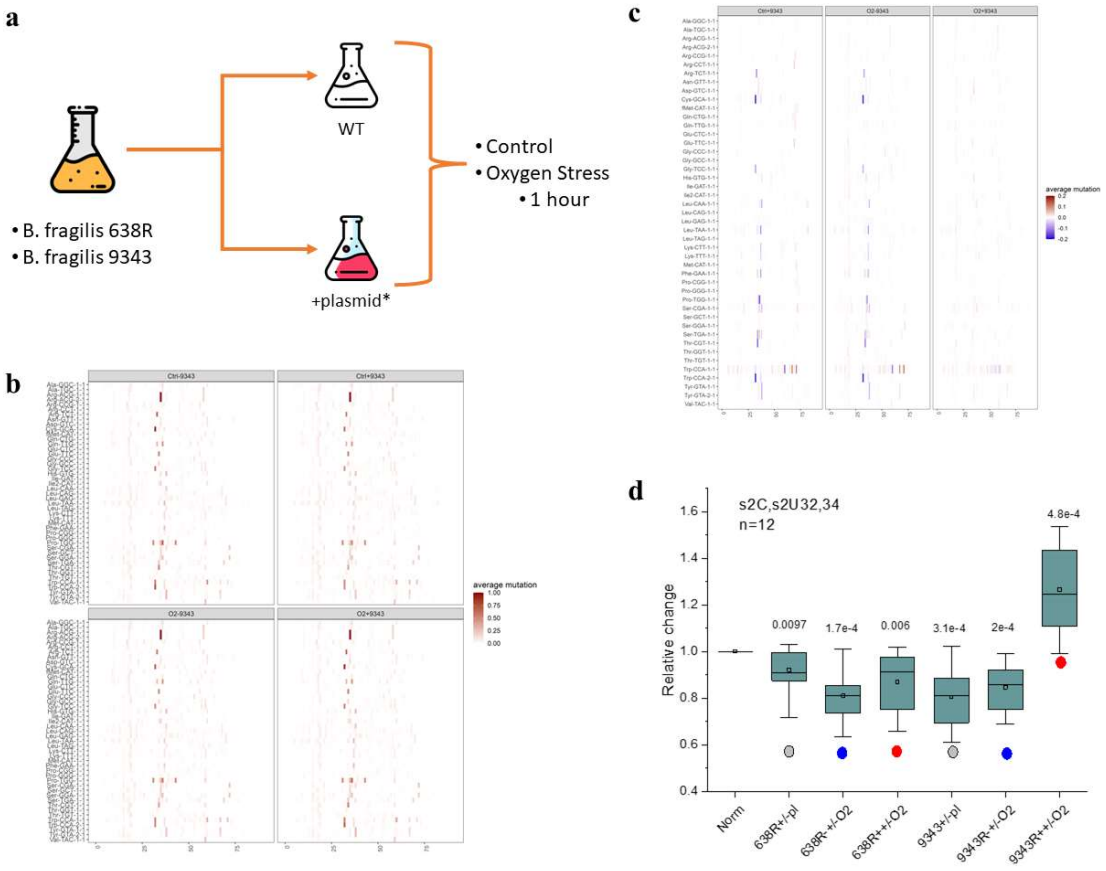

Figure 4.1: *Bacteroides fragilis* 9343 modification profile reveals plasmid dependent thio-modification changes under oxygen stress.

(a) Experimental setup of *Bacteroides fragilis* stress cultures.
(b) MSR-seq positional mutation rate for all *B. fragilis* 9343 samples. Heatmap scale is from 0-1. Samples are as follows, clockwise starting in top left quadrant: Control (-plasmid), Control (+plasmid), Oxygen Stress (+plasmid) & Oxygen Stress (-plasmid)
(c) MSR-seq results of Control *B. fragilis* 9343 (-plasmid) versus all other B. fragilis 9343 samples. Heatmaps show differential mutation rate: Ctrl-9343 (-plasmid) – Treatment, mean of n=2 biological replicates for each sample. Heatmap scale is from -0.2 to +0.2. Blue: decreased mutation rate in Ctrl compared to Treatment, red: increased mutation rate.
(d) Relative change in modification rate ratios for the 12 thio-containing tRNA modifications relative to their respective sample controls for both *B. fragilis* strains. White: +/-plasmid, Control, blue: +/-Oxygen, -plasmid, red: +/-Oxygen, +plasmid.

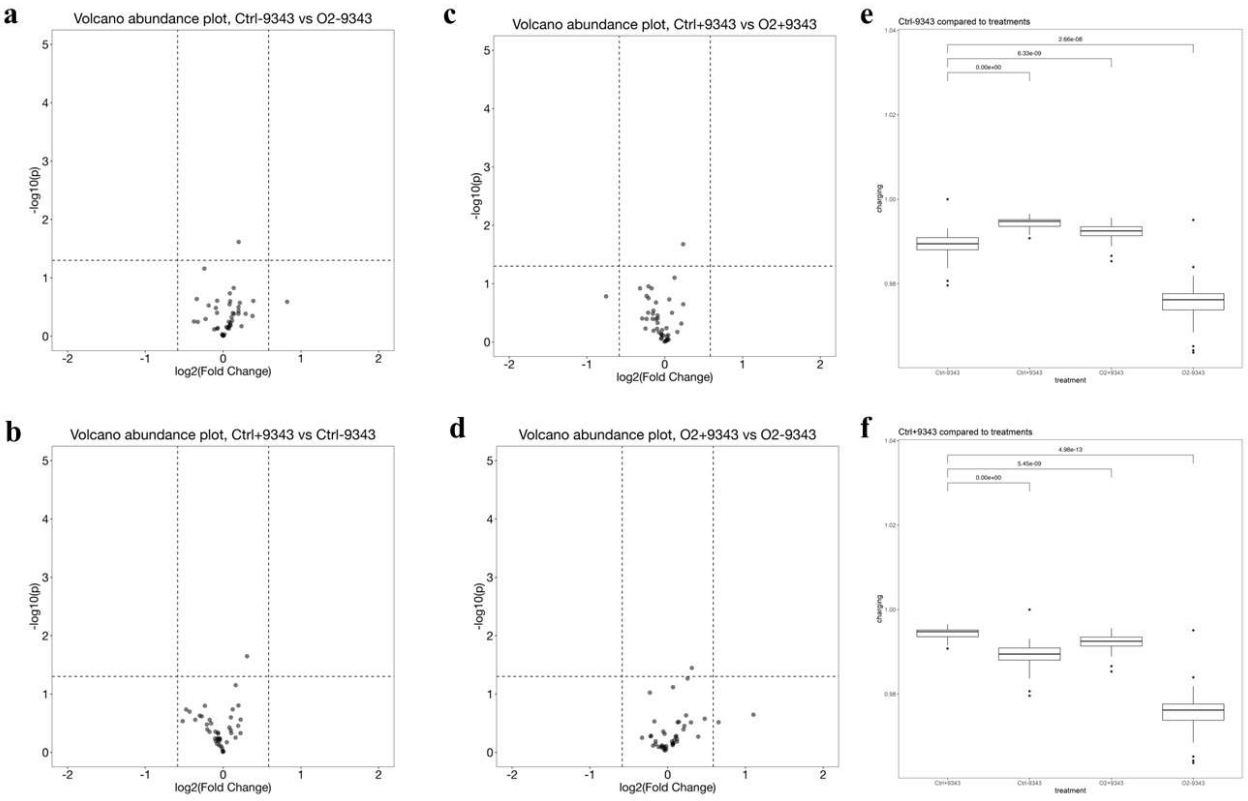

Figure 4.2: *Bacteroides fragilis* 9343 tRNA abundance and charging profile comparisons.

(a) tRNA abundance changes of Control (-plasmid) vs Oxygen (-plasmid)
(b) tRNA abundance changes of Control (-plasmid) vs Control (+plasmid)
(c) tRNA abundance changes of Control (+plasmid) vs Oxygen (+plasmid)
(d) tRNA abundance changes of Oxygen (-plasmid) vs Oxygen (+plasmid)
(e) tRNA charging changes of Control (-plasmid) compared to all other treatments
(f) tRNA charging changes of Control (+plasmid) compared to all other treatments

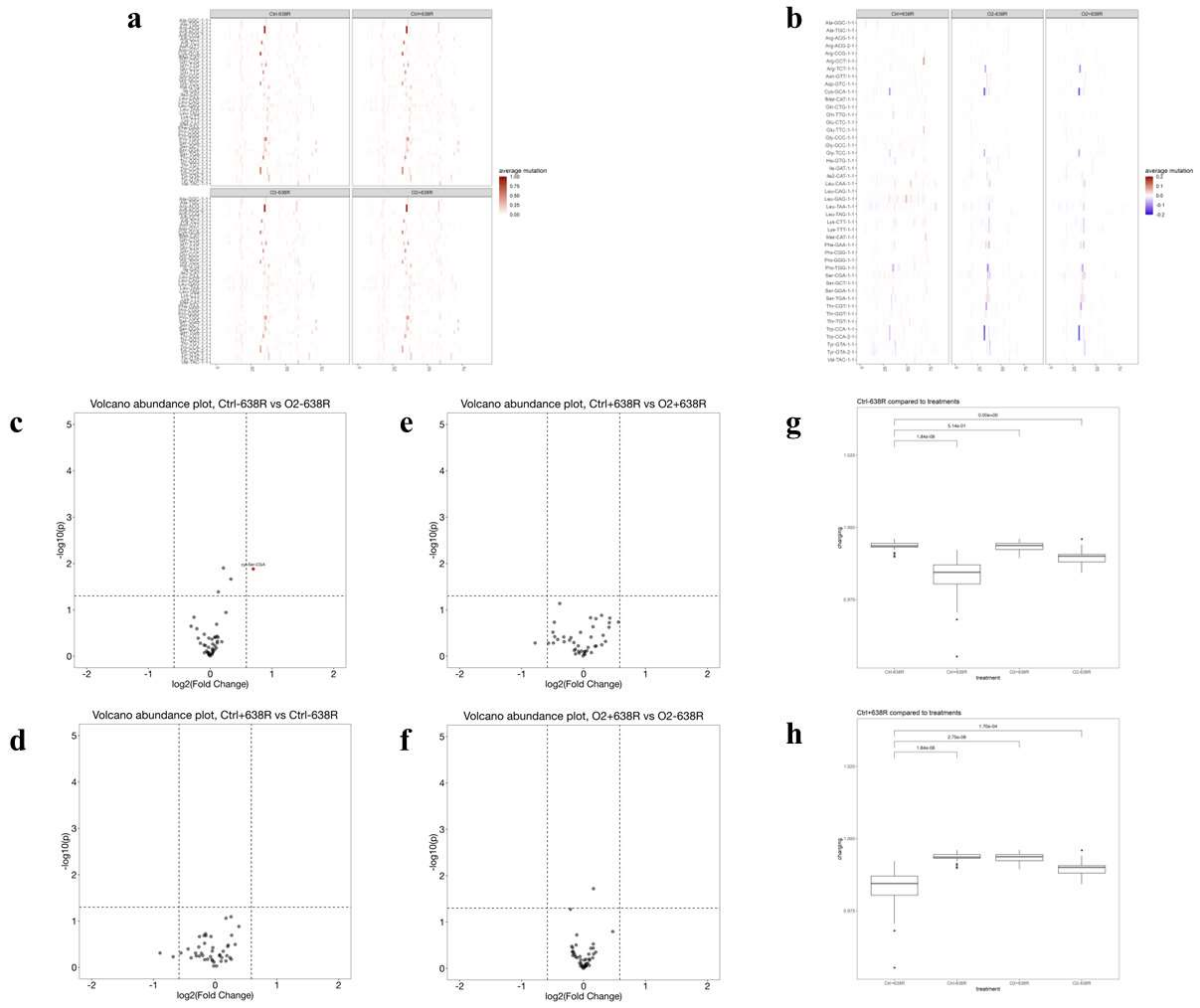

Figure 4.3: *Bacteroides fragilis* 638R tRNA modifications, abundance and charging profile comparisons.

(a) MSR-seq positional mutation rate for all *B. fragilis* 638R samples. Heatmap scale is from 0-1. Samples are as follows, clockwise starting in top left quadrant: Control (-plasmid), Control (+plasmid), Oxygen Stress (+plasmid) & Oxygen Stress (-plasmid).
(b) MSR-seq results of Control *B. fragilis* 638R (-plasmid) versus all other *B. fragilis* 638R samples. Heatmaps show differential mutation rate: Ctrl-638R (-plasmid) – Treatment, mean of n=2 biological replicates for each sample. Heatmap scale is from -0.2 to +0.2. Blue: decreased mutation rate in Ctrl compared to Treatment, red: increased mutation rate.
(c) tRNA abundance changes of Control (-plasmid) vs Oxygen (-plasmid).
(d) tRNA abundance changes of Control (-plasmid) vs Control (+plasmid).
(e) tRNA abundance changes of Control (+plasmid) vs Oxygen (+plasmid).
(f) tRNA abundance changes of Oxygen (-plasmid) vs Oxygen (+plasmid).
(g) tRNA charging changes of Control (-plasmid) compared to all other treatments. tRNA charging changes of Control (+plasmid) compared to all other treatments.

4.4 Discussion

This work provides an opening look into the tRNA profile changes of *B. fragilis* under oxidative stress, achieved through the use of our high throughput tRNA sequencing technology. As expected of tRNA populations, we see a highly decorated modification profile in *B. fragilis* with thio-modifications being sensitive to the introduction of pBI143 or induction of oxygen stress. This brings to focus the concept of tRNA modification tunable mRNA transcripts, where translation regulation of these select group can be affected through tRNA modifications. Further examination into the associated writers, beyond that of just TtcA, the s^2C32 writer, will provide better understanding of how the modification installation can be regulated. Analysis of the codon usage and bias for proteins related to mounting a proper oxidative stress response could also provide insight into the role of these modifications, given the changes observed.

4.5 Materials and Methods

Bacteroides fragilis oxygen stress strains and growth conditions

All bacterial growth experiments were performed in a Coy Anaerobic Chamber with a chamber atmosphere of H_2(7.5%), CO_2(10%), N_2(82.5%) and grown in Brain-Heart Infusion broth [BD, Sparks, MD] supplemented with yeast extract, cysteine, hemin and Vitamin K (BHIS media). An overnight starter culture of each strain was grown from a freezer stock in BHIS at 37°C. Starter cultures were diluted 1:10 in fresh BHIS and outgrown to mid-log phase (A600 nm of 0.5–0.65) where the culture was split into two and treated as follows: (1) a 5-mL culture continued anaerobic incubation for 1-hour at 37°C as the “no stress treatment control”; (2) a 5-mL culture was exposed to oxygen

stress by incubated shaking in atmospheric O2 at 250 r.p.m. for 1-hour at 37°C. The experiment was conducted twice (two different day using two different starter cultures) to generate biological replicates.

Oxygen stress RNA extraction.

After 1-hour exposure timeframe, 3-mL of the culture volume were used for total RNA extraction via Qiazol lysis following the manufacturer's protocol. Sample clean-up was performed with the Oligo Clean & Concentrator kit [Zymo Research], eluted into RNase-free water and RNA quantified by Qubit (Invitrogen). An equal volume of an acid-buffered elution buffer (10mM NaOAc, 1mM EDTA, pH 4.8) was then added to maintain charging of tRNAs.

MSR-seq library preparation

Total RNA MSR-seq libraries including AlkB treatment for library construction was prepared as described by Watkins et al[1]. Major steps of the protocol were followed as stated in the general protocol for MSR-seq, in sequential order: One-pot diacylation and β-elimination for tRNA charging; first barcode ligation; binding to dynabeads; dephosphorylation; reverse transcription; RNase H digestion; periodate oxidation; second ligation; PCR; TBE-PAGE gel extraction. 1 µg of total RNA were used as the starting input material. Each replicate was assigned a different first barcode with the strains being assigned as the second ligation index. Following the gel run of PCR products, lanes were cut at the expected tRNA band sizes of ~150 bp to 250 bp for extraction and recovery of the final library preparation samples.

Sequencing data analysis

The data analysis followed the MSR-seq data processing pipeline[1]. Libraries were sequenced from Illumina NovaSeq 6000 platforms. The resulting paired-end reads were then demultiplexed by the identification of barcode sequences using Je demultiplex[75] with the following parameters: BPOS=BOTH BM=READ_2 LEN=6:4 FORCE=true C=false. These options are optimized for samples where the barcode sequence is likely present on read 2. Barcode sequences are as previously described18. Following demultiplexing, data was aligned using bowtie233 (version 2.3.3.1) with the following parameters: -q -p 10 --local --no-unal. These reads were aligned to a curated *Bacteroides fragilis* tRNA reference which was curated from the *Bacteroides fragilis* reference from the GtRNAdb: Genomic tRNA Databse[76]. This reference contained sequences of tRNA that was curated for non-redundancy, to add 5S rRNA, to remove low scoring tRNA so that all tRNA-scan SE score > 46, removing intron sequences, and appending 3' "CCA". The Bowtie2 output sam files were converted to bam files, which were then sorted by samtools' sort function[77]. IGVtools count was used to collapse reads into 1nt windows using the following parameters: -z 5 -w 1 -e 250 —bases. The resulting IGV output wig files were reformatted using a custom Python script to obtain mutation rate and read coverage compatible with R for data visualization and analysis.

To obtain a more accurate position for mutation rates, alignment was done from the 3’ CCA end. R was then used for visualization of the data and Welch’s unequal variances t-test was used for the abundance volcano plots and the global charging box and whiskers plots.

Chapter 5. *De novo* discovery of regulatory small non-coding RNA in *Bacteroides fragilis*

Chapter 5 is a collaboration between the Pan lab and the Eren lab. Marcus Foo, A. Murat Eren and Tao Pan conceived the project. M. F. built B. fragilis MSR-seq libraries and performed initial data analysis. Sihao Huang analyzed MSR-seq data for sRNA. Karen Lolans grew B. fragilis cultures and extracted total RNA. The E. coli sRNA data was from the published data in a previous publication[102] and analyzed by Chris Katanski.

5.1 Abstract

Small ncRNA discovery using biochemical approaches typically has limitations related to the known protein binding partners of these transcripts that are needed for experimental enrichment and downstream analysis. Sequencing based approaches cast a wider net for capturing ncRNAs of interest, but also impose selection criteria based on previously characterized sRNA species. Here we introduce MSR-seq, a high throughput multiplex sequencing technology, as a generalizable approach for ncRNA discovery and characterization. We first benchmark this method using *E. coli* stress samples and document associated changes in specific regulatory sRNAs upon induction of stress. We then apply this method to *Bacteroides fragilis* stress cultures to discover new sRNA species with and without oxygen stress. This method allows for parallel investigation into ncRNA profiling, providing a better understanding of how different RNA species may underlie cellular regulation and activity.

5.2 Introduction

More recent biochemical approaches to regulatory sRNA discovery and characterization usually hinges on the mechanistic property that many sRNAs base pair with their target mRNAs[103]. In addition to mRNA binding, many sRNAs also interact with specific proteins, called chaperones[21, 104], to facilitate their regulatory functions. For example, in *E. coli*, the sRNA chaperone Hfq, is a prominent protein cofactor that bind groups of sRNAs, allowing for co-purification experiments followed by high-throughput analysis of bound RNAs[105, 106] as well as cross-linking of RNAs associated with Hfq[107, 108]. Hfq has been characterized to stabilize sRNAs by facilitating annealing to the target sequence, typically within the 5' UTR region and the start codon of an mRNA, overlapping the Shine-Dalgarno sequence to inhibit translation initiation[103]. The enrichment process of using Hfq binding is beneficial in the identification of sRNAs of low abundance while allowing for easier identification of sRNAs processed from mRNAs. This approach has also been applied to other types of RNA-binding proteins, such as ProQ[109], another RNA chaperone that binds sRNAs, and CsrA[110], a posttranscriptional regulator that tunes the cellular availability of sRNAs to interact with mRNAs.

However, in Bacteroidetes species, without the obvious presence of a chaperone homolog, as well as the lack of classical Shine-Dalgarno sequence present in Proteobacteria[22], global computational search approaches prove to be a more valuable alternative. Initial approaches from this angle relied on the particular assumptions on sRNA properties[103], such as the expression from dedicated promoters and from intergenic regions, of predicted RNA sizes (~100 to 200 nt), containing the presence of a Rho-independent terminator, and the lack of open reading frames. These predicted sRNAs

could then be followed up by validation using Northern analysis for *in vivo* confirmation. As an example, Ryan et al.[23] utilized differential RNA-seq[111] to refine the transcriptome annotation of *Bacteroides thetaiotamicron* leading to the identification of 269 ncRNA elements, and most prominently, the characterization of GibS as a new regulatory sRNA that activates an operon related to one of the bacterium's polysaccharide utilization loci. Another trans-encoded *B. thetaiotamicron* sRNA that has been characterized includes RteR, a 90 nt transcript that promotes discoordinate expression of the tra operon that is involved with the proper assembly of the mating apparatus for the transfer of CTnDOT[16, 112, 113].

Here, we apply MSR-seq[1], a high throughput sequencing technology, as an approach to *de novo* sRNA discovery. While this workflow is agnostic to the organism from which input RNA sample is derived from, we utilize *Bacteroides fragilis* as an organism of interest for our study due to its relevance in the human gut microbiome as prevalent commensal bacterium and opportunistic pathogen. Given the few identified regulatory sRNAs in Bacteroides, we first benchmark this approach using *E. coli* samples under stress conditions, identifying changes in the relevant, known sRNAs. Sequencing analysis of the Bacteroides samples identifies previously 9 groups of characterized sRNAs and other ncRNA, along with other unannotated ncRNAs of interest.

5.3 Results

MSR-seq captures *E. coli* sRNA changes under stress responses.

The MSR-seq method has been utilized in multiple applications[1, 102, 114] with regards to tRNA biology investigations, given its high throughput approach and modularity in dealing with RNA modifications that may interfere with sequencing read through quality. As compared to other tRNA focused sequencing technologies, MSR-seq is not limited to just tRNA specific input, expanding its potential use to other types of RNA, particularly for those in the size range between 30 – 300 nt. We sought to first benchmark its versatility for sRNA identification by first preparing libraries using *E. coli* total RNA under laboratory growth conditions. As expected, a majority of reads aligned to mature tRNA (92%), while the remaining reads consisted of rRNA, mRNA, and ncRNA (**Figure 5.1**). This proportion of reads approximately reflects the molar ratios of cellular RNA transcripts in each category where tRNA makes up 80 – 90% by moles. Under regular growth conditions, ncRNA reads consisted of a few abundant sRNA species that have been well characterized, such as ffs (SRP RNA), ssrS (6S RNA), and rnpB (RNase P RNA) (**Figure 5.1a**). Application of demethylase treatment during library preparation remains a staple when utilizing MSR-seq, particularly for tRNA specific investigations. In this case, the demethylase step did not appear to affect transcript abundances in any significant way, as compared to the untreated samples prepared in parallel (**Figure 5.1b**).

Among the non-tRNA ncRNAs present in the sample, approximately 50 types were identified that varied by ~2000-fold in expression levels (**Figure 5.1c**). In the absence of stress, these were dominated by several conserved bacterial RNA species such as SRP RNA, tmRNA (ssrA), and RNase P RNA, while the remaining majority were expressed at much lower levels, consistent with the regulatory expectations of many of these sRNAs.

Given the ability of MSR-seq to detect these ncRNAs of varying low abundances, we reasoned that this sequencing approach could quantify changes of these ncRNA populations upon induction of stress. We subjected *E. coli* to three acute stress conditions: H_2O_2 corresponding to oxidative stress, 2,2'-dipyridyl (DIP) causing iron starvation[115], and α-methyl glucoside-6-phosphate (aMG) to induce glucose starvation[116]. One major bacterial response to stress is the upregulation of specific non-coding RNAs[117, 118]. For each stress tested, we detected significant increases in the expression of the respective specific RNAs known to respond to each stress: ~75-fold increase in oxyS for oxidative stress, ~10-fold increase in ryhB for iron starvation, and ~60-fold increase in sgrS for glucose starvation (**Figure 5.1c, f**). As a control, the ffs (SRP RNA) levels remained unchanged in each condition tested (**Figure 5.1d, e**).

Application of MSR-seq for de novo *Bacteroides fragilis* sRNA discovery

Given that MSR-seq could be applied for ncRNA detection and analysis, we then applied it for de novo sRNA identification and characterization. Using *Bacteroides fragilis*, we provide a generalized workflow for utilizing MSR-seq in identification and preliminary analysis of sRNA from RNA samples (**Figure 5.2a**). Several *Bacteroides* regulatory sRNAs have been previously described and characterized[23, 119] using other sequencing based approaches, providing known sRNA species as benchmarks of our method. We took a more unbiased approach in the sequencing mapping by removing all annotated rRNA, tRNA, and mRNA genes in the *B. fragilis* genome, focusing on only the remaining genome regions which can also include sRNAs that are adjacent to dedicated promoters, the presence of terminator sequences or the lack of long open reading frames[103].

Using type-strained *Bacteroides fragilis* 638R, we subjected the growth culture to anaerobic control condition and the oxygen stress condition described in Chapter 4. An additional strain containing a patient derived plasmid, pBI143, was also included in the experimental testing. This plasmid is a cryptic, widespread plasmid found in the human gut microbiomes of most developed nations[101] that only carries two annotated genes: a mobilization protein (mobA) and a replication protein (repA). Preliminary library analysis of cultures with this plasmid showed RNA transcript reads mapping to a specific region of this plasmid. Reasoning that this could be a stress related regulatory RNA transcript, we included them in the preparation of our cultures (**Figure 5.2b**).

In order to identify ncRNAs, reads that mapped to rRNA, tRNA, and mRNA were removed, followed by filtering of remaining reads that had a minimum coverage of 50 reads. Using the set of sequences obtained from each sample, we then sought to concentrate on those that were of certain length (30 – 200 nt) and consolidated them into a reference set of unique peaks based on their genomic location. In addition, we compiled additional information for the new sequences obtained in our MSR-seq data, included the previously identified sRNA or ncRNA species from *B. fragilis*[22], genomic context and RNA secondary structural prediction.

Overview analysis of the *B. fragilis* 638R samples reveal a total of 94 unique unannotated peaks, each corresponding to a transcript mapped to specific genomic locations. Using Rfam searches to identify potential homology of these transcripts with previously characterized RNA families reveal that 9 of these peaks corresponded with known *Bacteroides* species ncRNAs, including SRP, 6S RNA, RteR, GibS and RNAs with the Bacteroides-1 RNA motif (**Table 4.1**). As an exemplar, Sequencing alignment of the

6S RNA in our data to the *B. fragilis* seed sequence indicated a buildup of reads with high coverage spanning the 3' end of the transcript, followed by a drop off towards the 5' end (**Figure 5.2e**). Comparing the presence of these unique peaks under control versus oxygen stress conditions, we note that while most of the sRNAs are found in both conditions, multiples were only found in the oxygen conditions, suggesting that their presence is associated with the onset of the stress (**Figure 5.2c**). As an exemplar, one unannotated peak, BF-638R-90, shows a predicted transcript structure with hairpin features in its secondary structure, along with high sequence conservation among 6 other species of Bacteroides (**Figure 5.2d**). Given this preliminary analysis of sequencing library, we expect that further detailed inquiry into the data set will reveal additional insight into these new sRNA peaks through comparison with known Bacteroides sRNA properties, such as the genomic context and location affecting regulatory aspects of the transcript, conserved motifs and structural elements that impact mRNA or sRNA-associated protein binding, and others.

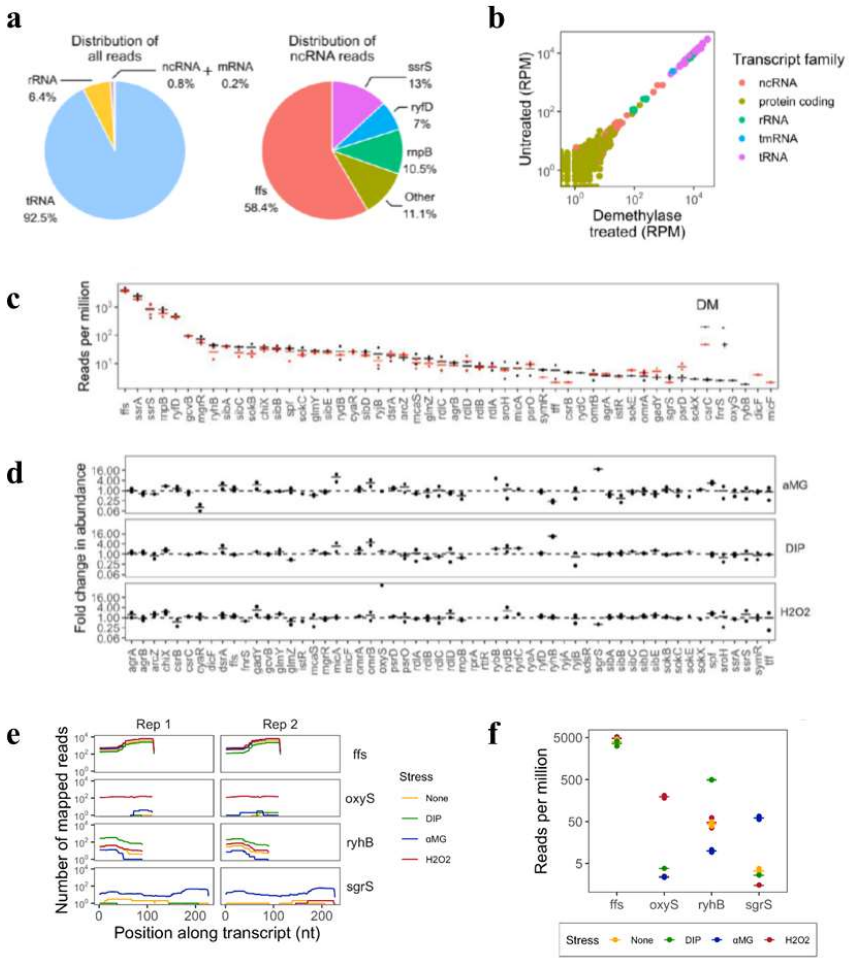

Figure 5.1: MSR-seq reveals ncRNA changes during *E. coli* stress response.

(a) Distribution of MSR-seq results were mapped to the *E. coli* genome showing a majority mapping to ncRNA. ncRNA mapped reads were examined for their distribution with most mapped to several abundant transcripts.
(b) Correlation of reads from demethylase treated samples to untreated samples.
(c) Abundance of ncRNA transcripts at rpm>1 under control conditions.
(d) Changes in ncRNA transcript abundances under different stress conditions: glucose starvation (aMG), iron starvation (DIP), & oxidative stress (H_2O_2).
(e) Coverage density of the 3 stress-responsive sncRNA and SRP (ffs) under control (none) and stress conditions.
(f) Abundance of three stress-responsive sncRNAs and SRP (ffs) under control (none) and stress conditions.

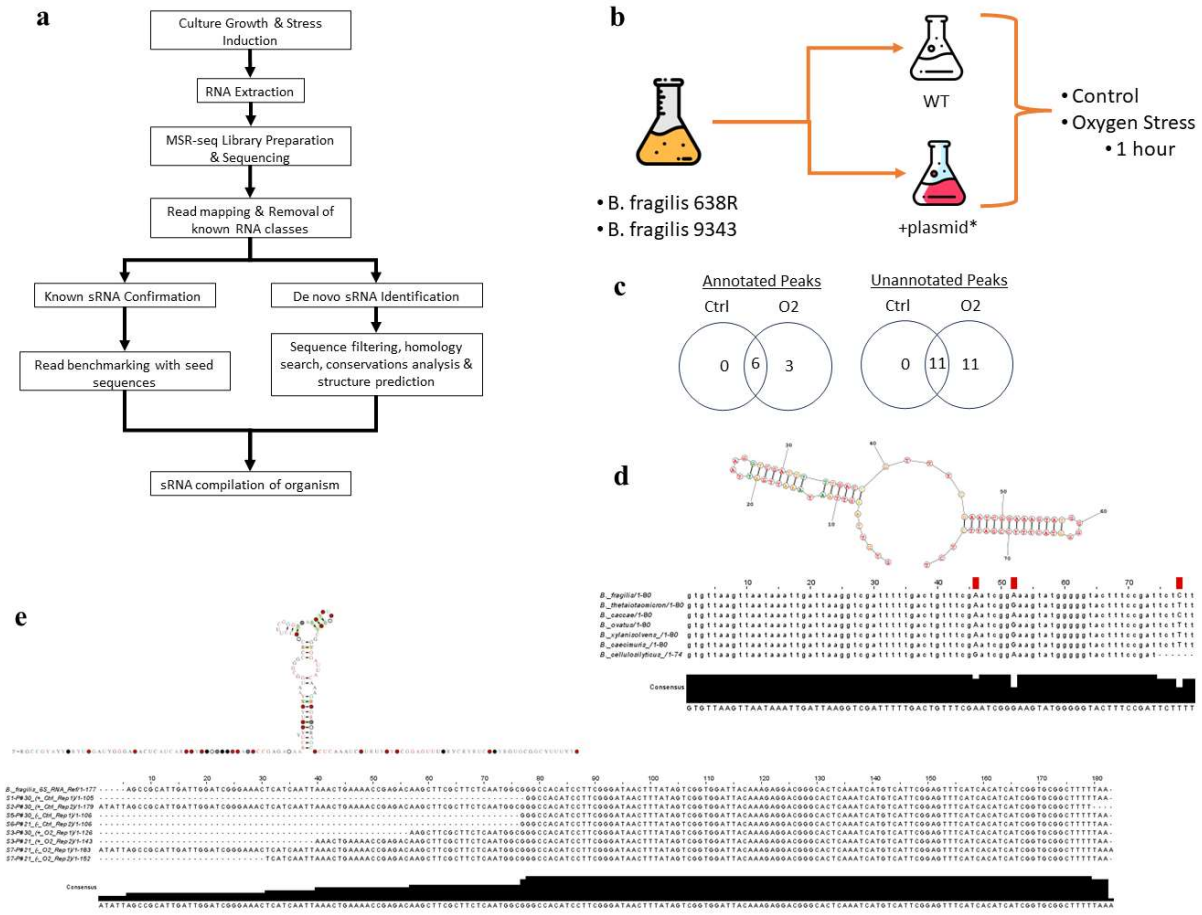

Figure 5.2: Application of MSR-seq as ncRNA discovery workflow in *Bacteroides fragilis*.

(a) Workflow for utilizing MSR-seq in identification and preliminary analysis of sRNA from RNA samples.
(b) Experimental setup of *Bacteroides fragilis* stress cultures.
(c) Venn diagram indicating *Bacteroides fragilis* 638R sRNA peaks of interest found in either Control samples, Oxygen stress samples, or both.
(d) Predicted structure and sequence conservation of unannotated unique peak BF-638R-90 compared to 6 other *Bacteroides* species. Red squares indicate single nucleotide differences between species.
(e) Representative structure and coverage of annotated 6S sRNA peaks from all samples aligned to the consensus reads.

Unique Peak #	Annotation	Genomic start	Genomic end	Length (nt)	Strand	Gene Upstream	Gene Downstream	Avg Coverage (%)	All Ctrl Samples?	All O2 Samples?
5	RteR*	225723	225804	82	-	conserved hypothetical protein (BF638R_0208)	exc (putative transposon excision protein)	-	No	Yes
13	sRNA BTnc231	1055168	1055322	155	-	hypothetical protein (BF638R_0867)	conserved hypothetical protein (BF638R_0868)	-	Yes	Yes
21	small SRP	1447179	1447285	107	Both	putative transmembrane protein (BF638R_1201)	tRNA-Thr (BF638R_tRNA0014)	70	Yes	Yes
28	sRNA BTnc231	1860280	1860326	47	+	hypothetical protein (BF638R_1593)	conserved hypothetical protein (BF638R_1594)	-	Yes	Yes
30	6S RNA	1982580	1982756	177	+	rplT (putative 50S ribosomal protein L20)	putative ferredoxin (BF638R_1702)	77	Yes	Yes
41	GibS	2754464	2754605	142	+	conserved hypothetical protein (BF638R_2346)	conserved hypothetical protein (BF638R_2347)	68	No	Yes
45	5S rRNA, intergenic	2987231	2987278	48	-			-	No	Yes
49	Bacteroides-1 RNA	3066680	3066873	194	-	putative lipoprotein (BF638R_2606)	galE (UDP-glucose 4-epimerase)	-	Yes	Yes
57	Bacteroides-1 RNA	3392931	3393094	164	+	hypothetical protein (BF638R_2901)	conserved hypothetical protein (BF638R_2902)	-	Yes	Yes

Table 5.1: 9 unique, annotated peaks were identified from the MSR-seq results.

5.4 Discussion

Applications of high throughput sequencing technology targeting specific RNA classes has allowed for in depth studies into intricate mechanisms of RNA biology. Using MSR-seq, we are not just limited to studying a single type of RNA species when using total RNA as input samples, allowing for parallel investigations into how cellular activity and regulation is impacted by different types of RNA, especially the ncRNA categories like tRNA and regulatory sRNAs.

Starting with *E. coli* cultures to benchmark the technique, we are able to identify and quantify changes in sRNA transcript abundance associated with specific stress conditions. While a majority of the sequencing reads are mapped to tRNAs as a reflection of general RNA content in cells, we are still able to utilize the remaining ncRNA reads for further analysis in stress response. We follow up this investigation with an exploratory foray into sRNA discovery using *Bacteroides fragilis*, a commensal bacterium in the human gut microbiome. Preliminary analysis reveals a large amount of unannotated RNA transcripts in the expected size range of sRNAs, each associated with specific genomic locations. Our method was able to identify 9 annotated transcripts that have been previously characterized in Bacteroides species.

Overall, the MSR-seq method shows much potential in applications for generating ncRNA profiles that underlie stress responses. While our focus has been on prokaryotic organisms, we intend this workflow to be generalizable to a variety of different organisms or sample types. As compared to other target investigations, our methodology provides a broad overview of ncRNA characterization in RNA samples, allowing for parallel analysis of multiple species of RNA types to examine if and how they might influence one another.

5.5 Materials and Methods

E. coli growth, stress, RNA extraction

E. coli MG1655 cells were grown in LB to a A_{600} of 0.4 before subjecting to the stress conditions. Mock treated cells, 25 mL, were left to grow for 10 min. Hydrogen peroxide stress was done by adding H_2O_2 to 25 mL cells to a final concentration of 0.5% for 10 min. Glucose phosphate stress was done by adding to 25 mL cells α-methyl glucoside-6-phosphate (αMG) to a final concentration of 1 mM for 10 min. Iron depletion stress was done by adding to 25 mL cells 2,2'-dipyridl (DIP) to 250 μM final concentration for 10 min. Cells were harvested by spinning 25 mL culture for 1 min at 12 000 rcf and decanting media. Cells were resuspended in 0.5 mL ice cold lysis buffer (150 mM KCl, 2 mM EDTA, 20 mM HEPES pH 7.5) then flash frozen in liquid nitrogen. RNA was extracted by a hot acid-phenol protocol. Briefly, 0.5 mL of acid-buffer phenol (pH 4.5 citrate) was added to frozen samples. Samples were incubated in a heat block with shaking at 50 °C for 30 min. The aqueous phase was extracted for another round of phenol extraction and 2 rounds of chloroform extraction before ultimately precipitating with glycoblue, 300 mM sodium acetate, and 3 volumes of ethanol. Samples were incubated for 1 hour at -80 °C, then spun at maximum speed (20k RCF) for 45 min to pellet RNA. Pellets were washed twice with 70% ethanol then resuspended in water.

Bacteroides fragilis oxygen stress strains and growth conditions

All bacterial growth experiments were performed in a Coy Anaerobic Chamber with a chamber atmosphere of H2(7.5%), CO2(10%), N2(82.5%) and grown in Brain-Heart Infusion broth [BD, Sparks, MD] supplemented with yeast extract, cysteine, hemin and Vitamin K (BHIS media). An overnight starter culture of each strain was grown from a freezer stock in BHIS at 37°C. Starter cultures were diluted 1:10 in fresh BHIS and outgrown to mid-log phase (A600 nm of 0.5–0.65) where the culture was split into two and treated as follows: (1) a 5-mL culture continued anaerobic incubation for 1-hour at 37°C as the “no stress treatment control”; (2) a 5-mL culture was exposed to oxygen stress by incubated shaking in atmospheric O2 at 250 r.p.m. for 1-hour at 37°C. The experiment was conducted twice (two different day using two different starter cultures) to generate biological replicates.

Oxygen stress RNA extraction.

After 1-hour exposure timeframe, 3-mL of the culture volume were used for total RNA extraction via Qiazol lysis following the manufacturer’s protocol. Sample clean-up was performed with the Oligo Clean & Concentrator kit [Zymo Research], eluted into RNase-free water and RNA quantified by Qubit (Invitrogen). An equal volume of an acid-buffered elution buffer (10mM NaOAc, 1mM EDTA, pH 4.8) was then added to maintain charging of tRNAs.

MSR-seq library preparation

Total RNA MSR-seq libraries including AlkB treatment for library construction was prepared as described by Watkins et al[1]. Major steps of the protocol were followed as stated in the general protocol for MSR-seq, in sequential order: One-pot diacylation and β-elimination for tRNA charging; first barcode ligation; binding to dynabeads; dephosphorylation; reverse transcription; RNase H digestion; periodate oxidation; second ligation; PCR; TBE-PAGE gel extraction. 1 µg of total RNA were used as the starting input material. Each replicate was assigned a different first barcode with the strains being assigned as the second ligation index. Following the gel run of PCR products, lanes were cut at the expected tRNA band sizes of ~150 bp to 250 bp for extraction and recovery of the final library preparation samples.

Read processing and mapping

Libraries were sequenced on Illumina Hi-Seq or NEXT-seq platform. Paired-end reads were combine with bbmerge from the JGI BBtools toolset. Reads were merged such that the sample barcode was oriented at the start of a read: for libraries constructed with the read-2 barcodes, the order of read1 and read2 were flipped for bbmerge inputs. Next, merged reads - one file for each index, were split by barcode using fastX toolkit barcode splitter. Barcode sequences are listed in Table S1. Custom python scripts (available on GitHub) were used to remove the barcode sequence (first 7 nt) and to collapse reads using the UMI, then remove the UMI (last 6 bases). Next reads were mapped using bowtie2 with the “local” parameter. Samples were mapped to either a curated list of non-redundant tRNAs from tRNA-scan SE with score >40 and CCA added where needed, or a combined genome including ensemble ORFs, and ensemble

ncRNAs. Bowtie2 output sam files were converted to bam files, then sorted using samtools. Next IGV was used to collapse reads into 1 nt window. IGV output .wig files were reformatted using custom python scripts (available GitHub). The bowtie2 output Sam files were also used with eXpress from pachter lab to sum all reads that mapped to each gene. Data was visualized with custom R scripts (GitHub). For the *B. fragilis* samples, reads mapping to mRNA, rRNA and tRNA were filtered accordingly before generating the unique list of unannotated transcripts of interest. Counts for these reads were normalized to the total number of aligned reads in each sample and sites with relative abundance greater than $3*10^7$ were retained for analysis.

Chapter 6 Conclusions and Perspectives

6.1 Applications of high throughput sequencing technology on non-coding RNA

Advancement in overall sequencing technology and library construction methods has been a boon for investigations into ncRNA biology. For example, in-depth studies into tRNA properties, including modification fraction, charging states, full-length abundances, and fragment formation, has allowed for a better understanding in how they impact cell proliferation[61], disease progression[120] and stress responses[1]. Further profiling of these tRNA and other ncRNA states allows for better understanding the complex networks of interactions between ncRNAs and other cellular components. Integration of these investigations into multi-omics approaches provides an even wider examination of regulatory elements in the cell, incorporating the entire central dogma of biology, starting from changes to DNA and gene expression, production of RNA transcripts of interest, and finally the physiological changes achieved through proper translation and protein production. On the extreme, scaling down to the single cell level likely provides a new avenue in detection and characterization of ncRNA species that might be rarer or context specific. These approaches provide value for future clinical applications by examining the larger regulatory networks that potential therapeutic targets may adversely or inadvertently affect. While our work here only scratches the surface in elucidating the ncRNA modification profile of two specific bacteria, it serves as examples in which ncRNA analysis techniques and workflows can be applied to other organisms as to better understand differences and providing comparative references to existing and future studies.

6.2 Follow up studies to m^1A tRNA modification study in *Streptomyces venezuelae*

The results in Chapters 2 and 3 provide multiple avenues for further investigation in the role of prokaryotic m^1A tRNA modifications. With respect to the maintenance of the modification, identification of the secondary writer responsible for the installation of m^1A59 remains of interest. Currently, the only other m^1A58 adjacent modification that has been described is the m^1A57 modification only found in archaea and serves as a precursor of m^1I57. Homology searches using the archaeal writer might prove as another starting point for this search, given its ability to target an adjacent site to position 58. This would be in addition to screening other unannotated *S. venezuelae* genes for sequence or structural conservation to previously characterized m^1A58 writers. Expression and purification of the putative writer will allow for in vitro methylation assays to confirm any methylation activity using unmodified tRNA substrates. These would include the set of 7 tRNAs that are expected to be modified, as well as another set of tRNAs to serve as negative controls. This would provide some insight into the speculation of the additional bases that may serve as a distinguishing identifier for the 7 tRNAs.

Further research into the mechanistic downstream effects of m^1A58 modulation is a broad next step to tackle. Changes in tRNA methylation levels in human cells have been shown to affect association of tRNAs to ribosomes[6]. One possible approach to investigating this in the *S. venezuelae* strains under the same stress conditions is through ribo-seq and analyzing the associated tRNAs to mRNAs within the ribosomes. I hypothesize that similar observations to the human study would be observed, in which tRNA association to ribosomes decreases upon demethylation of the m^1A58. Given the

profiled m^1A58 changes upon induction of stress as presented in chapter 5, this existing data set can be used to pinpoint isodecoders of interest to first examine.

Finally, the regulation of these maintenance enzymes remains a mystery in any organism. Having identified the two prominent m^1A58 demethylase enzymes, along with the previously annotated writer enzyme, conducting RNA-seq and identifying potential expression changes of these transcripts under stress conditions will be the first step in confirming if general changes in expression correlate with the documented m^1A profile changes.

6.3 Follow up studies to ncRNA discovery investigation in *Bacteroides fragilis*

The investigation into *Bacteroides fragilis* in Chapter 4 identify thio-modifications as biomarkers of interest under the concept of modification tunable transcripts. These modifications, unlike m^1A in *Streptomyces venezuelae*, are only found in a subset of tRNAs, providing greater focus in the decoding of these specific codons. Given the documented changes in transcript expression under oxidative conditions[98], codon usage analysis of oxidative stress response enzymes would be a suitable starting point for this study. I hypothesize that some of these enzymes show some bias in codon usage that require decoding by these thio-containing tRNAs. Generation of strains that lack these thio-modifications would then likely lead to reduced survival in such a stress condition, given the expected reduced efficiency in transcript decoding.

As for the *de novo* small RNA discovery workflow proposed in Chapter 5, continued work on further characterization of the unannotated sRNAs of interest remains a priority. Basic characterizations of these transcripts of interest would already include

predicted structures, genomic context and any conserved motifs as compared to existing Bacteroides sRNAs. Conservation analysis of these transcripts in other related species also provides useful supporting information in confirming them as relevant sRNAs. Once a concise list of transcripts has been selected, biochemical confirmation of their presence in samples from stress cultures through northern blot analysis will serves as a baseline standard in confirming the results seen from the sequencing approach.

References

(1) Watkins, C. P.; Zhang, W.; Wylder, A. C.; Katanski, C. D.; Pan, T. A multiplex platform for small RNA sequencing elucidates multifaceted tRNA stress response and translational regulation. *Nat Commun* **2022**, *13* (1), 2491. DOI: 10.1038/s41467-022-30261-3 From NLM Medline.
(2) Puerta-Fernandez, E.; Barrick, J. E.; Roth, A.; Breaker, R. R. Identification of a large noncoding RNA in extremophilic eubacteria. *Proc Natl Acad Sci U S A* **2006**, *103* (51), 19490-19495. DOI: 10.1073/pnas.0607493103 From NLM Medline.
(3) Shirahama, S.; Miki, A.; Kaburaki, T.; Akimitsu, N. Long Non-coding RNAs Involved in Pathogenic Infection. *Front Genet* **2020**, *11*, 454. DOI: 10.3389/fgene.2020.00454 From NLM PubMed-not-MEDLINE.
(4) Gupta, R. K.; Luong, T. T.; Lee, C. Y. RNAIII of the Staphylococcus aureus agr system activates global regulator MgrA by stabilizing mRNA. *Proc Natl Acad Sci U S A* **2015**, *112* (45), 14036-14041. DOI: 10.1073/pnas.1509251112 From NLM Medline.
(5) Endres, L.; Dedon, P. C.; Begley, T. J. Codon-biased translation can be regulated by wobble-base tRNA modification systems during cellular stress responses. *RNA Biol* **2015**, *12* (6), 603-614. DOI: 10.1080/15476286.2015.1031947 From NLM Medline.
(6) Liu, F.; Clark, W.; Luo, G.; Wang, X.; Fu, Y.; Wei, J.; Wang, X.; Hao, Z.; Dai, Q.; Zheng, G.; et al. ALKBH1-Mediated tRNA Demethylation Regulates Translation. *Cell* **2016**, *167* (7), 1897. DOI: 10.1016/j.cell.2016.11.045 From NLM PubMed-not-MEDLINE.
(7) Kramer, G. F.; Baker, J. C.; Ames, B. N. Near-UV stress in Salmonella typhimurium: 4-thiouridine in tRNA, ppGpp, and ApppGpp as components of an adaptive response. *J Bacteriol* **1988**, *170* (5), 2344-2351. DOI: 10.1128/jb.170.5.2344-2351.1988 From NLM Medline.
(8) Galvanin, A.; Vogt, L. M.; Grober, A.; Freund, I.; Ayadi, L.; Bourguignon-Igel, V.; Bessler, L.; Jacob, D.; Eigenbrod, T.; Marchand, V.; et al. Bacterial tRNA 2'-O-methylation is dynamically regulated under stress conditions and modulates innate immune response. *Nucleic Acids Res* **2020**, *48* (22), 12833-12844. DOI: 10.1093/nar/gkaa1123.
(9) Gehrig, S.; Eberle, M. E.; Botschen, F.; Rimbach, K.; Eberle, F.; Eigenbrod, T.; Kaiser, S.; Holmes, W. M.; Erdmann, V. A.; Sprinzl, M.; et al. Identification of modifications in microbial, native tRNA that suppress immunostimulatory activity. *J Exp Med* **2012**, *209* (2), 225-233. DOI: 10.1084/jem.20111044.
(10) Teng, Y.; Ren, Y.; Sayed, M.; Hu, X.; Lei, C.; Kumar, A.; Hutchins, E.; Mu, J.; Deng, Z.; Luo, C.; et al. Plant-Derived Exosomal MicroRNAs Shape the Gut Microbiota. *Cell Host Microbe* **2018**, *24* (5), 637-652 e638. DOI: 10.1016/j.chom.2018.10.001.
(11) Lei, C.; Teng, Y.; He, L.; Sayed, M.; Mu, J.; Xu, F.; Zhang, X.; Kumar, A.; Sundaram, K.; Sriwastva, M. K.; et al. Lemon exosome-like nanoparticles enhance stress survival of gut bacteria by RNase P-mediated specific tRNA decay. *iScience* **2021**, *24* (6), 102511. DOI: 10.1016/j.isci.2021.102511.
(12) Masuda, I.; Matsubara, R.; Christian, T.; Rojas, E. R.; Yadavalli, S. S.; Zhang, L.; Goulian, M.; Foster, L. J.; Huang, K. C.; Hou, Y. M. tRNA Methylation Is a Global Determinant of Bacterial Multi-drug Resistance. *Cell Syst* **2019**, *8* (4), 302-314 e308. DOI: 10.1016/j.cels.2019.03.008.
(13) Chionh, Y. H.; McBee, M.; Babu, I. R.; Hia, F.; Lin, W.; Zhao, W.; Cao, J.; Dziergowska, A.; Malkiewicz, A.; Begley, T. J.; et al. tRNA-mediated codon-biased translation in mycobacterial hypoxic persistence. *Nat Commun* **2016**, *7*, 13302. DOI: 10.1038/ncomms13302.
(14) Thongdee, N.; Jaroensuk, J.; Atichartpongkul, S.; Chittrakanwong, J.; Chooyoung, K.; Srimahaeak, T.; Chaiyen, P.; Vattanaviboon, P.; Mongkolsuk, S.; Fuangthong, M. TrmB, a tRNA m7G46 methyltransferase, plays a role in hydrogen peroxide resistance and positively modulates the translation of katA and katB mRNAs in Pseudomonas aeruginosa. *Nucleic Acids Res* **2019**, *47* (17), 9271-9281. DOI: 10.1093/nar/gkz702.

(15) Withey, J. H.; Friedman, D. I. A salvage pathway for protein structures: tmRNA and trans-translation. *Annu Rev Microbiol* **2003**, *57*, 101-123. DOI: 10.1146/annurev.micro.57.030502.090945 From NLM Medline.
(16) Waters, J. L.; Salyers, A. A. The small RNA RteR inhibits transfer of the Bacteroides conjugative transposon CTnDOT. *J Bacteriol* **2012**, *194* (19), 5228-5236. DOI: 10.1128/JB.00941-12 From NLM Medline.
(17) Peluso, P.; Herschlag, D.; Nock, S.; Freymann, D. M.; Johnson, A. E.; Walter, P. Role of 4.5S RNA in assembly of the bacterial signal recognition particle with its receptor. *Science* **2000**, *288* (5471), 1640-1643. DOI: 10.1126/science.288.5471.1640 From NLM Medline.
(18) Wassarman, K. M. 6S RNA, a Global Regulator of Transcription. *Microbiol Spectr* **2018**, *6* (3). DOI: 10.1128/microbiolspec.RWR-0019-2018 From NLM Medline.
(19) Vogel, J.; Argaman, L.; Wagner, E. G.; Altuvia, S. The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. *Curr Biol* **2004**, *14* (24), 2271-2276. DOI: 10.1016/j.cub.2004.12.003 From NLM Medline.
(20) Sharma, C. M.; Darfeuille, F.; Plantinga, T. H.; Vogel, J. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev* **2007**, *21* (21), 2804-2817. DOI: 10.1101/gad.447207 From NLM Medline.
(21) Zhang, A.; Altuvia, S.; Tiwari, A.; Argaman, L.; Hengge-Aronis, R.; Storz, G. The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein. *EMBO J* **1998**, *17* (20), 6061-6068. DOI: 10.1093/emboj/17.20.6061 From NLM Medline.
(22) Ryan, D.; Prezza, G.; Westermann, A. J. An RNA-centric view on gut Bacteroidetes. *Biol Chem* **2020**, *402* (1), 55-72. DOI: 10.1515/hsz-2020-0230 From NLM Medline.
(23) Ryan, D.; Jenniches, L.; Reichardt, S.; Barquist, L.; Westermann, A. J. A high-resolution transcriptome map identifies small RNA regulation of metabolism in the gut microbe Bacteroides thetaiotaomicron. *Nat Commun* **2020**, *11* (1), 3557. DOI: 10.1038/s41467-020-17348-5 From NLM Medline.
(24) Ryvkin, P.; Leung, Y. Y.; Silverman, I. M.; Childress, M.; Valladares, O.; Dragomir, I.; Gregory, B. D.; Wang, L. S. HAMR: high-throughput annotation of modified ribonucleotides. *RNA* **2013**, *19* (12), 1684-1692. DOI: 10.1261/rna.036806.112.
(25) Cozen, A. E.; Quartley, E.; Holmes, A. D.; Hrabeta-Robinson, E.; Phizicky, E. M.; Lowe, T. M. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat Methods* **2015**, *12* (9), 879-884. DOI: 10.1038/nmeth.3508.
(26) Zheng, G.; Qin, Y.; Clark, W. C.; Dai, Q.; Yi, C.; He, C.; Lambowitz, A. M.; Pan, T. Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods* **2015**, *12* (9), 835-837. DOI: 10.1038/nmeth.3478.
(27) Clark, W. C.; Evans, M. E.; Dominissini, D.; Zheng, G.; Pan, T. tRNA base methylation identification and quantification via high-throughput sequencing. *RNA* **2016**, *22* (11), 1771-1784. DOI: 10.1261/rna.056531.116.
(28) Behrens, A.; Rodschinka, G.; Nedialkova, D. D. High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Mol Cell* **2021**, *81* (8), 1802-1815 e1807. DOI: 10.1016/j.molcel.2021.01.028.
(29) Goodenbour, J. M.; Pan, T. Diversity of tRNA genes in eukaryotes. *Nucleic Acids Res* **2006**, *34* (21), 6137-6146. DOI: 10.1093/nar/gkl725.
(30) Spenkuch, F.; Motorin, Y.; Helm, M. Pseudouridine: still mysterious, but never a fake (uridine)! *RNA Biol* **2014**, *11* (12), 1540-1554. DOI: 10.4161/15476286.2014.992278.
(31) Bakin, A.; Ofengand, J. Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry* **1993**, *32* (37), 9754-9762. DOI: 10.1021/bi00088a030.
(32) Schwartz, S.; Bernstein, D. A.; Mumbach, M. R.; Jovanovic, M.; Herbst, R. H.; Leon-Ricardo, B. X.; Engreitz, J. M.; Guttman, M.; Satija, R.; Lander, E. S.; et al. Transcriptome-wide mapping

reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **2014**, *159* (1), 148-162. DOI: 10.1016/j.cell.2014.08.028.
(33) Carlile, T. M.; Rojas-Duran, M. F.; Zinshteyn, B.; Shin, H.; Bartoli, K. M.; Gilbert, W. V. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **2014**, *515* (7525), 143-146. DOI: 10.1038/nature13802.
(34) Song, J.; Zhuang, Y.; Zhu, C.; Meng, H.; Lu, B.; Xie, B.; Peng, J.; Li, M.; Yi, C. Differential roles of human PUS10 in miRNA processing and tRNA pseudouridylation. *Nat Chem Biol* **2020**, *16* (2), 160-169. DOI: 10.1038/s41589-019-0420-5.
(35) Khoddami, V.; Yerra, A.; Mosbruger, T. L.; Fleming, A. M.; Burrows, C. J.; Cairns, B. R. Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc Natl Acad Sci U S A* **2019**, *116* (14), 6784-6789. DOI: 10.1073/pnas.1817334116.
(36) Dai, Q.; Zhang, L. S.; Sun, H. L.; Pajdzik, K.; Yang, L.; Ye, C.; Ju, C. W.; Liu, S.; Wang, Y.; Zheng, Z.; et al. Quantitative sequencing using BID-seq uncovers abundant pseudouridines in mammalian mRNA at base resolution. *Nat Biotechnol* **2023**, *41* (3), 344-354. DOI: 10.1038/s41587-022-01505-w From NLM Medline.
(37) Zhang, M.; Jiang, Z.; Ma, Y.; Liu, W.; Zhuang, Y.; Lu, B.; Li, K.; Peng, J.; Yi, C. Quantitative profiling of pseudouridylation landscape in the human transcriptome. *Nat Chem Biol* **2023**, *19* (10), 1185-1195. DOI: 10.1038/s41589-023-01304-7 From NLM Medline.
(38) Edelheit, S.; Schwartz, S.; Mumbach, M. R.; Wurtzel, O.; Sorek, R. Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet* **2013**, *9* (6), e1003602. DOI: 10.1371/journal.pgen.1003602.
(39) Schaefer, M.; Pollex, T.; Hanna, K.; Lyko, F. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res* **2009**, *37* (2), e12. DOI: 10.1093/nar/gkn954.
(40) Squires, J. E.; Patel, H. R.; Nousch, M.; Sibbritt, T.; Humphreys, D. T.; Parker, B. J.; Suter, C. M.; Preiss, T. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res* **2012**, *40* (11), 5023-5033. DOI: 10.1093/nar/gks144.
(41) Gogakos, T.; Brown, M.; Garzia, A.; Meyer, C.; Hafner, M.; Tuschl, T. Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Rep* **2017**, *20* (6), 1463-1475. DOI: 10.1016/j.celrep.2017.07.029.
(42) Marchand, V.; Ayadi, L.; Ernst, F. G. M.; Hertler, J.; Bourguignon-Igel, V.; Galvanin, A.; Kotter, A.; Helm, M.; Lafontaine, D. L. J.; Motorin, Y. AlkAniline-Seq: Profiling of m(7) G and m(3) C RNA Modifications at Single Nucleotide Resolution. *Angew Chem Int Ed Engl* **2018**, *57* (51), 16785-16790. DOI: 10.1002/anie.201810946.
(43) Cui, J.; Liu, Q.; Sendinc, E.; Shi, Y.; Gregory, R. I. Nucleotide resolution profiling of m3C RNA modification by HAC-seq. *Nucleic Acids Res* **2021**, *49* (5), e27. DOI: 10.1093/nar/gkaa1186.
(44) Lin, S.; Liu, Q.; Jiang, Y. Z.; Gregory, R. I. Nucleotide resolution profiling of m(7)G tRNA modification by TRAC-Seq. *Nat Protoc* **2019**, *14* (11), 3220-3242. DOI: 10.1038/s41596-019-0226-7.
(45) Pang, Y. L.; Abo, R.; Levine, S. S.; Dedon, P. C. Diverse cell stresses induce unique patterns of tRNA up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy number. *Nucleic Acids Res* **2014**, *42* (22), e170. DOI: 10.1093/nar/gku945.
(46) Shigematsu, M.; Honda, S.; Loher, P.; Telonis, A. G.; Rigoutsos, I.; Kirino, Y. YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Res* **2017**, *45* (9), e70. DOI: 10.1093/nar/gkx005.
(47) Erber, L.; Hoffmann, A.; Fallmann, J.; Betat, H.; Stadler, P. F.; Morl, M. LOTTE-seq (Long hairpin oligonucleotide based tRNA high-throughput sequencing): specific selection of tRNAs with 3'-CCA end for high-throughput sequencing. *RNA Biol* **2020**, *17* (1), 23-32. DOI: 10.1080/15476286.2019.1664250.

(48) Pinkard, O.; McFarland, S.; Sweet, T.; Coller, J. Quantitative tRNA-sequencing uncovers metazoan tissue-specific tRNA regulation. *Nat Commun* **2020**, *11* (1), 4104. DOI: 10.1038/s41467-020-17879-x.
(49) Hu, J. F.; Yim, D.; Ma, D.; Huber, S. M.; Davis, N.; Bacusmo, J. M.; Vermeulen, S.; Zhou, J.; Begley, T. J.; DeMott, M. S.; et al. Quantitative mapping of the cellular small RNA landscape with AQRNA-seq. *Nat Biotechnol* **2021**. DOI: 10.1038/s41587-021-00874-y.
(50) Smith, A. M.; Abu-Shumays, R.; Akeson, M.; Bernick, D. L. Capture, Unfolding, and Detection of Individual tRNA Molecules Using a Nanopore Device. *Front Bioeng Biotechnol* **2015**, *3*, 91. DOI: 10.3389/fbioe.2015.00091.
(51) Lucas, M. C.; Pryszcz, L. P.; Medina, R.; Milenkovic, I.; Camacho, N.; Marchand, V.; Motorin, Y.; Ribas de Pouplana, L.; Novoa, E. M. Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing. *Nat Biotechnol* **2024**, *42* (1), 72-86. DOI: 10.1038/s41587-023-01743-6 From NLM Medline.
(52) Boccaletto, P.; Machnicka, M. A.; Purta, E.; Piatkowski, P.; Baginski, B.; Wirecki, T. K.; de Crecy-Lagard, V.; Ross, R.; Limbach, P. A.; Kotter, A.; et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* **2018**, *46* (D1), D303-D307. DOI: 10.1093/nar/gkx1030.
(53) Kimura, S.; Dedon, P. C.; Waldor, M. K. Comparative tRNA sequencing and RNA mass spectrometry for surveying tRNA modifications. *Nat Chem Biol* **2020**, *16* (9), 964-972. DOI: 10.1038/s41589-020-0558-1 From NLM Medline.
(54) Foo, M.; Frietze, L. R.; Enghiad, B.; Yuan, Y.; Katanski, C. D.; Zhao, H.; Pan, T. Prokaryotic RNA N1-Methyladenosine Erasers Maintain tRNA m1A Modification Levels in Streptomyces venezuelae. *ACS Chemical Biology* **2024**. DOI: 10.1021/acschembio.4c00278.
(55) Dunn, D. B. The occurrence of 1-methyladenine in ribonucleic acid. *Biochim Biophys Acta* **1961**, *46*, 198-200. DOI: 10.1016/0006-3002(61)90668-0 From NLM Medline.
(56) Dominissini, D.; Nachtergaele, S.; Moshitch-Moshkovitz, S.; Peer, E.; Kol, N.; Ben-Haim, M. S.; Dai, Q.; Di Segni, A.; Salmon-Divon, M.; Clark, W. C.; et al. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **2016**, *530* (7591), 441-446. DOI: 10.1038/nature16998 From NLM Medline.
(57) Li, X.; Xiong, X.; Zhang, M.; Wang, K.; Chen, Y.; Zhou, J.; Mao, Y.; Lv, J.; Yi, D.; Chen, X. W.; et al. Base-Resolution Mapping Reveals Distinct m(1)A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. *Mol Cell* **2017**, *68* (5), 993-1005 e1009. DOI: 10.1016/j.molcel.2017.10.019 From NLM Medline.
(58) Li, X.; Xiong, X.; Wang, K.; Wang, L.; Shu, X.; Ma, S.; Yi, C. Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat Chem Biol* **2016**, *12* (5), 311-316. DOI: 10.1038/nchembio.2040 From NLM Medline.
(59) Kanazawa, H.; Baba, F.; Koganei, M.; Kondo, J. A structural basis for the antibiotic resistance conferred by an N1-methylation of A1408 in 16S rRNA. *Nucleic Acids Res* **2017**, *45* (21), 12529-12535. DOI: 10.1093/nar/gkx882 From NLM Medline.
(60) Safra, M.; Sas-Chen, A.; Nir, R.; Winkler, R.; Nachshon, A.; Bar-Yaacov, D.; Erlacher, M.; Rossmanith, W.; Stern-Ginossar, N.; Schwartz, S. The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **2017**, *551* (7679), 251-255. DOI: 10.1038/nature24456 From NLM Medline.
(61) Liu, Y.; Zhou, J.; Li, X.; Zhang, X.; Shi, J.; Wang, X.; Li, H.; Miao, S.; Chen, H.; He, X.; et al. tRNA-m(1)A modification promotes T cell expansion via efficient MYC protein synthesis. *Nat Immunol* **2022**, *23* (10), 1433-1444. DOI: 10.1038/s41590-022-01301-3 From NLM Medline.
(62) Zhang, C.; Jia, G. Reversible RNA Modification N(1)-methyladenosine (m(1)A) in mRNA and tRNA. *Genomics Proteomics Bioinformatics* **2018**, *16* (3), 155-161. DOI: 10.1016/j.gpb.2018.03.003 From NLM Medline.
(63) Schwartz, M. H.; Wang, H.; Pan, J. N.; Clark, W. C.; Cui, S.; Eckwahl, M. J.; Pan, D. W.; Parisien, M.; Owens, S. M.; Cheng, B. L.; et al. Microbiome characterization by high-throughput

transfer RNA sequencing and modification analysis. *Nat Commun* **2018**, *9* (1), 5353. DOI: 10.1038/s41467-018-07675-z.
(64) Helm, M.; Brule, H.; Degoul, F.; Cepanec, C.; Leroux, J. P.; Giege, R.; Florentz, C. The presence of modified nucleotides is required for cloverleaf folding of a human mitochondrial tRNA. *Nucleic Acids Res* **1998**, *26* (7), 1636-1643. DOI: 10.1093/nar/26.7.1636 From NLM Medline.
(65) Anderson, J.; Phan, L.; Cuesta, R.; Carlson, B. A.; Pak, M.; Asano, K.; Bjork, G. R.; Tamame, M.; Hinnebusch, A. G. The essential Gcd10p-Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes Dev* **1998**, *12* (23), 3650-3662. DOI: 10.1101/gad.12.23.3650 From NLM Medline.
(66) Droogmans, L.; Roovers, M.; Bujnicki, J. M.; Tricot, C.; Hartsch, T.; Stalon, V.; Grosjean, H. Cloning and characterization of tRNA (m1A58) methyltransferase (TrmI) from Thermus thermophilus HB27, a protein required for cell growth at extreme temperatures. *Nucleic Acids Res* **2003**, *31* (8), 2148-2156. DOI: 10.1093/nar/gkg314 From NLM Medline.
(67) Masuda, M.; Nishihira, T.; Itoh, K.; Mizugaki, M.; Ishida, N.; Mori, S. An immunohistochemical analysis for cancer of the esophagus using monoclonal antibodies specific for modified nucleosides. *Cancer* **1993**, *72* (12), 3571-3578. DOI: 10.1002/1097-0142(19931215)72:12<3571::aid-cncr2820721205>3.0.co;2-9 From NLM Medline.
(68) Sharma, S.; Watzinger, P.; Kotter, P.; Entian, K. D. Identification of a novel methyltransferase, Bmt2, responsible for the N-1-methyl-adenosine base modification of 25S rRNA in Saccharomyces cerevisiae. *Nucleic Acids Res* **2013**, *41* (10), 5428-5443. DOI: 10.1093/nar/gkt195 From NLM Medline.
(69) Cao, X.; Limbach, P. A. Enhanced detection of post-transcriptional modifications using a mass-exclusion list strategy for RNA modification mapping by LC-MS/MS. *Anal Chem* **2015**, *87* (16), 8433-8440. DOI: 10.1021/acs.analchem.5b01826 From NLM Medline.
(70) Alberti, F.; Corre, C. Editing streptomycete genomes in the CRISPR/Cas9 age. *Nat Prod Rep* **2019**, *36* (9), 1237-1248. DOI: 10.1039/c8np00081f From NLM Medline.
(71) Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* **2002**, *419* (6903), 174-178. DOI: 10.1038/nature00908 From NLM Medline.
(72) Doull, J. L.; Singh, A. K.; Hoare, M.; Ayer, S. W. Conditions for the production of jadomycin B by Streptomyces venezuelae ISP5230: effects of heat shock, ethanol treatment and phage infection. *J Ind Microbiol* **1994**, *13* (2), 120-125. DOI: 10.1007/BF01584109 From NLM Medline.
(73) Sekurova, O. N.; Zhang, J.; Kristiansen, K. A.; Zotchev, S. B. Activation of chloramphenicol biosynthesis in Streptomyces venezuelae ATCC 10712 by ethanol shock: insights from the promoter fusion studies. *Microb Cell Fact* **2016**, *15*, 85. DOI: 10.1186/s12934-016-0484-9 From NLM Medline.
(74) Shirling, E. B.; Gottlieb, D. Methods for characterization of Streptomyces species. *International Journal of Systematic Bacteriology* **1966**, *16* (3), 313-340. DOI: 10.1099/00207713-16-3-313.
(75) Girardot, C.; Scholtalbers, J.; Sauer, S.; Su, S. Y.; Furlong, E. E. Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. *BMC Bioinformatics* **2016**, *17* (1), 419. DOI: 10.1186/s12859-016-1284-2 From NLM Medline.
(76) Chan, P. P.; Lowe, T. M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* **2016**, *44* (D1), D184-189. DOI: 10.1093/nar/gkv1309 From NLM Medline.
(77) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Genome Project Data Processing, S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25* (16), 2078-2079. DOI: 10.1093/bioinformatics/btp352 From NLM Medline.
(78) Motorin, Y.; Helm, M. RNA nucleotide methylation. *Wiley Interdiscip Rev RNA* **2011**, *2* (5), 611-631. DOI: 10.1002/wrna.79 From NLM Medline.

(79) Anantharaman, V.; Koonin, E. V.; Aravind, L. SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J Mol Microbiol Biotechnol* **2002**, *4* (1), 71-75. From NLM Medline.
(80) Tkaczuk, K. L.; Dunin-Horkawicz, S.; Purta, E.; Bujnicki, J. M. Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC Bioinformatics* **2007**, *8*, 73. DOI: 10.1186/1471-2105-8-73 From NLM Medline.
(81) Oerum, S.; Degut, C.; Barraud, P.; Tisne, C. m1A Post-Transcriptional Modification in tRNAs. *Biomolecules* **2017**, *7* (1). DOI: 10.3390/biom7010020 From NLM Medline.
(82) Gupta, A.; Kumar, P. H.; Dineshkumar, T. K.; Varshney, U.; Subramanya, H. S. Crystal structure of Rv2118c: an AdoMet-dependent methyltransferase from Mycobacterium tuberculosis H37Rv. *J Mol Biol* **2001**, *312* (2), 381-391. DOI: 10.1006/jmbi.2001.4935 From NLM Medline.
(83) Guelorget, A.; Barraud, P.; Tisne, C.; Golinelli-Pimpaneau, B. Structural comparison of tRNA m(1)A58 methyltransferases revealed different molecular strategies to maintain their oligomeric architecture under extreme conditions. *BMC Struct Biol* **2011**, *11*, 48. DOI: 10.1186/1472-6807-11-48 From NLM Medline.
(84) Barraud, P.; Golinelli-Pimpaneau, B.; Atmanene, C.; Sanglier, S.; Van Dorsselaer, A.; Droogmans, L.; Dardel, F.; Tisne, C. Crystal structure of Thermus thermophilus tRNA m1A58 methyltransferase and biophysical characterization of its interaction with tRNA. *J Mol Biol* **2008**, *377* (2), 535-550. DOI: 10.1016/j.jmb.2008.01.041 From NLM Medline.
(85) Takuma, H.; Ushio, N.; Minoji, M.; Kazayama, A.; Shigi, N.; Hirata, A.; Tomikawa, C.; Ochi, A.; Hori, H. Substrate tRNA recognition mechanism of eubacterial tRNA (m1A58) methyltransferase (TrmI). *J Biol Chem* **2015**, *290* (9), 5912-5925. DOI: 10.1074/jbc.M114.606038 From NLM Medline.
(86) Chen, Z.; Qi, M.; Shen, B.; Luo, G.; Wu, Y.; Li, J.; Lu, Z.; Zheng, Z.; Dai, Q.; Wang, H. Transfer RNA demethylase ALKBH3 promotes cancer progression via induction of tRNA-derived small RNAs. *Nucleic Acids Res* **2019**, *47* (5), 2533-2545. DOI: 10.1093/nar/gky1250 From NLM Medline.
(87) Wei, J.; Liu, F.; Lu, Z.; Fei, Q.; Ai, Y.; He, P. C.; Shi, H.; Cui, X.; Su, R.; Klungland, A.; et al. Differential m(6)A, m(6)A(m), and m(1)A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Mol Cell* **2018**, *71* (6), 973-985 e975. DOI: 10.1016/j.molcel.2018.08.011 From NLM Medline.
(88) Cobb, R. E.; Wang, Y.; Zhao, H. High-efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system. *ACS Synth Biol* **2015**, *4* (6), 723-728. DOI: 10.1021/sb500351f From NLM Medline.
(89) Bochner, B. R.; Gadzinski, P.; Panomitros, E. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res* **2001**, *11* (7), 1246-1255. DOI: 10.1101/gr.186501 From NLM Medline.
(90) Rokytskyy, I.; Koshla, O.; Fedorenko, V.; Ostash, B. Decoding options and accuracy of translation of developmentally regulated UUA codon in Streptomyces: bioinformatic analysis. *Springerplus* **2016**, *5* (1), 982. DOI: 10.1186/s40064-016-2683-6 From NLM PubMed-not-MEDLINE.
(91) Koshla, O.; Vogt, L. M.; Rydkin, O.; Sehin, Y.; Ostash, I.; Helm, M.; Ostash, B. Landscape of Post-Transcriptional tRNA Modifications in Streptomyces albidoflavus J1074 as Portrayed by Mass Spectrometry and Genomic Data Mining. *J Bacteriol* **2023**, *205* (1), e0029422. DOI: 10.1128/jb.00294-22 From NLM Medline.
(92) Roovers, M.; Wouters, J.; Bujnicki, J. M.; Tricot, C.; Stalon, V.; Grosjean, H.; Droogmans, L. A primordial RNA modification enzyme: the case of tRNA (m1A) methyltransferase. *Nucleic Acids Res* **2004**, *32* (2), 465-476. DOI: 10.1093/nar/gkh191 From NLM Medline.
(93) S Brooks, M.; J Burdock, T.; E Ghaly, A. Changes in Cell Structure, Morphology and Activity of Streptomyces venezuelae during the Growth, Shocking and Jadomycin Production Stages.

Journal of Microbial & Biochemical Technology **2012**, *04* (03), 063-075. DOI: 10.4172/1948-5948.1000073.
(94) Zhu, Y.; Zhang, P.; Zhang, J.; Wang, J.; Lu, Y.; Pang, X. Impact on Multiple Antibiotic Pathways Reveals MtrA as a Master Regulator of Antibiotic Production in Streptomyces spp. and Potentially in Other Actinobacteria. *Appl Environ Microbiol* **2020**, *86* (20), e01201-01220. DOI: 10.1128/AEM.01201-20 From NLM Medline.
(95) Sprouffske, K.; Wagner, A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics* **2016**, *17*, 172. DOI: 10.1186/s12859-016-1016-7 From NLM Medline.
(96) Rocha, E. R.; Herren, C. D.; Smalley, D. J.; Smith, C. J. The complex oxidative stress response of Bacteroides fragilis: the role of OxyR in control of gene expression. *Anaerobe* **2003**, *9* (4), 165-173. DOI: 10.1016/S1075-9964(03)00118-5 From NLM PubMed-not-MEDLINE.
(97) Rocha, E. R.; Selby, T.; Coleman, J. P.; Smith, C. J. Oxidative stress response in an anaerobe, Bacteroides fragilis: a role for catalase in protection against hydrogen peroxide. *J Bacteriol* **1996**, *178* (23), 6895-6903. DOI: 10.1128/jb.178.23.6895-6903.1996 From NLM Medline.
(98) Sund, C. J.; Rocha, E. R.; Tzianabos, A. O.; Wells, W. G.; Gee, J. M.; Reott, M. A.; O'Rourke, D. P.; Smith, C. J. The Bacteroides fragilis transcriptome response to oxygen and H2O2: the role of OxyR and its effect on survival and virulence. *Mol Microbiol* **2008**, *67* (1), 129-142. DOI: 10.1111/j.1365-2958.2007.06031.x From NLM Medline.
(99) Romsang, A.; Duang-Nkern, J.; Khemsom, K.; Wongsaroj, L.; Saninjuk, K.; Fuangthong, M.; Vattanaviboon, P.; Mongkolsuk, S. Pseudomonas aeruginosa ttcA encoding tRNA-thiolating protein requires an iron-sulfur cluster to participate in hydrogen peroxide-mediated stress protection and pathogenicity. *Sci Rep* **2018**, *8* (1), 11882. DOI: 10.1038/s41598-018-30368-y From NLM Medline.
(100) Imlay, J. A. Iron-sulphur clusters and the problem with oxygen. *Mol Microbiol* **2006**, *59* (4), 1073-1082. DOI: 10.1111/j.1365-2958.2006.05028.x From NLM Medline.
(101) Fogarty, E. C.; Schechter, M. S.; Lolans, K.; Sheahan, M. L.; Veseli, I.; Moore, R. M.; Kiefl, E.; Moody, T.; Rice, P. A.; Yu, M. K.; et al. A cryptic plasmid is among the most numerous genetic elements in the human gut. *Cell* **2024**, *187* (5), 1206-1222.e1216. DOI: 10.1016/j.cell.2024.01.039 From NLM.
(102) Katanski, C. D.; Watkins, C. P.; Zhang, W.; Reyer, M.; Miller, S.; Pan, T. Analysis of queuosine and 2-thio tRNA modifications by high throughput sequencing. *Nucleic Acids Res* **2022**, *50* (17), e99. DOI: 10.1093/nar/gkac517 From NLM Medline.
(103) Hör, J.; Matera, G.; Vogel, J.; Gottesman, S.; Storz, G. Trans-Acting Small RNAs and Their Effects on Gene Expression in Escherichia coli and Salmonella enterica. *EcoSal Plus* **2020**, *9* (1). DOI: 10.1128/ecosalplus.ESP-0030-2019 From NLM.
(104) Wassarman, K. M.; Repoila, F.; Rosenow, C.; Storz, G.; Gottesman, S. Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* **2001**, *15* (13), 1637-1651. DOI: 10.1101/gad.901001 From NLM Medline.
(105) Zhang, A.; Wassarman, K. M.; Rosenow, C.; Tjaden, B. C.; Storz, G.; Gottesman, S. Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol* **2003**, *50* (4), 1111-1124. DOI: 10.1046/j.1365-2958.2003.03734.x From NLM Medline.
(106) Chao, Y.; Papenfort, K.; Reinhardt, R.; Sharma, C. M.; Vogel, J. An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* **2012**, *31* (20), 4005-4019. DOI: 10.1038/emboj.2012.229 From NLM Medline.
(107) Melamed, S.; Peer, A.; Faigenbaum-Romm, R.; Gatt, Y. E.; Reiss, N.; Bar, A.; Altuvia, Y.; Argaman, L.; Margalit, H. Global Mapping of Small RNA-Target Interactions in Bacteria. *Mol Cell* **2016**, *63* (5), 884-897. DOI: 10.1016/j.molcel.2016.07.026 From NLM Medline.
(108) Holmqvist, E.; Wright, P. R.; Li, L.; Bischler, T.; Barquist, L.; Reinhardt, R.; Backofen, R.; Vogel, J. Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed

by UV crosslinking in vivo. *EMBO J* **2016**, *35* (9), 991-1011. DOI: 10.15252/embj.201593360 From NLM Medline.
(109) Holmqvist, E.; Li, L.; Bischler, T.; Barquist, L.; Vogel, J. Global Maps of ProQ Binding In Vivo Reveal Target Recognition via RNA Structure and Stability Control at mRNA 3' Ends. *Mol Cell* **2018**, *70* (5), 971-982 e976. DOI: 10.1016/j.molcel.2018.04.017 From NLM Medline.
(110) Potts, A. H.; Vakulskas, C. A.; Pannuri, A.; Yakhnin, H.; Babitzke, P.; Romeo, T. Global role of the bacterial post-transcriptional regulator CsrA revealed by integrated transcriptomics. *Nat Commun* **2017**, *8* (1), 1596. DOI: 10.1038/s41467-017-01613-1 From NLM Medline.
(111) Sharma, C. M.; Vogel, J. Differential RNA-seq: the approach behind and the biological insight gained. *Curr Opin Microbiol* **2014**, *19*, 97-105. DOI: 10.1016/j.mib.2014.06.010 From NLM Medline.
(112) Jeters, R. T.; Wang, G. R.; Moon, K.; Shoemaker, N. B.; Salyers, A. A. Tetracycline-associated transcriptional regulation of transfer genes of the Bacteroides conjugative transposon CTnDOT. *J Bacteriol* **2009**, *191* (20), 6374-6382. DOI: 10.1128/JB.00739-09 From NLM Medline.
(113) Whittle, G.; Shoemaker, N. B.; Salyers, A. A. Characterization of genes involved in modulation of conjugal transfer of the Bacteroides conjugative transposon CTnDOT. *J Bacteriol* **2002**, *184* (14), 3839-3847. DOI: 10.1128/JB.184.14.3839-3847.2002 From NLM Medline.
(114) Katanski, C. D.; Alshammary, H.; Watkins, C. P.; Huang, S.; Gonzales-Reiche, A.; Sordillo, E. M.; van Bakel, H.; Mount Sinai, P. S. P. s. g.; Lolans, K.; Simon, V.; et al. tRNA abundance, modification and fragmentation in nasopharyngeal swabs as biomarkers for COVID-19 severity. *Front Cell Dev Biol* **2022**, *10*, 999351. DOI: 10.3389/fcell.2022.999351 From NLM PubMed-not-MEDLINE.
(115) Porcheron, G.; Dozois, C. M. Interplay between iron homeostasis and virulence: Fur and RyhB as major regulators of bacterial pathogenicity. *Vet Microbiol* **2015**, *179* (1-2), 2-14. DOI: 10.1016/j.vetmic.2015.03.024 From NLM Medline.
(116) Bobrovskyy, M.; Vanderpool, C. K. The small RNA SgrS: roles in metabolism and pathogenesis of enteric bacteria. *Front Cell Infect Microbiol* **2014**, *4*, 61. DOI: 10.3389/fcimb.2014.00061 From NLM Medline.
(117) Spitale, R. C.; Flynn, R. A.; Zhang, Q. C.; Crisalli, P.; Lee, B.; Jung, J. W.; Kuchelmeister, H. Y.; Batista, P. J.; Torre, E. A.; Kool, E. T.; et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **2015**, *519* (7544), 486-490. DOI: 10.1038/nature14263 From NLM Medline.
(118) Rouskin, S.; Zubradt, M.; Washietl, S.; Kellis, M.; Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **2014**, *505* (7485), 701-705. DOI: 10.1038/nature12894 From NLM Medline.
(119) Adams, A. N. D.; Azam, M. S.; Costliow, Z. A.; Ma, X.; Degnan, P. H.; Vanderpool, C. K. A Novel Family of RNA-Binding Proteins Regulate Polysaccharide Metabolism in Bacteroides thetaiotaomicron. *J Bacteriol* **2021**, *203* (21), e0021721. DOI: 10.1128/JB.00217-21 From NLM Medline.
(120) Liu, Y.; Zhang, S.; Gao, X.; Ru, Y.; Gu, X.; Hu, X. Research progress of N1-methyladenosine RNA modification in cancer. *Cell Commun Signal* **2024**, *22* (1), 79. DOI: 10.1186/s12964-023-01401-z From NLM Medline.

Appendix – Supplemental Information

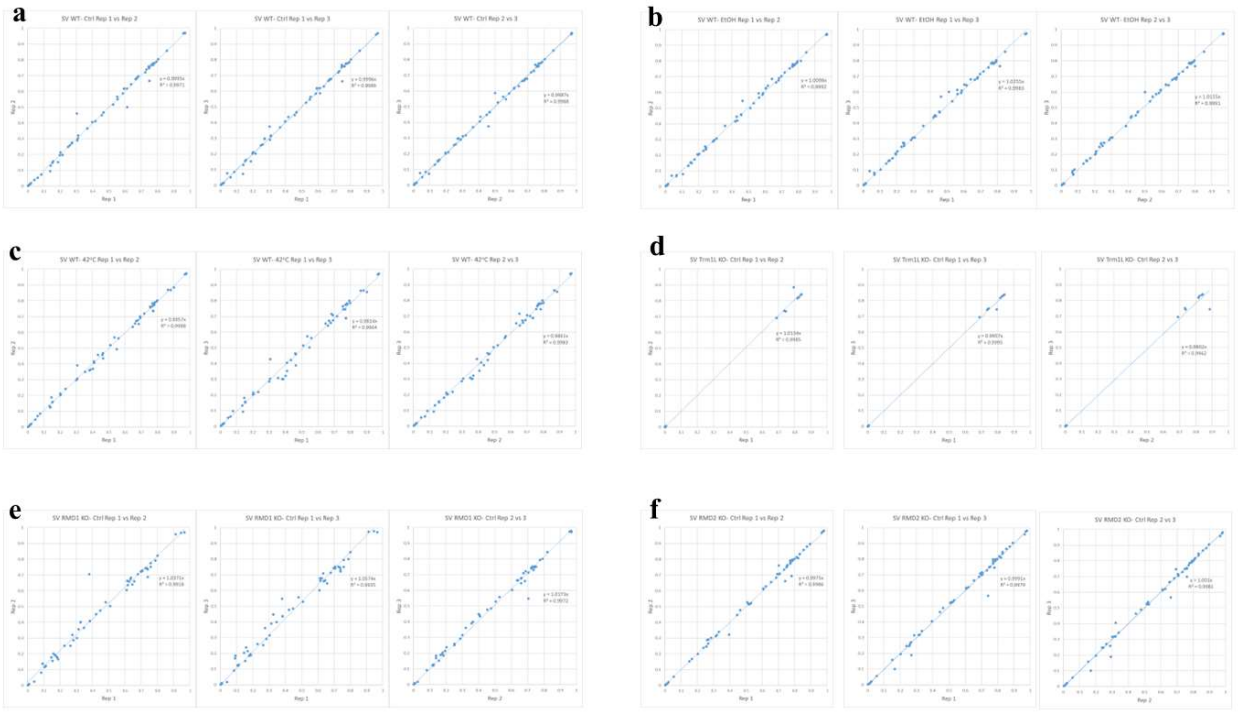

Figure S3.1: Correlation of $m^1A58/59$ mutation rates between replicates used in MSR-seq stress libraries.

(a) Correlation of mutational rates of WT under control conditions between replicate 1 vs 2, 1 vs 3 and 2 vs 3, respectively.
(b) Same as (a) but for WT under ethanol stress.
(c) Same as (a) but for WT under heat stress.
(d) Correlation of mutational rates of Trm1L KO under control conditions between replicate 1 vs 2, 1 vs 3 and 2 vs 3, respectively.
(e) Same as (d) but for RMD1 KO under control conditions.
(f) Same as (d) but for RMD12KO under control conditions.

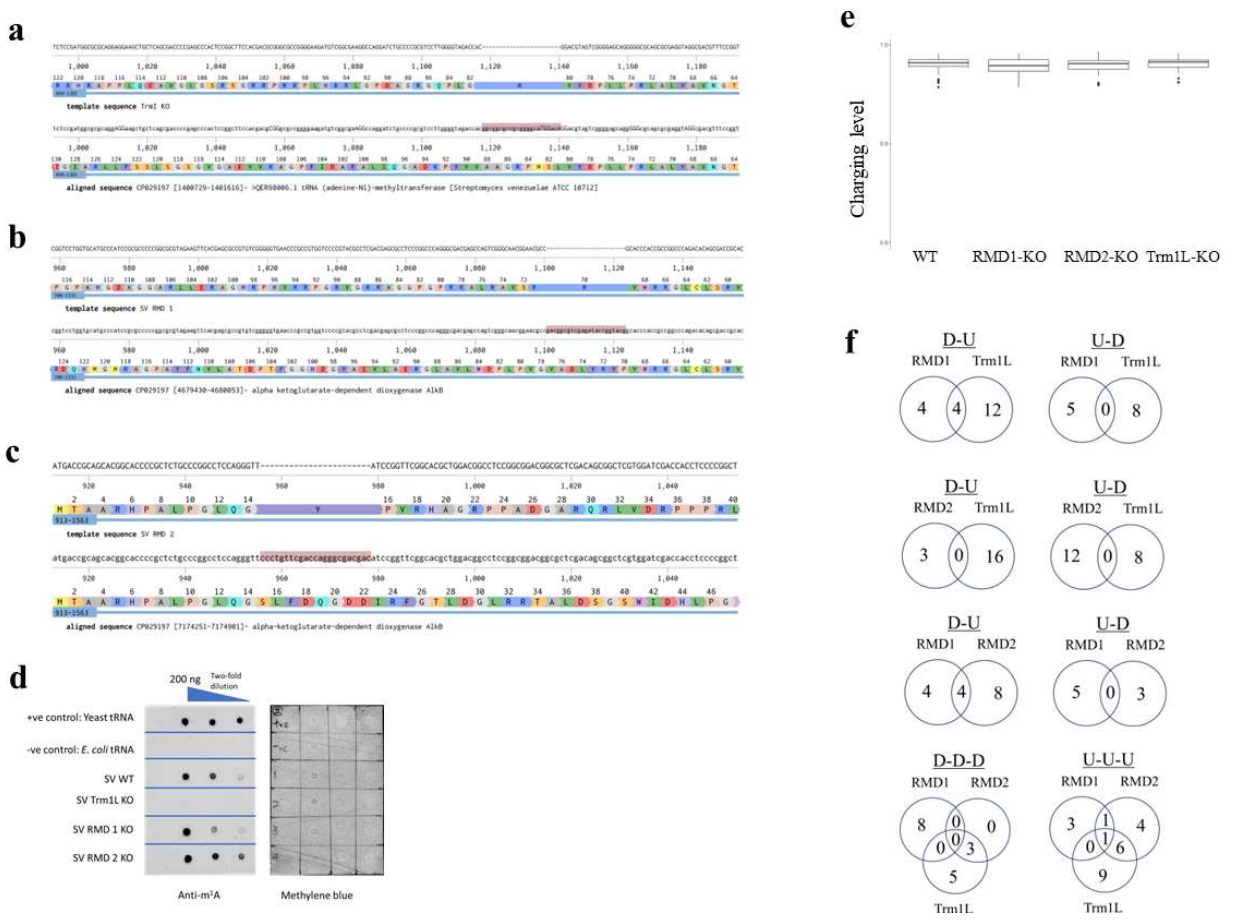

Figure S3.2: Additional data for Fig. 3.2.

(a) Saenger sequencing result of TrmI-like gene KO. Gene deletion was targeted at bases 242-264 (corresponding to deletion of codons 81-88) resulting in a frameshift mutation at codon position 81.
(b) Saenger sequencing result of RMD1 gene KO. Gene deletion was targeted at bases 209-231 (corresponding to deletion of codons 70-77), resulting in a frameshift mutation at codon position 70.
(c) Saenger sequencing result of RMD2 gene KO. Gene deletion was targeted at bases 43-66 (corresponding to deletion of codons 15-22), resulting in a frameshift mutation starting at codon 15.
(d) Dot blot result using m1A antibody of total RNA from WT and KO strains.
(e) tRNA charging levels in WT, RMD1-KO, RMD2-KO, and Trm1L-KO strains.
(f) Venn diagram of the number of tRNA isoacceptor abundance changes in the 3 KO strains compared to the wild-type strain. D: decrease in KO; U: increase in KO.

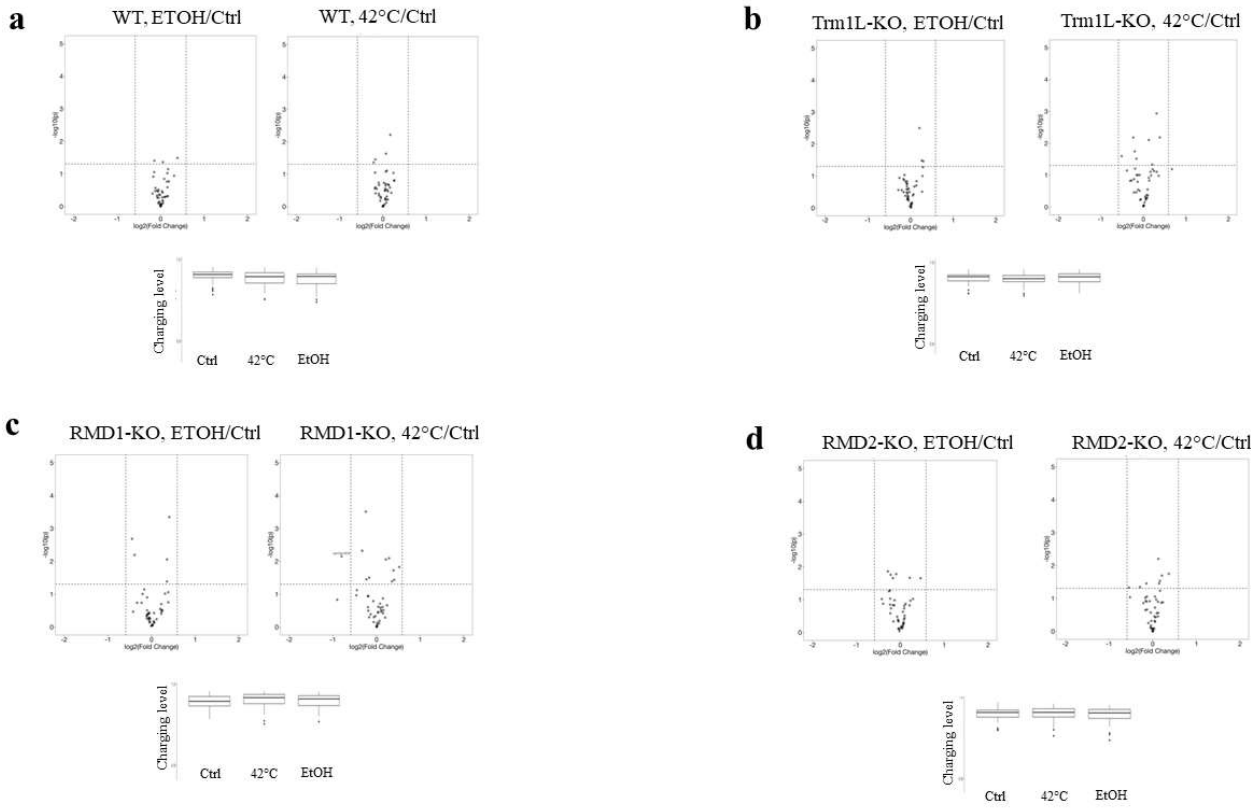

Figure S3.3: Additional results for Fig. 3.3.

(a) tRNA abundance (top two graphs) and charging changes of WT, stress vs control.
(b) tRNA abundance (top two graphs) and charging changes of Trm1L-KO, stress vs control.
(c) tRNA abundance (top two graphs) and charging changes of RMD1-KO, stress vs control.
(d) tRNA abundance (top two graphs) and charging changes of RMD2-KO, stress vs control.

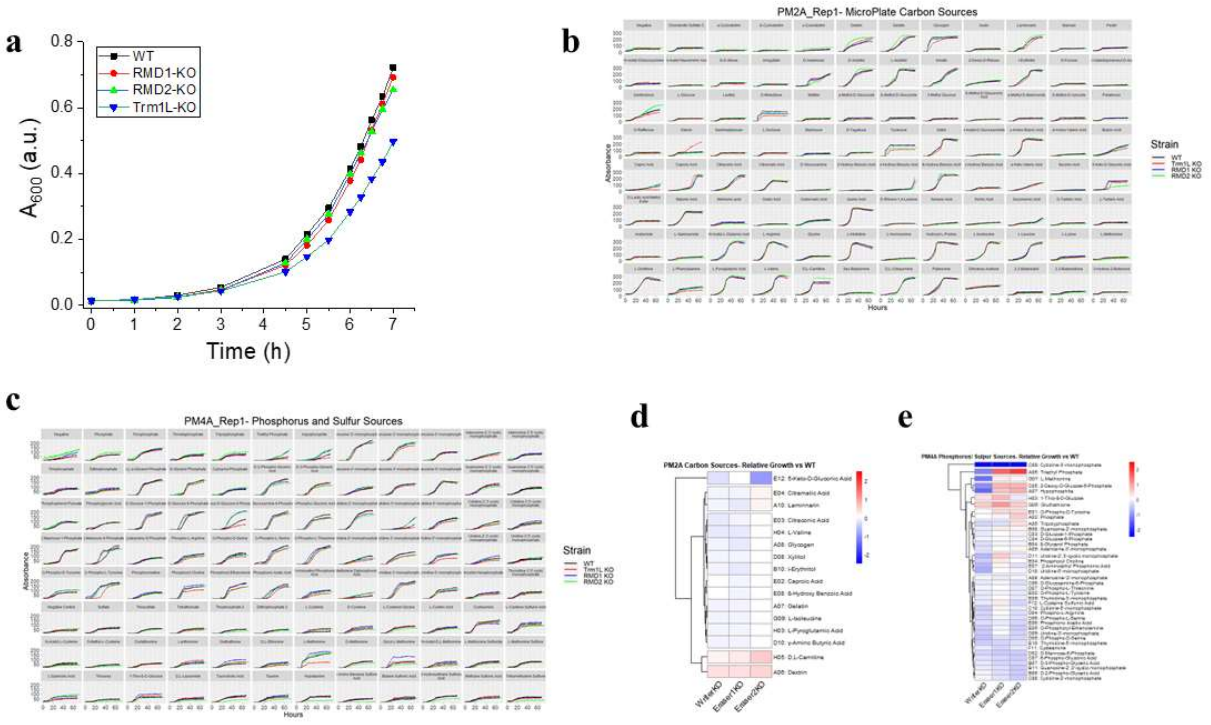

Figure S3.4: Additional data for Fig. 3.4.

(a) Growth difference of WT and the 3 KO strains in rich medium.
(b) Biolog PM 2A (Carbon sources 2) microarray plate growth of WT and the 3 KO strains.
(c) Biolog PM 4A (Phosphorus & sulfur sources) microarray plate growth of WT and the 3 KO strains
(d) Relative growth comparison of Biolog PM 2A (Carbon sources 2) microarray plate of WT vs 3 KO strains, quantified using Growthcurver.
(e) Relative growth comparison of Biolog PM 4A (phosphorus & sulfur sources) microarray plate of WT vs 3 KO strains, quantified using Growthcurver.

Table S3.1: Bacterial strains and plasmids used in this study

Strain or plasmid	Description	Source or reference
Strains		
S. venezuelae		
ATCC 10712	Wild type	ATCC
RMD1 mutant	SV RMD1 disrupted strain	This study
RMD2 mutant	SV RMD2 disrupted strain	This study
Trm1L mutant	SV Trm1L disrupted strain	This study
E. coli		
T7 Express Competent (High Efficiency)	Strain used for protein expression	NEB
Plasmids		
pET-SV-RMD1	SV RMD1 expression plasmid	This study
pET-SV-RMD2	SV RMD2 expression plasmid	This study