2020

IT'S THE LITTLE THINGS: AN EXPLORATION OF SMALL RNAS AND SELFISH GENETIC ELEMENTS OF THE HUMAN BACTERIAL PATHOGENS COXIELLA BURNETII AND BARTONELLA BACILLIFORMIS

Shaun Gregory Wachter

IT'S THE LITTLE THINGS: AN EXPLORATION OF SMALL RNAS AND SELFISH GENETIC ELEMENTS OF THE HUMAN BACTERIAL PATHOGENS *COXIELLA BURNETII* AND *BARTONELLA BACILLIFORMIS*

By

SHAUN GREGORY WACHTER

BS, University of Illinois at Urbana-Champaign, Champaign, IL, 2010
MS, Northern Illinois University, Dekalb, IL, 2013

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in Cellular, Molecular, and Microbial Biology, Microbiology

The University of Montana
Missoula, MT

August 2020

Approved by:

Scott Whittenburg,
Graduate School Dean

Michael Minnick, Chair
Division of Biological Sciences

Scott Samuels
Division of Biological Sciences

Steve Lodmell
Division of Biological Sciences

Dan Drecktrah
Division of Biological Sciences

Kent Sugden
Department of Chemistry and Biochemistry

Wachter, Shaun, Doctor of Philosophy, Summer 2020 Microbiology

It's the little things: An exploration of small RNAs and selfish genetic elements of the human bacterial pathogens *Coxiella burnetii* and *Bartonella bacilliformis*

Chairperson: Dr. Michael Minnick

Coxiella burnetii is a Gram-negative gammaproteobacterium and zoonotic agent of Q fever in humans. Previous work in our lab has demonstrated that *C. burnetii* codes for several small RNAs (sRNAs) that are differentially expressed between *in vivo* and *in vitro* growth conditions. sRNAs serve as post-transcriptional regulatory effectors involved in the control of nearly all biological processes. We demonstrated that several of the identified sRNAs, namely ***C****oxiella* ***b****urnetii* **s**mall **R**NA **3** (CbsR3), Cbsr13, and CbsR16, represent members of two novel families of miniature inverted-repeat transposable elements (MITEs), termed QMITE1 and QMITE2. Furthermore, we have characterized a highly expressed, infection-specific sRNA, CbsR12, and have determined that it is necessary for expansion of the *C. burnetii* intracellular niche in a human monocyte-derived alveolar macrophage cell line. We have determined that CbsR12 may participate in broad gene regulation by acting as an "RNA sponge" for the global regulatory RNA-binding protein CsrA. Additionally, CbsR12 is a *trans*-acting sRNA that targets transcripts of the *carA*, *metK*, and *cvpD* genes *in vitro* and *in vivo*.

Bartonella bacilliformis is a Gram-negative alphaproteobacterium and the etiological agent of Carrión's disease in humans. *B. bacilliformis* is spread between humans through the bite of female phlebotomine sand flies. As a result, the pathogen encounters significant environmental shifts during its life cycle, including changes in pH and temperature. Bacterial sRNAs can serve as a means of rapid regulation under shifting environmental conditions. We therefore performed total RNA-sequencing analyses on *B. bacilliformis* grown *in vitro* then shifted to one of ten distinct conditions that simulate various environments encountered by the pathogen during its life cycle. From this, we identified 160 sRNAs significantly expressed under at least one of the conditions tested. Northern blot analysis was used to confirm the expression of eight novel sRNAs. We also characterized a ***B****artonella* ***b****acilliformis* **g**roup **I** intron (BbgpI) that disrupts an un-annotated $tRNA_{CCU}^{Arg}$ gene and determined that the intron splices *in vivo* and self-splices *in vitro*. Furthermore, we verified the predicted molecular targeting of a sand fly-specific sRNA, ***B****artonella* ***b****acilliformis* **s**mall **R**NA **9** (BbsR9), to transcripts of the *ftsH*, *nuoF*, and *gcvT* genes, *in vitro*.

Acknowledgments

I would like to first and foremost thank my amazing wife, Jenny Wachter, for her incredible support throughout the last five years of our lives. We fell in love over gonorrhea so many years ago, and our passion for microbiology continues to bring us together.

For my two beautiful children and native Montanans: Jesse and Elise.

I'd like to thank all of my colleagues who have made this research possible. First and foremost, Mike Minnick, my wonderful PI and mentor. The intellectual freedom you have given me, combined with little pushes when I started slacking, have helped me grow into the scientist I am today. Thank you for the funding throughout my PhD, the countless revisions on my manuscripts and grant/fellowship applications, and for maintaining a constant flow of coffee every single day.

Thank you to Linda Hicks for being the best lab technician and sympathetic ear a guy could have.

Thank you to Indu Warrier, Jim Battisti, Sarah Bidwell, and the Secor and Samuels lab for various support throughout the years, whether through research advice, instrument usage, or TA support.

Thank you to Jill Burke, Zooey Zephyr, and other members of the DBS office for various support and patience throughout this grad school experience.

Thank you to frequent collaborators of the Minnick lab, Dr. Rahul Raghavan and Dr. Matteo Bonazzi. Much of the success in my research is due to them.

Thank you to my Mom and Dad, my Stepmom Penny, Stepdad TJ, my cousins Vince, RJ, and Nate, my brother Bill, and good friend Dan, who have served as constant supports and sources of entertainment throughout this entire Montana adventure.

Thank you to the Simpson family for being the background noise to almost every study / writing session since 2006.

Big thanks to my car for surviving 100 mile round trips 5 days a week for 4 years, and thank you Stephen King, Brandon Sanderson, George R. R. Martin, Patrick Rothfuss, and Robert Jordan for the entertainment on all those trips.

Thank you to the Montana Academy of Sciences for providing the opportunities that led to receipt of a graduate student grant.

Finally, thank you to the University of Montana and the Department of Biological Sciences for providing the vehicle and setting for this amazing PhD experience. Thank you especially for the various scholarship/fellowship/grant opportunities that have

provided me with funding over the years, including the Linda Philips Knoblock Fellowship, the Toelle-Bekken Memorial Fund Fellowship, the Dr. Mitsuru J. Nakamura Graduate Scholarship, and the Robert J. and Carol Anthony Seim Scholarship.

Table of Contents

List of Figures

Chapter 5 **Page No.**

List of Tables

List of Abbreviations

Abbreviation	Description
sRNA	small RNA
IGR	intergenic region
TSS	transcription start site
RACE	rapid amplification of cDNA ends
SD	Shine-Dalgarno
RBS	ribosome-binding site
UTR	untranslated region
ORF	open reading frame
TCS	two-component system
CCV	*Coxiella*-containing vacuole
T4BSS	type IV-B secretion system
LPS	lipopolysaccharide
LCV	large-cell variant
SCV	small-cell variant
SAM	S-adenosyl methionine
qRT-PCR	quantitative reverse transcription polymerase chain reaction
RNase III	ribonuclease III
IHF	integration host fact
HGT	horizontal gene transfer
TPM	transcripts per million
FDR	false discovery rate
OF	Oroya fever
VP	verruga peruana
HIBB	Bacto heart infusion blood agar
EMSA	electrophoretic mobility shift assay
TE	transposable element
Tn	transposon
IS	insertion sequence
MITE	miniature inverted-repeat transposable element
IVS	intervening sequence
REP	repetitive extragenic palindrome
TIR	terminal inverted-repeat
DR	direct repeat

Chapter 1: Introduction

As published, in parts, in:

BMC Genomics. 2018 Apr 11;19(1):247. doi: 10.1186/s12864-018-4608-y

J Bacteriol. 2019 Oct 21;201(22):e00524-19. doi: 10.1128/JB.00524-19

bioRxiv. 2020 doi: 10.1101/2020.08.04.235903

Bacterial small RNAs as a means of rapid gene regulation

Overview of small RNA functions

Bacterial small RNAs (sRNAs) are small (<500 nts) transcripts that usually do not code for functional proteins. Instead, they serve as *cis*- and/or *trans*-acting regulators through a variety of mechanisms (reviewed in [1]). For example, *cis*-acting sRNAs are often coded antisense to a functional gene target. Upon transcription, the sRNA binds to the mRNA with perfect complementarity, usually culminating in ribonuclease degradation of the target. This effectively limits the free mRNA molecules available for translation (reviewed in [2]). Alternatively, *trans*-acting sRNAs are often coded in distant intergenic regions (IGRs) and bind to a variety of mRNAs through a more limited base-pairing mechanism involving a seed region of around ~7 - 12 nts (**Figure 1.1**). Many *trans*-acting sRNAs have been discovered in bacteria since *Escherichia coli* MicF was first described in 1984 [3]. These regulatory RNAs have been implicated in a variety of processes, including virulence [4], global regulation of transcription [5], iron homeostasis [6], protein degradation [7], and stress response [8, 9].

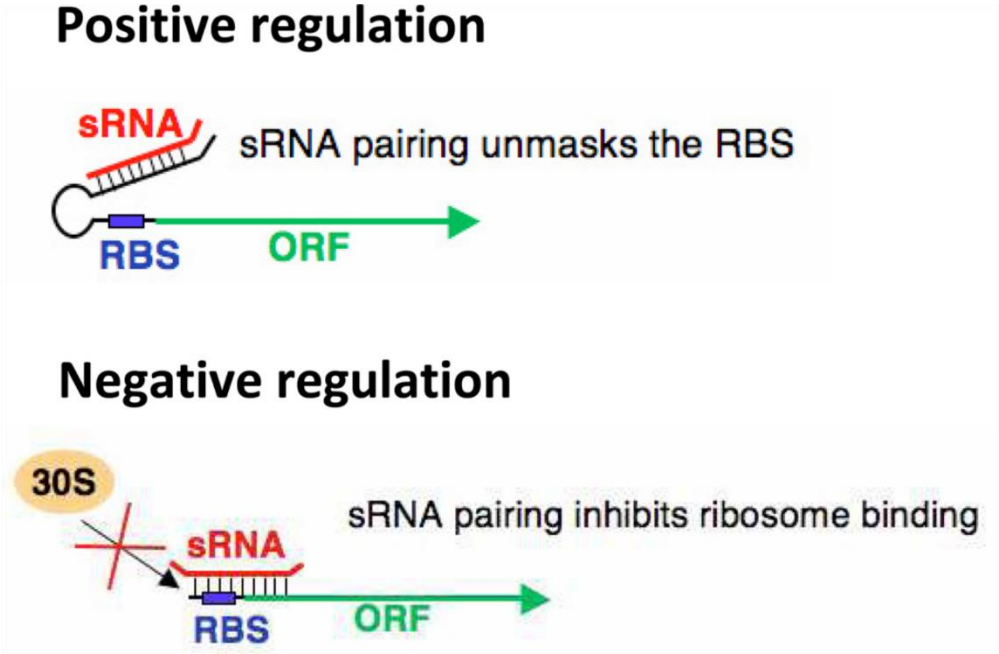


Figure 1.1: Generalized mechanism of sRNA regulation. sRNAs can positively or negatively regulate translation of the mRNAs to which they bind by freeing or creating a ribosome-binding site occlusion, respectively. Adapted from: Vanderpool *et al*. 2011.

Some sRNAs are widely conserved in bacteria. These sRNAs often serve an essential function, the loss of which may result in cell death. For example, 6S RNA is widely found in all bacterial phyla except for *Deinococcus*, *Thermus*, *Thermotogae*, *Tenericutes*, *Elusimicrobia*, and *Fibrobacteres*, although even in these phyla, highly diverged copies of 6S RNA may be found [10]. 6S RNA functions by binding and sequestering RNA polymerase complexed with sigma 70, leading to global inhibition of transcription under certain conditions [11]. 6S RNA was first discovered in *E. coli* in 1967 [12], but it wasn't until more than 30 years later that it was found to regulate RNA polymerase activity during stationary phase [5]. Indeed, an *E. coli* strain lacking 6S RNA was found to be at a survival disadvantage during stationary phase [13].

Another example of a highly conserved sRNA is the tmRNA, a bifunctional sRNA that acts as both a tRNA and mRNA. tmRNA primarily functions by binding to and rescuing stalled ribosomes. Unlike an actual tRNA, though, tmRNA lacks an anticodon so instead functions as an mRNA by coding for a short peptide that is added to the C-terminal end of the nascent polypeptide chain [14]. This peptide targets the incomplete nascent protein for degradation. Like 6S RNA, tmRNA was first discovered and described in *E. coli* [15], although unlike 6S RNA, a mutant strain of *E. coli* unable to express tmRNA does not have a significant growth phenotype [7]. The *in vivo* stability of tmRNA is much like that of a tRNA, leading to a generally high abundance of the sRNA in bacterial cells [14]. High expression or conservation of a sRNA, then, is not necessarily linked to necessity. This stresses the need for functional characterization of sRNAs based on factors other than expression and conservation. Indeed, an important additional characteristic of a sRNA may be its reliance on RNA chaperones for function.

The Hfq chaperone

Typically, *trans*-acting sRNAs require assistance in "finding" their respective mRNA targets. In most bacteria, this is accomplished by the RNA chaperone Hfq, which binds to both sRNAs and mRNAs and plays the role of a molecular matchmaker (reviewed in [16]). Hfq is essential for the function of many *trans*-acting sRNAs that rely on limited base pairing to regulate their target mRNAs [17]. Hfq was first identified more than 50 years ago as a **h**ost **f**actor required for replication of the **Q**β bacteriophage in *E. coli* (reviewed in [18]). Since then, *hfq* homologs have been discovered in many bacteria, and mutants of these *hfq* genes expectedly cause pleiotropic phenotypes specific to the

repertoire of sRNAs they are associated with (reviewed in [19]). Hfq binds to its target sRNAs and mRNAs via its proximal face, distal face, rim, and C-terminal regions, which are solvent-exposed motifs with unique architectures able to bind RNA molecules [19]. For example, the proximal face of Hfq binds to poly-U stretches immediately following a hairpin loop structure, such as those found in Rho-independent terminators [20]. All known Hfq-binding sRNAs have been found to bind Hfq via this structure [20]. Meanwhile, the distal face of Hfq binds to both sRNAs and mRNAs via a sequence of ribonucleotides that varies between bacteria. For example, the *E. coli* Hfq distal face binds to $(A\text{-}A\text{-}N)_n$ repeat motifs, while *Staphylococcus aureus* Hfq binds to $(A\text{-}L)_n$ repeat motifs, where L is a linker ribonucleotide [21]. The rim region of *E. coli* Hfq is a secondary binding site for UA-rich regions in sRNAs and mRNAs [22], while the C-terminal disordered region of Hfq seems to function as a stabilizing force in the binding of some sRNAs [23]. Generally speaking, Hfq-binding sRNAs can be divided into two classes: Class I sRNAs bind to the proximal and rim domains of Hfq and target mRNAs that bind the distal face, and Class II sRNAs bind the proximal and distal domains of Hfq and target mRNAs with rim domain binding sites (**Figure 1.2**) [16].

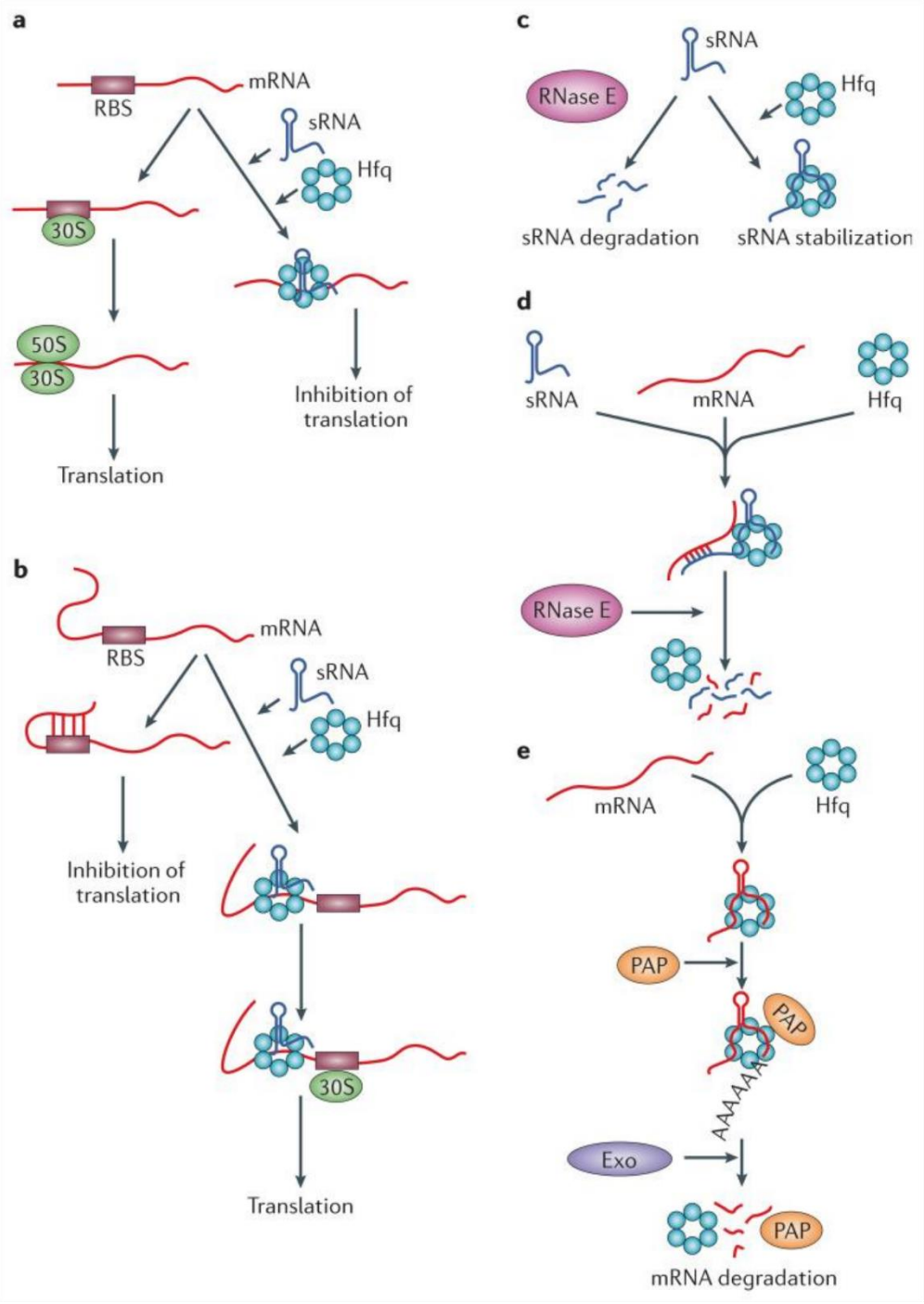


Figure 1.2: Mechanisms of Hfq binding and gene regulation. Negative and positive regulation via sRNA-Hfq complexes is shown in (a) and (b), respectively. Hfq-binding sRNAs can lead to stabilization (c) or degradation (d) of the sRNA. Hfq-sRNA

complexes targeting mRNAs can also lead to mRNA degradation (e). Adapted from: Vogel J *et al*. 2011.

Hfq is not obligatory, however. For example, *S. aureus* has several sRNAs but does not require Hfq protein for their activities [24]. Similarly, *Coxiella burnetii* does not have a readily apparent *hfq* gene. However, this doesn't rule out the possibility of an atypical Hfq or some other novel RNA chaperone in these bacteria. *Bartonella bacilliformis* strain KC583 encodes a single *hfq* gene, although its function in the bacterium has yet to be elucidated.

CsrA and RsmY/Z

Some sRNAs act by binding to and titrating RNA-binding proteins, effectively sequestering them away from regulatory activities. For instance, *C. burnetii* codes for two homologs (CsrA-1, CsrA-2) of the RNA-binding protein CsrA (also referred to as **r**epressor of **s**tationary phase **m**etabolites, RsmA), which has been shown to regulate metabolism, biofilm formation, and Type IV secretion in other bacteria [25-27]. CsrA functions as a homo-dimer where each monomer binds to a Shine-Dalgarno (SD)-like motif (AGGA or ANGGA), leading to inhibition and, in some cases, upregulation, of translation (**Figure 1.3**) [28]. CsrA is regulated by CsrA-binding sRNAs, termed CsrB/C (also called RsmY/Z). Classical CsrB/C sRNAs consist of a series of stem-loops containing exposed AGGA or ANGGA motifs that bind and sequester CsrA, effectively limiting its mRNA regulatory capabilities [29]. Some RsmY/Z sRNAs, however, differ in the number of stem-loop regions containing CsrA-binding sites, and can harbor far fewer

motifs than the classical CsrB/C *E. coli* counterparts [30, 31]. The CsrA regulatory cascade has not been studied in *C. burnetii*, in large part due to the absence of readily-discernible RsmY/Z sRNAs, although the CsrA regulon in *Legionella pneumophila*, a close relative of *C. burnetii*, has been extensively studied [32, 33].

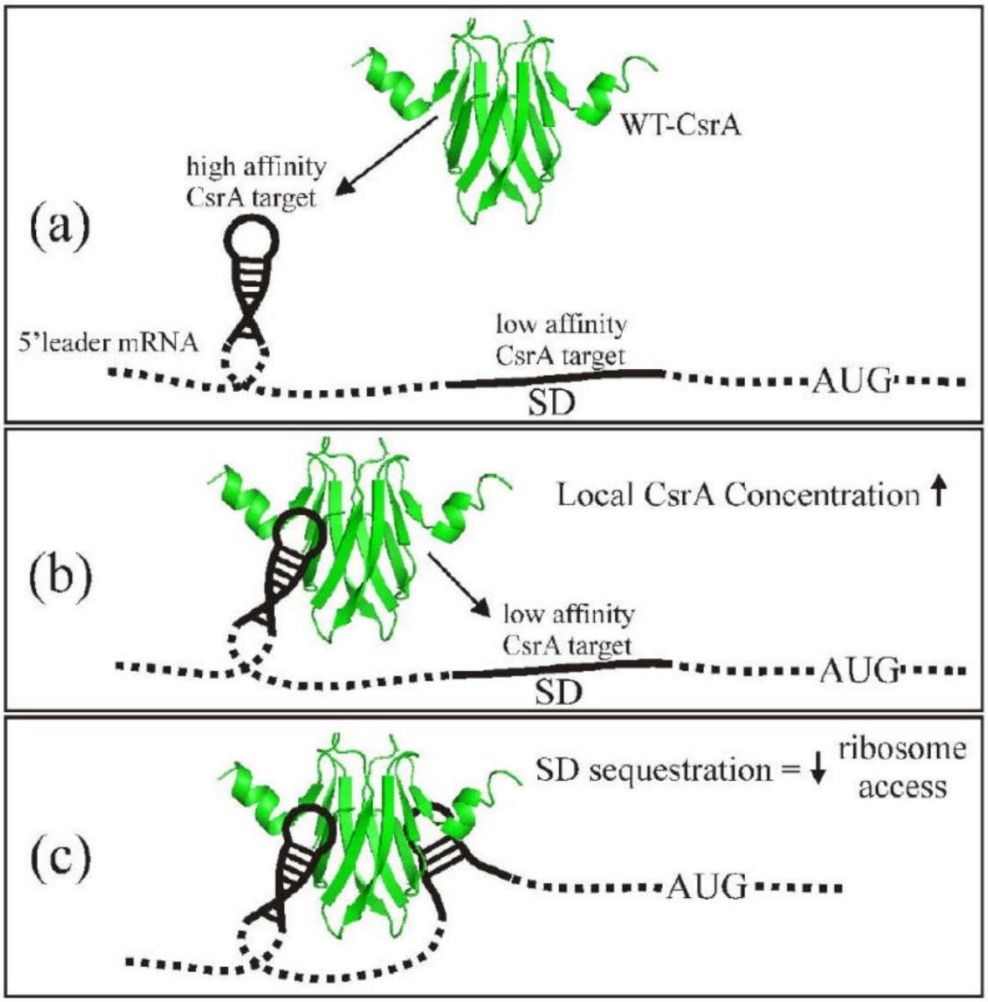


Figure 1.3: Generic mechanism of CsrA-mediated negative regulation. CsrA targets SD-like motifs of stem-loop structures. CsrA first binds to a high affinity target upstream of a transcript's ribosome-binding site (a). Then, the proximity of CsrA to the ribosome-binding site (b) leads to SD sequestration (c), preventing ribosome binding and inhibiting translation. Adapted from: Mercante J *et al*. 2009.

The regulatory cascade leading to CsrA production varies between bacteria. This is due in some part to CsrA being involved in adaptation to host infection conditions, which differs between pathogenic bacteria depending on the niche they inhabit. For example, *L. pneumophila* occupies an intracellular niche in which it is essential for the bacterium to avoid lysosomal degradation. CsrA is a repressor of *L. pneumophila* transmission phenotypes and an activator of intracellular replication [32]. Thus, a successful infection relies on production of CsrA within the intracellular niche, although tight regulation is required for the activation of transmission phenotypes during the later stages of infection. In *L. pneumophila* and many other bacteria, CsrA is regulated by a two-component system (TCS) referred to as LetA/S, where LetS represents a sensor histidine kinase that senses some environmental stimuli that marks the necessity for transmission phenotypes. Concurrently, the RpoS sigma factor is also produced during *L. pneumophila* stationary phase. RpoS, along with the aid of the LetA response regulator, transcribes the RsmY/Z sRNAs, which bind to and sequester CsrA, allowing for the activation of transmission phenotypes and the production of effectors necessary for survival within the "next" intracellular niche (**Figure 1.4**) [27]. Like Hfq, though, CsrA is not ubiquitious among pathogens. *B. bacilliformis*, for example, does not code for any known CsrA homologs.

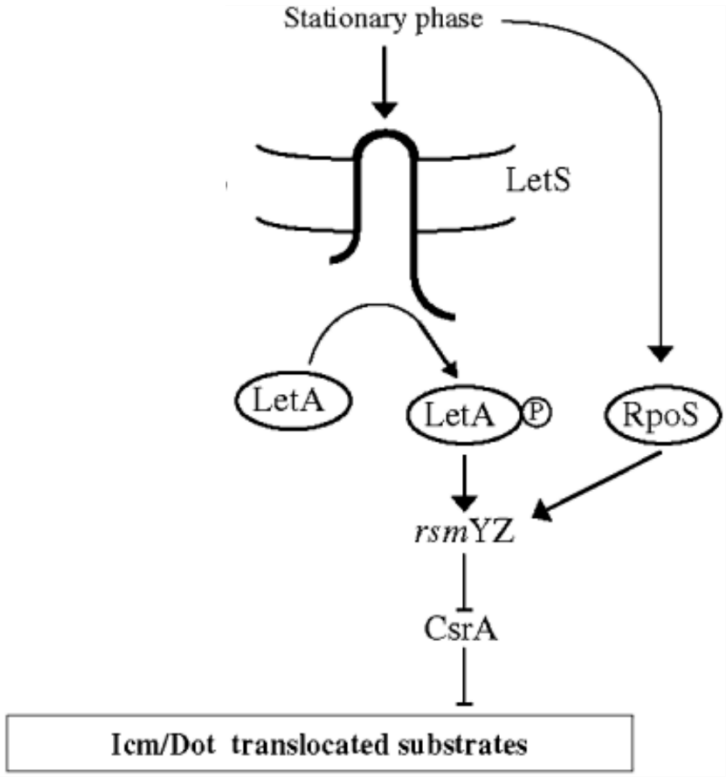
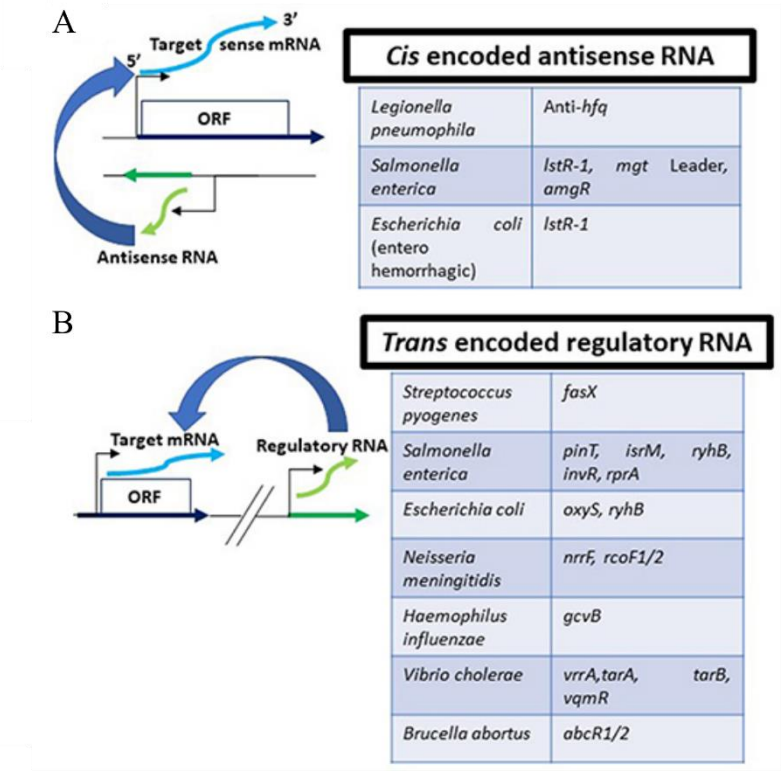


Figure 1.4: The *L. pneumophila* CsrA regulatory cascade. LetA/S and the RpoS sigma factor are indirect regulators of CsrA via the RsmY/Z sRNAs. Adapted from: Rasis M *et al*. 2009.

sRNA-dependent regulation of virulence

sRNAs have been implicated in the regulation of most bacterial processes, from transcription (6S RNA) and housekeeping (tmRNA), to the translation of specific gene subsets (RsmY/Z). Importantly, these sRNAs (6S RNA, tmRNA, and RsmY/Z) all perform their regulatory functions indirectly via protein binding. Most sRNAs function by directly binding mRNAs in *cis* or in *trans*, with or without the assistance of Hfq (**Figure 1.5**) [34]. For example, one of the most extensively studied *trans*-acting sRNAs is RNAIII of *S. aureus*. Firstly, while it is very uncommon for *trans*-acting sRNAs to

contain an open reading frame (ORF), RNAIII encodes the δ-hemolysin at its 5' end that aids in virulence [35]. Furthermore, RNAIII targets multiple mRNAs *in trans*, leading to their up- or down-regulation. Among the down-regulated virulence factors are coagulase (*coa* gene) [36] and peptidoglycan hydrolase (*lytM* gene) [37], while the α-hemolysin (*hla* gene) is up-regulated [38]. This regulatory scheme facilitates *S. aureus* dissemination. Another example is the PapR sRNA of uropathogenic *E. coli* strains, which binds to the coding sequence of *papI* mRNA, thereby causing translational repression [39]. The product of *papI* is itself involved in the activation of P-fimbriae biosynthesis. P-fimbriae is an essential virulence factor involved in the attachment of uropathogenic *E. coli* to renal tissue [40]. Regulation by the PapR sRNA, then, prevents activation of P-fimbriae synthesis, when required, during infection.



***Cis* encoded antisense RNA**

Legionella pneumophila	Anti-*hfq*
Salmonella enterica	*IstR-1, mgt* Leader, *amgR*
Escherichia coli (entero hemorrhagic)	*IstR-1*

***Trans* encoded regulatory RNA**

Streptococcus pyogenes	*fasX*
Salmonella enterica	*pinT, isrM, ryhB, invR, rprA*
Escherichia coli	*oxyS, ryhB*
Neisseria meningitidis	*nrrF, rcoF1/2*
Haemophilus influenzae	*gcvB*
Vibrio cholerae	*vrrA,tarA, tarB, vqmR*
Brucella abortus	*abcR1/2*

Figure 1.5: Mechanisms and examples of sRNAs affecting virulence. Numerous examples are given of *cis*-encoded (A) and *trans*-encoded (B) sRNAs that affect the virulence of pathogenic bacteria. Adapted from: Chakravarty and Massé. 2019.

Means of identifying sRNA targets

Perhaps the most challenging step in determining the roles of sRNAs in virulence is identifying the mRNA targets to which a particular sRNA binds. As an initial step, it is common to employ algorithms that can scan a given genome for mRNAs that may contain seed regions able to be bound by a particular sRNA. Although several of these algorithms exist, each prioritizes a certain metric, and thus, some tend to perform better than others when comparing the predictions to known *in vivo* sRNA targets [41]. For example, the TargetRNA2 algorithm prioritizes conservation of the sRNA, followed by the accessibility of the sRNA seed region, the accessibility of the mRNA target region, and the energy of hybridization [42]. As a result, the algorithm is best suited for sRNAs with known homology to other sRNAs. Meanwhile, the IntaRNA 2.0 algorithm prioritizes the energy of hybridization, followed by accessibility of the sRNA and mRNA as determined by RNA secondary structure prediction. Conservation of the sRNA/mRNA, though, is not taken into consideration [43]. Meanwhile, the CopraRNA algorithm rigorously prioritizes conservation of the sRNA and mRNA, while its methods for computing free energy and sRNA/mRNA accessibility are more outdated [44]. It is helpful, then, to have some *a priori* knowledge of the sRNA of interest when considering the results of these algorithms. For example, the expression pattern of the sRNA in various stages of growth, in axenic culture vs. *in vivo* conditions, etc., are all valuable in

determining which *in vitro* targets "make sense". That said, these *in silico* predictions need to be confirmed through other methods such as *in vitro* binding assays and/or *in vivo* cross-linking experiments.

RNA-RNA electrophoretic mobility shift assays (EMSAs) are valuable for the determination of specific sRNA-mRNA interactions. When combined with *in vitro* mutagenesis of predicted seed regions, one can even identify the specific RNA bases mediating these interactions [45, 46]. There are also several methods for the *in vivo* cross-linking of sRNAs to their targets. These sRNA-mRNA complexes are usually captured and sequenced in order to determine the biologically relevant repertoire of mRNA targets for any given sRNA. For example, **c**ross-linking, **l**igation, **a**nd **s**equencing of **h**ybrids (CLASH) is predicated on the affinity purification of UV cross-linked RNA-protein complexes [47]. Some of these RNA-protein complexes will contain multiple RNAs that are then ligated to each other before RNA-sequencing [47]. For example, CLASH has been used to determine the repertoire of *E. coli* sRNA-mRNA interactions mediated by Hfq during nutrient-limiting conditions [48].

Bacterial selfish genetic elements

Overview of selfish genetic elements

A selfish genetic element is a DNA segment that enhances its own transmission and maintenance at the expense of other genes in a given genome. This is regardless of whether such an element would enhance the fitness of an organism or not [49]. The concept of a selfish genetic element was first noted in 1928, when it was found that the female X chromosome in the fruit fly, *Drosophila obscura*, seemed to be over-abundant

in fruit fly populations, indicating an uneven female:male sex ratio that could lead to extinction of the species [50]. In 1950, Barbara McClintock discovered the existence of transposable elements (TEs) in maize [51]. This represented the first characterization of a selfish genetic element. Despite this, it wasn't until 1980 that Leslie Orgel and Francis Crick popularized the concept of selfish DNA, showing that these elements spread in a population regardless of their effect on fitness [52].

In the context of bacteria, selfish genetic elements constitute anywhere from 0-21% of a genome and seems to vary with ecology rather than phylogeny [53]. Notably, bacteria undergoing reductive evolution harbor larger amounts of selfish genetic elements when compared to other bacteria [54]. The impetus for reductive evolution seems to be the transition of an organism from free-living to obligate parasite [54]. As this transition occurs and some genes become unnecessary due to the nutrient replete environment of the host, large quantities of DNA may be lost. This leads to an increase in the number of pseudogenes (pseudogenization) and the multiplication and spread of selfish genetic elements, in part due to successive bottleneck events [55]. As an obligate intracellular pathogen, *C. burnetii* represents a bacterium undergoing reductive evolution. As such, it contains a high number of pseudogenes (10.1% of all ORFs) and a variety of different selfish genetic elements [56].

Transposons, insertion sequences, MITEs, and Group I introns

Bacterial TEs include transposons (Tns), insertion sequences (IS), miniature inverted-repeat transposable elements (MITEs), and group I introns, among others [57]. TEs are considered intracellular mobile genetic elements in that they rely on other mobile genetic elements for inter-cellular spread [57]. While Tns, IS, MITEs, and group I introns all

have similarities in function, there are some differences that distinguish one from another. Tns represent a broad term for TEs and can be divided into two major categories: Class I (retrotransposons) and Class II (DNA transposons). Retrotransposons are most often found in eukaryotes and so won't be discussed. Class II Tns can be found in both eukaryotes and prokaryotes [57]. Furthermore, each class can be divided into autonomous Tns, which are able to catalyze their own transfer, and non-autonomous Tns, which require some other element for transposition [57]. Broadly, Tns represent large DNA segments that contain a gene coding for a transposase enzyme. That gene is flanked by terminal inverted-repeat (TIR) sequences that aid in targeting the Tn to specific DNA sequences in a cut-and-paste mechanism that is facilitated by the transposase enzyme [57].

Bacterial Tns usually encode other genes that may provide some adaptive benefit for an organism, such as an antibiotic resistance gene [58]. When no gene other than the transposase gene is present, a Tn is referred to as an IS. As a result, IS elements are smaller than Tns. IS elements are widely distributed in bacteria and may have large effects on genome evolution [57]. For example, the phenomenon of IS expansion is seen in bacteria that have recently adopted a host-restricted lifestyle [59]. IS expansion is thought to be an early consequence of host adaptation brought on by the sudden enhancement of genetic drift due to successive population bottlenecks in a host nutrient replete environment [60]. Ultimately, as a bacterium becomes adapted to the host, deletion of IS elements and adjacent DNA leads to genome reduction. As IS elements become non-autonomous due to loss of transposase function, these elements are eventually lost as the bacterial genome becomes more streamlined [57].

MITEs are non-autonomous class II TEs with defective or missing transposase genes. As such, they can only be mobilized in *trans* by transposases expressed from related Tns [61]. Most bacterial MITEs consist of 4-30 bp TIRs with a TA dinucleotide at their termini. MITEs are typically small (100-400 bp) and do not encode proteins; rather, their transcripts generate highly stable stem-loop structures [62]. MITE insertions have been implicated in virulence by fostering a plastic genome that enhances acquisition of virulence traits [63] and through physical insertions that alter ORFs and directly lead to virulence phenotypes [64]. Promoter regions and ORFs are common features of bacterial MITEs [65-68]. Moreover, integration host factor (IHF)-binding sites and methyltransferse binding domains have been reported [66, 69]. While most MITEs integrate into IGRs, some have been reported: a) in structural RNA genes [70], b) in protein-encoding genes to create in-frame protein fusions [71], and c) proximal to genes whose transcripts are regulated by the corresponding MITE RNA [72, 73]. Thus, MITEs can potentially interact at DNA, RNA or protein levels in a host bacterium, depending upon their structure and genomic sites of integration.

Group I introns are considered to be autonomous class II TEs that insert into tRNAs, rRNAs, and protein-coding genes. Generally speaking, group I introns are ribozymes that, upon transcription, catalyze their own splicing. Although this RNA splicing is auto-catalytic, they sometimes require protein co-factors for self-splicing *in vitro*, and it is presumed that all group I introns require protein co-factors to some extent for splicing *in vivo* [74]. Furthermore, some group I introns harbor a gene coding for a homing endonuclease, which is thought to further facilitate their transposition [75].

Coxiella burnetii is a zoonotic, obligate intracellular human pathogen

Overview of pathogenicity

C. burnetii is a Gram-negative, obligate intracellular bacterium and etiological agent of Q (query) fever in humans. Q fever most often manifests as an acute, flu-like illness, which in rare cases progresses to potentially life-threatening endocarditis [76]. *C. burnetii* undergoes a biphasic life cycle in which it alternates between a metabolically-active, replicative large-cell variant (LCV) and a non-replicative, spore-like small-cell variant (SCV) [77]. Upon aerosol transmission of SCVs to a mammalian host, *C. burnetii* is primarily endocytosed by alveolar macrophages, after which it survives acidification of the host phagolysosome and metamorphoses to LCVs. *C. burnetii* then utilizes the fusion of its *Coxiella*-containing vacuole (CCV) with lysosomes and autophagosomes in order to expand the intracellular niche (**Figure 1.6**) [78, 79]. CCV expansion is dependent on *C. burnetii* protein synthesis, but independent of replication, so expansion of the CCV is facilitated by a repertoire of Dot/Icm effector proteins secreted by a Type IV-B secretion system (T4BSS) [80, 81]. Many Dot/Icm substrates have been identified in recent years [82] and shown to modulate the host inflammasome [83], influence autophagosomal/lysosomal fusion with the CCV by various mechanisms [84-88], and regulate the host transcriptome after localizing to the nucleus [89, 90]. Little is known about regulation of *C. burnetii*'s T4BSS, although the PmrA response regulator has been shown to enhance synthesis of the T4BSS apparatus as well as certain Dot/Icm substrates [91].

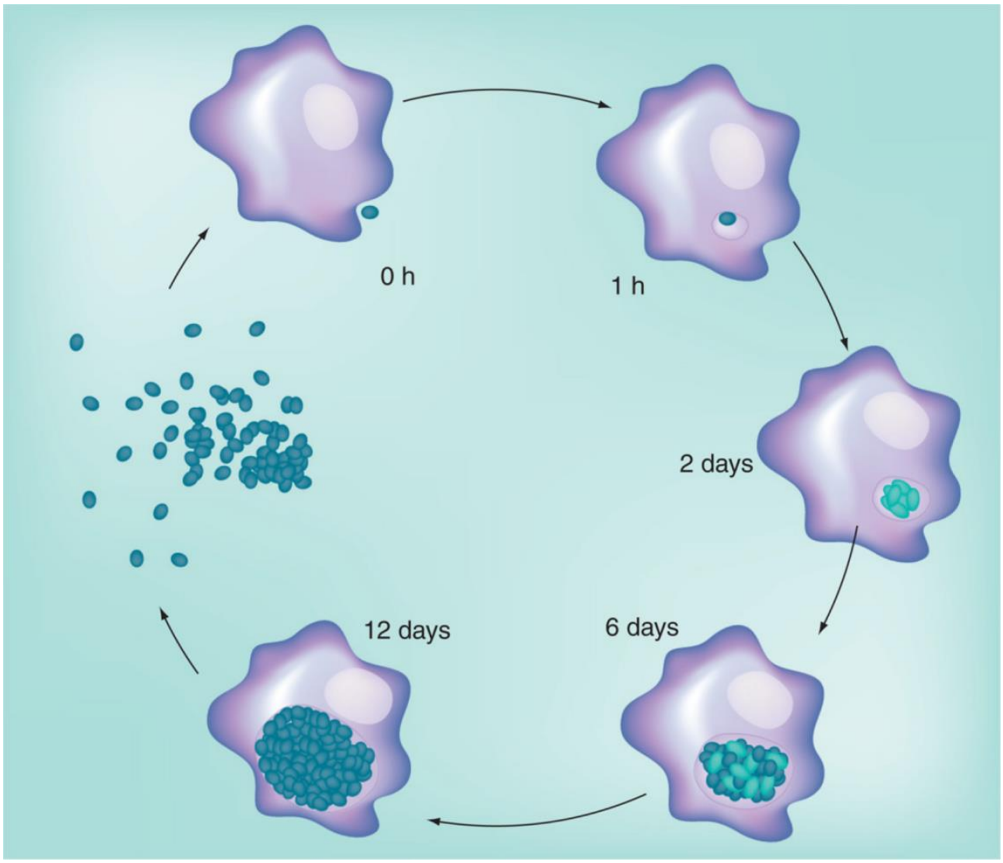


Figure 1.6: Developmental cycle of *C. burnetii*. The developmental cycle of *C. burnetii* in alveolar macrophages is shown. The LCV (light blue) and SCV (dark blue) morphotypes are indicated as the infection proceeds. Adapted from: Minnick and Raghavan, 2011.

The only confirmed virulence factor in *C. burnetii* is lipopolysaccharide (LPS). *C. burnetii* LPS undergoes a process called phase variation where its length and molecular properties may change, leading to altered virulence [92]. For example, expression of phase I LPS (full-length LPS) leads to virulence in mammal hosts, while expression of phase II LPS (deep rough LPS) renders the LPS immunogenic, leading to clearance [93]. It has also been shown that repeated passages of *C. burnetii* in embryonated hen's eggs

with phase I LPS (~10 passages) leads to a phase II LPS phenotype [94]. The noted difference in immunogenicity is due to the steric hindrance of antibody binding to phase I LPS compared to phase II LPS [95]. Recently, a series of *C. burnetii* genes was implicated in successful synthesis of phase I LPS, and it was shown that an accumulation of mutations in several genes leads to truncation and formation of phase II LPS *in vivo* [96].

C. burnetii small RNAs

A previous study revealed 15 novel C. burnetii sRNAs that were differentially transcribed either in LCVs vs. SCVs, or in host cell infections vs. growth in ACCM-2 medium [97, 98]. Of special interest were ***C****oxiella* ***b****urnetii* **s**mall **R**NA **1** (CbsR1) and CbsR12, which, along with being upregulated during infection compared to growth in axenic medium, also have predicted Rho-independent terminators and strong predicted promoter elements [97]. **We hypothesize that these highly expressed, "infection-specific" sRNAs play important roles in *C. burnetii* infection of mammalian cell lines.**

C. burnetii selfish genetic elements

C. burnetii's genome suggests that it is a relatively recent obligate intracellular pathogen, based upon the high number of pseudogenes and selfish genetic elements [99]. Among these elements are an intein [100], two group I introns [101], an intervening sequence (IVS) [102], and TEs, including multiple copies of IS1111 [103]. The IS1111 transposon has been studied extensively and found to preferentially insert into a palindromic DNA sequence that is widely distributed throughout the *C. burnetii* genome [104]. This palindromic DNA sequence has been described as a GTAG repetitive extragenic palindrome (REP), although the nature and distribution of the REP has not been

described [105]. There is little information on other families of TEs in the *C. burnetii* genome. **We hypothesize that non-autonomous IS elements such as MITEs also exist within the *C. burnetii* genome.**

***Bartonella bacilliformis* is a vector-borne, facultative intracellular human pathogen**

Overview of pathogenicity and disease

B. bacilliformis is a Gram-negative, facultative intracellular bacterium and the etiological agent of Carrión's disease in humans. Carrión's disease often manifests as a biphasic illness characterized by acute hemolytic anemia followed by eruptions of blood-filled hemangiomas of the skin [106]. Timely antibiotic administration restricts the fatality rate of Carrión's disease to ~10%, although if left untreated, the rate has been reported to be as high as 88% [107, 108]. *B. bacilliformis* is transmitted between humans through the bite of female phlebotomine sand flies, specifically *Lutzomyia* spp. [109, 110]. The endemic region of Carrión's disease has historically been limited to arid, high-altitude valleys (600 – 3200m) in the Andes Mountains of Peru, Colombia, and Ecuador, reflecting the habitat of the sand fly vector [111, 112].

The initial, acute stage of Carrión's disease is referred to as Oroya fever (OF), and it is characterized by colonization of the entire circulatory system, leading to infection of ~61% of all circulating erythrocytes [112, 113]. This bacterial burden typically leads to severe anemia, fever, jaundice, and hepatomegaly, among other symptoms [114]. Weeks or months following OF, *B. bacilliformis* seemingly invades endothelial cells, where it triggers cell proliferation and angiogenesis. This event leads to formation of hemangiomas of the skin, referred to as verruga peruana (VP) (**Figure 1.7**). The VP stage

is chronic and lasts about one month to a year [106, 111]. Although Carrión's disease can present as a severe illness, there are many documented cases with relatively milder symptoms and/or the onset of VP without having presented with OF symptoms [115]. In consideration of reports involving less virulent *B. bacilliformis* strains and the possibility that other *Bartonella* spp. can cause mild symptoms resembling Carrión's disease, the incidence of the disease is likely underreported [116-118].

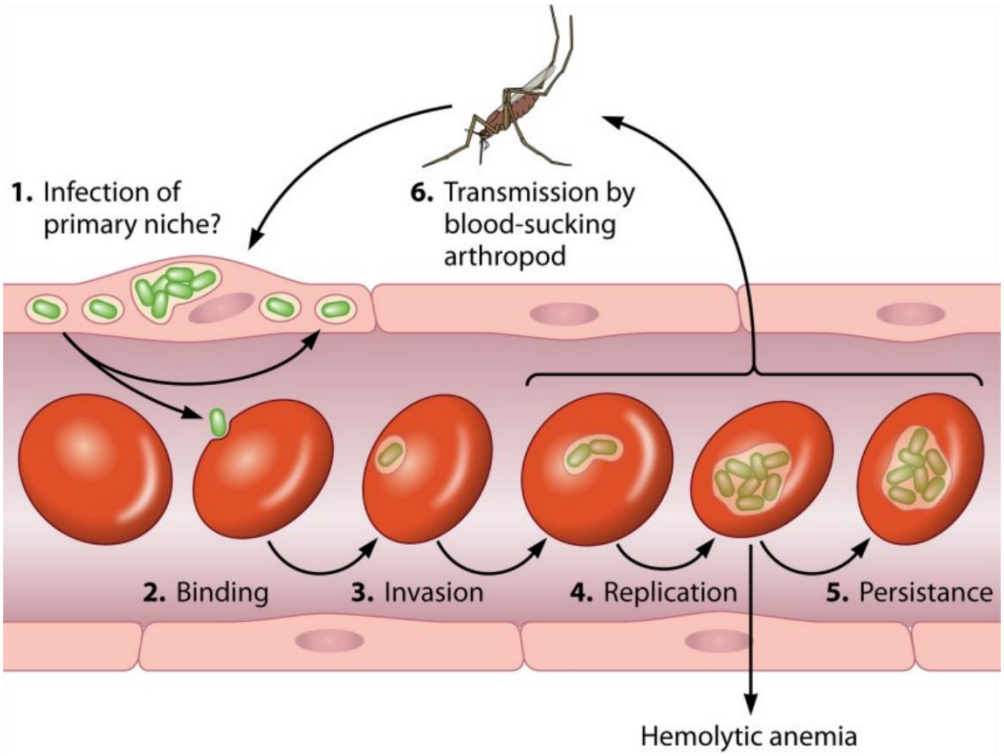


Figure 1.7: Model of an acute *B. bacilliformis* infection. Upon transmission by a phlebotomine sand fly (not depicted), *B. bacilliformis* may first invade endothelial cells, followed by dissemination to the blood stream and infection of erythrocytes, leading to cell death and hemolytic anemia. Subsequently, endothelial cells may provide a secondary niche during the chronic VP stage of infection. Adapted from: Gomes and Ruiz, 2017.

Infection cycle and virulence factors

The *B. bacilliformis* infection cycle is strikingly under-studied compared to other vector-borne pathogens. It is clear that the bacterium is transmitted by female *L. verrucarum* sand flies, although artificial feeding experiments showed that *L. longipalpis* can also "vector" the pathogen in the laboratory [110]. These studies also revealed that *B. bacilliformis* colonized and persisted in the lumen of the abdominal midgut of *L. verrucarum* but were digested along with the blood meal in *L. longipalpis* [110]. Despite this, viable bacteria were retrieved from both insects following a 7-d colonization period [110]. Other *Lutzomyia* spp. have been found to contain *B. bacilliformis* DNA, but colonization experiments have not been performed [119]. It has also been suggested that other mammals may serve as reservoir hosts for *B. bacilliformis*. However, serosurveys of animals that came into contact with infected humans were negative for *B. bacilliformis* DNA [120]. Interestingly, in various attempts to establish an animal model of *B. bacilliformis* infection, the bacterium was only able to infect rhesus macaques [121] and owl monkeys [122]. These results suggest that other primates could conceivably serve as natural reservoir hosts for *B. bacilliformis*, although there is a paucity of non-human primate species in *L. verrucarum*'s geographic range. Regardless, the lack of a small animal model severely limits the prospects of laboratory studies examining *B. bacilliformis* infections *in vivo*.

A number of virulence attributes are involved in *B. bacilliformis* pathogenesis, including erythrocyte attachment [123], invasion [124-126] and hemolysis [127]. Similarly, several factors have been implicated in endothelial cell invasion [128] and proliferation [129-131]. However, regulatory mechanisms that facilitate the pathogen's

virulence, colonization, and persistence in the sand fly have not been explored, to date. The disparate environments encountered by *B. bacilliformis* during transmission from sand fly vector to human host, and back again, suggest that genetic regulatory mechanisms are used to rapidly adapt to prevailing conditions. For example, the temperature of the sand fly vector would be comparable to ambient temperatures in the geographical range of the insect. The competent vector, *L. verrucarum*, is endemic to high-elevation ranges of the Occidental and Inter-Andean valleys of Peru, Colombia, and Ecuador [132], where temperatures range from 17^0 C - 22^0 C; fairly consistent with laboratory "room temperature" [133]. Upon transmission to the human host, the bacterium would need to adjust to a human body temperature of ~37^0 C. Similarly, human blood has a pH of ~7.4, while the pH of the sand fly (*L. longipalpis*) abdominal midgut after a blood meal is ~8.2, lowers to ~7.7 as the blood meal is digested, and decreases to ~6.0 after digestion [134, 135]. In contrast, the thoracic midgut is maintained at pH ~6.0, regardless of digestion status [134]. A rapid means of regulating virulence and stress-related factors to counteract sudden shifts in temperature and pH would be clearly adaptive for *B. bacilliformis*. **We hypothesize that sRNAs play a role in the rapid gene regulation necessary for *B. bacilliformis* adaptation to different aspects of its infection cycle.**

Chapter 2

Identification of novel MITEs (miniature inverted-repeat transposable elements) in *Coxiella burnetii*: implications for protein and small RNA evolution

As published in: BMC Genomics. 2018 Apr 11;19(1):247. doi: 10.1186/s12864-018-4608-y

Abstract

Coxiella burnetii is a Gram-negative gammaproteobacterium and zoonotic agent of Q fever. *C. burnetii*'s genome contains an abundance of pseudogenes and numerous selfish genetic elements. MITEs are non-autonomous Tns that occur in all domains of life and are thought to be ISs that have lost their transposase function. Like most TEs, MITEs are thought to play an active role in evolution by altering gene function and expression through insertion and deletion activities. However, information regarding bacterial MITEs is limited. Here, we describe two MITE families discovered during research on small non-coding RNAs (sRNAs) of *C. burnetii*. Two sRNAs, Cbsr3 and Cbsr13, were found to originate from a novel MITE family, termed QMITE1. Another sRNA, CbsR16, was found to originate from a separate and novel MITE family, termed QMITE2. Members of each family occur ~50 times within the strains evaluated. QMITE1 is a typical MITE of 300-400 bp with short (2-3nt) direct repeats (DRs) of variable sequence and is often found overlapping annotated ORFs. Additionally, QMITE1 elements possess sigma-70 promoters and are transcriptionally active at several loci, potentially influencing expression of nearby genes. QMITE2 is smaller (150-190 bps), but has longer

(7-11nt) DRs of variable sequences and is mainly found in the 3' untranslated region (UTR) of annotated ORFs and IGRs. QMITE2 contains a GTAG REP that serves as a target for IS1111 TE insertion. Both QMITE1 and QMITE2 display inter-strain linkage and sequence conservation, suggesting that they are adaptive and existed before divergence of *C. burnetii* strains.

Introduction

C. burnetii is a Gram-negative, obligate intracellular gammaproteobacterium and the etiologic agent of Q fever in humans. Q fever is an acute, flu-like illness that can present with pneumonitis, hepatitis and malaise. In less than 5% of cases, chronic infection can develop with potentially life-threatening endocarditis as the most common manifestation [76]. *C. burnetii* undergoes a biphasic life cycle in which it alternates between a metabolically-active, replicative large-cell variant (LCV) and a dormant, spore-like small-cell variant (SCV) [136]. Upon inhalation of SCV's by a mammalian host, alveolar macrophages internalize the bacteria and trap them within a highly acidic (pH ~4.5) parasitophorous vacuole that has features of a mature phagolysosome [78]. *C. burnetii* has adapted to survive in this acidic environment, where it forms a replicative niche. Dot/Icm effectors are translocated to the host cell in a T4BSS-dependent manner in order to establish and maintain the vacuole [81]. LPS is another critical virulence determinant in *C. burnetii* [137], although it has been found to be truncated (rough) in some strains, including the Nine Mile phase II laboratory strain, RSA 439 [138]. Interestingly, the Dugway 5J108-111 strain has a full-length LPS, but is avirulent [139]. Dugway is considered to be the most primitive of the sequenced *C. burnetii* strains based on a larger

genome with apparently less reductive evolution than virulent strains, such as RSA 493 [56]. It is hypothesized that Dugway either contains a gene(s) that impedes infection in humans, or that the virulent RSA 493 strain has some altered virulence gene(s) rendering it infective [140].

C. burnetii was recently shown to produce at least 15 sRNAs [97]. In this report, we show that ***C****oxiella* ***b****urnetii* **s**RNA **3** (Cbsr3), CbsR13, and a newly defined sRNA, CbsR16, arose from two novel MITE families of the pathogen. Furthermore, we demonstrate how these novel MITE families can serve as a timeline for IS1111 transposition based upon their linkage and sequence conservation between strains. Finally, we show that although MITE copies show linkage and sequence conservation, an indel in a potential virulence-associated gene (*enhC*) affected by QMITE2 has created a truncated version of the gene in the virulent RSA 493 strain as compared to the avirulent Dugway strain.

Materials and Methods

Discontiguous MegaBLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used as a local alignment program using default parameters to identify regions of homology to CbsR13 and CbsR16 in the *C. burnetii* genome (strains RSA 493; GenBank accession number AE016828.3 and Dugway 5J108-111; GenBank accession number CP000733.1). In order to compare the various QMITE loci in the RSA 493 genome, multiple sequence alignments of QMITE copies were performed using MUSCLE alignments via Geneious version 11.0.2 software with the default settings [141] (https://www.geneious.com/download/). Phylogenetic analyses of various groups of

QMITE insertions were carried out by first trimming the MUSCLE alignments utilizing Gblocks version 0.91b software [142] (http://molevol.cmima.csic.es/castresana/Gblocks.html). This served to eliminate poorly aligned and highly divergent regions in the various alignments. The default parameters are exceptionally stringent and are catered towards longer input sequences. Thus, the minimum block length was reduced to four, and gap positions were allowed for half of the input sequences at each aligned position in order to accommodate the relatively shorter input sequences. Phylogenetic trees of these trimmed alignments were then constructed using FastTree version 2.1 [143] (http://www.microbesonline.org/fasttree/#FAQ). The generalized time-reversible model of nucleotide evolution was used and phylogeny was inferred using maximum likelihood. The resulting Newick tree file was visualized using FigTree version 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). To support our designations of QMITEs as MITEs, supplemental MITE predictions of the *C. burnetii* RSA 493 genome were performed using MUSTv2 software [144] (http://www.healthinformaticslab.org/supp/resources.php). Predicted RNA secondary structures used to confirm the presence of TIRs were generated using mfold [145] (http://unafold.rna.albany.edu/?q=mfold). In order to demonstrate the potential for transcription of QMITE inserts, prediction of sigma-70 consensus promoter elements and Rho factor-independent terminators in QMITE inserts was performed using BPROM (http://www.softberry.com/berry.phtml?topic=bprom&group =programs&subgroup=gfindb) and ARNold (http://rna.igmors.u-psud.fr/toolbox/arnold/), respectively. CIRCOS software [146] (http://circos.ca/software/ download/ circos/) was

used to visualize and depict positions of QMITEs on the *C. burnetii* chromosome. RNA-Seq data [Sequence Read Archive (SRA) database under accession number SRP041556] were analyzed using a custom pipeline, although various nesoni version 0.128 applications for processing high-throughput sequence data were also used (http://www.vicbioinformatics.com/software.nesoni.shtml). Transcripts per million (TPM) were calculated using custom perl and python scripts that can be accessed through GitHub (https://github.com/shawachter/TPM_Scripts). The Artemis genome browser was used to visualize alignment files generated from ambiguous and unambiguous read data (http://www.sanger.ac.uk/science/tools/artemis) [147]. Other figures were created using Powerpoint 2010 software (Microsoft, Redmond, WA).

Results

CbsR3 and CbsR13 loci are members of a novel MITE family

CbsR13 was originally identified as a *C. burnetii* sRNA by RNA-Seq analysis of the transcriptome [97]. It is often helpful to analyze both ambiguous and unambiguous reads associated with any RNA-Seq data. Ambiguous reads refer to those reads that can't be aligned to one specific area of the genome because multiple copies of that sequence exist in the genome. Unambiguous reads refer to those that could only be mapped to one region of the genome. Upon visualization of ambiguous and unambiguous reads that map to the CbsR13 locus, we discovered that there were many ambiguous reads associated with it (**Figure 2.1A**). We also found that CbsR13 RNA produced a stable predicted secondary structure resembling a very long palindromic sequence (**Figure 2.1B**). Although a megaBLAST search produced several hits of high homology, the divergent

nature of the CbsR13 sequences necessitated use of a discontiguous megaBLAST search, which identified dozens of sequences with significant homology to CbsR13 in the genome. Specifically, the search identified 44 ranges, with E values of 8e-11 to 3e-123. Of these hits, 21 were at least 75% of the length of CbsR13 (>232 bp). It was noted upon alignment of the regions flanking these sequences that some of the ranges contained truncated 5′ ends and elongated 3′ ends. An artificial sequence combining the native CbsR13 sequence and the 3′ extension (see **Figure S2.1**) was thus used as an input for another discontiguous megaBLAST search. This search revealed 45 ranges, with E values from 9e-10 to 5e-123. Twenty-three of these hits were at least 75% of the input sequence length (>350 bp). A multiple alignment and phylogenetic analysis of these 23 sequences is shown in **Figure 2.2A** and **Figure S2.2**, respectively. The remaining 22 elements ranged in size from 39-321 bp (not shown), possibly representing degenerate forms of the original nucleotide sequences. One megaBLAST hit for the extended-CbsR13 corresponded to a large portion of the CbsR3 gene sequence (i.e., nt 481609-481806) (see **Figure 2.2A**, range 2) [97]. This result suggests that the two sRNAs share a common ancestor, although unambiguous TPM values from RNA-Seq show that CbsR13 is expressed at a markedly higher level relative to CbsR3 (**Figure S2.3**). Confirming what is seen in **Figure 2.1A**, the ambiguous TPMs associated with CbsR3 and CbsR13 are much higher than the unambiguous TPMs, indicating that additional CbsR13 loci are transcriptionally active (**Figure S2.3**). Indeed, a sigma-70 promoter search using BPROM predicts a promoter in the forward strand and two promoters in the reverse strand of the input sequence (**Figure S2.1**).

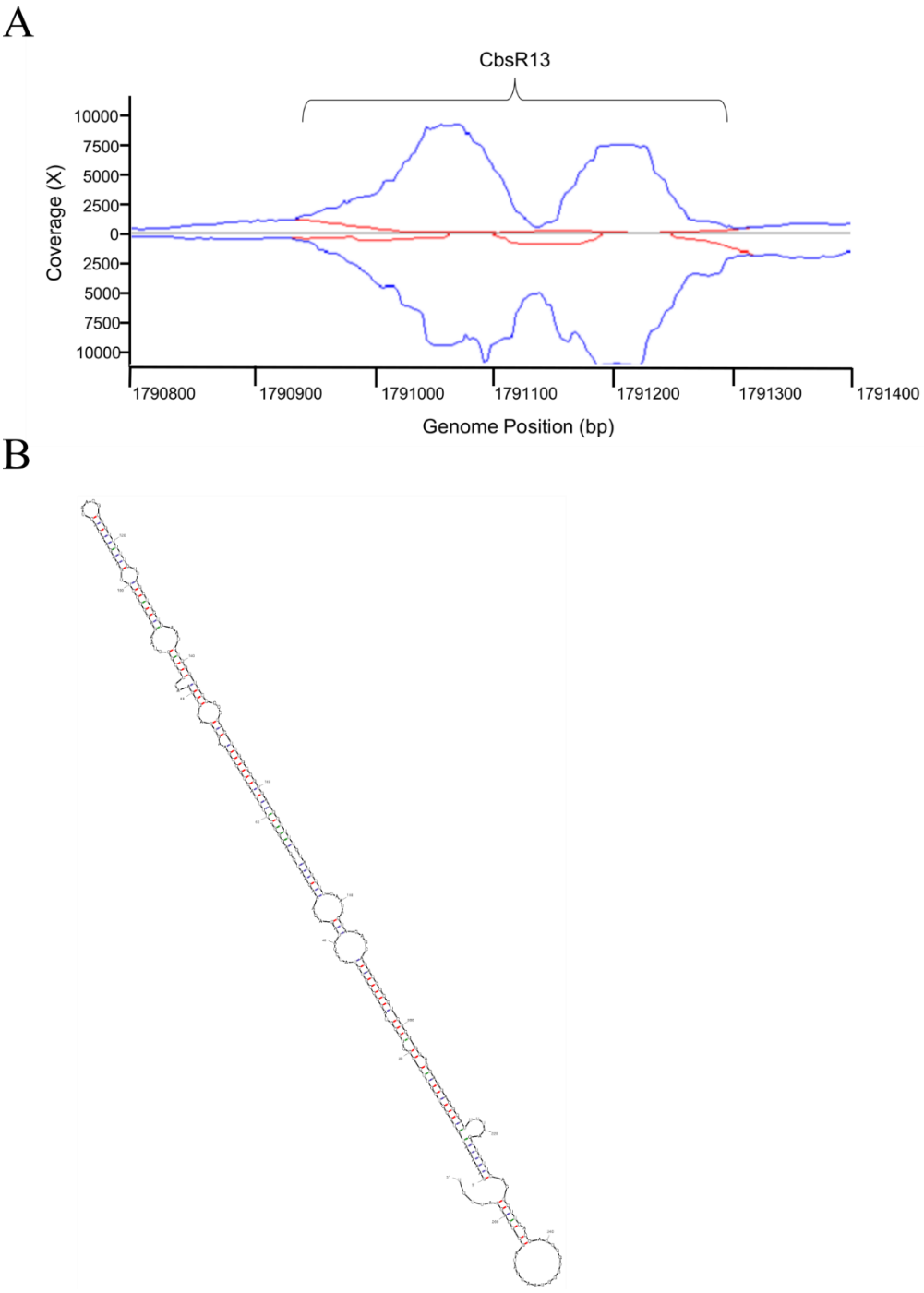


Figure 2.1: Ambiguous and unambiguous reads map to the CbsR13 locus. (**A**) Artemis view of reads mapping to the CbsR13 locus (RSA 439 genome). The x-axis indicates the location (bp) on the chromosome and the y-axis indicates coverage of reads mapping to that location. Reads above the y-axis indicate antisense reads, whereas reads

below the y-axis indicate sense reads mapping to that genomic location. Blue lines signify ambiguous reads mapping to this locus, while the red lines denote unambiguous reads. (**B**) mfold secondary structure prediction of the CbsR13 sRNA. Red, blue, and green lines forming stem structures indicate G-C, A-U, and G-U base-pairing, respectively (predicted ΔG = -128.5 kcal/mol).

A

B

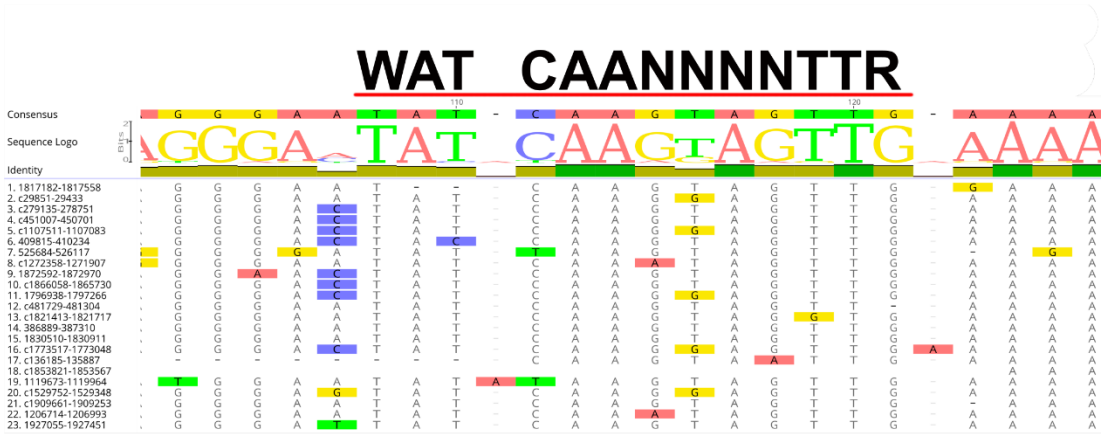


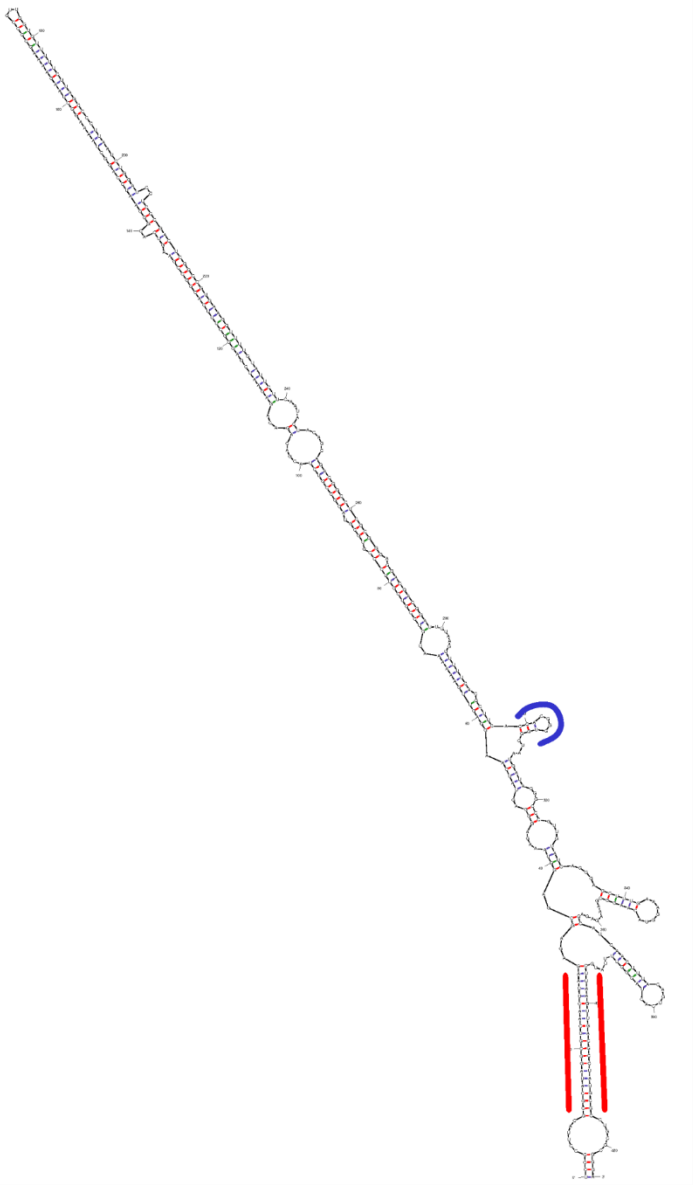
Figure 2.2: CbsR13 loci contain a canonical IHF-binding site. (**A**) MUSCLE sequence alignment of discontiguous megaBLAST hits (>75% of input sequence) associated with the extended-CbsR13 input sequence. Conserved bases appear as gray blocks, while unaligned bases appear as green, yellow, blue, and red bands, corresponding to T, G, C, and A bases, respectively. An identity indicator is shown above the sequence alignment,

where height signifies conservation of bases at that position, with a color indicator for overall identity between aligned ranges (green: 100%, yellow: 20-99%, red: 0-19%). The consensus sequence is shown above the identity indicator as colored bands indicating bases as described above. (**B**) The same alignment as shown in (A), focusing on the potential IHF-binding site. The sequence above the red line indicates the consensus IHF-binding site utilizing nucleotide notation, and above the alignment is a sequence logo where the height of the displayed bases indicates the relative identity of the aligned base at that position.

A common motif associated with bacterial TEs is an IHF-binding site [67]. IHF is a bacterial DNA-binding protein that binds to a specific DNA motif and facilitates bending of the DNA. It is thought that this bending aids in transposition of the locus [148]. The consensus IHF-binding nucleotide sequence is WATCAANNNNTTR [149]. Although IHF-binding sites are common in bacterial TEs, they are not always present in MITEs [67]. A manual search through the aligned ranges in **Figure 2.2A**, though, led to the discovery of a well-conserved IHF-binding site (**Figure 2.2B**). We chose Range 5 (**Figure 2.2A**) as a representative for this repeated sequence due to its completeness, and utilized mfold to visualize where this IHF-binding site was located and to see if the sequence had a TIR that could aid in the element's categorization as a MITE. As shown in **Figure 2.3A**, it is clear that the element has a TIR of 21 bp in length. Based on the length of the element (~400 bp), the TIR, and the multiple loci scattered throughout the *C. burnetii* RSA 493 genome, we conclude that this element is a *bona fide* MITE. Moreover, no similar MITEs have been previously described, and BLASTn searches

found no orthologues in other genomes. Thus, we can conclude that this is a novel MITE that we designate as QMITE1. Other ranges in **Figure 2.2A** generated similar predicted secondary structures, with corresponding TIRs ranging from 21-28 nts (not shown). MUSTv2 software was also employed to confirm QMITE1 as a MITE (**Figure S2.4**) [144]. Using stringent parameters, MUSTv2 identified eight of the top ten most homologous ranges to the extended-CbsR13 input sequence and also identified 2-4 bp DRs of nucleotide compositions WW, SS, or GAAG. From this information, a model of QMITE1 was generated and is shown in **Figure 2.3B**.

A



B

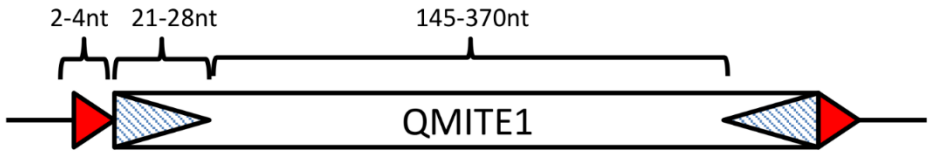


Figure 2.3: CbsR13 loci represent a novel MITE, called QMITE1. (**A**) mfold secondary structure prediction of a selected QMITE1 (range 5; predicted ΔG = -192.72 kcal/mol). Red bars bracket the TIRs and the blue line indicates the location of the potential IHF-binding site. (**B**) Model of QMITE1 depicting DRs as red arrow heads and the TIRs as hatched arrow heads. Length ranges for these features are also shown.

QMITE1 copies encode basic peptides and overlap with annotated genes

Along with being transcriptionally active, 19 QMITE1 copies fully contain short, annotated ORFs that encode predicted peptides with an average isoelectric point (pI) of 12.4. These basic peptides can be divided into three major groups based on sequence similarity (**Figure S2.5**), and they constitute the entire DUF1658 family of small, uncharacterized *C. burnetii* proteins in the Pfam database [150].

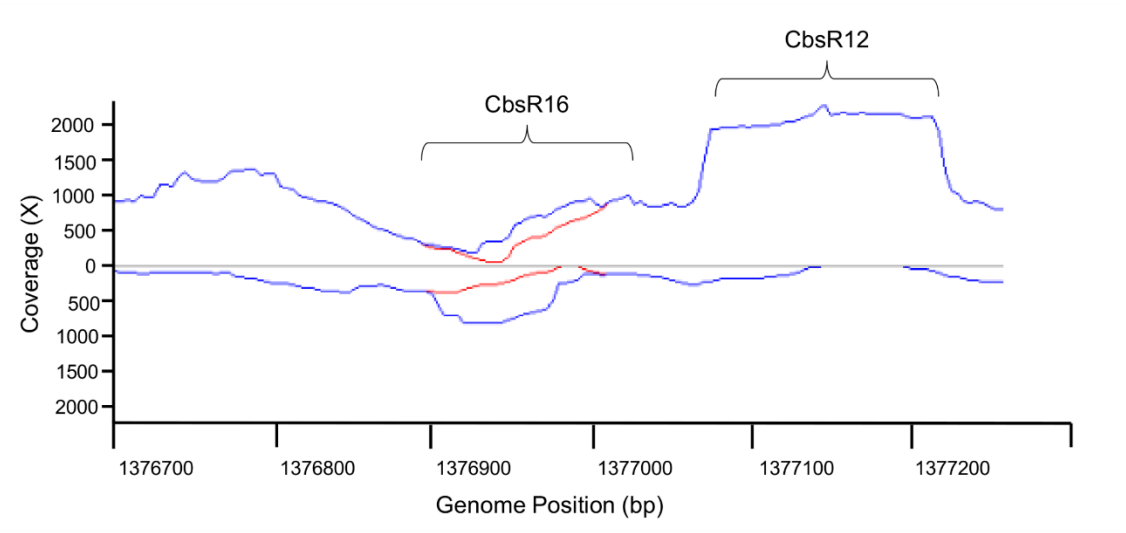
Other annotated genes that are affected by QMITE1 insertions mainly encode hypothetical proteins of unknown function. However, QMITE1 copies also overlap with several functional genes, including: *ubiB* C-terminal 2-bp overlap, *pntAA* C-terminal 42-bp overlap, *mutT* C-terminal 26-bp overlap, CBU_2058 proline/betaine transporter C-terminal 49-bp overlap, *nagZ* C-terminal 50-bp overlap, and CBU_2020 glutamate transporter C-terminal 3-bp overlap. The effect of these QMITE1 insertions in the 3′ UTRs of these genes could not be determined, although other MITE insertions in 3′ UTRs have been observed to translationally repress the affected genes [151].

The CbsR16 locus is a member of a second novel MITE family

We recently identified a new sRNA termed CbsR16 while analyzing CbsR12; a sRNA that is significantly upregulated during *C. burnetii*'s intracellular infection of host cells

[97]. The *cbsR16* gene is located immediately downstream of the *cbsR12* gene, which shares a bi-directional Rho-independent terminator with *cbsR16* (data not shown). When viewing the CbsR16 locus with the Artemis genome browser, it was clear that there was minor differential expression of the locus when taking ambiguous reads into consideration (**Figure 2.4A**). Additionally, when we analyzed CbsR16 using mfold, the predicted secondary structure was highly stable (**Figure 2.4B**). Moreover, although QMITE1 is significantly transcribed at more than one location in the *C. burnetii* genome, CbsR16 is transcribed at a considerably lower level (**Figure S2.3**), with very minor TPM differences between mapped unambiguous and ambiguous transcripts. This indicates that although other sequences homologous to CbsR16 may exist in the RSA 493 genome, only the locus adjacent to CbsR12 is transcribed to any significant level. The strong secondary structure and minor presence of ambiguously mapped reads of CbsR16, though, warranted a genome-wide search for similar sequences.

A



B

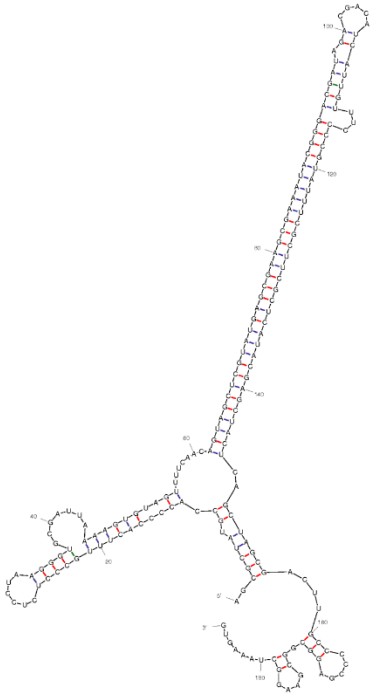


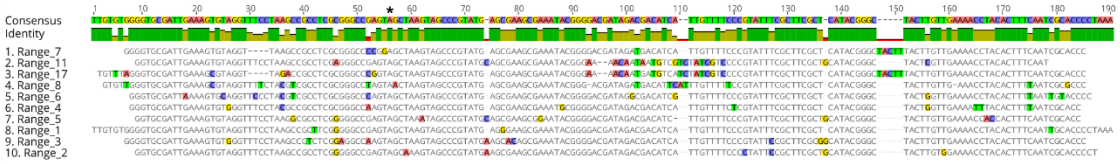
Figure 2.4: CbsR16 is lowly transcribed, with some ambiguous reads mapping to it. (**A**) Artemis view of reads mapping to the CbsR16 locus (RSA 439 genome). The x-axis shows the location (bp) on the chromosome and the y-axis indicates coverage of reads mapping to that location. Reads above the y-axis indicate antisense reads, whereas reads

below the y-axis indicate sense reads mapping to that genomic location. Blue lines signify ambiguous reads mapping to this locus, while red lines signify unambiguous reads. (**B**) mFold prediction of the CbsR16 sRNA secondary structure (ΔG = -85.24 kcal/mol). Red, blue, and green lines forming stem structures indicate G-C, A-U, and G-U base-pairing, respectively.

A discontiguous megaBLAST search with CbsR16 resulted in 78 hits with E-values ranging from 1e-07 to 4e-33. We initially divided these 78 hits into two groups: full-size sequences and smaller sequences. From these pools, we aligned those that covered at least 75% of the input CbsR16 sequence. The full-size versions (**Figure 2.5A**) have a 5′ sequence of ~40 nts that is apparently missing in smaller versions of the element (**Figure 2.5B**). Phylogenetic trees for these full-size and smaller versions were constructed and can be seen in **Figure S2.6** and **Figure S2.7**, respectively. As with QMITE1, we generated representative predicted secondary structures for the full-size (Range 7, **Figure 2.6A**) and small ranges (Range 9, **Figure 2.6B**). Although there are no IHF-binding sites in the CbsR16-like sequences, the full-size ranges have TIRs and are flanked by unique DRs of 7-9 bp, while the smaller ranges are essentially REP elements. Interestingly, these REP elements were previously reported in *C. burnetii*, although their status as a truncated MITE was not recognized [105]. Taken as a whole, the size (~190 bp), presence of TIRs and DRs, and their distribution across the RSA 493 genome suggest that the CbsR16-like loci are indeed MITEs. We therefore propose to designate this family of elements as QMITE2. A model of QMITE2 is shown in **Figure 2.6C**. The

smaller QMITE2 copies strongly resemble a REP element; i.e., they do not contain TIRs nor do they have discernible DRs in flanking genomic regions.

A



B

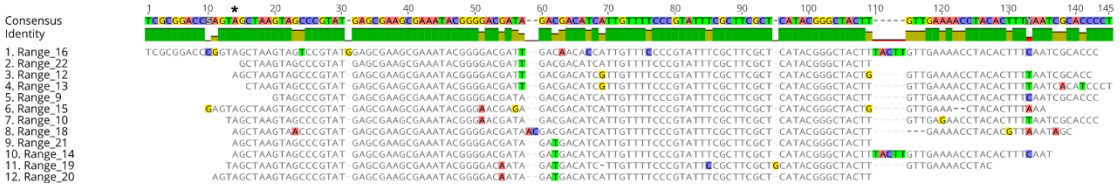


Figure 2.5: CbsR16 loci have full-size and small versions. (**A**) MUSCLE sequence alignment of discontiguous megaBLAST hits that returned full-size versions of the CbsR16 locus. Conserved bases appear as gray blocks, while unaligned bases appear as green, yellow, blue, and red bands, corresponding to T, G, C, and A bases, respectively. An identity indictor is shown above the sequence alignment, where the height signifies conservation of bases at that position, with a color indicator for overall identity between aligned ranges (green: 100%, yellow: 20-99%, red: 0-19%). Above this identity indicator is the consensus sequence, appearing as colored bands indicating bases as described above. (**B**) As in (A), except the MUSCLE alignment displays the top discontiguous megaBLAST hits (>75% of input sequence) associated with the CbsR16 locus, excluding all full-size hits. An asterisk indicates equivalent positions in the full-size and small versions of QMITE2.

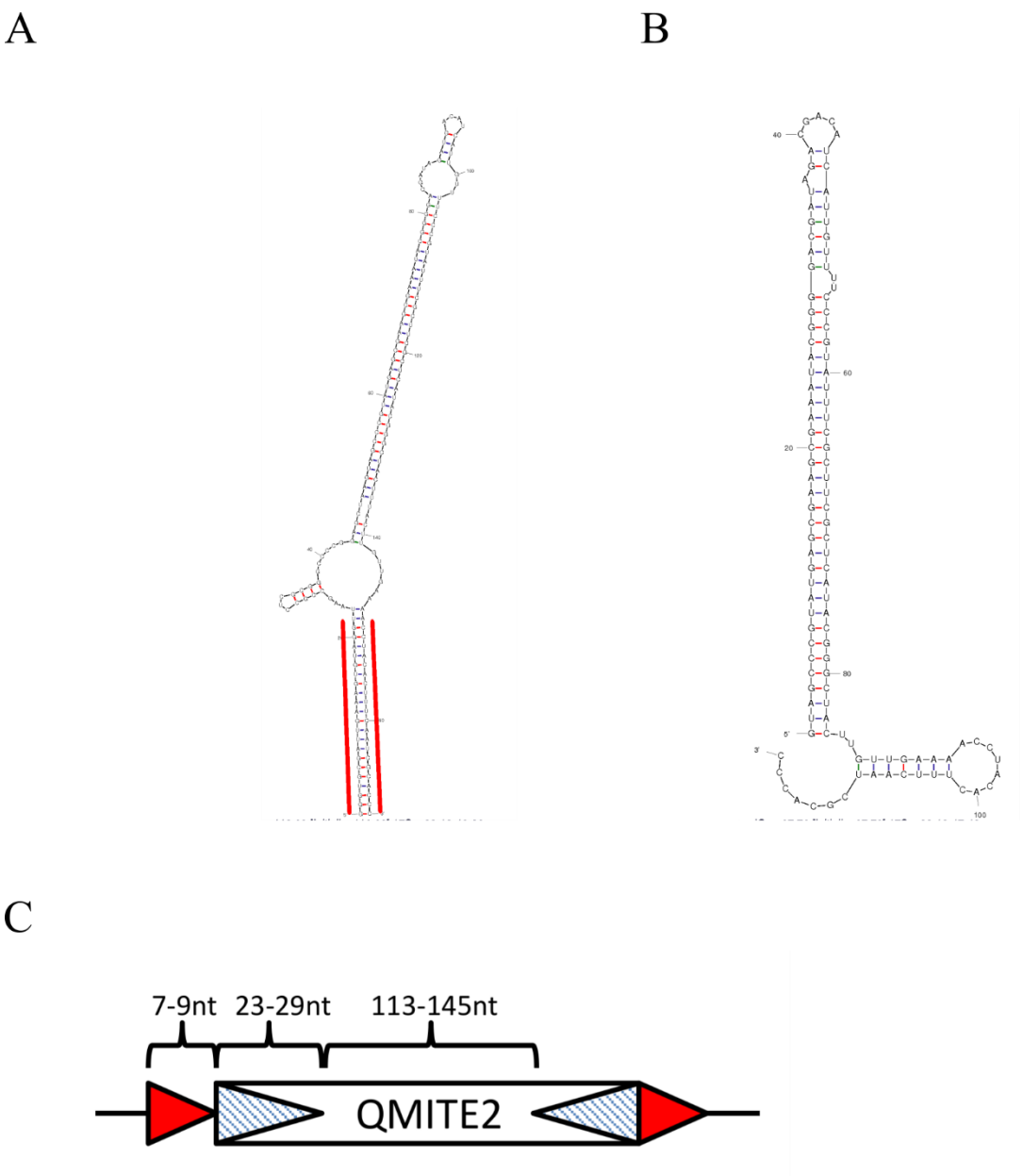


Figure 2.6: CbsR16 loci comprise another novel MITE family, termed QMITE2. (**A**) mFold prediction of the RNA secondary structure of a full-size version of the CbsR16 repeated locus (range 7; predicted ΔG = -113.09 kcal/mol). Red, blue, and green lines forming stem structures indicate G-C, A-U, and G-U base-pairing, respectively. Red lines bracket the identified TIR. (**B**) As in (A), but depicting the secondary structure prediction of a small version of the CbsR16 repeated locus (range 9; predicted ΔG = -67.7

kcal/mol). (**C**) Model of QMITE2 depicting DRs as red arrow heads and TIRs as hatched arrow heads. Length ranges for these features are also shown.

As observed with QMITE1, QMITE2 copies may also affect certain annotated ORFs. Although they do not encode annotated genes like some QMITE1 copies, there is some overlap with neighboring functional genes, including a C-terminal 1-bp overlap with *kdgK*, a C-terminal 8-bp overlap with *ogt*, a C-terminal 7-bp overlap with *recN*, a C-terminal 10-bp overlap with CBU_2078 Fic-Family protein, and a C-terminal 6-bp overlap with *ruvB*. Additionally, although MUSTv2 identified QMITE1 in the RSA 493 genome, it was unable to find QMITE2 under stringent parameters. However, a full-size QMITE2 copy was identified using less stringent parameters (data not shown). The inability for MUSTv2 to identify QMITE2 most likely reflects the filtering parameters of the program itself. Namely, the program searches for copies of the MITE with similar DR's. If a copy with a similar DR is not found, it will filter it out. QMITE2 has unique DR's for each copy, making it difficult to detect.

QMITE2 loci are hot-spots for IS1111 insertion

While parsing various QMITE2 ranges, we found that 20 of the 21 annotated IS1111 TEs in the RSA 493 genome possessed a small QMITE2 located ~400 bp downstream of their stop codons. These small QMITE2 ranges were aligned and shown in **Figure S2.8**. A phylogenetic tree of these transposon-associated QMITE2 insertions was created and is shown in **Figure S2.9**. These ranges are nearly identical to the other small QMITE2 ranges (**Figure 2.5B**), except that they are missing 10-20 bp at the 5′ end. Upon closer inspection, these "missing" bases are actually located 5′ of the IS1111, indicating that the

transposon inserted into this region of QMITE2. Indeed, this has been described before, although the insertion site was not previously recognized as a MITE [104]. It is worth noting that these QMITE2 copies are more divergent than their IS1111-free counterparts, implying neutral selection while they are associated with IS1111. Interestingly, of the twenty IS1111 insertions in QMITE2 copies, only one clearly inserted into a full-length QMITE2 locus, as the TIR is still discernible up- and down-stream of the Tn insertion. The other QMITE2 loci may also have been full-length once, but their flanking sequences presumably diverged rapidly after insertion.

QMITE2 is not specific to *C. burnetii*

Unlike QMITE1, QMITE2 is apparently not unique to *C. burnetii*. A discontiguous megaBLAST search using the CbsR16 sequence yielded hits in multiple alphaproteobacteria, including *Bradyrhizobium* spp. and *Rhodobacter* spp. These hits had sizes of 83-100 nucleotides in length with E-values ranging from 1E-04 to 8E-07. QMITE2 also appeared in one location in *Lacimicrobium alkaliphilum*, a gammaproteobacteria. These sequences were aligned to the small version of QMITE2 (**Figure 2.7**) and a phylogenetic tree was constructed (**Figure S2.10)**. The alignment indicates that although the majority of the sequence corresponding to the predicted stem structure of the small QMITE2 is conserved, the palindromic "tip" (see **Figure 2.6B**, bases 31-53) is more divergent among the alphaproteobacteria shown in the alignment. These results suggest that a majority of the palindromic stem structure may serve some function in *Bradyrhizobium* and *Rhodopseudomonas* spp., while the entirety of this stem is under purifying selection in *C. burnetii*. It's also worth noting that the 3′ portion of QMITE2 is missing from the alphaproteobacterial MITEs. The 3′ end of QMITE2

comprises half of the TIR formed in the full-length QMITE2 suggesting that full-length QMITE2 never existed in the alphaproteobacterial species or was present further back in their evolutionary histories.

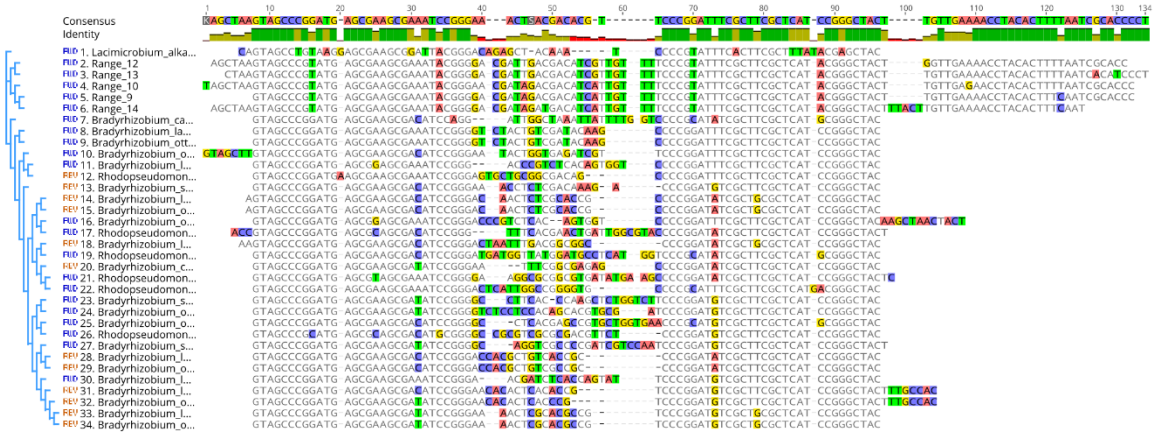


Figure 2.7: QMITE2 is not unique to *C. burnetii*. MUSCLE sequence alignment of discontiguous megaBLAST hits that returned QMITE2 sequences in other organisms. Conserved bases appear as gray blocks, while unaligned bases appear as green, yellow, blue, and red bands, corresponding to T, G, C, and A bases, respectively. An identity indicator is shown above the sequence alignment, where the height signifies conservation of bases at that position with a color indicator for overall identity between aligned ranges (green: 100%, yellow: 20-99%, red: 0-19%). Above this identity indicator is the consensus sequence, appearing as colored bands indicating bases as described above. Ranges 12, 13, 10, 9, 14 in the sequence alignment refer to small QMITE2 ranges included in the BLAST as shown in Figure 2.5B.

Full-length QMITE2 displays inter-strain linkage and sequence conservation in *Coxiella*

Due to the unique DRs produced by individual full-length QMITE2 insertions, we were interested to see if these DRs displayed inter-strain linkage conservation. To accomplish this, full-length QMITE2 ranges were found in the *C. burnetii* Dugway strain and the DRs produced by these inserts were compared to those produced by QMITE2 inserts in the RSA 493 strain. If there were two DRs that were identical in sequence between strains, we determined if the associated QMITE2 copies were linked by observing syntenic genome blocks that were produced via genome rearrangements as the strains diverged [139]. We discovered that the Dugway strain contains 12 full-length QMITE2 copies versus 10 in RSA 493 (**Table 2.1**). Furthermore, seven of the nine discernible DRs in the RSA 493 strain had perfect homologs in the Dugway strain and displayed perfect linkage and sequence conservation. The single unique DR in RSA 493 without a counterpart in Dugway resulted from an IS1111 insertion in the corresponding position in Dugway's genome. Likewise, 11 of the 12 full-length QMITE2 copies in Dugway had unique DRs associated with them and seven of these had perfect homologs in RSA 493, two had a IS1111 inserted into the corresponding position in RSA 493, one position belonged to a genomic segment unique to Dugway, and the final position displayed a QMITE2 inversion in RSA 493, leaving no discernible DR (**Table 2.1**). In summary, most DRs are conserved in both strains with a few lost via deletion, IS1111 insertion, or genome inversion events.

Table 2.1: Full-size QMITE2 copies exhibit inter-strain linkage conservation.

Strain	**Range**	**TIR Length**	**DR Length**	**DR Sequence**	**Homolog?**

RSA 493	c1006608-1006428	25	*	*	No**
	1066751-1066922	29	*	*	*
	1380514-1380685	26	7	TCAGRGG	No***
	c1168547-1168380	24	9	CCGTCAATA	Yes
	c1360856-1360689	23	9	CACATCGAT	Yes
	1988089-1988258	23	7	CAACATTW	Yes
	1586332-1586502	23	9	GTTGGCGCG	Yes
	220015-220188	25	8	GGGGTGTT	Yes
	c970302-970140	24	7	GCTACTT	Yes
	1252325-1252500	24	9	TTCTGTTTA	Yes
Dugway	c334562-334393	23	9	GTTGGCGCG	Yes
	c1836762-1836594	25	8	GGGGTGTT	Yes
	2151397-2151569	23	8	CAACATTW	Yes
	117745-117908	22	*	*	*
	c1299129-1298960	23	9	CCGTCAATA	Yes
	c374053-373882	31	9	AATTTTAAC	No**
	1295396-1295566	26	9	GTATCRTCC	No***
	1561569-1561721	21	13	CCTTCTTCTTTSA	No****
	1384775-1384900	23	9	TTCTGTTTA	Yes
	1261463-1261626	17	9	GGGCTTTCA	No***
	c565652-565819	25	9	CACATCGAT	Yes
	c1003105-1002901	24	7	GCTACTT	Yes

* No discernible DR; ** QMITE2 inversion in other strain; *** IS1111A insertion in other strain; **** Genomic segment deleted in other strain

QMITE1 and QMITE2 copies in the RSA 493 and Dugway genomes

QMITE1 and QMITE2 (full-size and small) copies were mapped against the RSA 493 genome using Circos software (**Figure 2.8**) [146]. We identified 45 copies of QMITE1 and 78 copies of QMITE2 in the RSA 493 genome that in total affect 60 annotated ORFs, with 19 of these ORFs being completely contained within QMITE1 copies and encoding the DUF1658 family of proteins (see **Figure S2.5**). When combined, QMITE1 and QMITE2 copies make up 0.93% of the RSA 493 genome. Interestingly, our analysis revealed that there were generally higher concentrations of QMITE insertions in the second "half" of the genome (~1,000,000 – 1,995,488 bp), with small QMITE "deserts". Accordingly, the first half of the genome was found to contain lower concentrations of QMITEs, with larger deserts (e.g., 570,000 – 690,000 bp) bearing no QMITE inserts.

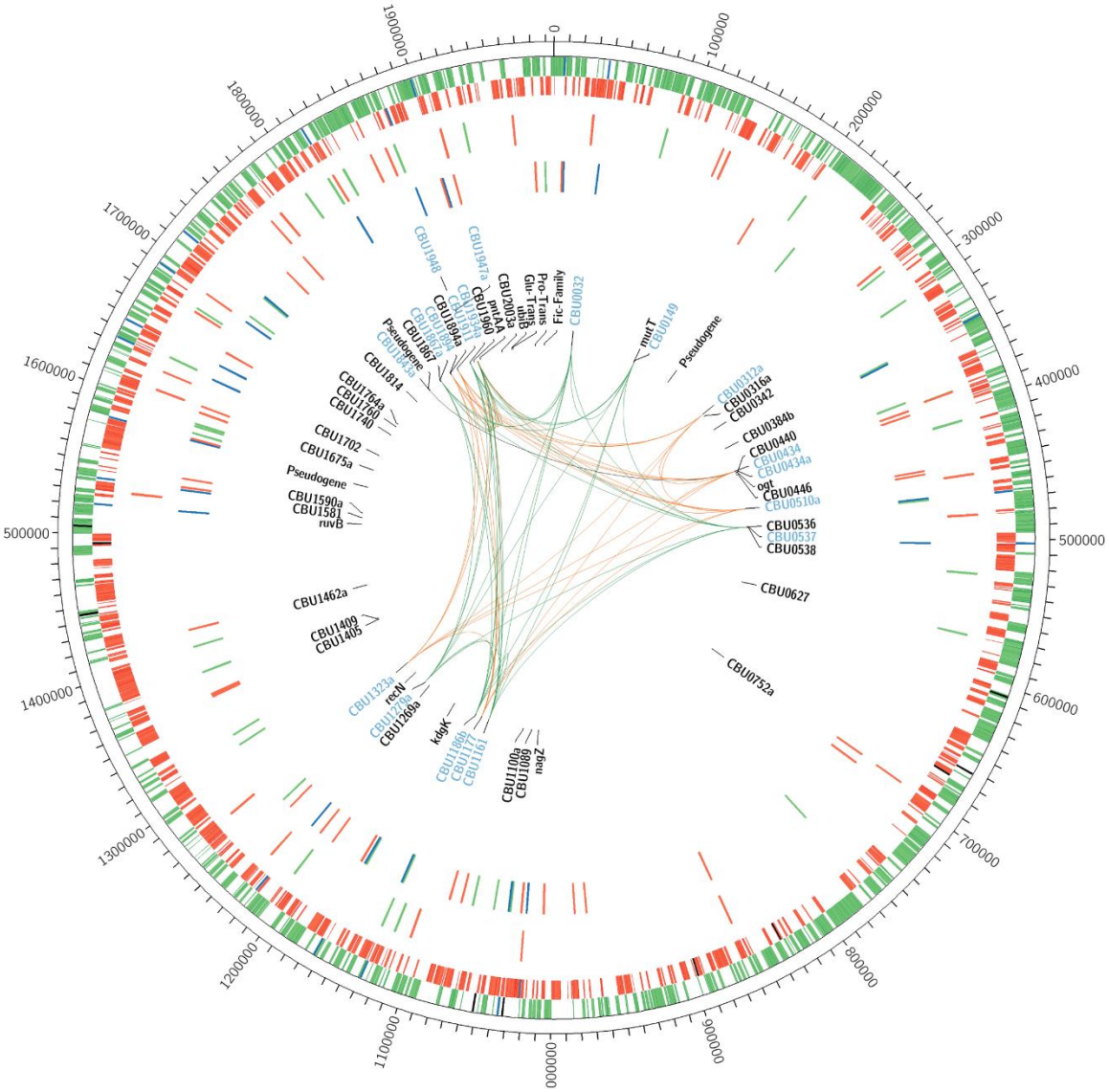


Figure 2.8: Locations of QMITE1 and QMITE2 insertions in the *C. burnetii* RSA 493 genome. The outer ring depicts the RSA 493 chromosome in 100,000 bp increments. The next ring depicts locations of forward strand ORFs in green, IS1111 locations in blue, and non-IS1111 TEs in black, followed by reverse strand ORFs in red on the next ring also featuring IS1111 in blue and non-IS1111 TEs in black. The next ring depicts all chromosomal QMITE1 locations. Green ticks indicate QMITE1 insertions oriented in the forward, while red ticks indicate QMITE1 insertions in the reverse orientation. The next

ring depicts QMITE2 insertions, with green ticks indicating forward insertions, red ticks indicating reverse insertions, and blue ticks indicating IS1111-associated QMITE2 inserts. The following ring labels all of the locus tags for ORFs that have some overlap with either QMITE1 or QMITE2 insertions. ORFs labeled in blue are those that are encoded by QMITE1 insertions and represent the DUF1658 family of proteins. Finally, the colored links between blue-labeled ORFs are indicative of groupings of the proteins coded by these genes (see **Figure S2.5**).

The distribution of QMITE1 and QMITE2 in the Dugway genome is displayed in **Figure 2.9**. Due to linkage conservation of QMITE1 and QMITE2 copies between strains, the genomic locations of the QMITE copies are generally the same as RSA 493, although due to divergence between strains, there are some differences in the number of QMITE copies. Specifically, there are 53 copies of QMITE1 and 62 copies of QMITE2 that together comprise 0.91% of the Dugway genome. There are also 56 ORFs affected by MITEs in the Dugway strain. All of the functional annotated genes affected are the same in the two strains, except for the *enhC* gene, which shows a 3′ extension due to an indel linking the gene to a QMITE2 copy. Interestingly, a C-terminally extended EnhC protein has been previously described for the Dugway strain [139].

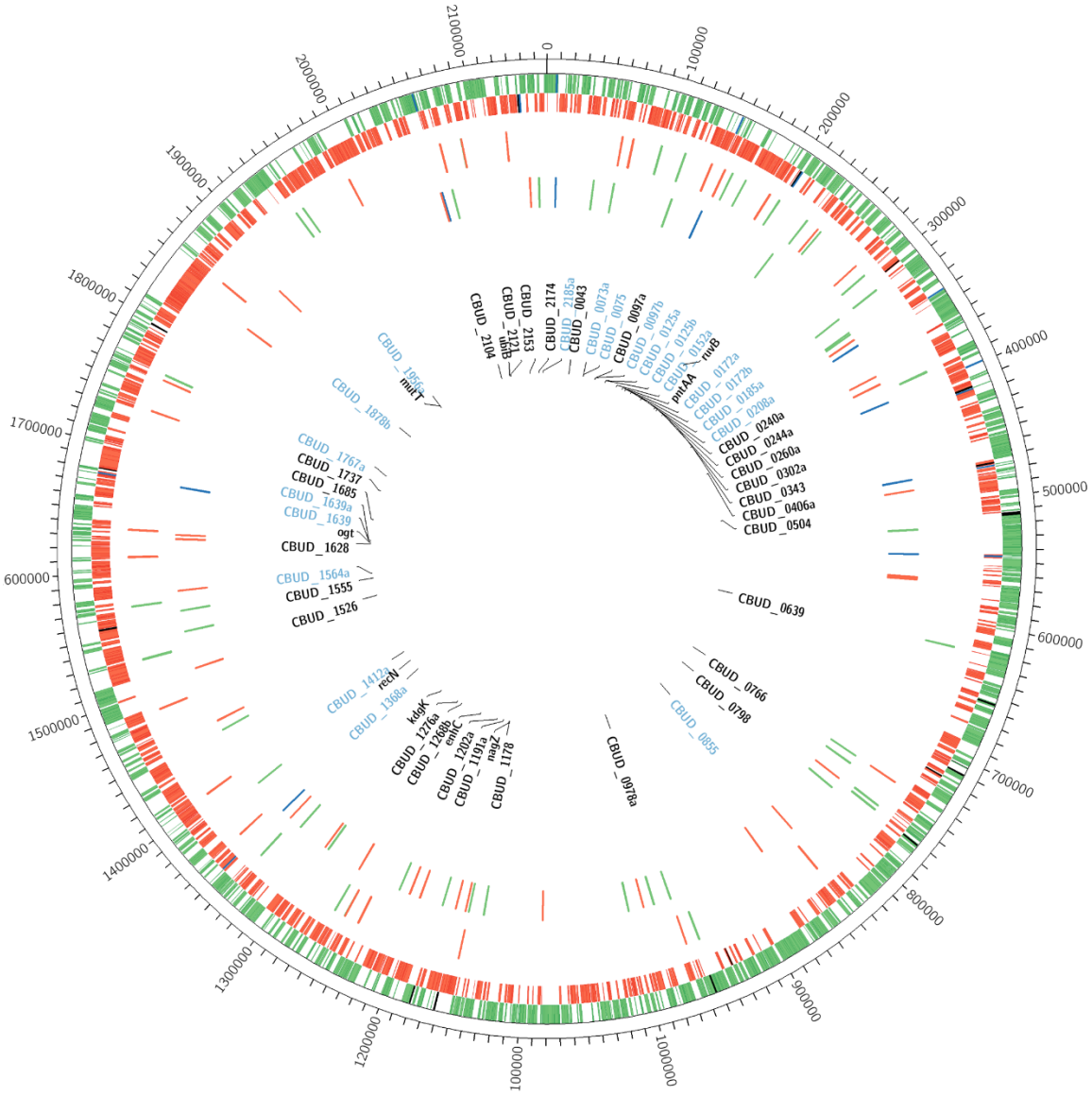


Figure 2.9: Locations of QMITE1 and QMITE2 insertions in the *C. burnetii* Dugway 5J108-111 genome. The outer ring depicts the RSA 493 chromosome in 100,000 bp increments. The next ring depicts locations of forward strand ORFs in green, IS1111 locations in blue, and non-IS1111 TEs in black, followed by reverse strand ORFs in red on the next ring also featuring IS1111 in blue and non-IS1111 TEs in black. The next ring depicts all chromosomal QMITE1 locations. Green ticks indicate QMITE1 insertions oriented in the forward, while red ticks indicate QMITE1 insertions in the

reverse orientation. The next ring depicts QMITE2 insertions, with green ticks indicating forward insertions, red ticks indicating reverse insertions, and blue ticks indicating IS1111-associated QMITE2 inserts. The following ring labels all of the locus tags for ORFs that have some overlap with either QMITE1 or QMITE2 insertions. ORFs labeled in blue are those that are encoded by QMITE1 insertions and represent the DUF1658 family of proteins. Colored links are omitted because the DUF1658 protein products remain the same between strains and largely depend on how the genome was annotated.

QMITE copies affect sRNA genes

New bacterial sRNAs can arise from degraded bacteriophage genes [152]. Similarly, we show that three sRNAs of *C. burnetii* are derived from MITEs. These results suggest that, as shown in eukaryotes [153], genomic parasitic elements can serve as a source for the generation of novel non-coding RNAs of bacteria. For instance, QMITE1 copies have inserted directly downstream of promoter elements for CbsR3 and CbsR13. Moreover, a QMITE2 copy has apparently provided the -10 promoter element for CbsR16, while the -35 promoter element is located directly upstream of the QMITE2 insert (**Figure S2.11**). All of these sRNAs show varying levels of expression (see **Figure S2.3**), indicating that they are being actively transcribed. Furthermore, previously published Northern blots have confirmed that CbsR3 and CbsR13 are transcribed and produce sRNA molecules of the expected size [102].

Discussion

We have described two novel MITE elements in *C. burnetii*, termed QMITE1 and QMITE2. Although their structures and distribution are clear, the nature of their transposition and origin remains indeterminate. Several lines of evidence suggest that QMITE copies are ancient and likely lost the ability to transpose before divergence of present-day *C. burnetii* strains. First, *C. burnetii* RSA 493 contains a plasmid called QpH1 that encodes type 4 secretion system substrates involved in virulence [154]. We could not detect QMITE copies in QpH1, or other *C. burnetii* plasmid types, indicating that either *Coxiella* gained the plasmid after the QMITEs lost the ability to transpose or that the plasmid is too gene-rich to contain stable QMITE copies. Second, the fact that QMITE copies show linkage conservation between strains suggests that they were present before the rearrangement of chromosomes that occurred during divergence of strains. Finally, the presence of QMITE deserts in *C. burnetii* chromosomes (see **Figures 2.8 and 2.9**), especially between CBU_0664 and CBU_0715, which code for non-IS1111 TEs, implies that horizontal gene transfer (HGT) was involved in forming these regions. Indeed, a recent report has shown that this region (608,000 – 660,000 bp; **Figure 2.8)**, is rich in genes that were acquired via HGT, including some LPS biosynthesis genes that are essential to *C. burnetii*'s virulence [155]. The lack of QMITEs in this region indicates that it was acquired after QMITE1 and QMITE2 lost the ability to transpose, but before divergence of strains, since this region displays inter-strain linkage conservation. It is also worth noting that the chromosomal region harboring the *icm/dot* genes involved in type IV secretion display a paucity of QMITE inserts and is flanked by IS1111 TEs that have inserted into QMITE2 copies (see 1,540,000 to 1,580,000 bp in **Figure 2.8**). This

suggests that QMITE2 copies indirectly affected the evolution of *C. burnetii* from a free-living to an obligate parasite by fostering genome plasticity.

Interestingly, QMITE insertions can also be used as a marker for the transposition of certain IS1111 TEs. For example, it is likely that the IS1111 transposons at CBU_1217a and CBU_1186 in the RSA 493 strain inserted into these positions after divergence from the Dugway strain, because in Dugway there are full-size QMITE2 copies with discernible DRs that have no IS1111 TEs in these positions. Similarly, the CBUD_0567a IS1111 of Dugway inserted into that position after divergence, since there is a full-size QMITE2 copy at this position in the RSA 493 genome.

The uniqueness of the QMITE1 insert sequence suggests that it may have utility as a molecular signature for detecting *C. burnetii* in clinical or environmental samples. A current detection protocol utilizes PCR to amplify the so-called *htpAB*-repetitive element, which is part of the IS1111 TE [156]. Recent reports, however, have expressed concerns regarding this method due to the existence of IS1111 TEs in *Coxiella*-like endosymbionts, which may confound results obtained from environmental samples [157]. The QMITE1 sequence has variable ends, although it maintains a conserved core across insertions in the *C. burnetii* genome (see **Figure 2.2A**) that could easily serve as a sizeable DNA template for PCR amplification. Also, the abundance of insertion sites in the *C. burnetii* genome should ensure sensitivity of the assay.

Although results suggest that QMITE1 is unique to *C. burnetii*, a relative of QMITE2 was observed in very distantly-related alphaproteobacteria. However, these QMITE2 copies are not full-length and strongly resemble transposon-associated QMITE2 copies (see **Figure S2.8**). These alphaproteobacteria, including several *Bradyrhizobium*

and *Rhodopseudomonas* spp., are root nodule-associated, free-living bacteria that also encode several copies of the IS1111 TE in their genomes. There are several possible scenarios that could help explain the occurrence of QMITE2 between these distantly-related organisms. First, *C. burnetii* may have acquired QMITE2 from root nodule-associated bacteria via HGT (or vice versa) during its free-living past. Indeed, *C. burnetii*'s genome contains relics of competence, including an almost-complete type IV pilus system that could have facilitated uptake of foreign DNA [99]. Second, QMITE2 may be ancient, existing long before divergence of alpha- and gamma-proteobacteria. Finally, it is entirely possible that these alphaproteobacteria acquired QMITE2 copies via cut-and-paste transposition of IS1111 following HGT, which in turn left relics of QMITE2 scattered across their respective genomes. This latter explanation is certainly possible since full-size QMITE2 copies are absent in these species and the shorter QMITE2 copies they harbor are highly divergent.

All functional annotated genes affected by QMITE contain insertions located at their 3′ ends. The reason for this preference is unknown but may reflect the general tractability of the C-terminus of proteins to a change in amino acid composition. Indeed, when comparing these protein products to counterparts in *L. pneumophila*, there is no significant difference in the overall masses of the proteins, indicating that QMITE insertions neither extend nor truncate the proteins to a significant degree, although the amino acid composition is altered. These alterations are summarized in **Table 2.2** below. In general, QMITE insertions into these genes increase the hypothetical pI of the encoded protein relative to predicted products lacking the QMITE insert. Such a chimera could have conceivably provided a subtle, adaptive advantage to *C. burnetii* as it transitioned

from a free-living bacterium to an obligate intracellular pathogen, as high pI proteins could potentially serve as proton sinks in an acidic host cell phagolysosome. In fact, many *C. burnetii* proteins have been described as having a very high pI, comparable to those found in the human stomach pathogen, *Helicobacter pylori* [99]. This may have been adequate to confer a selective advantage, but the alternative possibility is that QMITE insertions are simply under neutral selection with little to no effect on the fitness of the gene in question. Unfortunately, among the genes listed in Table 2.2, only orthologues for *recN* and *ruvB* are found in *H. pylori*. Similar to *C. burnetii*, these *H. pylori* (strain 26695) proteins have a theoretical pI of 5.84 and 5.86, respectively. This suggests that maintenance of an acidic pI was necessary and the minor change caused by the QMITE2 insert in these genes had little effect on fitness. It is also worth noting that there seems to be a preference for QMITE insertions at the 3′ end of DNA-binding genes and genes involved in DNA repair, such as *ogt*, *recN*, *mutT*, and *ruvB*. It is possible that these insertions are simply due to their proximity to these genomic locations during transposon-induced DNA repair. In fact, it has been found that transposition of TEs is increased upon genotoxic stress in bacteria [158]. There also does not appear to be any QMITE elements that affect the 5′ end of genes with known functions. This is most likely due to the necessity for a promoter element upstream; a feature that may not be provided by the QMITE insertion. Alternatively, there may be a transcription factor binding site(s) upstream of the ORF that is necessary for regulation of that gene. In contrast, there seems to be no locational preference for QMITE insertion into annotated hypothetical proteins, wherein QMITE insertions sometimes appear in-frame in the middle of the ORF (e.g., QMITE2 insertions in CBU_0752a and CBU_1269a).

Table 2.2: QMITE effects on functional gene products.

Gene	QMITE Type	Overlap length (bp)	Amino acids conferred	pI without insert	pI with insert	Gene function
ubiB	1	2	(STOP)	N/A	N/A	Ubiquinone Biosynthesis
CBU_2020	1	3	(STOP)	N/A	N/A	Glutamate antiporter
pntAA	1	42	AQTHRRQLKGAR(STOP)	6.93	8.79	Redox, proton transport
mutT	1	26	LQQDIITQ(STOP)	5.1	4.96	Mutational DNA repair
CBU_2058	1	49	LVVPAQTHRRQLKGAR(STOP)	9.97	10.15	Proline/Betaine transporter
nagZ	1	50	ESQQRLLSFSRFTTGG(STOP)	5.76	5.88	Mureine tripeptide recycling
kdgK	2	1	(STOP)	N/A	N/A	Pentose phosphate pathway
ogt	2	8	TK(STOP)	7.67	8.32	DNA alkylation repair
CBU_2078	2	10	SAK(STOP)	6.16	6.29	Regulation of cell division
recN	2	7	SV(STOP)	6.05	6.05	DNA repair
ruvB	2	6	E(STOP)	5.85	5.73	Holliday Junction resolution; DNA repair

An intriguing aspect of QMITE inserts is the influence they can have on sRNAs, depending on where they insert into the genome. It has been suggested that a class of MITEs in *Neisseria* spp., termed the Correia repeats, may insert near sRNA genes and alter their functions [66, 159]. This is similar to what is observed with QMITE1, especially those inserts that give rise to CbsR3 and CbsR13, two confirmed sRNAs

harboring their own promoters upstream of the QMITE1 inserts and terminating within the confines of the insert itself (see **Figure S2.11**) [97]. When taking the unambiguous reads associated with all QMITE1 loci into account, the TPMs associated with QMITE1 loci reach approximately 9,342, or 0.93% of all transcripts expressed by *C. burnetii*. The fact that these promoter elements still exist after divergence of *C. burnetii* into separate strains speaks to the potential utility of the transcripts they produce, whether they: a) act *in trans* on mRNA target(s), b) affect expression of neighboring genes, or c) are actively translated to produce the high pI proteins listed in **Figure S2.5**.

The truly unique aspect of QMITE inserts is the sRNAs they may produce wherever they insert into the genome. It has been shown that the Correia repeats of *N. meningitidis* give rise to transcripts that are produced at varying levels depending on the specific repeat in question [160]. Here, we confirm this notion by showing that a QMITE2 insert in the coding region of the lowly transcribed sRNA CbsR16 provides the -10 promoter element for the sRNA (see **Figure S2.11**). Additionally, this seems to be one, if not the only, QMITE2 insert that is transcribed with near-equivalence of the ambiguous and unambiguous TPM data (see **Figure S2.3**). Additionally, although sRNAs arising from internal QMITE1 promoters have not been established, it is likely that transcripts are being produced by these inserts since many more ambiguous transcripts are associated with these loci than unambiguous transcripts (see **Figure S2.3**). As seen in **Figure S2.1**, these QMITE1 insertions also have identifiable promoters on both strands of DNA.

In general, ORFs that are affected by QMITE insertion events were the same between the two strains analyzed. One exception occurs in the Dugway strain's *enhC*

gene, which codes for a protein that is thought to inhibit release of peptide fragments during infection by *L. pneumophila*, *C. burnetii*'s closest pathogenic relative [161,162]. The function of EnhC in *C. burnetii*'s pathogenicity has not yet been established, although recent studies have speculated that it may play a similar role to the *L. pneumophila* counterpart [163]. In the Dugway strain, *enhC* is extended due to an in-frame QMITE2 insertion at the 3′ end of the gene. Thus, the C-terminal 33 amino acids are presumably provided by the QMITE2 insertion, and the stop codon occurs immediately downstream. This same QMITE2 insert also exists in RSA 493, although an indel has resulted in a stop codon immediately preceding the element. It is unclear whether the C-terminal extension in Dugway affects EnhC function when compared to the altered protein product expressed by RSA 493. Conceivably, as the Dugway EnhC mRNA is transcribed, the highly stable stem structure conferred by QMITE2 could serve as a substrate for ribonuclease III (RNase III) processing. This could create an mRNA lacking a stop codon, which would, in turn, lead to ribosome stalling and eventual targeting of the nascent polypeptide for degradation [7]. Whether this process occurs as hypothesized is currently under investigation.

A variety of TEs have been previously described in *C. burnetii*. Here, we have characterized two novel MITE families that exist as multiple copies in all annotated strains of *C. burnetii*. QMITE1 is of importance because its promoter elements could influence expression of nearby genes. QMITE2 is noteworthy due to unique DRs that could allow for identification of syntenic blocks and visualization of chromosomal rearrangements that have occurred between *C. burnetii* strains as they diverged. QMITE loci could also be used to identify chromosomal regions derived through HGT after the

QMITE copies became inactive but prior to divergence of strains. The linkage conservation between QMITE1 and QMITE2 elements has helped us establish a timeline that suggests that these elements helped influence the evolution of *C. burnetii* on its path towards becoming an obligate pathogen by serving as sites for IS1111 transposition and inserting into and influencing annotated ORFs and sRNA genes. Finally, we have described the influence that QMITE insertions have had on CbsR3, CbsR13, and CbsR16 sRNA's, the latter of which is produced from a promoter element within a QMITE2 insert.

Supplementary Material

GGAGAAGGCATATCATTACCCACAGTGGTAAGACTAAAAACTTAAAGCCCGTCATCCCCGC
GCAGGCGCATAGGCGTC**AACTTAAGG**GAGTGAGGTAAAGAGGG**AATCAA**GTAGTTGGAAAA
ATGTCCGTCGTTCCCGTGCGTAGGCCGTCTATCGATAGACAAAGAATCGAGGGCTTTCACG
GCCGAAGTCACGGGAATCCCGGCTAAGAGGGGCTTGAAGAACACTAACGGTGTTTTTCTTA
GCTCCTTAATC**TGGGTCC**CCCCGACTCGGCCGTGAAGGTT**TTGTATTCT**TCAATACTACAG
CAGACGGCCTCGCGGGAGGACGACGGATTTTAAGTTTTTCACGCTTCACCACGGGTGGGTA
ATGATAGGGGAATCCGTTGTATCAAGGA**ATTTTGAAA**GGAACAAGGCGAAAGAGAG**CATCA**
ACTTTACAGTCATTAGGGCTTTATGCTTAAGTTGACGCCTATGCGCGCAGGCGAGGACCAA
ACGCAA

Figure S2.1. Extended-QMITE1 sequence for discontiguous megaBLAST searches. Predicted sigma-70 promoter elements for: Forward -10 (red), Forward -35 (red); Reverse -10 (blue), Reverse -35 (blue).

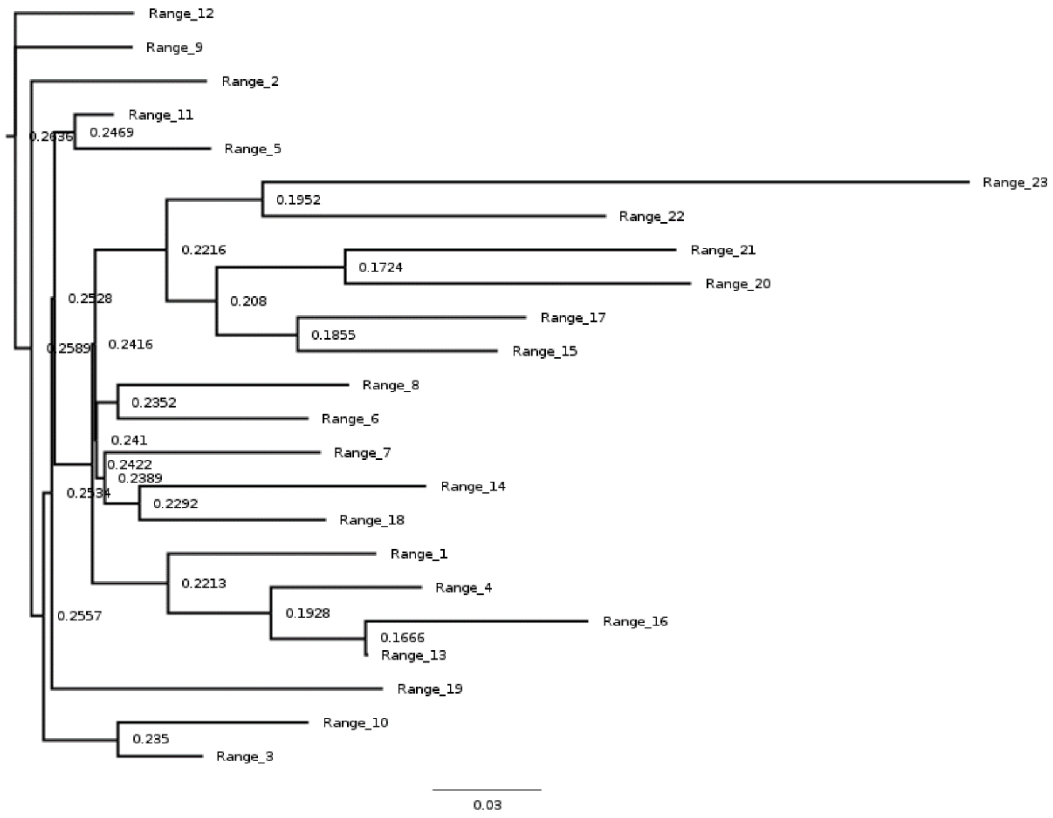


Figure S2.2. Maximum likelihood phylogenetic tree of QMITE1 inserts. Node labels are indicated at the corresponding locations, and a branch length legend is shown at the bottom of the figure.

Gene	Ambiguity	Average TPM (n=2)
CbsR3	Unambiguous	242
	Ambiguous	6586
CbsR13	Unambiguous	373
	Ambiguous	5875
CbsR16	Unambiguous	99
	Ambiguous	95

Figure S2.3. QMITE-associated TPMs obtained by RNA-Seq from *C. burnetii* LCVs grown in infected Vero cells (n=2 biological replicates).

Strand	Length	DR		DRIdent	DRLeft	DRRight	TIR	TIRIdent	TIRLeft	TIRRight	MITEAT
+	402	2		1	GC	GC	19	1	GCCTAGGCGTCAACTTAAG	CTTAAGTTGACGCCTAGGC	0.5012
+	411	2		1	GG	GG	10	0.9	CGCATAGACG	CGTTTATGCG	0.5
-	403	2		1	CG	CG	20	0.95	CGCATAGGCGTCAACTTAAG	CTTAAGTTGACGTCTATGCG	0.5149
+	394	2		1	GC	GC	19	1	GCATAGGCGTCAACTTAAG	CTTAAGTTGACGCCTATGC	0.5165
-	388	2		1	TA	TA	14	0.92857	GGGGTCAACTTAAG	CTTAAGTTGACGCC	0.5193
-	393	2		1	TA	TA	14	0.92857	GGCCTCAACTTAAG	CTTAAGTTGACGCC	0.5076
-	401	2		1	GC	GC	19	1	GCATAGGCGTCAACTTAAG	CTTAAGTTGACGCCTATGC	0.5
-	413	4		1	GAAG	GAAG	21	1	GCGCATAGGCGTCAACTTAAG	CTTAAGTTGACGCCTATGCGC	0.4952

Figure S2.4. MUSTv2 search results indicating identified QMITE1 elements in the *C. burnetii* RSA 493 genome. Attributes of individual MITES are shown.

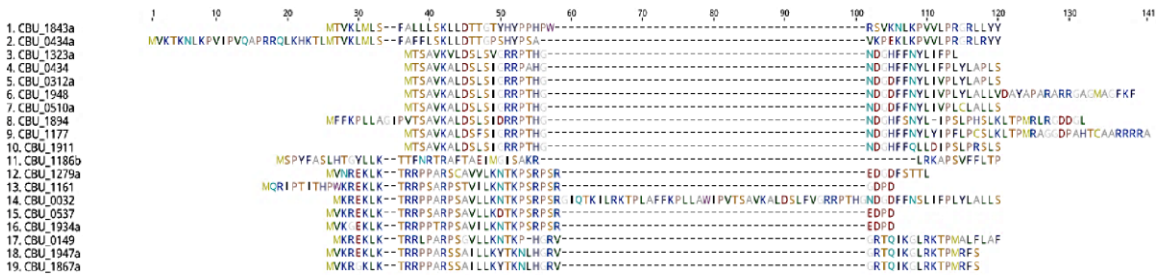


Figure S2.5. MUSCLE alignment of RSA 493 DUF1658 proteins.

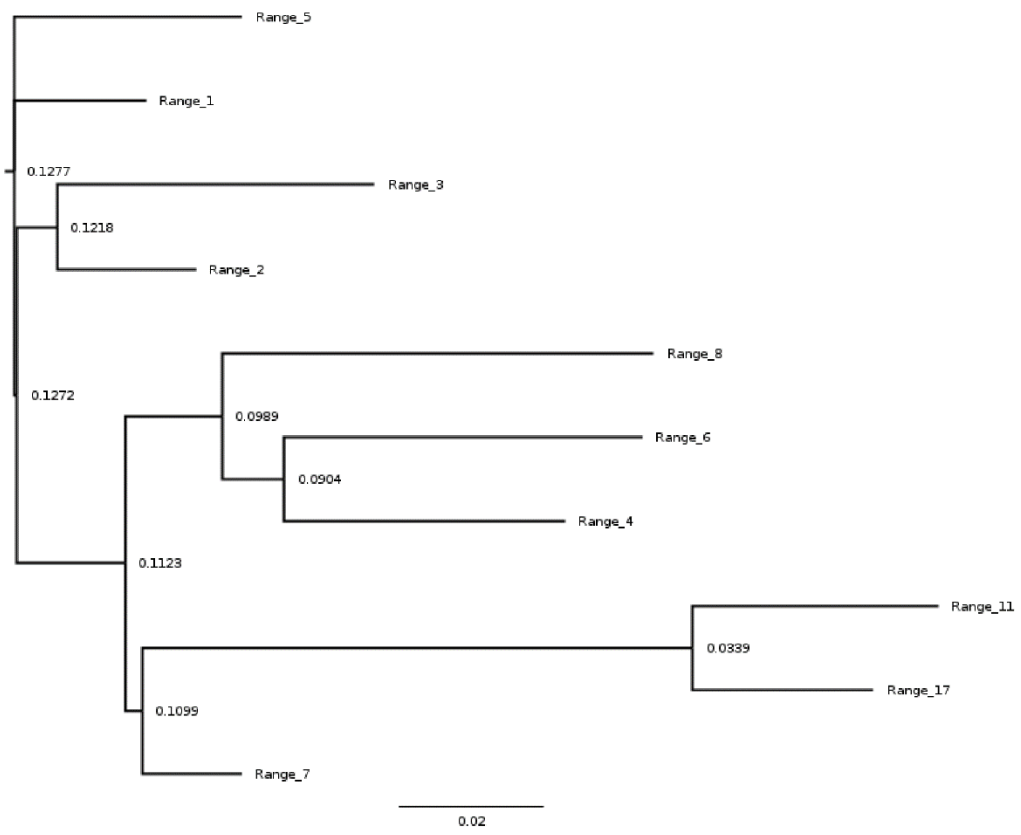


Figure S2.6. Maximum likelihood phylogenetic tree of full-size QMITE2 inserts. Node labels are indicated at the corresponding locations, and a branch length legend is shown at the bottom of the figure.

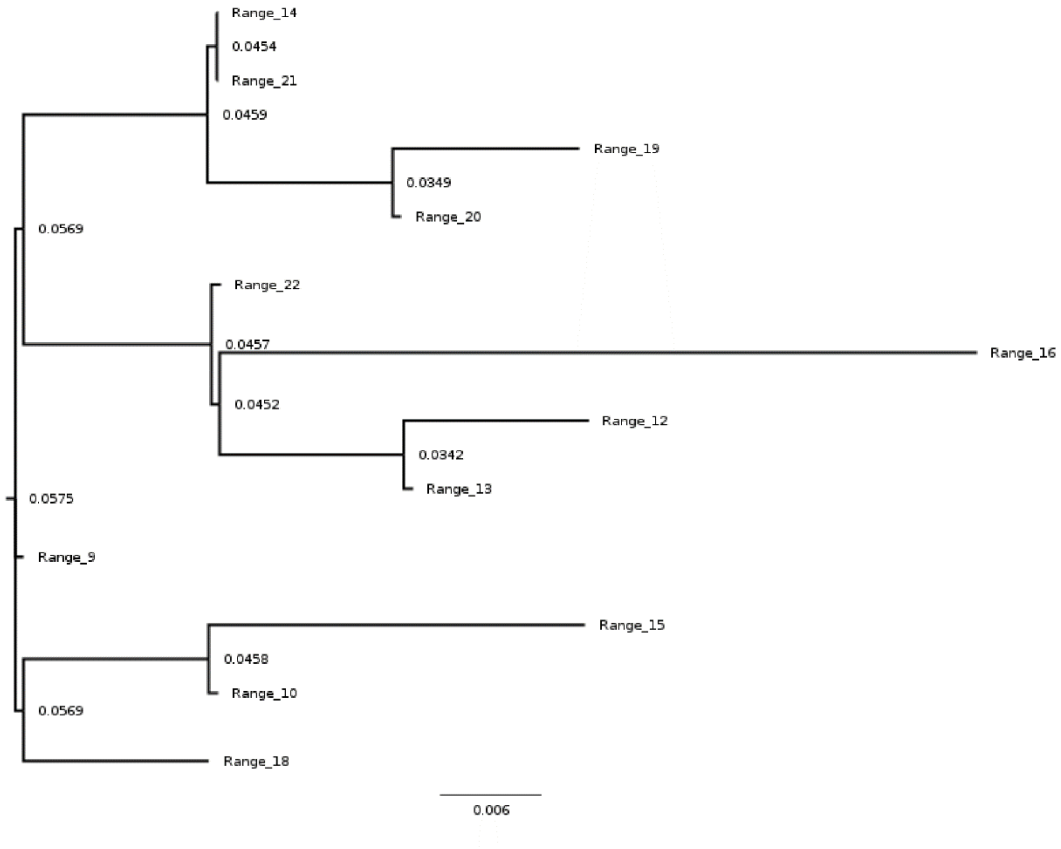


Figure S2.7. Maximum likelihood phylogenetic tree of small QMITE2 inserts. Node labels are indicated at the corresponding locations, and a branch length legend is shown at the bottom of the figure.

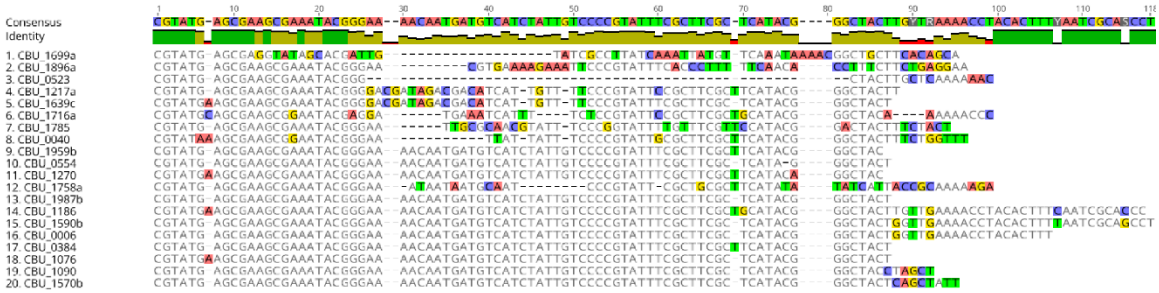


Figure S2.8. MUSCLE alignment of transposon-associated QMITE2 inserts.

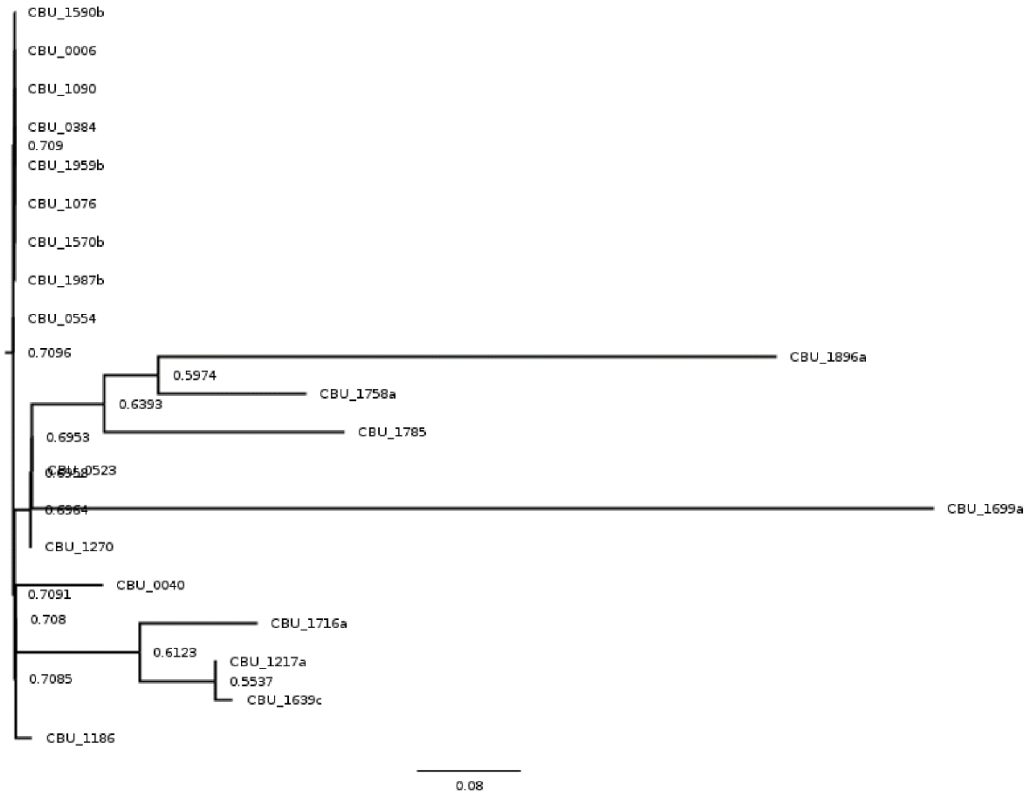


Figure S2.9. Maximum likelihood phylogenetic tree of transposon-associated QMITE2 inserts. Node labels are indicated at the corresponding locations, and a branch length legend is shown at the bottom of the figure.

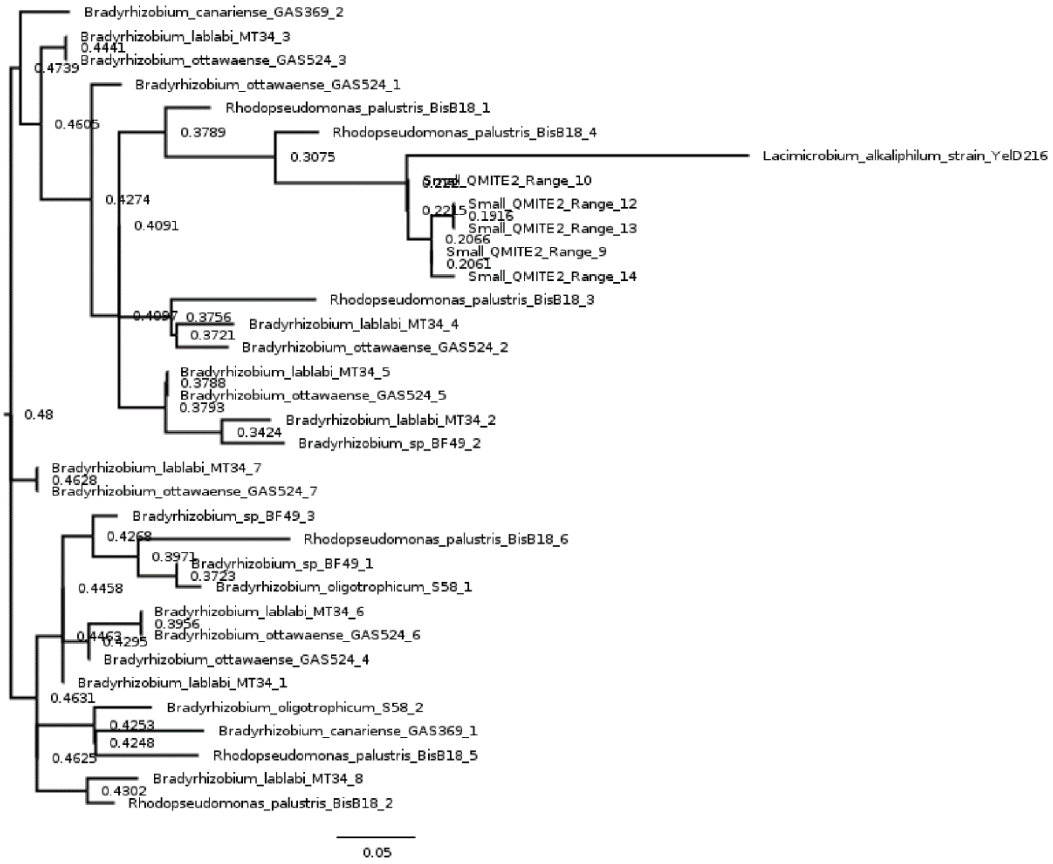


Figure S2.10. Maximum likelihood phylogenetic tree of QMITE2 inserts found in alphaproteobacteria. Node labels are indicated at the corresponding locations, and a branch length legend is shown at the bottom of the figure.

CbsR13 Coding Sequence (1817103-1817658):

GTCGATGATCGAGAGTTGAAGTCCTTT**TTGGAG**AATGGTTGAGAAAG**CGGTATAAT**TTCTTTTTATCTTC
CCGCTAGCG**GGAGAAGGCATATCATTACCCACAGTGGTAAGACTAAAAACTTAAAGCCCGTCATCCCCGC**
GCAGGCGCATAGGCGTCAACTTAAGGGAGTGAGGTAAAGAGGGAATCAAGTAGTTGGAAAAATGTCCGTC
GTTCCCGTGCGTAGGCCGTCTATCGATAGACAAAGAATCGAGGGCTTTCACGGCCGAAGTCACGGGAATC
CCGGCTAAGAGGGGCTTGAAGAACACTAACGGTGTTTTTCTTAGCTCCT**TAATCTGGGTCCCCCCGACTC**
GGCCGTGAAGGTTTTGTATTCTTCAATACTACAGCAGACGGCCTCGCGGGAGGACGACGGATTTTAAGTT
TTTCACGCTTCACCACGGGTGGGTAATGATAGGGGAGCATTGTGAAAGGGGGCGCTATTGCCAACACTTC
CACCTAGGTGGAAATTAAAATAATATTATTTTAAAAGTGCTATTAAAGTGAATTAGCTGATGGATA

CbsR3 Coding Sequence (c481204-481907):

TCTGAAATCAATAGTAAGGTTACTA**TATACA**AATTTTGTTCTAAA**ATCTATCAAT**AGGGGAGGGACAGGA
TCCTGTCGTTAGACTCCAGTCACAGCCACTTTAGGCCCCGTCACCGCGGCAGGGGGAATGAGGATTAAAA
ACAAACCTTGATAGAAAGCGCGCTAGATTCCCGCCTGC**GCGCATAGACGTCAACTTAAGGAGCGAGGTAA**
AGAGGGAATATCAAGTAGTTAAAAAATGTCCGTCGTTCCCGTGCGTAGGCCGTCTACCGATAGACAAAGA
ATCGAGGGCTTTCACGGCCGAAGTCACGGGAATCCCGGCTAAAAGCCTAAGAAAAGCGCCATCGGTGTTT
TTCTTAGCCCCC**TAATCCGGGTCCTCCCGACTCGGCCGTGAAGGTTTTGTATCTTTCAATAATACCGCAG**
ATGGCCTCGCGGAAGGACGACGGGTTTTAAGTTTTTCGCGCTTCACCACGGCTGGGTAATAATAGGAATC
CGTTGTATCAAGGAATTTTGAAAGGAACAAGGCGAAAGAGAGCATCAACTTTACAGTCATTAGGGCTTTA
TGCTTAAGTTGACGCCTATGCGCGCAGGCGAGGACCAAACGCAACGCGCTGAAAGCGCGCTTGTGATTTC
CGCTGCGCGGGATGAGGAGGCCAAAATTGAAGTGGCTGTGCGCTGGGGGTAATAATAGGGGCAGAATTGG
GGAA

CbsR16 Coding Sequence (1403050-1403350)

TGCATATGACATCCTCTCCATGCTGAAATCTGGCGCCATTCTAGCGGAATCAAAAAGATAACACAAAGCG
CTATGCCACCCCACT**TTGCCC**TCTCCTAA**GGGTGCG****ATTAAAA****GTGTAGTTTTCAACAGTAGCTCGTATG**
AGCGAAGCGAAATACGGGACGATAGACGACATCATTGTTTCCCCGTATTTCGCTTCGCTCATACGAGCTA
CTCAGCTAGCGACTTGCCCCCGAGGCGGCGAAGGCTAAAGTGAGAAAAAAAAAGGACCGCAAAAGCGGTCC
TCTTTATTAGTACAATCTGGA

-10 promoter element

-35 promoter element

QMITE Insertion

Rho-independent terminator

Figure S2.11. QMITE insertions in functional sRNAs of *C. burnetii*.

Chapter 3

A CsrA-binding, *trans*-acting sRNA of *Coxiella burnetii* is necessary for optimal intracellular growth and vacuole formation during early infection of host cells

As published in: Journal of Bacteriology. 2019 Oct 21;201(22):e00524-19. doi: 10.1128/JB.00524-19

Abstract

Coxiella burnetii is an obligate intracellular gammaproteobacterium and zoonotic agent of Q fever. We previously identified 15 small non-coding RNAs (sRNAs) of *C. burnetii*. One of them, CbsR12 (***C****oxiella* ***b****urnetii* **s**mall **R**NA **12**), is highly transcribed during axenic growth and becomes more prominent during infection of cultured mammalian cells. Secondary structure predictions of CbsR12 revealed four putative CsrA-binding sites in stem loops with consensus AGGA/ANGGA motifs. We subsequently determined that CbsR12 binds to recombinant *C. burnetii* CsrA-2, but not CsrA-1, proteins *in vitro*. Moreover, through a combination of *in vitro* and cell culture assays, we identified several in-*trans* mRNA targets of CbsR12. Of these, we determined that CbsR12 binds and upregulates translation of *carA* transcripts coding for carbamoyl phosphate synthetase A; an enzyme that catalyzes the first step of pyrimidine biosynthesis. In addition, CbsR12 binds and downregulates translation of *metK* transcripts coding for *S*-adenosyl methionine (SAM) synthetase, a component of the methionine cycle. Furthermore, we

found that CbsR12 binds to and downregulates the quantity of *cvpD* transcripts, coding for a type IVB effector protein, in mammalian cell culture. Finally, we found that CbsR12 is necessary for expansion of CCVs and affects growth rates in a dose-dependent manner in the early phase of infecting THP-1 cells. This is the first characterization of a *trans*-acting sRNA of *C. burnetii* and first example of a bacterial sRNA that regulates both CarA and MetK synthesis. CbsR12 is one of only a few identified *trans*-acting sRNAs that interacts with CsrA.

Importance

Regulation of metabolism and virulence in *C. burnetii* is not well understood. Here, we show that *C. burnetii* small RNA 12 (CbsR12) is highly transcribed in the metabolically active LCV compared to the non-replicative SCV. We show that CbsR12 directly regulates several genes involved in metabolism, along with a type IV effector gene, in *trans*. Additionally, we demonstrate that CbsR12 binds to CsrA-2 *in vitro* and induces autoaggregation and biofilm formation when transcribed ectopically in *E. coli*, consistent with other CsrA-sequestering sRNAs. These results implicate CbsR12 in the indirect regulation of a number of genes via CsrA-mediated regulatory activities. The results also support CbsR12 as a crucial regulatory component early on in a mammalian cell infection.

Introduction

A previous study by our group revealed 15 novel *C. burnetii* sRNAs that were differentially transcribed either in LCVs vs. SCVs, or in host cell infections vs. growth in ACCM-2 medium [97, 98]. Among these, ***C****oxiella* ***b****urnetii* **s**mall **R**NA **12** (CbsR12) was found to be markedly upregulated in the intracellular niche as compared to ACCM-2. Northern blots also showed that CbsR12 was upregulated in SCVs vs. LCVs in ACCM-2, and revealed two distinct sizes of the sRNA, suggesting that either an alternative transcription start site (TSS) or ribonuclease processing of the sRNA was responsible. In a subsequent study, CbsR3 and CbsR13 were found to originate from transcribed loci of a selfish genetic element, termed QMITE1 [164]. However, despite the identification and verification of several CbsRs, none has been functionally characterized, to date.

In this study, we describe activities of a highly transcribed, infection-specific sRNA of *C. burnetii*, termed CbsR12. Our analyses show that CbsR12 binds to CsrA-2, but not CsrA-1 *in vitro*. We also establish that CbsR12 binds and up-regulates *carA* (CBU_1282) and down-regulates *metK* (CBU_2030) transcripts, in *trans*. The bacterial *carA* gene codes for carbamoyl-phosphate synthetase (small) subunit A (CarA), which forms a heterodimer with carbamoyl-phosphate synthetase (large) subunit B (CarB). The CarAB complex catalyzes the first step in pyrimidine biosynthesis and is involved in arginine biosynthesis in some bacteria [165]. The bacterial *metK* gene codes for SAM synthetase, an enzyme responsible for catalyzing production of SAM, the major donor of methyl groups during metabolism in prokaryotic cells. As a methyl donor, SAM affects DNA methylation and thus global transcription [166]). It has also been implicated in virulence, being necessary for the production of N-acyl homoserine lactones involved in bacterial quorum sensing (reviewed in [167]). We also implicate CbsR12 in expansion of

the CCV, as its size is directly correlated with levels of the sRNA. Furthermore, we find that CbsR12 binds *ahcY* (CBU_2031) and *cvpD* (CBU_1818) transcripts, which are components of the methionine cycle and a T4BSS effector protein, respectively [84]. Overall, this study highlights CbsR12 as a crucial component in early stages of a *Coxiella* infection.

Results

CbsR12 is a principal non-rRNA/tRNA/tmRNA transcript during *C. burnetii* infection of Vero and THP-1 cells

CbsR12 was first described as a highly-transcribed, infection-specific sRNA that was upregulated in SCVs compared to LCVs when analyzed by Northern blots [97]. The impetus for our study came when we analyzed previous RNA-Seq data (SRP041556) [97] by converting raw read data into TPM, a normalized measure of gene expression [168]. These results showed that CbsR12 was the most highly transcribed non-tmRNA transcript in both LCVs and SCVs during *C. burnetii* infection of African green monkey kidney epithelial cells (Veros). Additional data from LCVs obtained from a *C. burnetii* infection of monocytic THP-1 cells corroborates the observation that CbsR12 is a principal transcript during infection (Rahul Raghavan, unpublished data). Moreover, we were surprised to find that CbsR12 was more abundant in LCVs, not SCVs (**Table 3.1**).

Table 3.1. Top ten expressed genes[a] across various *C. burnetii* growth conditions**.**

Rank	**ACCM-2[b] LCV (3dpi)**	**ACCM-2[b] SCV (21dpi)**	**Vero LCV (3dpi)**	**Vero SCV (21dpi)**	**THP-1 LCV (3dpi)**

1	tmRNA (104,383)	tmRNA (4,430)	tmRNA (153,422)	tmRNA (22,500)	tmRNA (128,928)
2	CBU_1183 (32,842)	Intergenic (2,632)	**CbsR12 (79,870)**	**CbsR12 (8,521)**	RNase P RNA (63,637)
3	CBU_0307 (13,902)	CBU_1538 (2,133)	RNase P RNA (25,715)	CBU_0089a (8,012)	**CbsR12 (33,331)**
4	CBU_1224a (11,023)	Intergenic (1,977)	CBU_0089a (13,882)	CBU_2034 (4,375)	6S RNA (22,540)
5	RNase P RNA (10,3923)	tRNA (1,902)	CbsR1 (11,445)	CBU_1170 (3,704)	CBU_0456 (17,603)
6	CBU_0311 (7,544)	CBU_0183 (1,593)	CBU_0718 (10,370)	CBU_0718 (3,703)	CBU_0474 (16,267)
7	CBU_0474 (7,120)	Intergenic (1,543)	CBU_1932 (9,734)	CBU_1280a (3,522)	CBU_1183 (13,349)
8	Intergenic (6,928)	CBU_0157 (1,390)	CBU_1272 (9,011)	RNase P RNA (3,429)	CBU_0307 (12,753)
9	CBU_0306 (6,666)	Intergenic (1,340)	CBU_0711 (5,530)	Intergenic (3,340)	CBU_0311 (12,146)
10	**CbsR12 (3,522)**	Intergenic (1,203)	CBU_1170 (4,633)	CBU_1272 (3,134)	CBU_0473 (11,449)

[a]Genes as annotated for RSA493 (NCBI Ref Seq: NC_002971.4) with corresponding TPM values in parentheses. [b]ACCM-2 is an axenic growth medium (37). CbsR12 is indicated in boldface. N = 2 biological replicates per condition.

CbsR12 is processed by ribonuclease III *in vitro*

Previous Northern blot analyses showed that CbsR12 produced two bands of approximately 170 and 50 nucleotides, regardless of *C. burnetii* growth conditions or developmental stage [97]. We therefore wished to determine whether these RNAs arose from alternative TSSs for the *cbsR12* gene or from RNase III processing of the full-length CbsR12 transcript. We first utilized 5' rapid amplification of cDNA ends (RACE) with total RNA derived from *C. burnetii* LCVs infecting Veros (3dpi) in order to determine

the TSS of CbsR12. This experiment revealed the full-size CbsR12 (~200 nucleotides) as expected, but also indicated that two potential alternative TSSs existed ~110 nucleotides upstream of the *cbsR12* gene's Rho-independent terminator (**Figure S3.1A**). To determine whether the TSSs were generated by RNase-mediated cleavage, we treated *in vitro*-transcribed CbsR12 with recombinant *C. burnetii* RNase III [102] and commercially-available *E. coli* RNase III (New England BioLabs). Results showed that CbsR12 was similarly processed by both kinds of RNase III into two RNA fragments with sizes that closely resembled those in the previously-reported Northern blot analysis (**Figure S3.1B**) [97]. These results strongly suggest that the two sites are not alternative TSSs but are instead generated by RNase III processing.

CbsR12 binds to *C. burnetii* recombinant CsrA-2, but not CsrA-1, *in vitro*

The predicted secondary structure of CbsR12 also revealed four conserved single-strand sequence motifs among the various stem-loop structures (**Figure S3.1A**). This motif, AGGA/ANGGA, corresponds exactly to the conserved CsrA-binding motif of many bacteria [27, 30]. *C. burnetii* contains two annotated and distinct types of CsrA (termed CsrA-1 [CBU_0024] and CsrA-2 [CBU_1050]) that share 65% primary sequence identity. To test the functionality of these domains, we employed an *in vitro* binding assay and a RNA-protein EMSA to determine if CbsR12 binds to natively-purified recombinant CsrA-1 (rCsrA1) and/or CsrA-2 (rCsrA-2). EMSA results clearly showed that CbsR12 binds to rCsrA-2, but not rCsrA-1, *in vitro* (**Figure 3.1**). In addition, a CbsR10 negative-control sRNA did not bind either rCsrA. Furthermore, the K_D for rCsrA-2 was determined to be 130 nM, consistent with published values for CsrA-binding sRNAs [169].

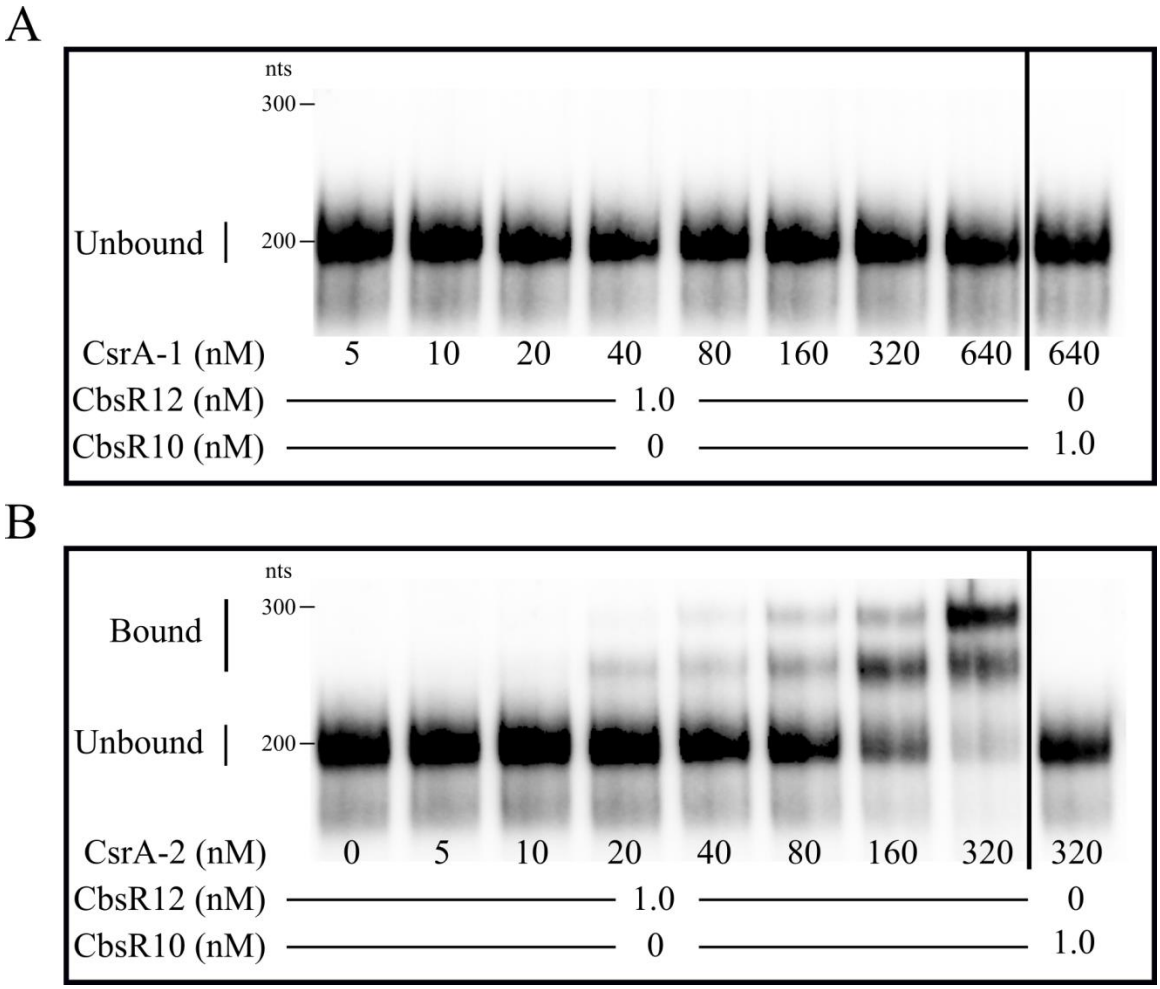


Figure 3.1: CbsR12 binds to CsrA-2, but not CsrA-1, protein *in vitro*. EMSAs showing RNA-protein interactions between biotin-labeled, *in vitro*-transcribed CbsR12 (0 or 1 nM) and increasing concentrations of purified, rCsrA-1 (**A**) or rCsrA-2 (**B**). CbsR10 (at 0 or 1 nM; right of vertical black line) is included as a negative control, as the sRNA contains only a single discernible CsrA-binding motif.

A *cbsR12* mutant shows prolonged lag phase in axenic media

A *cbsR12* mutant of *C. burnetii* (strain RSA439 with Tn 327, hereafter referred to as MB-*cbsR12*), as well as an otherwise isogenic Tn insertional control strain (strain RSA439 with Tn 1832, hereafter referred to as MB-WT), were previously generated using a

Himar1-based Tn system [86]. The location of the Tn insertion of strain MB-*cbsR12* is shown in **Figure S3.2A**. We also constructed a transposon-directed complement of strain MB-*cbsR12* (hereafter referred to as MB-*cbsR12*-Comp) containing the wild-type *cbsR12* gene plus ~100 bp of 5' and 3' flanking sequences, to include any potential transcriptional regulator element(s) that could influence *cbsR12* expression. PCR was used to confirm the Tn insertions in the MB-*cbsR12* and MB-*cbsR12*-Comp (**Figure S3.2B**). Furthermore, we confirmed that the *cbsR12* cassette inserted into an IGR between CBU_1788 and CBU_1789 (RSA493: accession number NC_002971.4), and we utilized copy-number quantitative PCR (qPCR) to confirm the single insertional event in MB-*cbsR12*-Comp (**Figure S3.2C**).

Next, we conducted growth curve analyses of MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp strains grown axenically in ACCM-2 (**Figures 3.2A, S3.3A, S3.3B**) and assayed production of CbsR12 at incremental time points from the LCV stage (~ 1-6 d) through the SCV stage (≥7 d) by quantitative real-time PCR (qRT-PCR) (**Figure 3.2B**). Growth curve results showed that MB-*cbsR12* displayed a prolonged lag phase from 1-3 d post-inoculation that was not observed in MB-WT or MB-*cbsR12*-Comp strains (**Figures 3.2A, S3.3A, S3.3B**). Following lag phase, MB-*cbsR12* grew at a slightly increased rate relative to the other strains (6-9 d post-infection), but failed to reach cell numbers seen in the other two strains throughout the assay. The "wild-type" (MB-WT) and complemented (MB-*cbsR12*-Comp) strains produced essentially indistinguishable growth curves. The qRT-PCR results showed that the Tn insertion in MB-*cbsR12* completely abrogated CbsR12 production (**Figure 3.2B**). The results also confirmed

cbsR12's increased expression in LCVs compared to SCVs as copies of CbsR12 per *C. burnetii* genome were highest at 3 d post-inoculation.

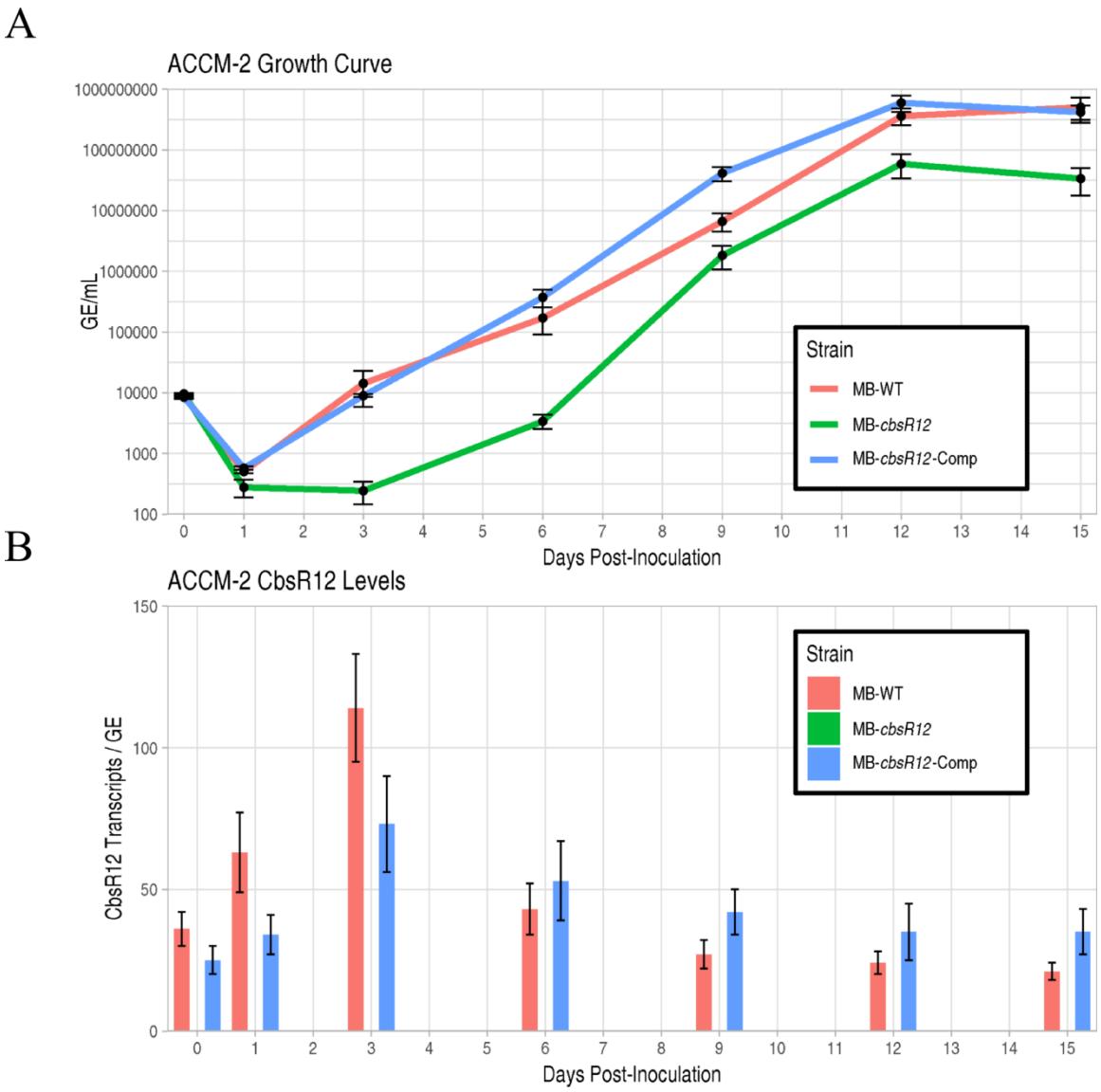


Figure 3.2: CbsR12 production and growth effects on *C. burnetii* grown in ACCM-2. (**A**). Growth curves for MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp in ACCM-2 as determined by qPCR. The 0dpi time point refers to the inoculum. Values represent the means ± standard error of means (SEM) of three technical replicates. Data are representative of three independent experiments with consistent and indistinguishable results. (**B**). CbsR12 production over time for MB-WT, MB-*cbsR12*, and MB-*cbsR12*-

Comp grown in ACCM-2 as determined by qRT-PCR. Values represent the means ± standard error of means (SEM) of three independent determinations.

CbsR12 impacts intracellular replication of *C. burnetii*

C. burnetii typically infects alveolar macrophages during human infection. We therefore infected a differentiated human monocyte cell line (THP-1) with MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp strains. Comparative growth curves showed that MB-*cbsR12* has a slower growth rate in exponential phase (1-3dpi) as compared to the two other strains, and never attained the bacterial cell numbers seen in infections with MB-WT or MB-*cbsR12*-Comp (**Figures 3.3A, S3.3C, S3.3D**). Furthermore, CbsR12 production in THP-1s correlated with replication efficiency of the individual strain. For example, production of CbsR12 in MB-WT and MB-*cbsR12*-Comp strains increased between 1dpi and 3dpi, and CbsR12 levels directly correlated to growth rates of the respective strains between these two time points. However, we observed a dysregulation of CbsR12 in MB-*cbsR12*-Comp infecting THP-1 cells that was strikingly different than what was seen during axenic growth. Specifically, we observed a maintenance of CbsR12 production throughout infection of THP-1s (**Figure 3.3B**), whereas in axenic growth there was a progressive drop-off in synthesis after 3dpi (see **Figure 3.2B**). These results suggest that *cbsR12* expression differs in this host-cell type and that a transcriptional regulatory motif may exist outside the bounds of the complementation insertion, resulting in dysregulation due to the genomic context *of cbsR12* in MB-*cbsR12*-Comp compared to MB-WT in a THP-1 infection.

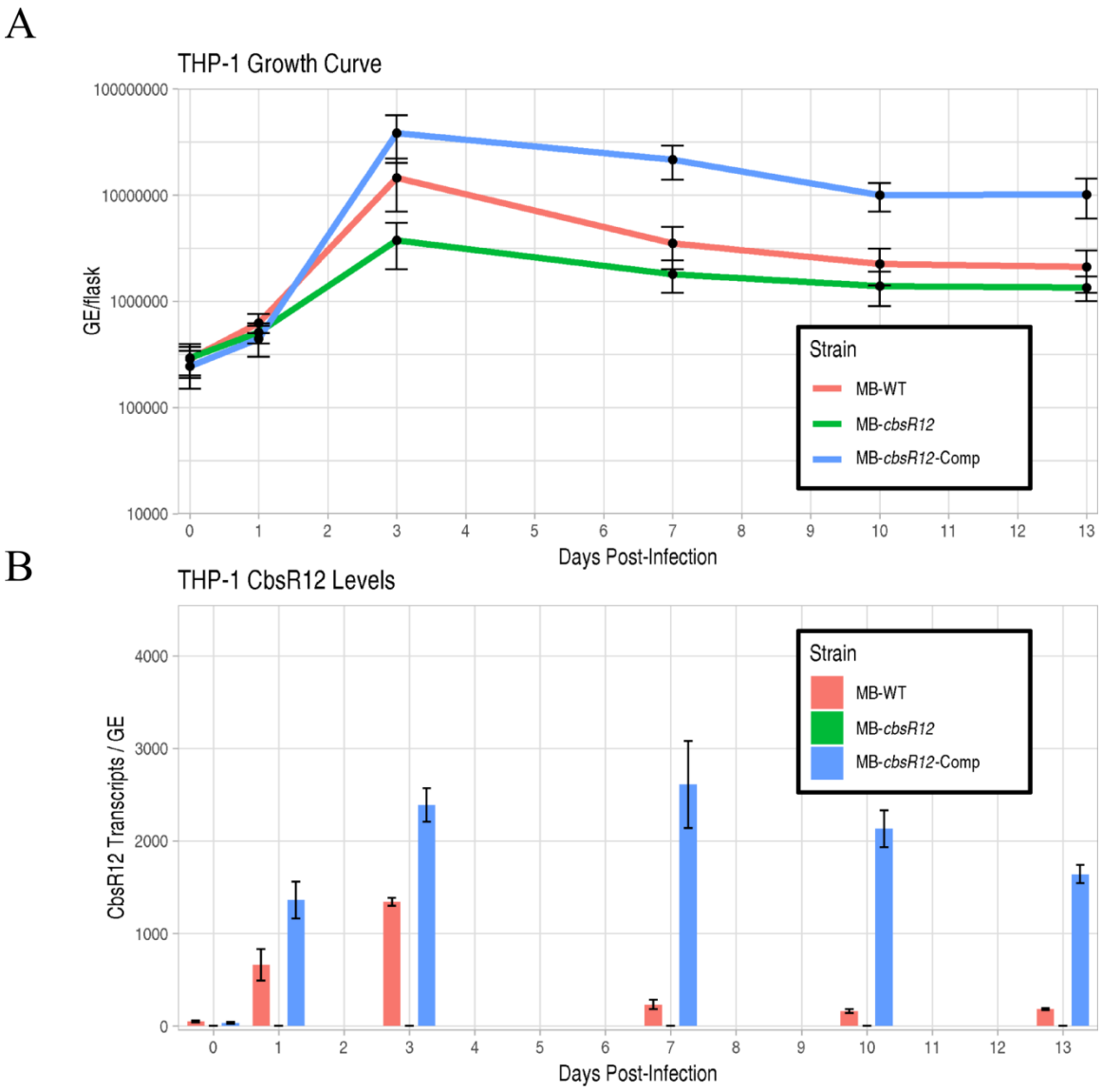


Figure 3.3: CbsR12 production and growth effects on *C. burnetii* infecting THP-1 cells. (**A**). Growth curves for MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp in THP-1 cells as determined by qPCR. The 0dpi time point refers to the inoculum. Values represent means ± standard error of means (SEM) of three technical replicates. Data are representative of three independent experiments with consistent and indistinguishable results. (**B**). CbsR12 production over time for MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp infecting THP-1 cells, as determined by qRT-PCR. Values represent means ± standard error of means (SEM) of three independent determinations.

CCV size correlates with CbsR12 production in THP-1 infection

To more closely examine bacterial-host cell interactions, we employed immunofluorescence assays (IFAs) of *C. burnetii* infecting THP-1 cells. *C. burnetii* colonies and CCV boundaries were visualized using anti-*Coxiella* (anti-Com1 [170]) and anti-LAMP1 antibodies at both 3dpi (late LCVs) and 7dpi (SCVs). Here, we define a *C. burnetii* colony as multiple *C. burnetii* inhabiting a LAMP1-decorated intracellular vacuole. LAMP1 is a host cell protein recruited to lysosomes and found on CCVs after lysosome fusion [171]. We observed a robust infection at 3dpi for MB-WT and MB-*cbsR12*-Comp strains, whereas the MB-*cbsR12* strain only produced a few, small CCVs with relatively unclear boundaries (**Figure 3.4A**). In contrast, the MB-*cbsR12*-Comp strain produced CCVs that were similar in size to those generated by the MB-WT strain, reflecting the trend observed in their respective growth curves (see **Figure 3.3A**). Quantitatively, the differences in CCV sizes between MB-*cbsR12* and the other two strains were significant at 3dpi (**Figure 3.4B**). However, by 7dpi the CCVs were of similar size in MB-*cbsR12* and MB-WT infections, indicating some compensatory or redundant mechanisms allowing for CCV expansion as the infection proceeded, even though MB-*cbsR12* genome counts never reached wild-type levels (see **Figure 3.3A**). Interestingly, the MB-*cbsR12*-Comp strain formed consistently larger CCVs at 3dpi (significantly greater than MB-WT and MB-*cbsR12*) and 7dpi (significantly greater than MB-*cbsR12*), which meshes well with the sustained expression of *cbsR12* evidenced throughout the course of infection (see **Figure 3.3B**). Taken as a whole, these results suggest that CbsR12 is important for optimum growth and establishment of CCVs early

in the course of infection of THP-1 cells, and CbsR12 can influence CCV expansion throughout a THP-1 infection.

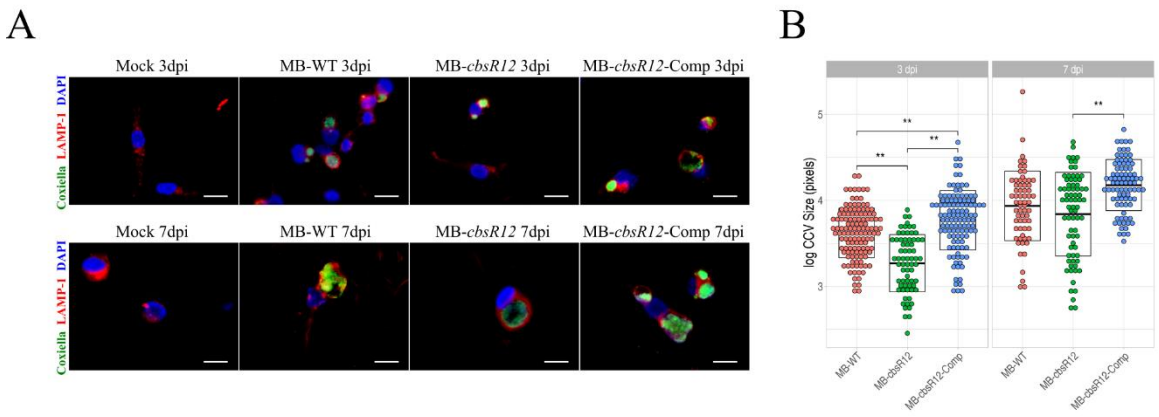


Figure 3.4: CbsR12 affects CCV expansion in infected THP-1 cells. (**A**). Representative IFAs of MB-WT, MB-*cbsR12* and MB-*cbsR12*-Comp infecting THP-1 cells at 3dpi and 7dpi. *C. burnetii* was probed with anti-Com1 antibodies coupled to Alexa Fluor 488 (green), CCV boundaries were labeled with anti-LAMP1 antibodies coupled to rhodamine (red), and host cell nuclei were labeled with DAPI (blue). (**B**). Sizes of individual CCVs in $\log_{10}$(pixels) for MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp. Measurements were taken from 46 individual images of random fields of view spanning three different experiments for each *C. burnetii* strain. Crossbars represent means ± standard error of means (SEM) (** = $P < 0.01$, one-way ANOVA). Scale bars = 20 µm.

CbsR12 binds to *carA*, *metK*, and *cvpD* transcripts *in vitro*

Although CbsR12 was identified as a CsrA-binding sRNA, nothing was known about the CsrA regulon in *C. burnetii*, making it difficult to ascribe intracellular phenotypes to regulation by CsrA. Therefore, we wanted to determine if CbsR12 could act by regulating

mRNAs in *trans*. To identify potential mRNA targets of CbsR12, we first employed three *in silico* sRNA target discovery algorithms. Each algorithm takes into consideration the extent of sRNA-mRNA hybridization, conservation of the sRNA, and the accessibility of both the sRNA and its target, although TargetRNA2 and IntaRNAv2 prioritize accessibility while CopraRNA prioritizes comparative interaction predictions among different strains of the indicated bacterium. From these search results, we omitted genes annotated as hypothetical and chose *cvpD*, *metK*, *carA*, *purH*, *rpsA*, and *dnaA* as potential targets based on conserved predictions (**Table 3.2**). To get a sense of CbsR12's ability to bind to these potential mRNA targets, we next performed a RNA-RNA hybridization followed by EMSAs. The results clearly showed that CbsR12 bound to *carA*, *metK*, and *cvpD* transcripts *in vitro*, but did not interact with *rpsA*, *purH*, or *dnaA* transcripts (**Figure 3.5**). We further tested CbsR12's specificity for these transcripts by performing dose-dependent and unlabeled-chase experiments. Results of the EMSA analyses showed that CbsR12 specifically bound *carA*, *metK*, and *cvpD* transcripts in a dose-dependent manner (**Figure S3.4**).

Table 3.2. CbsR12 target[a] prediction using various algorithms

Rank	TargetRNA2	IntaRNA	CopraRNA
1	CBU_1041 (0.001)	***cvpD* (0.000017)**	***cvpD* (0.00001)**
2	*prlC* (0.001)	CBU_0537 (0.00018)	CBU_0537 (0.000039)
3	*suhB* (0.001)	CBU_1161 (0.00018)	CBU_0922 (0.00053
4	*mutS* (0.002)	CBU_0922 (0.00091)	CBU_0103 (0.00074)
5	*bioD* (0.003)	CBU_2028 (0.0012)	CBU_2028 (0.00078)
6	*trmD* (0.003)	CBU_0103 (0.0013)	***rpsA* (0.0013)**
7	***dnaA* (0.005)**	***rpsA* (0.0022)**	***metK* (0.0014)**
8	*yciL* (0.007)	***metK* (0.0023)**	***purH* (0.0019)**
9	***metK* (0.008)**	***purH* (0.003)**	CBU_1741 (0.0025)
10	CBU_0558 (0.01)	***carA* (0.015)**	***carA* (0.0067)**

[a]Target genes identified through algorithms, ranked. Calculated p-values for each target are indicated in parentheses. Genes involved in subsequent experiments are indicated by bold-face type.

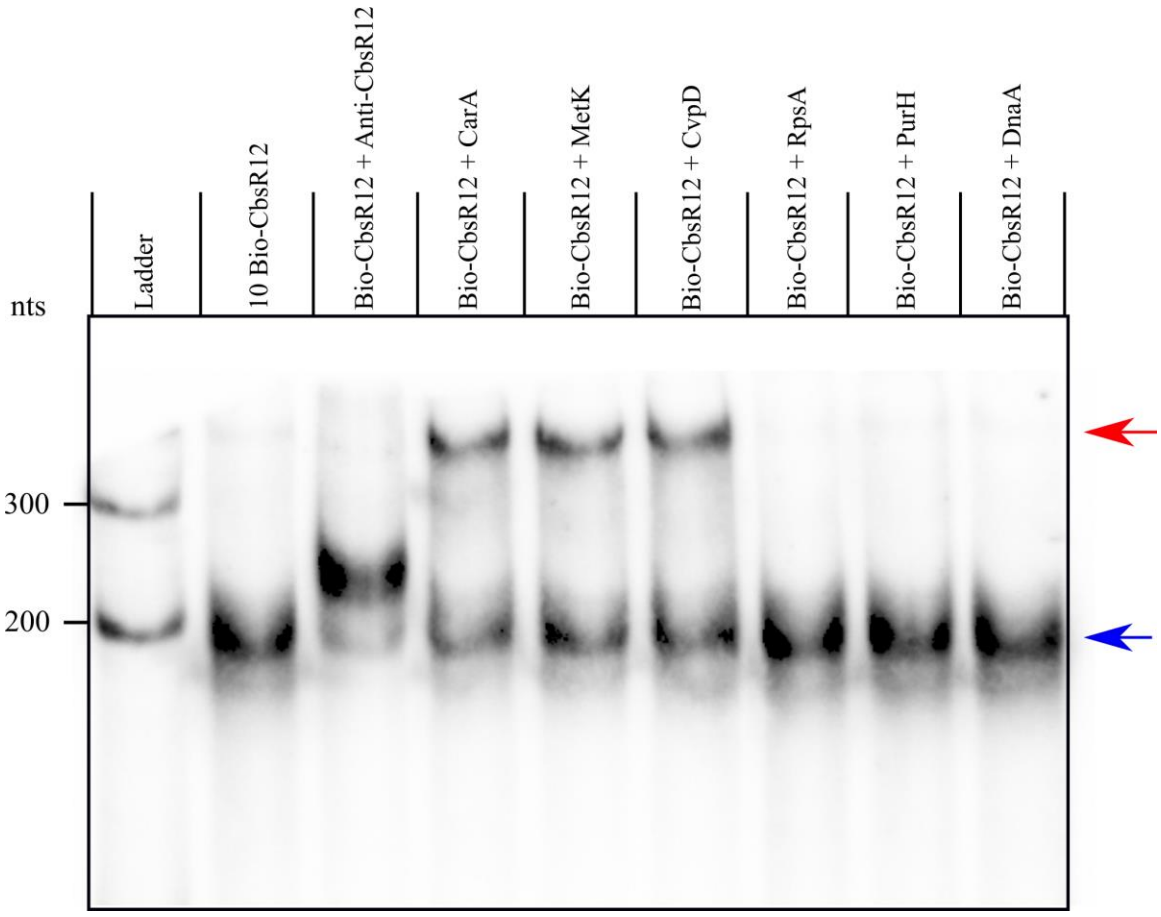


Figure 3.5: CbsR12 targets *carA*, *metK*, and *cvpD* transcripts *in vitro*. RNA-RNA EMSA showing hybridization reactions with 10 nM biotin-labeled CbsR12 and 5 nM *in vitro*-transcribed segments of *carA*, *metK*, *cvpD*, *purH*, *rpsA* or *dnaA*. Anti-CbsR12 represents a 10 nM positive control consisting of a transcript equal in size, but antisense, to the CbsR12 transcript. Arrows indicate un-bound bio-CbsR12 (blue) and bio-CbsR12 bound to RNA targets (red).

CbsR12 binds to *metK, carA, cvpD,* and *ahcY* transcripts in *C. burnetii* cells

To determine the CbsR12 targetome within *C. burnetii* cells, we employed a Crosslink-Seq technique previously used to detect intracellular mRNA targets of *E. coli* sRNAs [152]. For this procedure, we used *C. burnetii* LCVs grown in ACCM-2 to produce sufficient volumes to capture CbsR12 target RNAs for cDNA library preparation and RNA-Seq analysis. Hybridized RNAs in lysates from both MB-WT and MB-*cbsR12* strains were chemically cross-linked, captured by anti-CbsR12 probes, and analyzed by RNA-Seq in order to identify RNAs enriched in MB-WT compared to MB-*cbsR12*. Crosslink-Seq results confirmed that CbsR12 targeted *carA*, *metK*, and *cvpD* transcripts in *C. burnetii* cells (**Figure 3.6**), as demonstrated *in vitro* (see **Figure 3.5**). We also discovered an additional mRNA target, *ahcY*, coding for adenosylhomocysteinase, another component of the methionine cycle. Interestingly, *ahcY* was also predicted as an mRNA target by IntaRNA, although the p-value was not significant ($p = 0.13$). Additionally, *ahcY* is in an operon with and downstream of *metK*. To address whether *ahcY* is actually a target of CbsR12, or if it is a result of CbsR12's binding to a polycistronic mRNA, we used the Artemis genome browser to observe Crosslink-Seq reads aligned to the *C. burnetii* RSA439 genome. This analysis showed distinct segments of these genes to which the captured reads mapped, suggesting they are separate binding events (**Figure S3.5B**). We also confirmed the other identified targets by the same method (**Figures S3.5A, S3.5C**).

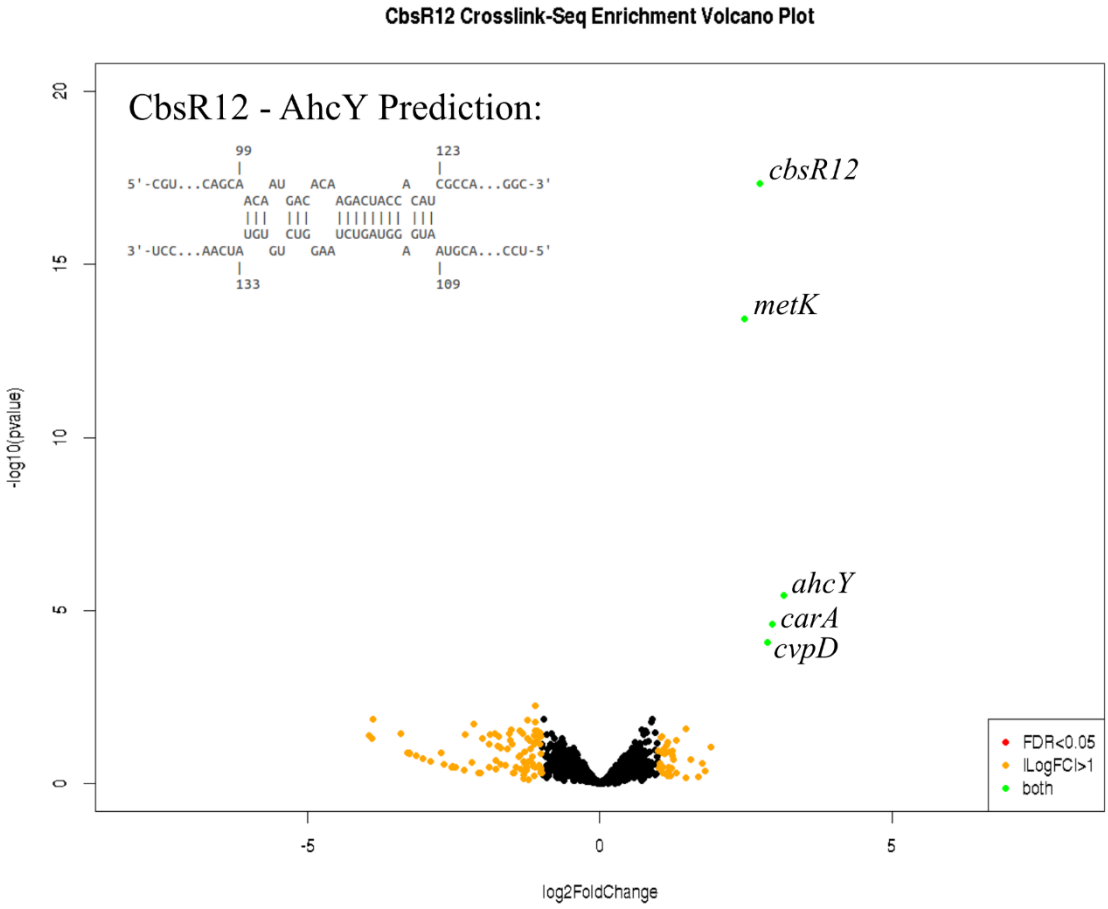


Figure 3.6: CbsR12 targets several *C. burnetii* transcripts, including those of *metK*, *carA* and *cvpD*. Volcano plot highlighting mRNAs that occur at different levels between Crosslink-Seq experiments with strains MB-WT and MB-*cbsR12*. Labeled transcripts are indicated by a green dot and are significantly enriched in MB-WT versus MB-*cbsR12*, identifying them as targets of CbsR12. Black dots represent transcripts not significantly different between the strains tested. Orange dots represent transcripts having a $\log_2$-fold change > 1, but a false discovery rate (FDR) > 0.05. There were no transcripts indicated by red dots, which would represent mRNAs with a FDR < 0.05 but a $\log_2$-fold change in levels < 1. Data shown are representative of two biological replicates each of MB-WT

and MB-*cbsR12* Crosslink-Seq experiments. The potential CbsR12-binding site in the coding region of the *ahcY* transcript is inset.

Next, we wanted to determine whether CbsR12-mediated regulation of predicted in-*trans* mRNA targets would occur independently of CsrA. To this end, we searched for potential CsrA-binding (AGGA/ANGGA) motifs within the 100 bases up and downstream of the start codons of *carA*, *metK*, *cvpD*, and *ahcY*. This search showed that *carA*, *metK* and *cvpD* contained single potential CsrA-binding sites whereas *cvpD* had none. Moreover, the motifs of *carA* and *metK* did not occur in predicted ribosome-binding sites (RBS), suggesting that CsrA is unlikely to regulate the corresponding transcripts [33] (**Table 3.3**). Thus, CbsR12 regulation of *carA*, *metK*, and *cvpD* transcripts likely occurs as a direct result of in-*trans* binding by CbsR12 and independently of CsrA. In contrast, the *ahcY* sequence has two potential regulatory ANGGA CsrA-binding sites, so we cannot exclude the possibility of indirect regulatory effects caused by CbsR12-mediated sequestering of CsrA. As such, we did not further explore CbsR12-mediated regulation of *ahcY*.

Table 3.3. Predicted CsrA motifs in CbsR12 targets

CbsR12 Target Gene	Locus Tag[a]	Base Range[b]	CsrA Motifs[c]	Position of Motif[d]
carA	CBU_1282	c(1234872-1235073)	1	(+71) – (+75)
metK	CBU_2030	1936983-1937183	1	(-12) – (-8)

cvpD	CBU_1818	1748916-1749116	0	N/A
ahcY	CBU_2031	1938166-1938366	2	(+1) – (+5), (+13) – (+17)

[a]Locus tags as annotated for RSA493 (NCBI Ref Seq: NC_002971.4). [b]Base range indicates 100 nts up and downstream of the annotated start codon for RSA493 (NCBI Ref Seq: NC_002971.4). [c]Consensus CsrA-binding motifs are AGGA or ANGGA. [d]Position of CsrA-binding motifs relative to the first nucleotide of the start codon (+1).

CbsR12 negatively affects the quantity of *cvpD* transcripts and regulates synthesis of CarA and MetK

Next, we set out to determine if CbsR12 regulates *carA*, *metK*, and *cvpD* transcripts in *C. burnetii*. First, we performed 5' RACE on the three transcripts in total RNA extracted from LCVs infecting mammalian cells (3dpi). 5' RACE results for the MB-WT *cvpD* gene indicated three apparent TSSs, including a TSS for the full-length transcript, several questionable "TSSs" within the CbsR12-binding site, and an alternative TSS for a short transcript downstream of the CbsR12-binding site and with its own predicted promoter element (**Figure S3.6A**). Interestingly, putative RBSs and start codons exist downstream of TSSs for both the full-length and short transcripts. Moreover, the two start codons are in-frame with each other and the existence of putative RBSs supports the possibility that translation occurs from both elements. The questionable "TSSs" within the CbsR12-binding region likely result from CbsR12-mediated RNase III degradation of *cvpD* mRNA, because 5' RACE results for MB-*cbsR12*, a strain that lacks CbsR12, did not produce TSSs in this region. We predict that CbsR12 down-regulates production of full-

length CvpD since the CbsR12-binding site occurs in the coding region. However, CbsR12 would predictably have no effect on production of the putative truncated CvpD, as the CbsR12-binding site occurs upstream of the alternative TSS (**Figures S3.6A, S3.6B**).

To determine if CbsR12 binds to and causes degradation of full-length *cvpD* transcripts, we performed qRT-PCR on MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp LCVs obtained from infected THP-1 cells. These results clearly showed that the absence of CbsR12 in strain MB-*cbsR12* led to a significant increase in full-length *cvpD* transcripts in LCVs (3dpi) (**Figure S3.6C**). At 7dpi, MB-WT and MB-*cbsR12* levels were not significantly different, presumably due to reduced CbsR12 production in MB-WT SCVs (see **Figure 3.3B**). However, *cvpD* expression was significantly lower in MB-*cbsR12*-Comp, most likely due to the maintained production of CbsR12 in the strain's SCVs (see **Figure 3.3B**). Whether or not two different forms of CvpD are produced from *cvpD* is unknown, although it appears that CbsR12 may negatively regulate the full-length *cvpD* transcript.

In order to determine if CbsR12 binds to and regulates *carA* and *metK* in a cellular environment, we devised a reporter assay in *E. coli* that measures the effects of CbsR12 production on translation of *carA*-*luc* or *metK*-*luc* fusion constructs. 5' RACE results for both MB-WT and MB-*cbsR12* strains revealed that *carA* has two potential TSSs, a finding that is consistent with transcription of *E. coli carA* [165]. Based on the position of the TSSs, CbsR12 could only regulate the full-length *carA* transcript and not the shorter mRNA, whose transcription starts immediately upstream of the RBS and downstream of the CbsR12-binding site (**Figure 3.7A**). From these results, we hypothesized that CbsR12

binds to the 5' UTR of *carA* and upregulates translation by relieving the secondary structure that occludes the predicted RBS (**Figures 3.7A, 3.7B**). Results of the *E. coli* reporter assay confirmed our hypothesis, because translation of luciferase enzyme from a *carA5'UTR-luc* fusion was significantly upregulated in the presence of CbsR12 relative to a strain lacking the sRNA (**Figure 3.7C**).

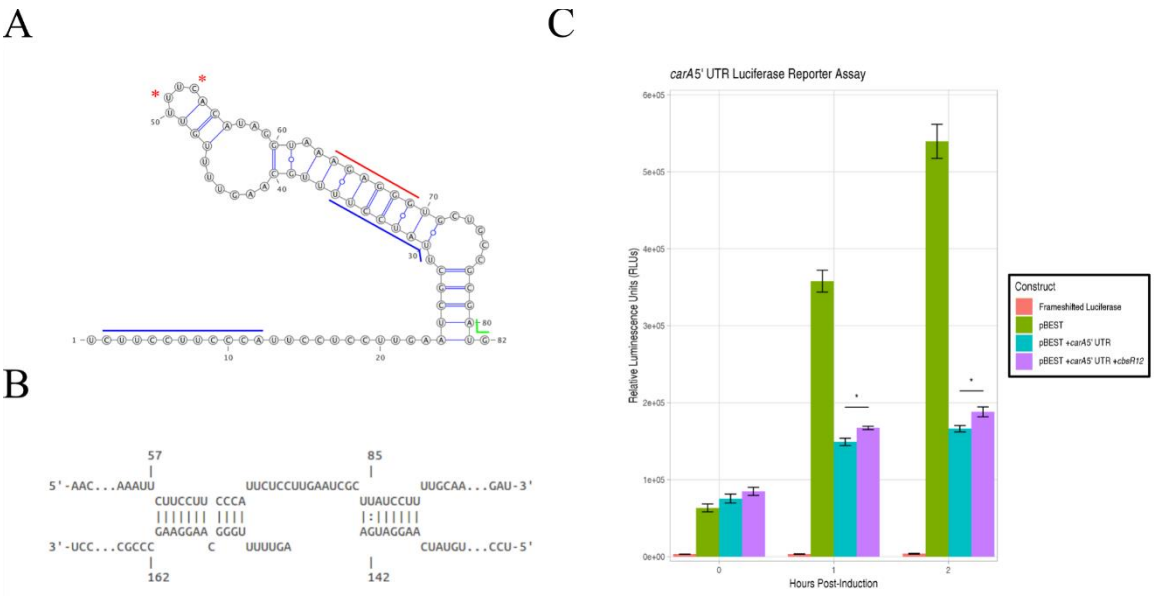


Figure 3.7: CbsR12 targets and upregulates translation of a *carA*-luciferase fusion construct. (**A**). Secondary structure of the *carA* 5' UTR as predicted by mFold. Red asterisks indicate TSSs for the shorter transcripts as determined by 5' RACE. (Nucleotide 1 was determined to be the TSS for the full-length transcript by 5' RACE). Colored lines represent the start codon (green), predicted RBS (red), and determined CbsR12-binding sites (blue). (**B**). Representation of CbsR12 binding to the *carA* transcript as determined by IntaRNA, with respective base numbers indicated. The top strand in the model represents the *carA* sequence, while the bottom strand represents the complementary CbsR12 sequence. (**C**). *carA-luc* reporter assay indicating relative luminescence units produced by pBESTluc constructs with: 1) no luciferase production (Frameshifted

Luciferase), 2) pBESTluc vector (pBEST), 3) pBESTluc with the *carA* 5' UTR upstream of *luc* but lacking *cbsR12* (pBEST + *carA* 5' UTR), and 4) *carA* 5' UTR upstream of *luc* plus the *cbsR12* gene driven by a P*tac* promoter (pBEST + *carA* 5' UTR + *cbsR12*). Values represent means ± standard error of means (SEM) of three independent determinations (* = $P < 0.05$, student's *t* test).

In contrast, CbsR12 was predicted to downregulate MetK translation by binding to the coding region of the transcript, immediately downstream of its start codon (**Figure 3.8A**). As is often the case with this type of sRNA-mediated regulation, RNase III would likely be recruited and the *metK* transcript cleaved, resulting in downregulation of the encoded protein product. Unexpectedly, 5' RACE analyses of *metK* mRNA also identified apparent alternative "TSSs" within the CbsR12-binding region, suggesting that the truncated mRNAs resulted from CbsR12-mediated RNase III processing (**Figures 3.8A, 3.8B**). Indeed, 5' RACE analysis of RNA from strain MB-*cbsR12* infecting THP-1 cells did not detect the "TSSs", suggesting they are a product of RNase III processing. Results of the reporter assays in *E. coli* confirmed our hypothesis, as the presence of CbsR12 significantly down-regulated translation of luciferase from the *metK-luc* fusion construct compared to a strain lacking the sRNA (**Figure 3.8C**).

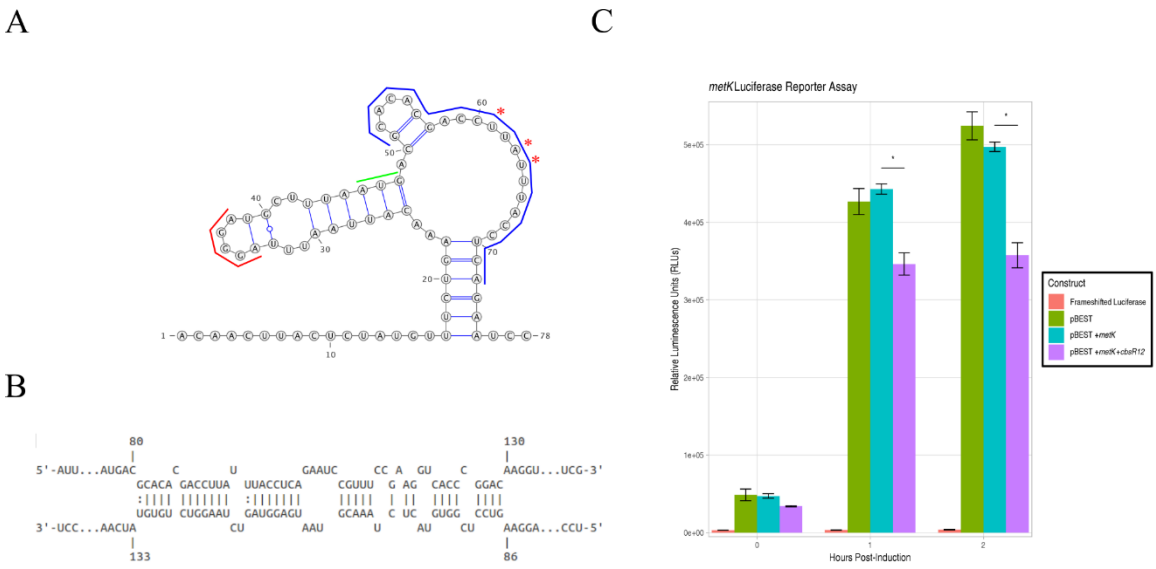


Figure 3.8: CbsR12 targets and downregulates translation of a *metK*-luciferase fusion construct. (**A**). Secondary structure of the *metK* 5' UTR and initial coding sequence as predicted by mFold. Red asterisks indicate apparent alternative "TSSs" determined by 5' RACE. (Nucleotide 1 was determined to be the TSS for the full-length transcript by 5' RACE). Colored lines represent the start codon (green), a predicted RBS (red), and the determined CbsR12-binding site (blue). (**B**). Representation of CbsR12 binding to the *metK* transcript as determined by IntaRNA with base numbers indicated. The top strand in the model represents the *metK* sequence, while the bottom strand represents the complementary CbsR12 sequence. (**C**). *metK-luc* reporter assay indicating relative luminescence units produced by pBESTluc constructs with: 1) no luciferase production (Frameshifted Luciferase), 2) pBESTluc vector (pBEST), 3) pBESTluc with the CbsR12 binding site cloned in-frame into *luc* but lacking the *cbsR12* gene (pBEST + *metK*) and 4) pBESTluc with the CbsR12 binding site cloned in-frame into *luc* plus the *cbsR12* gene driven by a P*tac* promoter (pBEST + *metK* + *cbsR12*). Values represent

means ± standard error of means (SEM) of three independent determinations (* = P < 0.05, student's *t* test).

Although we determined that CbsR12 targets *carA* and *metK* transcripts *in vitro* and in *E. coli* cells, we were curious whether the absence of CbsR12 would also result in differential amounts of CarA and MetK proteins in *C. burnetii*. To this end, we performed Western blots with polyclonal antibody generated against recombinant *C. burnetii* CarA and MetK. As predicted, when proteins from whole-cell lysates of MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp strains were compared, we found that CarA was synthesized in MB-WT and MB-*cbsR12*-Comp strains at comparable levels, but was undetectable in protein profiles of strain MB-*cbsR12* (**Figures 3.9A, S3.7A**). In sharp contrast, MetK was highly synthesized in strain MB-*cbsR12* but was produced at relatively lower and comparable levels in the MB-WT and MB-*cbsR12*-Comp strains (**Figures 3.9B, S3.7B**).

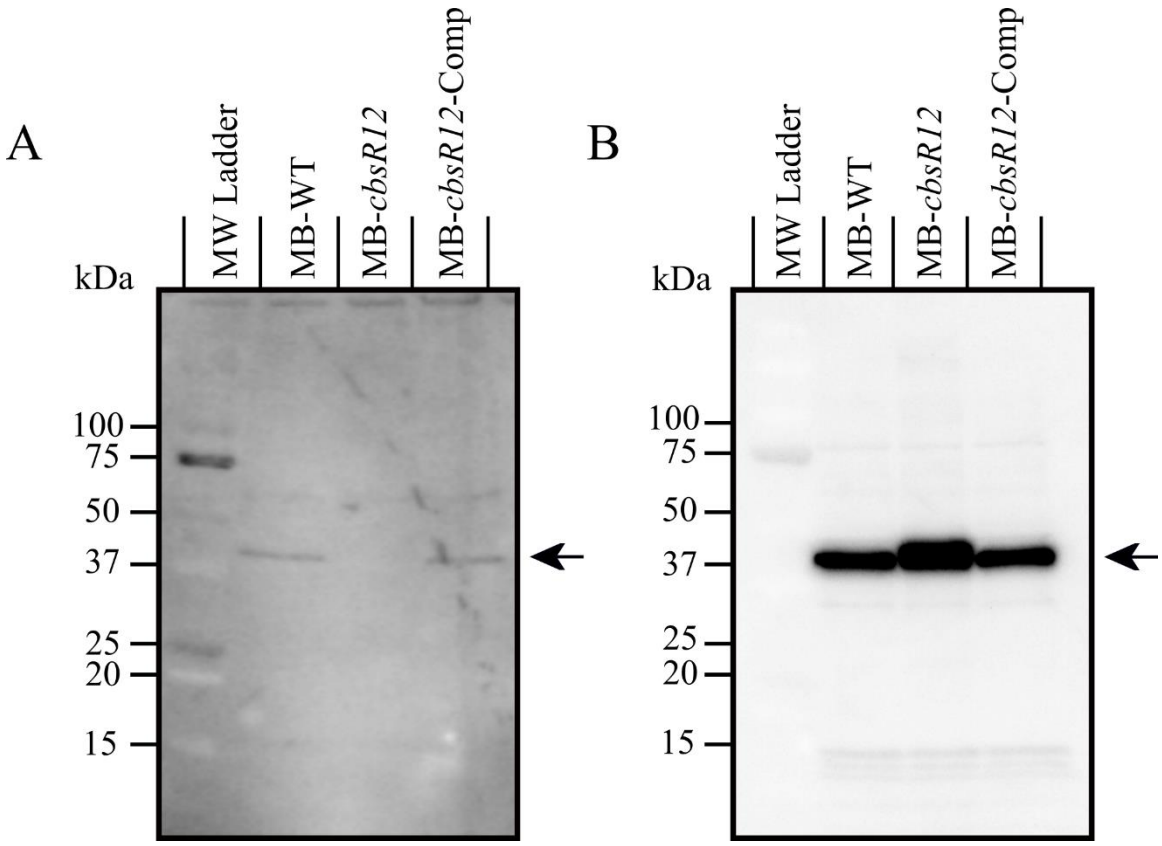


Figure 3.9: CarA and MetK proteins are differentially synthesized in MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp strains. (**A**). Proteins (30 µg total) from MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp LCVs (mid-log phase; 96h for MB-WT and MB-*cbsR12*-Comp and 144h for MB-*cbsR12*) grown in ACCM-2 were resolved on a 10-20% acrylamide gradient SDS-PAGE gel, blotted, probed with rabbit anti-CarA antibodies, and detected with chemiluminescence. The black arrow indicates CarA. (**B**). Proteins (60 µg total) from MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp LCVs (mid-log phase; 96h for MB-WT and MB-*cbsR12*-Comp and 144h for MB-*cbsR12*) grown in ACCM-2 were resolved on a 10-20% acrylamide gradient SDS-PAGE gels, blotted, probed with rabbit anti-MetK antibodies, and detected with chemiluminescence. The black arrow indicates MetK.

Discussion

In this report, we show that CbsR12, a multifunctional sRNA that binds to mRNAs and to the regulatory protein CsrA-2, is important for proper *Coxiella* replication and CCV expansion during infection of human macrophage-like THP-1 cells.

Induction of *cbsR12* expression in mammalian cell culture vs. *in vitro* conditions (see **Table 3.1**) led us to label CbsR12 as "infection-specific" and as a result we hypothesized that it played an important regulatory role in infection of host cells. The *cbsR12* sequence is conserved among all *C. burnetii* strains sequenced to date, underscoring the potential for an important regulatory role, but from an evolutionary viewpoint. Interestingly, though, the *cbsR12* gene is missing or degenerate in *Coxiella*-like endosymbionts [172], suggesting that the sRNA is important for a mammalian infection but is dispensable in endosymbionts that reside in arthropods.

Interestingly, CbsR12 binds rCsrA-2, but not rCsrA-1, in a dose-dependent manner, *in vitro.* Why CsrA-1 does not bind to consensus motifs present in CbsR12 is unclear, especially since both CsrA-1 and CsrA-2 maintain the critical L4 and R44 RNA-binding residues, although CsrA-1 has these residues at L4 and R46 [173]. There are several examples of pathogens harboring multiple copies of CsrA [174, 175]. For example, RsmF of *P. aeruginosa* is a homolog of RsmA (CsrA) and functions by binding a subset of mRNAs that RsmA also binds [175]. However, an *rsmF* mutant did not display a phenotype during infection [175]. Similarly, In *C. burnetii*, a transposon-mediated *csrA-1* mutant was shown to have no intracellular phenotype [87]. It is also conceivable that CsrA-1 diverged during *C. burnetii*'s adaptation to an intracellular lifestyle and is no longer functional. It is also possible that CsrA-1 binds to a non-

canonical motif not present in CbsR12, although such CsrA homologs have not been described, to our knowledge. Regardless, it is necessary to examine the CsrA-1 and CsrA-2 regulons in order to determine their respective roles during infection.

In *L. pneumophila*, a close relative of *C. burnetii*, successful infection depends on a LetAS-RsmYZ-CsrA regulatory cascade. LetAS is a TCS that regulates production of two sRNAs, RsmY and RsmZ, which in turn act as RNA "sponges" that soak up CsrA and modulate its activity [33]. We determined that CbsR12 possesses only four CsrA-binding sites, similar to the RsmY/Z sRNAs of *L. pneumophila*. Interestingly, *L. pneumophila* RsmY/Z was implicated in the formation of cell aggregates and biofilms when the sRNAs were ectopically overproduced in *E. coli*, mimicking the effects of *E. coli*'s own CsrA-binding sRNAs [31]. Likewise, when we overproduced CbsR12 in *E. coli* reporter assays (**Figures 3.7, 3.8**), we observed a similar autoaggregative phenotype (**Figure S3.8A**). Moreover, CbsR12 induced biofilm production in *E. coli*, reflecting a CsrA-depleting phenotype (**Figure S3.8B**). Together, these results suggest that CbsR12 is a CsrA-sequestering, RsmY/Z-like sRNA, although further research is necessary to determine the exact influence of CbsR12 on the regulatory roles of CsrA-2.

C. burnetii has several potential transcription factors that are known to upregulate bacterial expression, including IHF [149], response regulator PmrA [176], and transcription factor DksA [177]. However, only PmrA has been well-studied, to date [91]. It is interesting to note that the *cbsR12* gene contains a close approximation to a PmrA-binding site (consensus sequence with less conserved nucleotides in lowercase: cTTAA-N_2-TT-N_2-cTTAA) [178] immediately upstream of its predicted -10 promoter element (*cbsR12* sequence: gTTTA-N_2-TT-N_1-gTTAA). However, the presence of this sequence

does not explain the prolonged CbsR12 production observed during a THP-1 infection by MB-*cbsR12*-Comp since the putative PmrA-binding sequence is present in the *cbsR12* cassette that was inserted. We predict that expression of *cbsR12* is regulated by an unidentified TCS in a fashion similar to the *L. pneumophila* LetAS TCS regulation of RsmYZ sRNAs [33]. In fact, the LetAS TCS is not unlike the GacAS TCS involved in RsmYZ-CsrA cascades of other bacteria [179]. *C. burnetii* codes for four different GacA response regulators that could bind upstream elements of *cbsR12* and regulate its expression [99]. This upstream regulator may, in turn, help to explain the dysregulation of expression seen in MB-*cbsR12*-Comp during infection of THP-1 cells (see **Figure 3.3B**) that is not apparent during axenic growth (see **Figure 3.2B**). Alternatively, expression of *cbsR12* could be upregulated by PmrA, and some other regulator may be involved in its down-regulation, in conjunction with RNase III-mediated decay (see **Figure S3.1**). Together, these would aid in suppression of CbsR12 as the LCV-to-SCV transition occurs, effectively freeing sequestered CsrA-2 to regulate the fate of target transcripts.

It is worth noting that we have identified a second sRNA, ***C****oxiella* ***b****urnetii* **s**mall **R**NA **1** (CbsR1), that possesses 5 putative CsrA-binding sites with an ANGGA motif (**Figure S3.9B**) [97]. Similar to CbsR12, CbsR1 is also produced at high levels in LCVs infecting Vero cells (see **Table 3.1**). Furthermore, *cbsR1* harbors a putative LetA-binding site similar to that of *L. pneumophila* RsmY (**Figure S3.9A**) [31]. Together, these observations suggest that CbsR1 and CbsR12 may represent orthologs of RsmYZ, although further exploration of CbsR1 and its cooperativity with CbsR12 is required. If CbsR1 does indeed serve as a CsrA-binding sRNA, its potential, redundant regulatory

role may help to explain why MB-*cbsR12* CCVs expanded to wild-type sizes as the infection progressed (see **Figure 3.4B**).

We also found that CbsR12 binds *carA* transcripts (see **Figures 3.5, 3.7**) and upregulates production of *C. burnetii* CarA (see **Figure 3.9A**). Pyrimidine metabolism in *C. burnetii* presumably requires CarAB to catalyze the conversion of L-glutamine into carbamoyl phosphate and glutamate, since it is unable to shunt this process through the arginine dihydrolase pathway; *C. burnetii* apparently lacks the necessary enzymes [180]. CbsR12-mediated upregulation of CarA in LCVs would result in increased production of pyrimidines that the pathogen requires for robust intracellular growth.

In *E. coli*, *carA* expression is tightly controlled by a series of transcriptional regulators and the two distinct promoters that are regulated by feedback from arginine and pyrimidines [165]. 5' RACE analysis showed two distinct TSSs for *C. burnetii carA* mRNA, with the full-length transcript containing two CbsR12-binding sites and a shorter putative transcript lacking the site (see **Figure 3.7A**). We do not believe that the shorter, alternative TSS is due to RNase III-mediated degradation resulting from CbsR12 binding because this alternative TSS remained in 5' RACE analysis of the MB-*cbsR12* strain. We do not know conditions under which the shorter transcript is produced, but it may involve feedback from arginine/pyrimidine in accordance with *carA* regulation in *E. coli*.

We also showed that CbsR12 binds to *metK* transcripts and downregulates production of *C. burnetii* MetK protein (see **Figures 3.5, 3.8, 3.9B**). MetK is a key component of the methionine cycle, which converts methionine to SAM via MetK, SAM to *S*-adenosylhomocysteine (SAH) via various methylases, SAH to homocysteine via AhcY, and homocysteine to methionine via MetH/MetE. Cells produce homocysteine as

an input molecule through a series of reactions involving activated homoserines (reviewed in [181]). *C. burnetii* is a semi-auxotroph for methionine, since it can potentially grow without methionine in axenic media, albeit at a slower growth rate [182]. Interestingly, *C. burnetii* lacks several components of the methionine synthesis pathway, most notably the ability to produce activated homoserines. Most bacteria activate homoserine through addition of an O-succinyl group catalyzed by MetA or an O-acetyl group catalyzed by MetX (reviewed in [181]). *C. burnetii* apparently lacks genes coding for these enzymes. An ABC methionine transporter has been hypothesized [182] but not verified in *Coxiella*. If this is indeed a functional transporter, CbsR12's negative regulation of *metK* transcripts makes sense in the context of the sRNA's high level in LCVs, because any amount of scavenged methionine would be critical to growth. Shifting the equilibrium from SAM synthesis to methionine retention would be necessary as *C. burnetii* rapidly produces proteins to expand its intracellular niche.

SAM is a major methyl donor, is necessary for regulation of numerous enzymes, and has been implicated as a major contributor to virulence [183, 184]. Some bacteria lack *metK* and instead transport SAM directly [185]. There are many uncharacterized transporters encoded in the *C. burnetii* genome, so it is conceivable that a SAM transporter is present [99]. This would allow for SAM scavenging even when MetK production is downregulated by CbsR12. Furthermore, if SAM transport occurs, *C. burnetii* could synthesize methionine without having to scavenge it, since the amino acid can be synthesized from SAM without activated homoserine via the methionine cycle.

We also determined that CbsR12 actively targets *ahcY* transcripts (see **Figure 3.6**). AhcY is a component of the methionine cycle and catalyzes conversion of SAH into

homocysteine and adenosine. Based on the location of the CbsR12-binding site in the coding region of the *ahcY* transcript (**Figure 3.6 inset**), we predict that CbsR12 negatively regulates AhcY translation. The underlying reason for this negative regulation is unknown, although it could help to suppress adenosine and/or homocysteine accumulation in LCVs.

cvpD mRNA was also identified as a CbsR12 target through Crosslink-Seq (see **Figure 3.6**), and this was confirmed by RNA-RNA hybridization / EMSA and qRT-PCR analyses (see **Figures 3.5, S3.6C**). In this study, we found that CbsR12 was necessary for CCV expansion in early stages of a THP-1 infection. The mechanism for this is unclear, although it may involve regulation of *cvpD,* which is required for *C. burnetii*'s intracellular replication and CCV expansion in infected THP-1 and HeLa cells [84]. CbsR12 is predicted to target the coding region of the *cvpD* transcript and would negatively regulate translation. However, in the context of *cbsR12*'s expression pattern, this is an unclear association, as one would expect upregulation of CvpD synthesis at a time when CbsR12 is highly produced in LCVs. However, 5' RACE analysis of *cvpD* transcripts in MB-WT and MB-*cbsR12* provides a potential explanation, as an alternative *cvpD* promoter downstream of the CbsR12-binding site occurs that also possesses a putative RBS and start codon (see **Figure S3.6A**). From these results, we hypothesize that there are two gene isoforms of *cvpD* that are transcribed and differentially expressed depending on the *C. burnetii* morphotype. Due to high expression of *cbsR12* in LCVs, the longer *cvpD* transcript isoform would be downregulated by RNase III. As expression of *cbsR12* decreases as the infection proceeds, the longer transcript isoform would accumulate. qRT-PCR data support this explanation, since a lack of CbsR12 in MB-

cbsR12 significantly increased the quantity of long *cvpD* isoform transcripts (see **Figure S3.6C**). This hypothesis could be confirmed if the two putative CvpD products could be identified and distinguished.

The 2007-2010 Dutch outbreak involving *C. burnetii* yielded several newly annotated genomes specific to that epidemic [186]. Curiously, 7 of 13 strains analyzed contained a frameshift deletion in *cvpD*, leading to premature stop codons [186]. Among these, strains 18430 (NZ_CP014557.1), 14160-001 (NZ_CP014551.1), 701CbB1 (NZ_CP014553.1), and 2574 (NZ_CP014555.1) had single-base deletions that only affected the long *cvpD* isoform. These strains were isolated from aborted placentas of ruminants and cattle in the Netherlands and France [186]. Additional related strains include the Heizberg (NZ_CP014561.1), Henzerling (NZ_CP014559.1), and RSA 331 (NC_010117.1). These strains, which were isolated from patients with acute Q fever in northern Italy and Greece in the mid 1900's, harbored 4-bp frameshift deletions near the middle of the *cvpD* coding region, affecting both long and short *cvpD* isoforms and introducing premature stop codons [186, 187]. Apparently, CvpD was dispensable for virulence in the latter strains, whereas the Dutch outbreak strains harbored intact *cvpD* genes, or *cvpD* genes with a 1-bp frameshift deletion only affecting the longer gene isoform. Thus, in these Dutch isolate strains it appears that CbsR12 regulation of *cvpD* is dispensable. Granted, there are many genotypic differences between RSA439 and the Dutch isolates [186], and some compensatory mechanism(s) may exist for the absence of *cvpD*. Alternatively, *cvpD* may be necessary during infection of human cell lines and dispensable in host-animal infections. Regardless, the role and regulation of the CvpD effector requires further investigation.

cbsR12's high level of expression during infection likely facilitates and regulates the many functions we have described. CbsR12 not bound to CsrA-2 presumably acts in *trans* to facilitate efficient replication through translational up-regulation of CarA and down-regulation of MetK, and perhaps potentiates expansion of the CCV by means of *cvpD* transcript regulation. Furthermore, regulation is most likely independent of CsrA, because these genes lack multiple CsrA-binding sites necessary for regulation (see **Table 3.3**). It is worth noting that the CsrA-binding sites of CbsR12 do not overlap with *metK* and *cvpD* binding sites. Hence, it is feasible that CbsR12 may still regulate *metK* and *cvpD* in *trans* while bound to CsrA-2; in fact, a chaperone-like function such as this has recently been ascribed to CsrA [188].

CbsR12 is unique in that is one of only a few identified *trans*-acting sRNAs that also binds CsrA (reviewed in [189]). We hypothesize that CbsR12's role in regulating *C. burnetii* replication and CCV expansion is due to a combination of in-*trans* mRNA (*metK*, *carA*, *cpvD*) targeting and regulation of CsrA-2 function. Our lab is currently working to elucidate the CsrA-1/CsrA-2 regulons, along with regulation of the putative CbsR12-CsrA-2 cascade of *C. burnetii* and the nature of CsrA-CbsR1 binding to clarify the interplay between CbsR12's roles as a *trans*-acting and CsrA-sequestering sRNA.

Materials and Methods

Bacterial strains, cell lines and growth conditions

Strains, primers, and plasmids used in this study are listed in **Figure S3.10**. *E. coli* was grown in lysogeny broth (LB) supplemented with ampicillin (100 µg/mL) or kanamycin (50 µg/mL), as needed. When necessary, overnight cultures were expanded to 100 mL

LB, grown for 2 h, then supplemented with 1 mM IPTG for induction. *C. burnetii* Nine Mile phase II (strain RSA439, clone 4), MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp were grown in ACCM-2 medium [98] supplemented with ampicillin (5 µg/mL) or kanamycin (350 µg/mL) at 5% CO_2, 2.5% O_2, 92.5% N_2 37°C, and 100% humidity with continuous shaking at 75 RPM [98]. SCVs collected from Vero cells were used for all *C. burnetii* infections and growth curve experiments. Briefly, *C. burnetii* was used to infect Vero cell monolayers for 7d at 5% CO_2 and 37°C, after which the cultures were removed to room temperature and the flask lids tightened and covered for two additional weeks [190]. Following this, SCVs were harvested with digitonin, as previously described [100].

African green monkey kidney (Vero) epithelia (CCL-81; American Type Culture Collection; ATCC) and human monocytic leukemia (THP-1) cells (TIB-202; ATCC) cell lines were maintained in RPMI medium (Gibco) supplemented with 10% fetal bovine serum (FBS, RMBIO) in a humidified atmosphere with 5% CO_2 at 37°C. THP-1 cells were differentiated to macrophages by supplementing the growth medium with 200 nM phorbol myristate acetate (PMA, Sigma) overnight.

Plasmid construction

pBESTluc was used as a backbone for all reporter assay constructs and was included in the Luciferase Assay System kit (Promega). pBEST + *metK* was created by inserting nucleotides corresponding to the first 10 codons of *C. burnetii metK* immediately downstream of the *luc* start codon using a Q5 Site-Directed Mutagenesis Kit, as instructed (New England Biolabs). *cbsR12* was cloned into pBESTluc using primers containing XhoI and AfeI restriction sites on the forward and reverse primers,

respectively. The forward primer also encoded a P*tac* promoter and *lac* operator. The PCR product was cloned into pBEST + *metK* using unique XhoI and AfeI restriction sites in an irrelevant IGR. A frameshifted *luc* construct was created as a byproduct of the *metK* Q5 mutagenesis of pBESTluc and contained a 1-bp frameshift deletion in the 5' end of *luc*. PBEST + *carA* 5' UTR was created using primers specific to the 5' UTR of *carA* with HindIII and BamHI restriction sites on the forward and reverse primers, respectively. The forward primer also encoded the P*tac* promoter. Nucleotides corresponding to the *lac* operator were inserted using a Q5 Site-Directed Mutagenesis Kit to create the final pBEST + *carA* 5' UTR construct. The *cbsR12* gene was inserted into this construct in the same fashion as for pBEST + *metK* + *cbsR12*.

Recombinant CarA and MetK were generated by PCR amplification of *carA* and *metK* using forward and reverse primers containing BamHI and HindIII restriction sites, respectively. The resulting amplicons were cloned into compatible restriction sites of pQE30 (Qiagen) by standard protocol.

Axenic growth of *C. burnetii*

For growth curves and Crosslink-Seq experiments, 3.33 x 10^4 genomic equivalents (GE)/mL of MB-WT, MB-*cbsR12*, or MB-*cbsR12*-Comp were inoculated into 300 mL ACCM-2 in a 1-liter flask with either chloramphenicol (5 μg/mL), kanamycin (350 μg/mL) or both. GE/mL was initially determined from frozen cell stocks by qPCR as previously described [190], although different *dotA* primers were used (**Figure S3.10**). Cell viability was determined using a BacLight Bacterial Viability kit as instructed (Thermo Scientific).

***C. burnetii* infection of differentiated THP-1 cells**

THP-1 cells were seeded onto 4-well chambered glass slides (Labtek) or T-75 flasks. After 2d of growth, 200 nM PMA was added along with fresh medium and cells were allowed to differentiate overnight. The PMA-supplemented medium was removed and fresh medium restored, after which differentiated THP-1 cells were allowed to recover for 4 h prior to *C. burnetii* infection at an MOI of 10. Initial infections were rocked for 2 h at room temperature before returning cells to 5% CO_2 and 37°C. At 1dpi the supernatant was removed, extracellular *C. burnetii* was washed away with warmed 1X PBS, and fresh medium was added.

Total RNA and genomic DNA extraction and purification

C. burnetii grown in ACCM-2 was centrifuged at 15,000 x g at 4°C for 15 min, after which pellets were resuspended in 1 mL TRI Reagent (Ambion). The suspension was incubated for 1 h at room temperature, frozen for 2 h at -80°C, thawed for 30 min at room temperature, then pipetted vigorously until homogenized. 100 µL BCP (Acros Organics) was added, the solution vortexed for 30 sec, incubated for 5 min at room temperature, and centrifuged at 12,000 x g at 4°C for 10 min. The aqueous phase was then collected and 300 µL 100% ethanol added. The mixture was immediately vortexed for 10 sec, and a RiboPure RNA Purification kit (Ambion) was used to collect, concentrate, and wash the resulting RNA. RNA was collected in nuclease-free H_2O and treated with DNase I for 1 h at 37°C. After RNA precipitation in 100% ethanol, the purified RNA was run on a NanoDrop spectrophotometer (Thermo Scientific) to determine concentration and purity.

In order to purify total RNA from *C. burnetii* grown in THP-1 cell lines, growth medium was first removed and replaced with 1 mL TRI Reagent. Flasks containing TRI Reagent were rocked for 1 h at room temperature, after which cells were mechanically

scraped and collected into a 15-mL conical tube. The mixture was frozen overnight at -80°C, thawed to room temperature for 30 min, and the RNA purification procedure continued as described above.

Genomic DNA was purified from TRI Reagent mixtures according to manufacturer protocols (Ambion). The resulting DNA was purified using a Nucleotide Removal kit as instructed (Qiagen).

Quantitative PCR (qPCR) and quantitative real-time PCR (qRT-PCR)

qPCR and qRT-PCR experiments were performed as previously described [100] using 300 nM of primers specific to *cbsR12* and a volume of iQ SYBR Green Supermix (Bio-Rad). The resulting reactions were cycled on a MyiQ Single-Color Real Time PCR Detection System (version 1.0 software) (Bio-Rad). CbsR12 cDNA copy number was normalized against *dotA* copy number derived from *C. burnetii* genomic DNA to obtain copy numbers / GE values. Growth curve GE/mL and GE/flask values were obtained from genomic DNA purified from the same cells from which total RNA was purified.

For ACCM-2 growth curves, 30 mL aliquots of a 300-mL culture were removed at each time point. gDNA was extracted and resuspended in 30 μL nuclease-free H_2O. 1 μL gDNA was used in subsequent qPCR reactions producing GE/mL values. For THP-1 growth curves, 8 separate T-75 flasks for each *C. burnetii* strain tested were inoculated simultaneously. At the specified time point, one flask was taken and 30 μL gDNA was again extracted. 1 μL gDNA was used in subsequent qPCR reactions and the GE/flask values were calculated. Each growth curve and qPCR reaction were performed in triplicate.

RNase III assay

RNase III assays were performed as previously described [102] using 200 nM CbsR12 substrate and the *C. burnetii* IVS RNA as a positive control [102]. Resulting reactions were electrophoresed on a 7% denaturing polyacrylamide gels and stained with 2 μg/mL acridine orange to visualize bands.

Identification of transcription start sites

5' RACE analysis of *carA*, *metK*, *cvpD*, and *cbsR12* transcripts was performed on MB-WT and MB-*cbsR12* RNA extracted from infected Vero (*cbsR12*, *carA*, and *metK*) or THP-1 (*cvpD*) cells at 3dpi using a 5' RACE System kit (Invitrogen) according to manufacturer protocols and with gene-specific primers (see **Figure S3.10**). Resulting PCR products were cloned into pCR2.1-TOPO as instructed (Invitrogen) and then sequenced with M13 universal primers by Sanger automated sequencing. In general, three biological replicate total RNA samples were obtained from MB-WT and MB-*cbsR12* infections of Vero and THP-1 cells. From these pools, 5' RACE was performed as above and 4 clones from each replicate were sequenced, producing 12 total clones analyzed for each total RNA pool for each gene analyzed.

***In silico* and bioinformatics analyses**

RNA target predictions were carried out using TargetRNA2 [42], IntaRNA [43], and CopraRNA [44] algorithms with default settings. RNA was folded using mFold [145] and visualized with Visualization Applet for RNA [191]. Analyses of RNA-Seq data were carried out as previously described [164]. Briefly, raw fastq files were concatenated, quality-filtered with the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) then clipped, aligned, and filtered with Nesoni version 0.128 tools (http://www.vicbioinformatics.com/software.nesoni.shtml). TPM were calculated using

custom perl and python scripts that can be accessed through GitHub (https://github.com/shawachter/TPM_Scripts). Crosslink-Seq enrichment was accomplished by processing .bam files using featureCounts [192], followed by use of the DESeq2 package in R version 3.4.4 to obtain differentially expressed genes [193]. The Artemis genome browser was used to visualize generated alignment files (http://www.sanger.ac.uk/science/tools/artemis) [147].

All IFA images were processed and analyzed using Fiji [194] and Cell Profiler [195], respectively. Figures were created using R version 3.4.4, Inkscape (https://inkscape.org/release/inkscape-0.92.4/) and GIMP (https://www.gimp.org/downloads/).

RNA-RNA hybridization and EMSA

Regions of target genes were first selected for PCR amplification. The following regions were chosen based on inclusion of predicted 5' UTRs and CbsR12-binding sites (+1 nucleotide designation is the first nucleotide of the annotated start codon): *carA* (-143) – (-1), *metK* (-26) – (+110), *cvpD* (-41) – (+101), *purH* (-100) – (+44), *dnaA* (-97) – (+61), and *rpsA* (-65) – (+77). PCR products (1µg) of desired templates were *in vitro*-transcribed overnight at 37°C with 2.5 mM Ribonucleotide Solution Mix (New England Biolabs) and when needed, 0.5 mM Bio-16-UTP (Invitrogen) using a MAXIscript T7 Transcription kit (Invitrogen). Resulting reactions were incubated for 1 h at 37^{o}C with 1 µL TURBO DNase (Invitrogen), heated for 4 min at 85°C then immediately plunged in ice, and electrophoresed on a 7% polyacrylamide gel for 75 min at 100V. Gels were stained with a 2 µg/mL acridine orange solution and visualized bands were excised and eluted overnight into probe elution buffer (0.5M AmAc, 1mM EDTA, 0.2% SDS) at

37°C. The resulting solution was precipitated with ethanol overnight at -20 °C, washed with 70% ethanol, and resuspended in nuclease-free H_2O. RNA concentrations were determined using a NanoDrop spectrophotometer. Following this, 10 nM biotin-labeled CbsR12 and, unless otherwise noted, 5 nM target RNA were combined and heated for 5 min at 85°C. A high-salt TMN buffer (100 mM NaCl, 50 mM $MgCl_2$, 100 mM Tris-Cl, 0.05% Tween-20) was then added and reactions were immediately plunged on ice for 30 sec and then incubated for 30 min at 37°C. A non-denaturing loading dye (0.25% bromophenol blue) was added and the resulting RNA mixtures were resolved on a 7% polyacrylamide gel for 1 h 20 min at 100V. RNA was transferred to a BrightStar-Plus Positively Charged Nylon Membrane (Ambion) using an electro-blot transfer system (Bio-Rad) and cross-linked with short-wave UV light in a GS Gene Linker UV Chamber (Bio-Rad). A North2South Chemiluminescent Hybridization and Detection Kit (Thermo Scientific) was used to detect resulting bands. The blot was imaged on a LAS-3000 imaging system (Fujifilm).

RNA-protein electrophoretic mobility shift assay

CbsR12-CsrA1/2 EMSAs were performed as previously described for CsrA-binding RNAs [169]. Biotin-labeled CbsR12 was synthesized *in vitro* as above for RNA-RNA EMSAs. *C. burnetii csrA-1* and *csrA-2* genes were cloned into pQE30, expressed, and resulting proteins natively purified as previously described [102]. 1 nM biotin-labeled CbsR12 diluted in TE buffer (10mM Tris-HCL, 1mM EDTA) was heated at 75°C for 3 min and equilibrated to room temperature for 10 min. Purified CsrA1/2 diluted in CsrA-binding buffer (1 µl in 10mM Tris-HCl, 10mM $MgCl_2$, 100mM KCl, 10mM DTT, 10% glycerol, and 10 U RNasin [Promega]) was added, and reactions were incubated for 30

min at 37°C. Samples were immediately resolved on 10% non-denaturing polyacrylamide gels for 3h. Membrane transfer and imaging were performed as in RNA-RNA EMSAs described above. The K_D for CsrA-2 was determined as previously described [169].

Reporter assay

A Luciferase Assay System kit (Promega) was used. All pBESTluc constructs were transformed into *E. coli* Top 10 F'. Resulting *E. coli* strains were grown overnight at 30°C in 10 mL LB containing ampicillin (100 µg/mL) and 1% glucose in order to mitigate the autoaggregative effects of CbsR12. An aliquot (4.5 mL) of the overnight culture was inoculated into 40.5 mL LB with 100 µg/mL ampicillin and grown for 1.5 h at 30°C. IPTG was added (to 1 mM) and culture aliquots (100 µl) were removed at 0, 1, and 2 h time points. 80 µL LB and 20 µL CCLR lysis solution (1X CCLR, 25 mg BSA, 12.5 mg lysozyme, 7.5 mL water) were added to the aliquots and gently inverted until the solution clarified. 50 µL of the resulting lysate was aliquoted to a 96-well plate, 100 µL of luciferase assay substrate was added, and luminescence was immediately read with a SpectraMax M5 Plate Reader (Molecular Devices).

Crosslink-Seq analysis

RNA-RNA crosslinking was performed essentially as previously described [152, 196], except TRI Reagent was utilized for total RNA extraction as described above and 10 nmol of two distinct biotinylated *in vitro*-transcribed anti-CbsR12 RNAs were used as probes. The resulting captured RNA was sent to the Yale Center for Genomic Analysis for RNA-Seq analysis.

Immunofluorescence assay

IFAs on infected THP-1 cells were performed as previously described with modifications [86]. Briefly, 4-well chambered glass slides were coated for 30 min with a 0.2% solution of Sigmacote (Sigma). THP-1 cells were inoculated into chambered slides and incubated overnight or until 60% confluence was reached and then differentiated with 200 nM PMA. Confluent cells were then infected with MB-WT, MB-*cbsR12*, or MB-*cbsR12*-Comp strains at a MOI of 10. At 1dpi, infections were stopped by washing cells three times for 5 min in pre-warmed 1X PBS, after which fresh medium was added. At 3 or 7 dpi, the growth medium was removed and cells were fixed with ice-cold 100% methanol for 5 min at room temperature. Cells were washed three times for 5 min each with ice-cold 1X PBS, blocked for 1 h at room temperature with a 2% BSA solution in 1X PBS, and then incubated with anti-Com1 (1:1000) and anti-LAMP1 (1:50, H4A3 was deposited to the Developmental Studies Hybridoma Bank by August, J.T / Hildreth, J.E.K) antibodies for 2 h. Cells were washed and incubated with Alexa Fluor 488 (1:200, Thermo Scientific) and goat anti-mouse rhodamine antibodies (1:200, Thermo Scientific) along with DAPI (300 nM, Thermo Scientific) for 1 h. Cells were then washed three times for 5 min each in ice-cold 1X PBS and immediately imaged. Images were processed with Fiji [194]. Cell Profiler was used to measure CCV areas, as previously described [197]. Measurements were taken from 46 individual images of random fields of view spanning three different experiments for each *C. burnetii* strain.

Protein synthesis, purification and antibody production

Recombinant *Coxiella* RNase III was synthesized from a previously-generated pQE30 construct and purified as before [102]. *C. burnetii carA* and *metK* genes were cloned in-frame into pQE30 (Qiagen) and the resulting N-terminal His_6-tagged proteins synthesized

and purified as previously described for *C. burnetii* RNA helicase [101]. Purified recombinant CarA and MetK proteins were submitted to General Bioscience, Inc., for rabbit polyclonal antibody production.

Western blot

Western blots were performed as previously described [97] with minor modifications. MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp strains were grown to mid-log phase (4 d for MB-WT and MB-*cbsR12*-Comp, 6 d for MB-*cbsR12*) in ACCM-2. Proteins at 30 µg (CarA blot) or 60 µg (MetK blot) were resolved on 10% - 20% acrylamide gradient Tris-Glycine SDS-PAGE gels. Duplicate gels were run in parallel for Coomassie brilliant blue staining (CBB; 0.1% (w/v) in 50% methanol, 7% (v/v) acetic acid) in order to present a loading control. Blots were incubated with primary antibody solution (1X PBS pH7.4, 0.3% (v/v) Tween-20, 1:500 CarA / 1:5000 MetK primary antibody) for 2 h with rocking at room temperature. Blots were washed five times for 5 min in 1X PBS, then incubated for 1 h in secondary antibody solution (1X PBS + 1:2000 goat anti-rabbit::HRP). Blots were again washed five times for 5 min in 1X PBS and immediately developed using SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) according to the manufacturer's protocol. Imaging was performed on a ChemiDoc XRS+ System (Bio-Rad).

***E. coli* biofilm induction assay**

E. coli biofilm induction assays were performed as previously described [31]. 96-well plates were inoculated with an overnight culture of either pBEST or pBEST + *carA* 5' UTR + *cbsR12 E. coli* strains (see **Figure S3.10**). Cultures were allowed to grow for 3 h until induction with 1mM IPTG, after which cultures were allowed to grow an additional

21 h before subsequent staining with crystal violet. The average OD_{570} readings of 10 wells were obtained by spectrophotometry.

Generation of a CbsR12-complemented strain

MB-*cbsR12* was complemented as previously described, with modifications [198]. Briefly, wild-type *cbsR12* along with 100 bp of flanking sequences were PCR-amplified using primers containing EcoRI and BamHI restriction sites. The amplicon was cloned into compatible restriction sites of pMini-Tn7-KAN by standard protocol [199]. The resulting plasmid was transformed into electrocompetent *E. coli* PIR1 cells for propagation. The pMini-Tn7-CbsR12-KAN plasmid (20 μg), along with 10 μg of a second plasmid containing the transposase, pMini-TnS2-ABCD, were transformed into MB-*cbsR12* in a single electroporation reaction (25 kV, 500 ohms, 25 μF). Electroporated cells were allowed to recover 5 d in ACCM-2 supplemented with 1% FBS, then dilutions were plated onto ACCM-2 agar containing kanamycin (375 μg/mL). Isolated colonies were picked and re-cultured on ACCM-2 agar plates for several rounds. Colony-PCR was used to screen for the MB-*cbsR12*-Comp strain, and the location of the *cbsR12* cassette was determined by PCR and Sanger automated sequencing. qPCR of MB-*cbsR12*-Comp genomic DNA utilizing primers specific to *cbsR12* was used to ensure that a single Tn insertion event occurred.

Data availability

The sequencing reads from the Crosslink-Seq experiments are available at the NCBI sequencing read archive (accession number: PRJNA522455).

Acknowledgments

We thank Paul Beare for his generous donation of *E. coli* PIR1 cells, pMini-TnS2-ABCD plasmid, and the pMini-Tn7-KAN plasmid. We would thank Jenny Wachter for her contribution of the TPM calculator and Linda D. Hicks for her excellent technical assistance. The authors also wish to acknowledge the following grant support: RR was supported by NIH grants AI123464, AI126385 and AI133023. MFM was supported by NIH grants AI128575, AI123293 and AI119798. SW was supported by a research grant from the Montana Academy of Sciences.

Supplementary Material

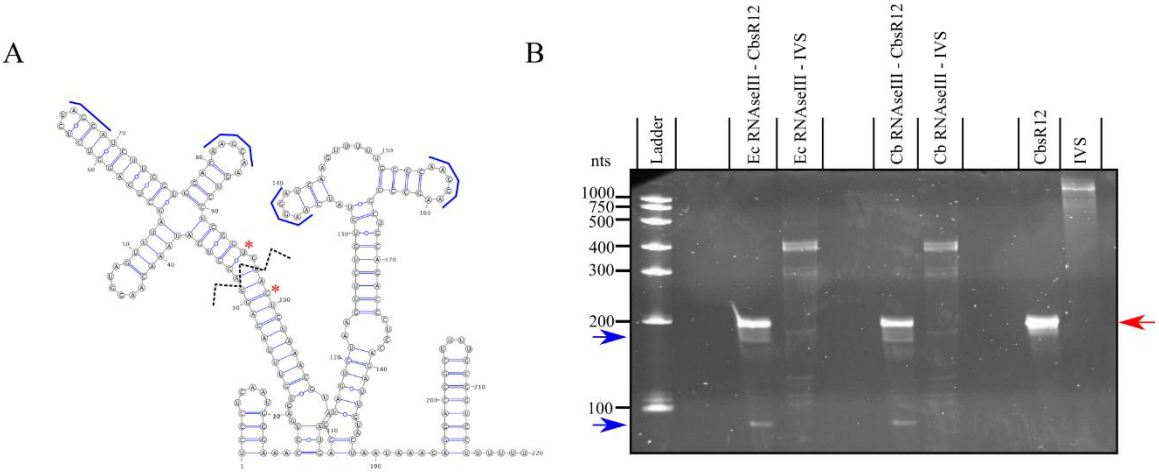


Figure S3.1: CbsR12 is processed by RNase III. (**A**). CbsR12 secondary structure as predicted by mFold. Nucleotide 1 was determined to be the TSS for the full-size transcript by 5' RACE. Red asterisks indicate apparent alternative TSSs by 5' RACE. Dotted line indicates the putative RNase III processing area. Blue solid lines indicate consensus CsrA-binding sites. (**B**). RNase III assay of in vitro-transcribed CbsR12 with the *C. burnetii* IVS RNA as a positive control. Results from treatment with *E. coli* (Ec) RNase III (New England BioLabs), recombinant *C. burnetii* (Cb) RNase III or no-

enzyme controls are shown. Arrows indicate RNase III-processed (blue) and un-processed (red) CbsR12 RNA.

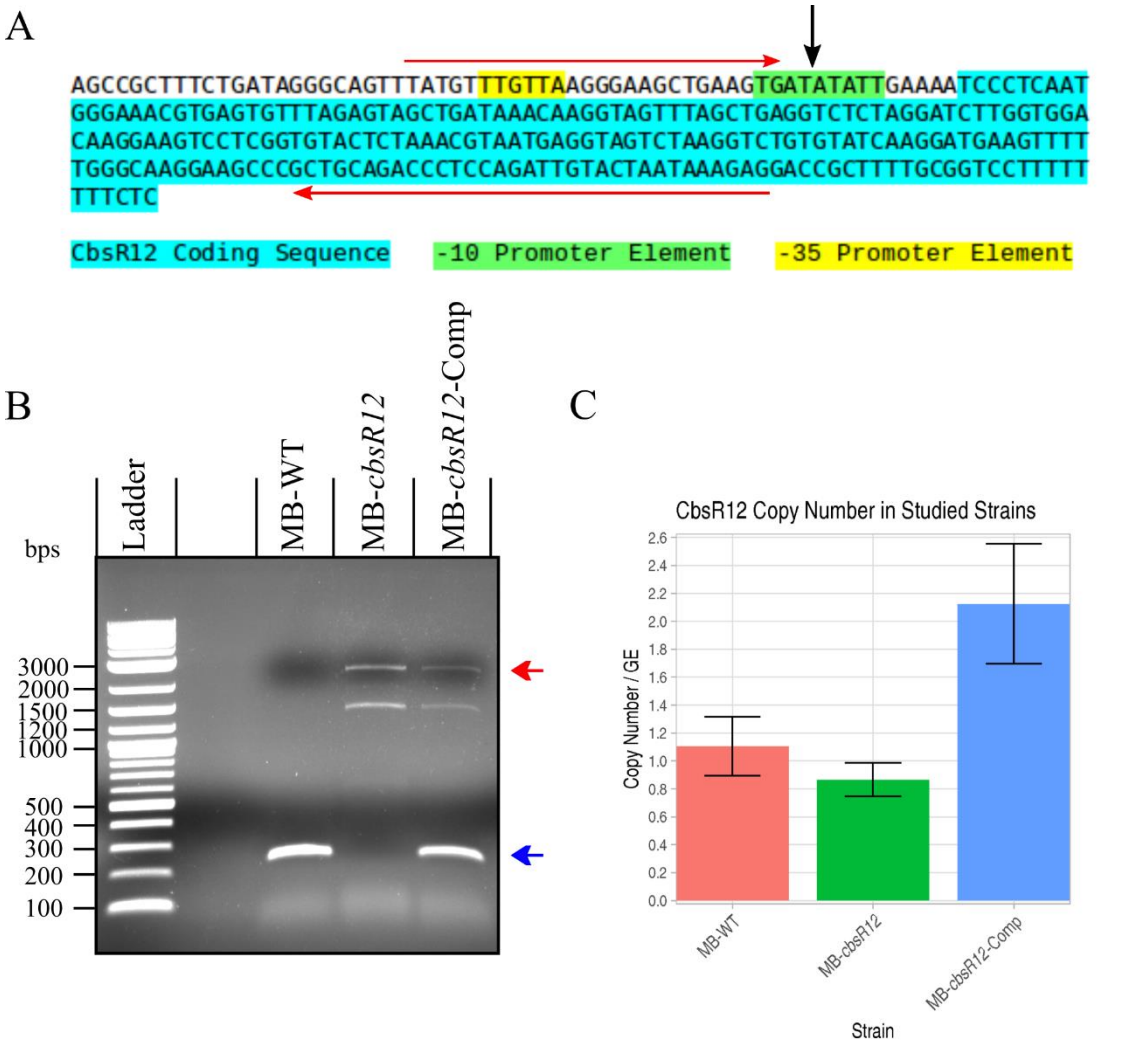


Figure S3.2: Location of the MB-cbsR12 transposon insertion to inactivate *cbsR12*. (**A**). The *cbsR12* gene and promoter elements are highlighted by the indicated colors, while the location of the Himar Tn insertion producing the MB-*cbsR12* strain is marked by a black arrow. Red arrows denote primer-binding sites for PCR confirmation of the lesion (forward and reverse primers above and below their annealing sequences, respectively). (**B**). PCR products confirming Tn insertion in *cbsR12* of MB-*cbsR12* (red

arrow) by loss of the ~250 bp amplicon and reintroduction of *cbsR12* in MB-*cbsR12*-Comp (blue arrow). (**C**). Copy number qPCR analysis confirming a single additional insertion of *cbsR12* in the MB-*cbsR12*-Comp strain. Values represent the means ± standard error of means (SEM) of three independent determinations.

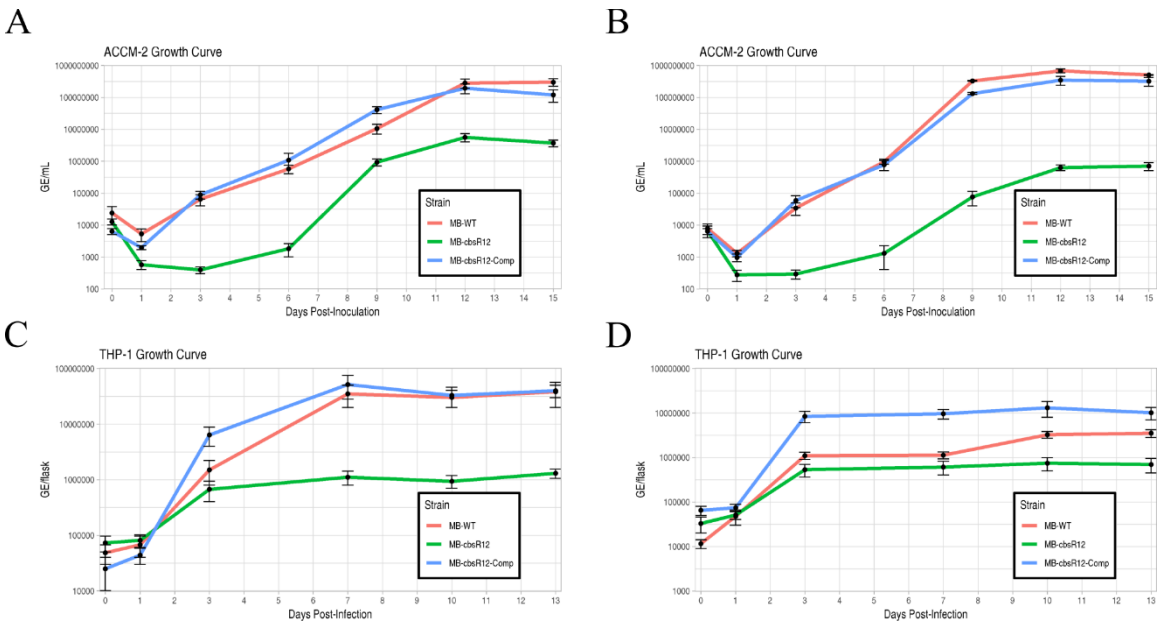


Figure S3.3: Additional biological replicates for ACCM-2 and THP-1 growth curves. Growth curves for MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp strains in ACCM-2 (**A**, **B**) or THP-1 cells (**C**, **D**) as determined by qPCR. The 0dpi time point refers to the inoculum. Values represent means ± standard error of means (SEM) of three technical replicates.

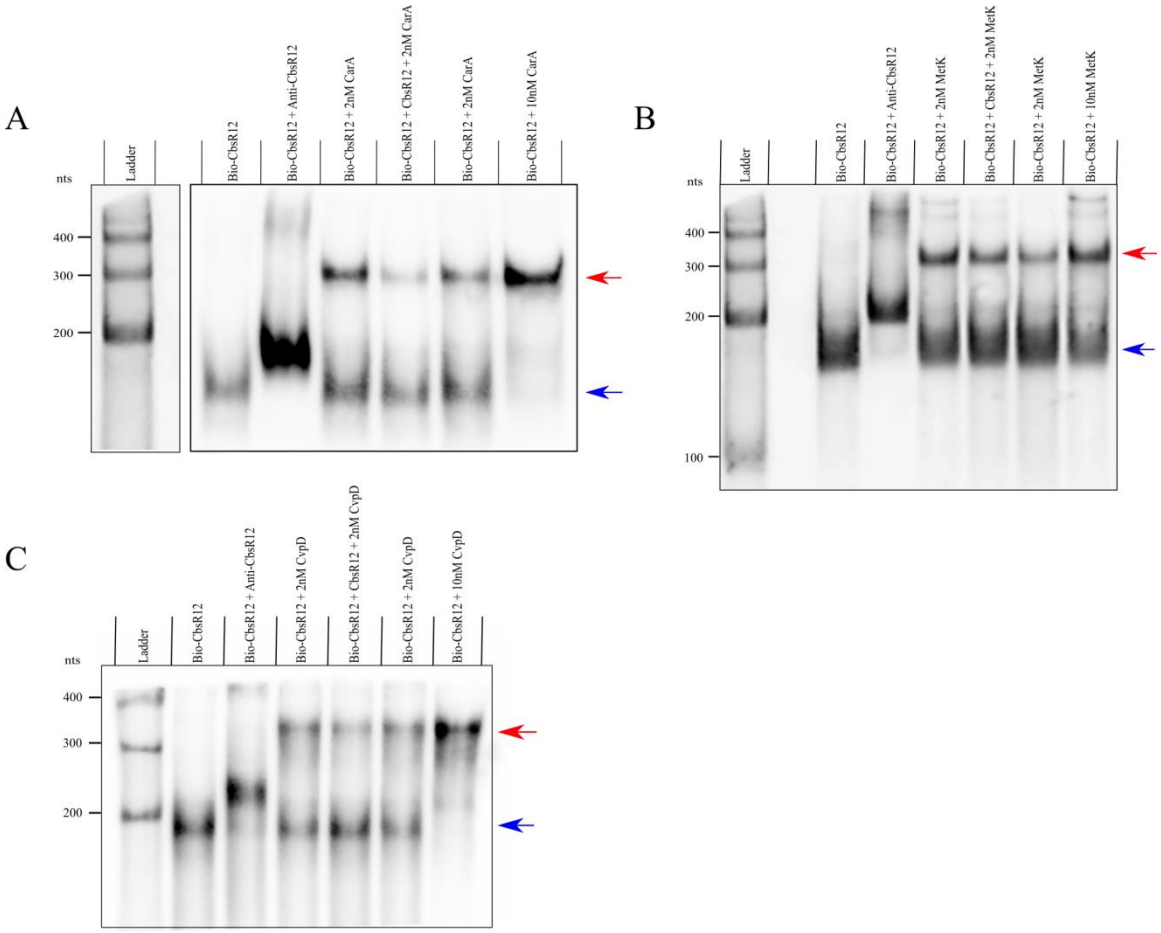
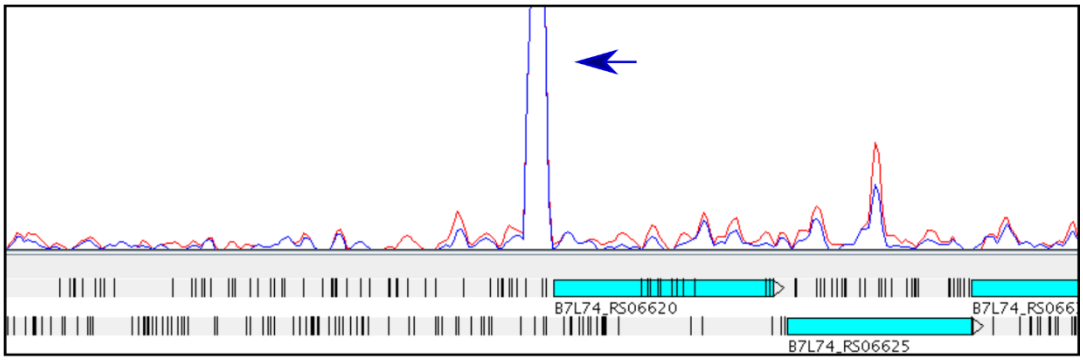
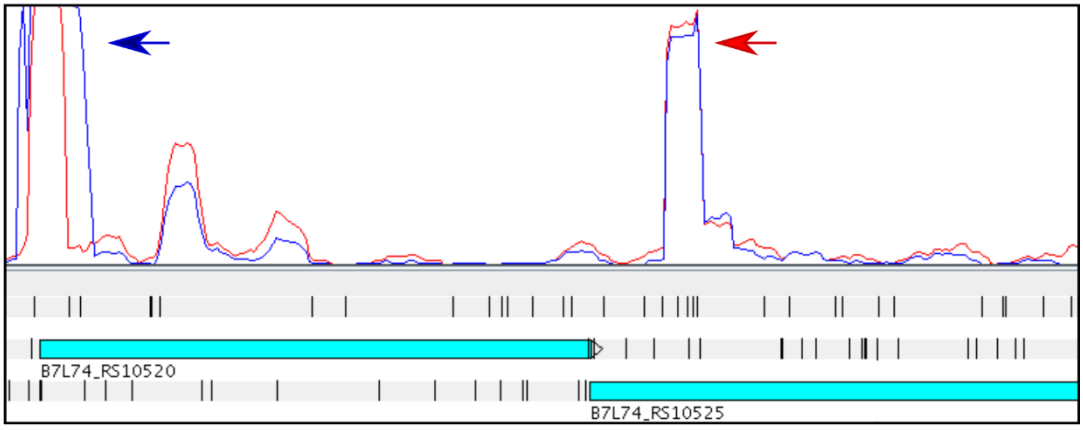


Figure S3.4: CbsR12 competitively binds *carA*, *metK*, and *cvpD* transcripts in a dose-dependent manner. RNA-RNA EMSAs showing hybridization reactions between biotin-labeled CbsR12 (Bio-CbsR12) and an *in vitro*-transcribed segment of *carA* (**A**), *metK* (**B**), or *cvpD* (**C**). Anti-CbsR12 represents a positive control consisting of a transcript equal in size but antisense to the CbsR12 transcript. A cold-chase sample containing Bio-CbsR12 + un-labeled CbsR12 + CarA/MetK/CvpD shows competitive (specific) binding relative to Bio-CbsR12 plus target alone, while increasing the dose of *carA*/*metK*/*cvpD* transcript (from 2 nM to 10 nM) increases the amount of retarded sample signal on the blot. Arrows indicate un-bound Bio-CbsR12 (blue) and Bio-CbsR12 bound to its RNA targets (red).

A



B



C

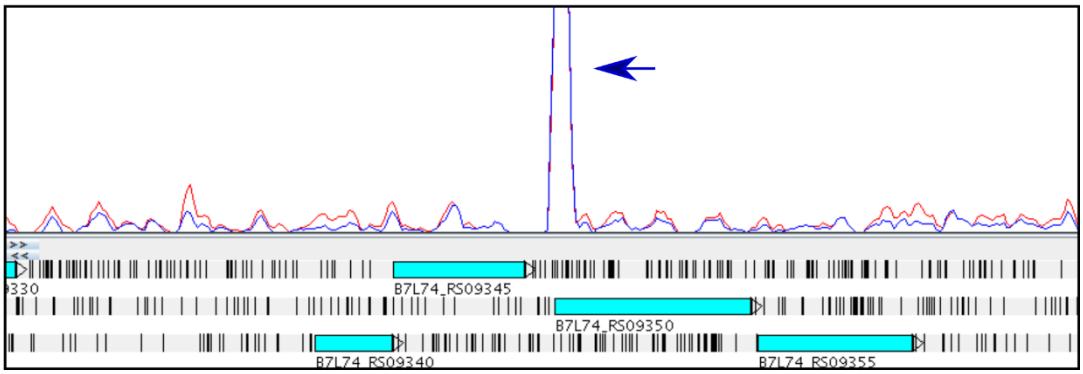
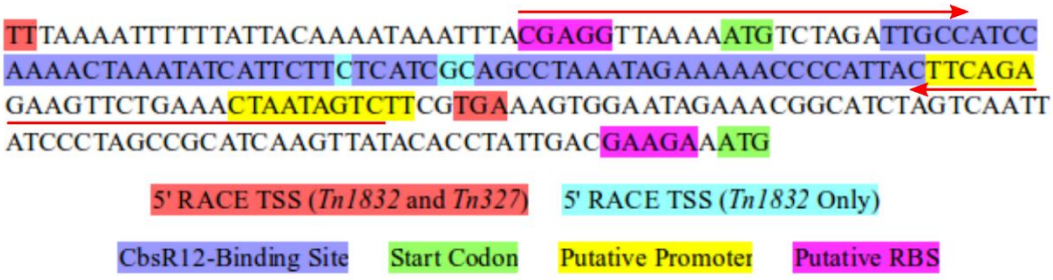


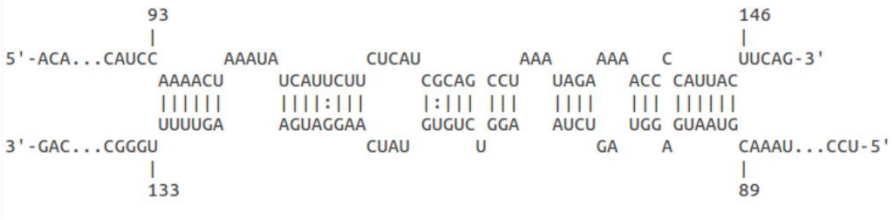
Figure S3.5: Artemis views of CbsR12 binding to *carA*, *metK*, *ahcY*, and *cvpD* transcripts. Artemis representation of Crosslink-Seq results for MB-WT. Red and blue lines represent the two biological replicates. (**A**). CbsR12 crosslinking with *carA* reads

(blue arrow). (**B**). CbsR12 crosslinking with *metK* reads (blue arrow) and *ahcY* reads (red arrow). (**C**). CbsR12 crosslinking with *cvpD* reads (blue arrow).

A



B



C

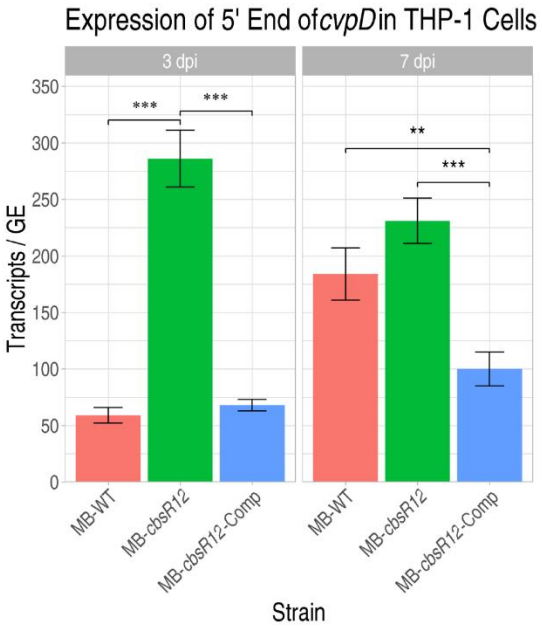


Figure S3.6: CbsR12 downregulates the quantity of transcripts arising from the 5' end of *cvpD* in LCVs from infected THP-1 cells. (**A**). *cvpD* gene sequence from the 5' TSS to the predicted downstream alternative start codon. Indicated colors highlight the TSSs, the CbsR12-binding site, the putative downstream promoter, putative RBSs, and start codons. Red arrows show primer annealing regions for qRT-PCR (forward and reverse primers above and below their respective annealing sequences). (**B**). Representation of CbsR12 binding to the *cvpD* transcript as determined by IntaRNA with base numbers indicated. The top strand in the model represents the *metK* transcript sequence, while the bottom strand represents the complementary CbsR12 sequence. (**C**). qRT-PCR of the 5' end of *cvpD* from MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp LCVs (3dpi) and SCVs (7dpi) infecting THP-1 cells. Values represent means ± standard error of means (SEM) of three independent determinations (** = $P < 0.01$, one-way ANOVA, *** = $P < 0.001$, one-way ANOVA).

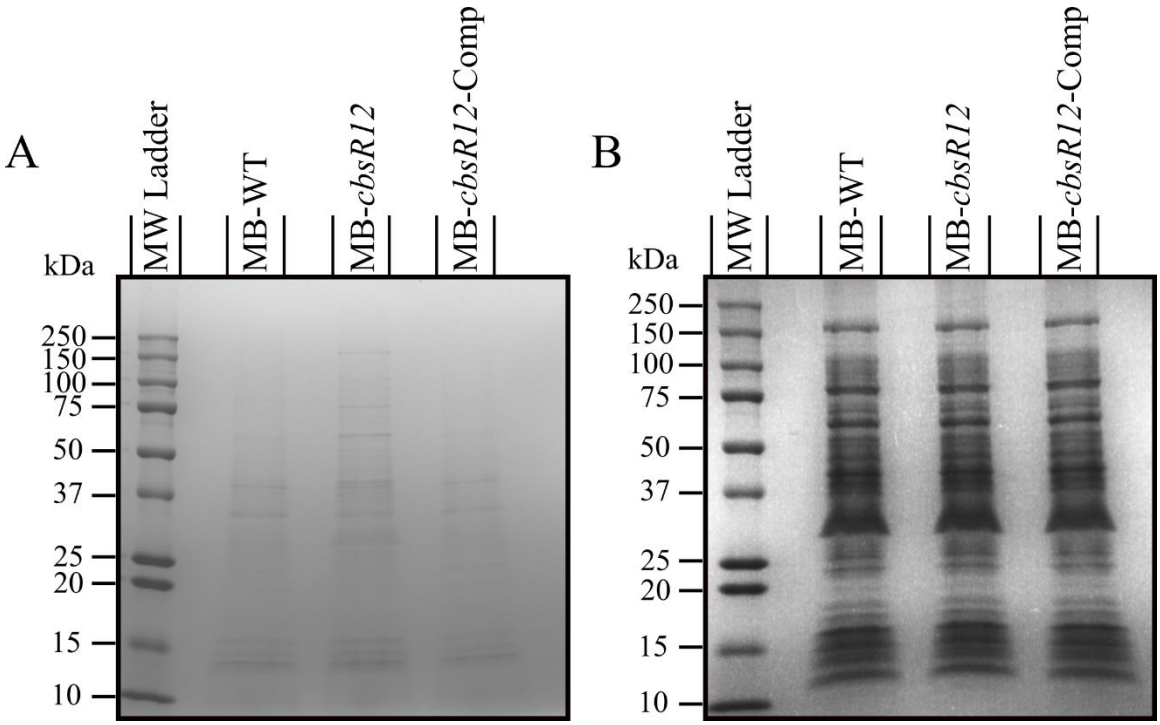


Figure S3.7: Coomassie blue-stained SDS-PAGE gels corresponding to Figure 3.9 to demonstrate loading consistency. (**A**). Proteins (30 µg total) from MB-WT, MB-*cbsR12*, and MB-*cbsR12*- Comp LCVs (mid-log phase; 96h for MB-WT and MB-*cbsR12*-Comp and 144h for MB-*cbsR12*) grown in ACCM-2 were resolved on a 10-20% acrylamide gradient SDS-PAGE gel and stained with Coomassie brilliant blue R. (**B**). Proteins (60 µg total) from MB-WT, MB-*cbsR12*, and MB-*cbsR12*-Comp LCVs (mid-log phase; 96h for MB-WT and MB-*cbsR12*-Comp and 144h for MB-*cbsR12*) grown in ACCM-2 were resolved on a 10-20% acrylamide gradient SDS-PAGE gel and stained with Coomassie brilliant blue R.

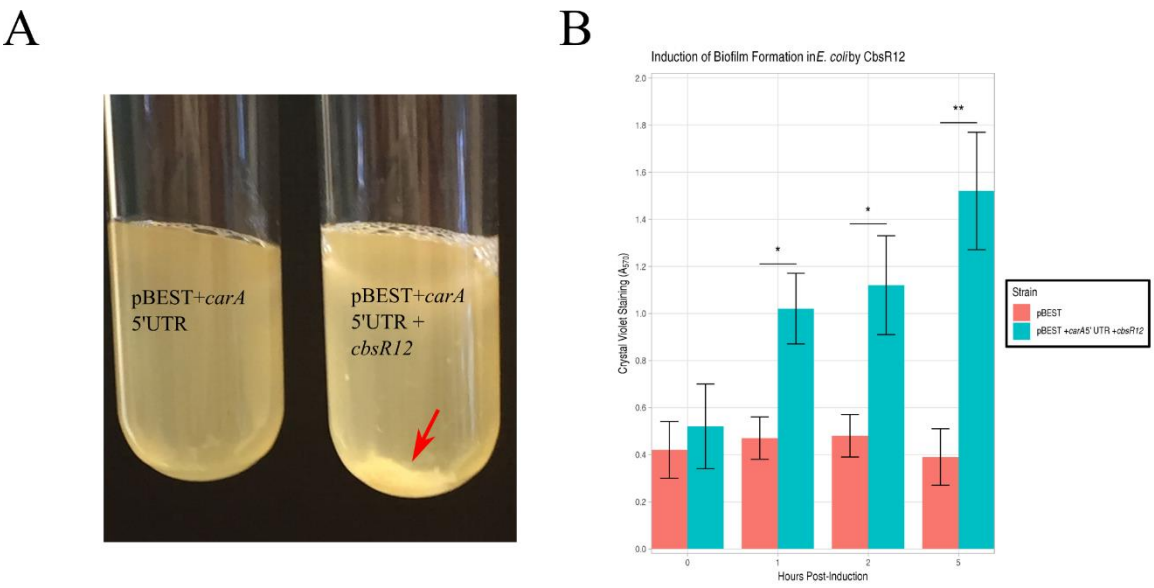


Figure S3.8: CbsR12 in *E. coli* leads to an autoaggregative phenotype and biofilm formation. (**A**). Overnight cultures of *E. coli* Top10 F' harboring pBEST + *carA*5'UTR or pBEST + *carA*5'UTR + *cbsR12* were inoculated into 3 mL LB supplemented with ampicillin (100 µg/mL), grown for 2 h at 37^0 C with shaking, then induced with 1 mM IPTG for 3 h before photography. The red arrow indicates autoaggregation of *E. coli* in the presence of CbsR12. (**B**). *in vitro* biofilm formation assay of *E. coli* Top10 F' harboring pBEST or pBEST + *carA*5'UTR + *cbsR12*. Crystal violet staining is indicative of adherence due to biofilm induction. Values represent the average OD570 readings of 10 wells ± standard error of means (SEM) of three independent determinations (* = $P < 0.05$, student's t test, ** = $P < 0.01$, student's t test).

A

B

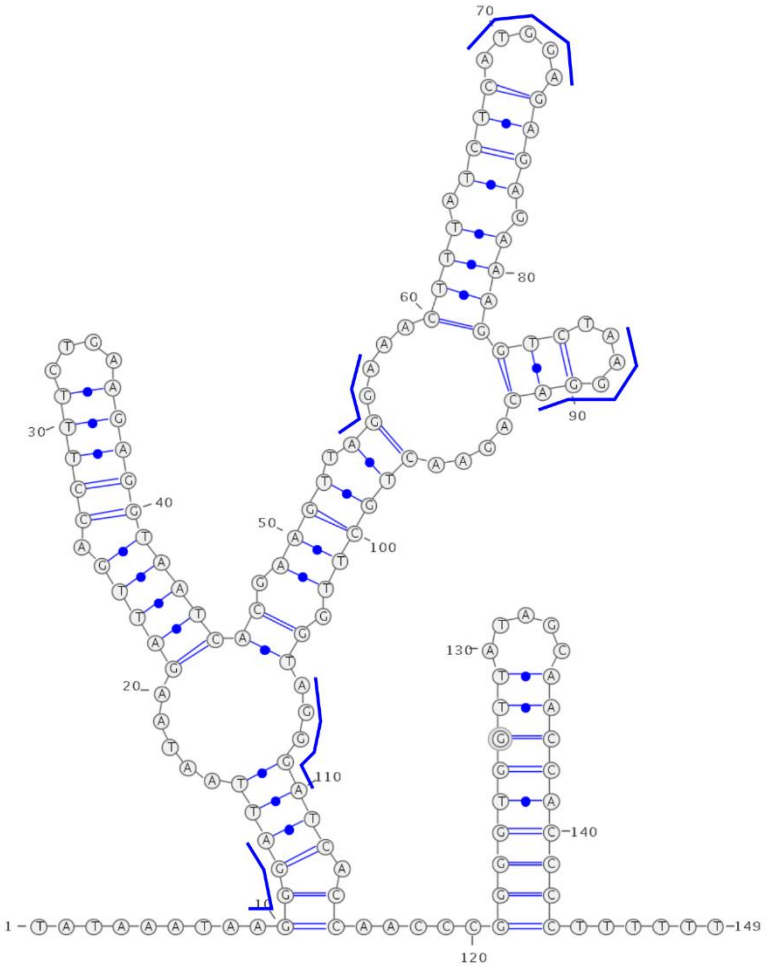


Figure S3.9: CbsR1 is an additional *C. burnetii* sRNA with RsmY/Z-like characteristics. (**A**). The *cbsR1* gene, predicted promoter elements, and putative LetA-binding site are highlighted by the indicated colors. (**B**). CbsR1 secondary structure as predicted by mFold. Nucleotide 1 was predicted to be the TSS for the full-size transcript by analysis of RNA-Seq datasets. Blue solid lines indicate putative CsrA-binding sites.

Strains	Description	Origin
TOP10F'	E. coli Chemically Competent Strain	Invitrogen
PIR1	*E. coli* Strain with PIR Origin of Replication	Paul Beare
MB-WT	*C. burnetii* "Control" Transposon Mutant (*Tn1832*)	[11]
MB-*cbsR12*	*C. burnetii* Transposon CbsR12 Mutant Clone (*Tn327*)	[11]
MB-*cbsR12*-Comp	Pmini-Tn7 Complement of MB-*cbsR12*	This Study

Name	Purpose	Background	Origin
Frameshifted Luciferase	Reporter Assay Negative Control	Top 10 F'	This Study
pBEST + carA 5'UTR	Reporter Assay	Top 10 F'	This Study
pBEST + carA 5'UTR _ CbsR12	Reporter Assay	Top 10 F'	This Study
pBEST + metK	Reporter Assay	Top 10 F'	This Study
pBEST + metK + CbsR12	Reporter Assay	Top 10 F'	This Study
carA_pQE30	CarA Expression Plasmid	Top 10 F'	This Study
metK_pQE30	MetK Expression Plasmid	Top 10 F'	This Study
pQE30_mc	RNase III Assay	Top 10 F'	[43]
csrA1_pQE30	CsrA-1 Expression Plasmid	Top 10 F'	This Study
csrA2_pQE30	CsrA-2 Expression plasmid	Top 10 F'	This Study
pMiniTnS2-ABCD	Tn327 Complementation	PIR1	[85]
pMiniTn7-CbsR12-KAN	Tn327 Complementation	PIR1	This Study
pCR2.1-TOPO	TA cloning vector	Top 10	Invitrogen

Primer Category	Primer Name	5'-3' Sequence	Reference
QRT-PCR	Q_CbsR12_F_L_qRT	GCTGATAAACAAGGTAGTTTAGCTGAGGTC	This Study
	Q_CbsR12_R_qRT	GTCTGCAGCGGGCTTCCTT	This Study
	Q_dotA_qRT_F	CTGGGAGAAGCTAAACAGGGGG	This Study
	Q_dotA_qRT_R	CCACAGCTAGCCCTGAAAAGGTATAC	This Study
	Q_cvpD_qRT_F	CGAGGTTAAAAATGTCTAGATTGCC	This Study
	Q_cvpD_qRT_R	GACTATTAGTTTCAGAACTTCTCTGAAG	This Study
EMSA	NEW_QCbsR12_F_T7	TAATACGACTCACTATAGGGGAAACGTGAGTGTTTAG	This Study
	Q_CbsR12_R_GOOD	CTCTTTATTAGTACAATCTGGAGGGTCTGCAGC	This Study
	QmetK_F+T7	TAATACGACTCACTATAGGTGAAACATTAATTTAGG	This Study
	QmetK_R	GGTCTTGGCCAATCAGGG	This Study
	QcarA_F+T7	TAATACGACTCACTATAGGCTCTAAAGTAACTCAACC	This Study
	QcarA_R	GCGGCAGCACCCTCTTTACCTA	This Study
	QCb1818_F+T7	TAATACGACTCACTATAGGTTTAAAATTTTTTATTAC	This Study
	QCb1818_R	CACGAAGACTATTAGTTTCAGAACTTC	This Study
	QpurH_F+T7	TAATACGACTCACTATAGGAGAGTGGTTATGCGCTAC	This Study
	QpurH_R	GCACGCTTAATCGGCCTTTCAGTA	This Study
	QdnaA_F+T7	TAATACGACTCACTATAGGAAAACTTTAATTTCTTTTCTTTCCA	This Study
	QdnaA_R	GCGGAATTTCATCGCGCAAATAACC	This Study
	QrpsA_F+T7	TAATACGACTCACTATAGGGAATCGTAAACAGACCCTAACC	This Study
	QrpsA_R	GCCTTGACCAAGGCTCCAGGAC	This Study
	QCbsR12A_F+T7	TAATACGACTCACTATAGGCTGGAGGGTCTGCAGCG	This Study
	QCbsR12A_R	GCTGATAAACAAGGTAGTTTAGCTGAGGTC	This Study
Reporter Assay	LucF_RepAss_carA	AAGCTTTGTTGACAATTAATCATCGGCTCGTATAATGTGCGAAATCGAGAAAGACTCTAAAG	This Study
	LucR_RepAssATG_carA	GGATCCCGCGGCAGCACCCTCTTTACCTAT	This Study
	RA_NEWEST_CbsR12_F	GACTCGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCCCTCAATGGGAAACGTG	This Study
	RA_CbsR12_Rev_G3	GAAGCGCTAAGTGAGAAAAAAAAAGGACCGCAAAAGCGG	This Study
	RA_LacO_Q5R	CTCACAATTCCACACATTATACGAGCCGATG	This Study
	RA_LacO_Q5F	CGGATAACAATTCGAAATCGAGAAAGACTC	This Study
	RA_NEW_MetK_Q5F	TTTACCTCAGAATCCGAAGACGCCAAAAACATAAAG	This Study
	RA_NEW_MetK_Q5R	TAAGGTCGTGTGCGTCATTTGGATCCTGTTTCC	This Study
XLINK_Seq	CbsR12_XLINK_1_F	TAATACGACTCACTATAGGGTTCCTTGTCCACCAAG	This Study
	CbsR12_XLINK_1_R	TCCCTCAATGGGAAACGTGAGTGTTTAG	This Study
	CbsR12_XLINK_2_F	TAATACGACTCACTATAGGGAGGACCGCAAAAGCG	This Study
	CbsR12_XLINK_2_R	GTTTTTGGGCAAGGAAGCCCGCTG	This Study
CarA and MetK Cloning	Q_CarA_Express_F	CAGGATCCAATCGCTTATCCTTTTGCAAG	This Study
	Q_CarA_Express_R	CAAAGCTTGGTGGAGTCCCTCATTAATTTAAC	This Study
	Q_MetK_Express_F	CAGGATCCACGCACACGACCTTATTTAC	This Study
	Q_MetK_Express_R2	CACTGCAGATTGGTTACGTTTGCAGC	This Study
CsrA-1/2 Cloning	Q_CsrA1_pQE_F	CAGGATCCTTAGTCTTAACACGAACAAATG	This Study
	Q_CsrA1_pQE_R	CAAAGCTTTTCAACTTCCTCAAGAATAGGTG	This Study
	Q_CsrA2_pQE_F	CAGGATCCTTAATACTAACCAGACGTATCGG	This Study
	Q_CsrA2_pQE_R	CAAAGCTTTTCAAATTCGTCAGTCTTCTCAC	This Study
5' RACE	CbsR12_GSP2	CTTCCTTGCCCAAAAACTTCATCC	This Study
	carA_GSP1	CCAAAATCGTAAACGACC	This Study
	carA_GSP2	CGTTGAAACAGCCTTTGCTAGATCTTTTCCTTTC	This Study
	metK_GSP1	CTTTGCCGGAGAAACAC	This Study
	metK_GSP2	CGGATTGACAAGGAAACGCGTGTGTTTATC	This Study
	cvpD_GSP1	GAAGGAGAGTGAGCGG	This Study
	cvpD_GSP2	GCCTACTATTAAGCGTCTCATGATATTCAAGGGC	This Study
RNAseIII Assay	IVS_Flank_F	TAATACGACTCACTATAGGCTGGTTCTCCTCG	[43]
	IVS_Flank_R	CTTTTCCTGGAAGCGTGG	[43]
CbsR12 Complement	CbsR12_EcoRI_F	GCCCCGAATTCGGCGAAGGCTAAAGTGAGAA	This Study
	CbsR12_BamHI_R	CCTCAGGATCCGTCGGTTTTTAGCCGCTTTC	This Study
	NM2_GlmS_F	CCTATTGCATACACGATTCCACTG	This Study
	NM2_Kan_F	ATGATTGAACAGATGGATTGCACGC	This Study
	Long_CbsR12_F	TATGTTTGTTAAGGGAAGCTGAAGTG	This Study

Figure S3.10: Strains, plasmids, and primers used in the study.

Chapter 4

Novel small RNAs expressed by *Bartonella bacilliformis* under multiple conditions reveal potential mechanisms for persistence in the sand fly vector and human host

As published in:

bioRxiv preprint server. 2020 doi: 10.1101/2020.08.04.235903

Abstract

Bartonella bacilliformis, the etiological agent of Carrión's disease, is a Gram-negative, facultative intracellular alphaproteobacterium. Carrión's disease is an emerging but neglected tropical illness endemic to Peru, Colombia, and Ecuador. *B. bacilliformis* is spread between humans through the bite of female phlebotomine sand flies. As a result, the pathogen encounters significant and repeated environmental shifts during its life cycle, including changes in pH and temperature. In most bacteria, small non-coding RNAs (sRNAs) serve as effectors that may post-transcriptionally regulate the stress response to such changes. However, sRNAs have not been characterized in *B. bacilliformis*, to date. We therefore performed total RNA-sequencing analyses on *B. bacilliformis* grown *in vitro* then shifted to one of ten distinct conditions that simulate various environments encountered by the pathogen during its life cycle. From this, we identified 160 sRNAs significantly expressed under at least one of the conditions tested. sRNAs included the highly-conserved tmRNA, 6S RNA, RNase P RNA component, SRP

RNA component, *ffH* leader RNA, and the alphaproteobacterial sRNAs αr45 and *speF* leader RNA. In addition, 153 other potential sRNAs of unknown function were discovered. Northern blot analysis was used to confirm the expression of eight novel sRNAs. We also characterized a ***B****artonella* ***b****acilliformis* **g**rou**p** **I** intron (BbgpI) that disrupts an un-annotated $tRNA_{CCU}^{Arg}$ gene and determined that the intron splices *in vivo* and self-splices *in vitro.* Furthermore, we demonstrated the molecular targeting of ***B****artonella* ***b****acilliformis* **s**mall **R**NA **9** (BbsR9) to transcripts of the *ftsH*, *nuoF*, and *gcvT* genes, *in vitro*.

Introduction

Bacteria often utilize sRNAs to rapidly and efficiently regulate gene products involved in multiple biological processes. sRNAs are small (< 500 nts) non-coding transcripts that typically serve to up- or down-regulate translation of proteins by binding to the respective mRNA in a *cis* or *trans* fashion [reviewed in [1]]. This fine-tuning of protein production can enhance tolerance to stressors, including temperature [8] and pH [200]. To our knowledge, sRNA research in *Bartonella* is represented by a single report on *B. henselae* [201]. We therefore utilized total RNA-Sequencing (RNA-Seq) to interrogate *B. bacilliformis* transcriptomes to identify sRNAs expressed under a variety of conditions, including temperatures and pH levels consistent with the sand fly vector and human host. In doing so, we discovered 153 novel sRNAs expressed under at least one of the conditions tested. Furthermore, we characterized two of the sRNAs. The first RNA is a group I intron related to similar elements found in other alphaproteobacteria, while the

other is a novel *Bartonella*-specific sRNA expressed only under conditions that simulate the sand fly vector.

Materials and Methods

Bacterial culturing

Bacterial strains, primers, and plasmids utilized in this study are described in **Table S4.1**. *B. bacilliformis* strain KC583 (passages #4-7) was cultivated on HIBB plates, comprised of Bacto heart infusion agar (Becton Dickinson; Franklin Lakes, NJ) containing 4% defibrinated sheep blood and 2% sheep serum (Quad Five, Ryegate, MT), by volume, as previously described [131]. Following 4 d of growth, 4 confluent *B. bacilliformis* plates per biological replicate were either shifted to different temperatures for 2 h, harvested and shifted to different pH levels in an HIBB liquid medium for 2 h, harvested and shifted to a human blood sample for 2 h, or harvested and used to infect cultured human umbilical vein endothelial cells (HUVECs; PCS-100-013; American Type Culture Collection; Manassas, VA) for 24 h. *Escherichia coli* (TOP10) was grown for 16 h at 37^0 C with shaking in lysogeny broth (LB), or on LB plates, supplemented with kanamycin (25 μg/ml) and ampicillin (100 μg/ml), when required.

HUVEC culturing and infection

HUVECs were cultured and maintained as previously described [131]. *B. bacilliformis* infections were carried out for 24 h after which the medium was removed and cells were treated with gentamicin (10 μg/ml) for 1 h. Remaining viable extracellular *B. bacilliformis* cells were removed by washing 5 times for 10 min with phosphate-buffered

saline (PBS; pH 7.4) solution. Finally, cells were harvested into TRI Reagent (Ambion; Austin, TX), as previously described for infected Vero cells [202].

Human blood infection

Blood was drawn into vials containing sodium citrate to prevent coagulation. 1-ml aliquots were dispensed into fresh tubes, after which the lids were replaced with gas-permeable membranes. Blood samples were equilibrated at 37^0 C (HB37 samples) or 37^0 C in a blood-gas atmosphere (HBBG samples) for 1 h. Four HIBB plates of confluent *B. bacilliformis* for each equilibrated blood vial were harvested into PBS, pelleted at 16,000 x g for 5 min at 4^0 C and washed twice in PBS with identical centrifugation steps. Cell pellets were resuspended into 300 µl equilibrated blood, then dispensed back into the corresponding tube. The tubes were incubated at the appropriate condition for 2 h, then 1 ml RNALater solution (Thermo Fisher; Waltham, MA) was immediately added. Total RNA extraction was done as described below.

Total RNA/genomic DNA isolation and preparation for RNA-Seq

Upon shifting *B. bacilliformis* for the designated time periods, cells were either harvested directly into one volume of RNAlater solution (Thermo Fisher) or centrifuged at 10,000 x g at room temperature for 2 min, after which the pellet was resuspended in a volume of RNAlater. The cells were incubated at room temperature for 1 h then frozen at -80^0 C for $\geq$ 2 h. The cells were thawed, centrifuged at 10,000 x g at 4^0 C for 10 min, and resuspended in 1 ml of TRI Reagent (Sigma-Aldrich; St. Louis, MO). The cells were incubated at room temperature for 1 h then frozen at -80^0 C for $\geq$ 2 h. Finally, cells were thawed and total RNA and genomic DNA isolation were done as previously described [202]. Total RNA pools from human blood infections were globin-depleted using a

GLOBINclear kit (Ambion) according to manufacturer's specifications. HUVE, HB37, and HBBG samples were enriched for bacterial RNA using a MICROBEnrich kit (Ambion). RNA (1 μg) from three independent biological replicates of each condition was sent to the Yale Center for Genomic Analysis (Pl25, Pl30, Pl37, pH06, pH07, and pH08 samples) or GENEWIZ (PlBG, HUVE, HB37, and HBBG samples) for bacterial rRNA depletion, stranded-library preparation, and HiSeq2500 (Illumina; San Diego, CA) 2x150 bp sequencing.

Data analysis

Raw reads were quality filtered and aligned as previously described [202]. Briefly, raw fastq files were concatenated, quality filtered with the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), and then clipped, aligned, and filtered with Nesoni version 0.128 tools (http://www.vicbioinformatics.com/software.nesoni.shtml). TPM were calculated using a custom Python script that can be accessed through GitHub (https://github.com/shawachter/TPM_Scripts). Stranded alignments were separated using SAMtools [203] and visualized using the Artemis genome browser [147].

sRNA identification was performed using the Artemis genome browser. RNA peaks were manually curated from IGRs and protein-coding gene regions. A read threshold for sRNA expression was devised for each condition tested based on reads that aligned to the *rpoD* gene (RpoD sigma factor), since TPM data suggest this gene is consistently expressed across all conditions. Using this method, putative sRNAs were identified, base ranges approximated, and sRNAs further characterized with putative promoters by manual searches using the conserved alphaproteobacterial sigma-70 promoter element, CTTGAC-N_{17}-CTATAT [204]. Rho-independent terminators were

identified using ARNold terminator prediction software (http://rna.igmors.u-psud.fr/toolbox/arnold/).

Since TPM calculations were done in the context of the total transcriptome and were not strand-specific, it was necessary to further refine TPM values for *cis*-anti sRNAs. This was done by considering the TPM value and proportion of the protein-coding gene to which the sRNA was antisense and subtracting the gene's approximate TPM contribution with the following formula:

$$sRNA\ Adjusted\ TPM = sRNA\ Total\ TPM\ - \left[\frac{Gene\ Total\ TPM}{\left(\frac{Gene\ Length}{Gene\ sRNA\ overlap\ Length}\right)}\right]$$

This was accomplished with a custom python script located in a GitHub repository (https://github.com/shawachter/TPM_Scripts). After this calculation, sRNAs whose TPMs were < 300 were considered not expressed for the purposes of the UpSet plot. This TPM value was chosen because it most accurately reflected the read threshold used in the initial manual sRNA search. All TPM values were included in the heatmap and determination of condition-specific sRNAs. Differentially-expressed sRNAs were determined by featureCounts [192] and the DESeq2 package of R version 3.4.4 [193]. For DESeq2 analysis, the p-value distribution of the significantly differentially expressed genes (DEGs) was re-sampled using fdrtools in order to more accurately achieve the desired null distribution [205]. This effectively made the analysis more stringent by providing fewer significant DEGs.

Infection-specific and sand fly-specific sRNAs were identified based on expression patterns as described in Results. The IntaRNA 2.0 sRNA target prediction algorithm was used to determine potential genes regulated by the sRNAs [43]. Only mRNA targets with

a predicted IntaRNA 2.0 p-value < 0.01 were included in the potential targets list. Further targets with FDR values of < 0.05 were given special indications since these predicted bindings were considered especially strong. GO enrichment was performed utilizing the biobam Blast2GO program in the OmicsBox program suite (https://www.blast2go.com/blast2go-pro/download-b2g) using functional annotation of the *B. bacilliformis* KC583 genome as the background [206]. KEGG enrichment was performed using DAVID Bioinformatics Resources [207].

Figures were made using R version 3.4.4 and various Bioconductor packages including UpSetR [208], gplots (https://cran.r-project.org/web/packages/gplots/index.html), ggplot2 [209], and fdrtool [205]. Raw PNG images were modified into figures using Inkscape (https://inkscape.org/release/inkscape-0.92.4/) and Gimp (https://www.gimp.org/downloads/).

Identification of transcription start sites

5' RACE analyses of BbgpI and BbsR9 were performed using total RNA from *B. bacilliformis* shifted to liquid medium at pH 7 using a 5' RACE System kit (Invitrogen; Carlsbad, CA) according to manufacturer's protocols and with gene-specific primers (**Table S4.1**). Resulting PCR products were cloned into pCR2.1-TOPO as instructed (Invitrogen), after which six arbitrary clones were sequenced with M13 universal primers by Sanger automated sequencing.

Northern blots

Northern blot analyses were carried out using total RNA extracted from *B. bacilliformis* under the noted conditions. Northern blot probes were synthesized *in vitro* by engineering probe-specific PCR primers to contain a T7 promoter then utilizing a MAXIscript T7

Transcription kit (Invitrogen) supplemented with 0.5 mM Bio-16-UTP (Invitrogen). *B. bacilliformis* total RNA (2 μg) was resolved on a 1% denaturing agarose gel for 130 min at 57 V in 1X MOPS running buffer (Quality Biological; Gaithersburg, MD). The gel was washed in nuclease-free H_2O for 10 min, followed by another wash in 20X SSC buffer (3M NaCl, 0.3M sodium citrate, pH 7.0) for 15 min. RNA was transferred overnight to a BrightStar-Plus nylon membrane (Ambion) in 20X SSC via upward capillary transfer. RNA was crosslinked to the membrane using a GS Gene Linker UV chamber (Bio-Rad; Hercules, CA) at 150 mJ. Membrane pre-hybridization and probe hybridization were done with a North2South Chemiluminescent Hybridization and Detection Kit (Thermo Fisher) according to manufacturer's protocol. 50 ng of the appropriate *in vitro*-transcribed biotin-labeled probe was hybridized to the membrane at 67°C overnight. Membranes were washed 3 times for 15 min at 67°C in 1X Hybridization Stringency Wash Buffer (Thermo Fisher), developed, and imaged with a ChemiDoc XRS+ system (Bio-Rad).

qRT-PCR

qRT-PCR was done on cDNA synthesized from 16 ng *B. bacilliformis* total RNA (for each 25 μl reaction) collected from various conditions using the Luna Universal One-Step RT-qPCR kit (New England BioLabs; Ipswich, MA) according to the manufacturer. *B. bacilliformis* total RNA was serially diluted and used as a standard curve, while primers targeting the *rpoD* housekeeping gene were used for normalization of gene expression between conditions. qRT-PCR was performed on a CFX Connect Real-Time System (Bio-Rad). cDNA from sRNAs of interest was analyzed for copy number, then divided by the copy number from the *rpoD* gene to achieve the sRNA transcripts / *rpoD*

transcript values.

Mutagenesis and RNA-RNA EMSAs

Mutagenesis of *gcvT*, *nuoF*, and *ftsH* target sequences was carried out *in vitro* using a Q5 mutagenesis kit (New England BioLabs) with specified primers (**Table S4.1**). Primers engineered with a T7 promoter sequence were used to amplify the *gcvT* (-100 to +100), *nuoF* (-86 to +100), *ftsH* (-100 to +105), RS02100 (-76 to +100), *trmD* (-50 to +100), and *hflK* (-70 to +100) target sequences, where nucleotide +1 represents the first nucleotide of the protein-coding sequence. PCR products were cloned into pCR2.1-TOPO as instructed (Invitrogen). Resulting plasmid DNA was used as the template for Q5 mutagenesis. Q5 clones were sequenced, re-amplified with T7-engineered primers, and *in vitro* transcribed using the MAXIscript T7 Transcription kit (Invitrogen) with or without 0.5 mM Bio-16-UTP (Invitrogen), as required. Dose-dependent RNA-RNA EMSAs were performed as previously described [202] using 2 nM biotin-labeled BbsR9 and varying concentrations of *in vitro*-transcribed, unlabeled target RNA.

Data availability

Aligned sequencing reads (BAM files) from all RNA-Seq experiments are available at the Sequencing Read Archive database (accession number PRJNA647605).

Ethics statement

The Institutional Biosafety Committee at the University of Montana granted approval for the experimental use of human blood (IBC 2019-05). Formal consent was obtained in verbal form from the blood donor (co-author MFM).

Results

Identification of *B. bacilliformis* sRNAs

We first analyzed the total transcriptomic profiles of *B. bacilliformis* following a timed shift from normal culture conditions (4-d incubation on HIBB plates at 30^0 C) to various *in vitro* conditions that mimic the sand fly vector and human host (**Table 4.1**). Specifically, we controlled for several environmental variables, including temperature, pH, solid vs. liquid media, and the presence of a human blood-gas atmosphere. Following quality control analysis of the resulting RNA-Seq datasets and correlation of variation analysis (**Table S4.2**), we discarded replicates that did not correlate well with others from the same condition. A principle component analysis (PCA) plot of the remaining RNA-Seq datasets confirmed statistical clustering of biological replicates (**Figure S4.1**).

Table 4.1: Conditions used to prepare *B. bacilliformis* cultures for RNA-Seq experiments.

Conditions	**Medium**	**Designation**	**Shift Time**	**Simulation**
pH 7.4, 25^0 C	HIBB plates	Pl25	2 hours	Sand fly ambient temperature
pH 7.4, 30^0 C	HIBB plates	Pl30	2 hours	Sand fly ambient temperature
pH 7.4, 37^0 C	HIBB plates	Pl37	2 hours	Human host
pH 7.4, 37^0 C with blood gas[a]	HIBB plates	PlBG	2 hours	Human host
pH 6.0, 30^0 C	HIBB liquid	pH06	2 hours	Sand fly post-blood meal
pH 7.4, 30^0 C	HIBB liquid	pH07	2 hours	Human host / sand fly blood meal mid-digestion
pH 8.2, 30^0 C	HIBB liquid	pH08	2 hours	Sand fly initial blood meal
pH 7.4, 37^0 C	HUVECs in	HUVE	24 hours	Human

with blood gas	EGM-Plus medium			endothelial cell infection
pH 7.4, 37^0 C	Human blood	HB37	2 hours	Human erythrocyte infection
pH 7.4, 37^0 C with blood gas	Human blood	HBBG	2 hours	Human erythrocyte infection

HIBB, Bacto heart infusion blood agar containing 4% defibrinated sheep blood and 2% sheep serum (vol/vol); HUVECs, human umbilical vein endothelial cells; EGM-Plus (Lonza), endothelial cell growth medium containing 2% fetal bovine serum and bovine brain extract.

[a] Blood gas is comprised of 5% CO_2, 2.5% O_2, and 92.5% N_2 at 100% humidity to simulate human blood.

Next, we visualized alignments for each RNA-Seq dataset and manually curated transcript peaks that could correspond to novel sRNAs. Peaks were found in IGRs, antisense to annotated genes (*cis*-anti) or as leader RNAs in 5' UTRs of annotated genes. The peaks were further refined based on proximity to neighboring peaks and a threshold read coverage based on expression of a housekeeping gene, *rpoD* (encoding sigma factor RpoD). We discovered 160 potential sRNAs, including seven highly-conserved sRNAs of other bacteria/alphaproteobacteria (**Table S4.3**). Of the 153 other potential sRNAs, 81 were located antisense to annotated genes, 57 were encoded in IGRs, and the remaining 15 were potential leader RNAs. Leader RNAs were included in the study because further analysis would be needed to determine if they are true leader RNAs, stand-alone sRNAs, or perhaps both. We also identified putative promoter elements for each identified sRNA based on approximated TSS's. Next, we constructed an UpSet plot to visualize the

numbers of significantly expressed sRNAs shared between various combinations of conditions (**Figure 4.1**) [210]. Results of this analysis suggested that, while 19 of the 160 identified sRNAs were expressed regardless of circumstance, the majority of sRNAs were expressed under specific conditions. Following this, we calculated TPM for each sRNA under all ten conditions in the context of the total transcriptomes (**Table S4.4**). TPM is a normalized measure of gene expression, and although it is not always appropriate to compare TPM values across different RNA-Seq experiments [168], we constructed a heatmap to get a broad sense of sRNA expression patterns (**Figure S4.2**). These results revealed three distinct clusters of conditions with similar sRNA expression patterns and allowed us to identify interesting sRNAs for further characterization.

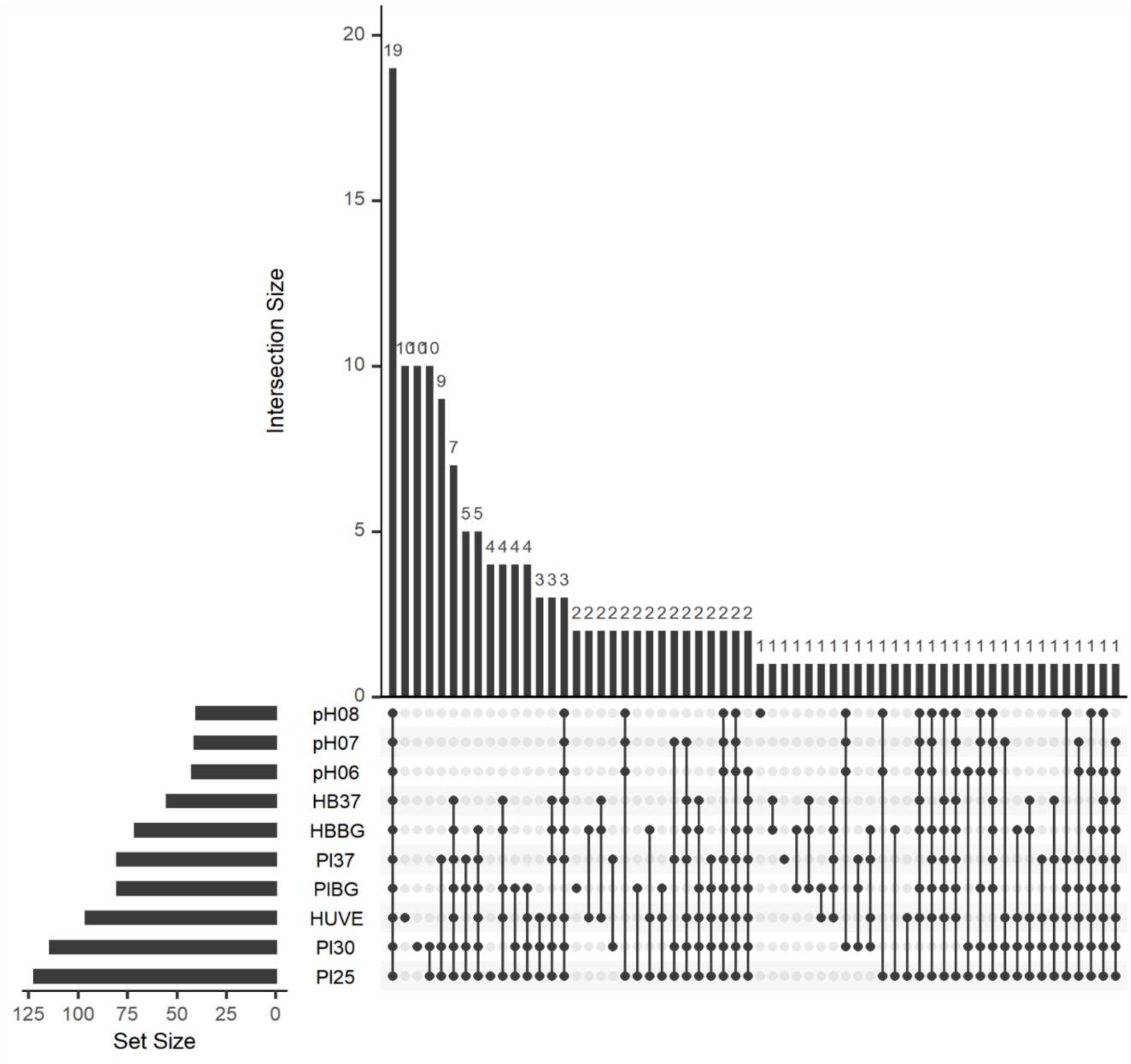


Figure 4.1: Most *B. bacilliformis* sRNAs are expressed under specific conditions. An UpSet plot is shown and displays the number of sRNAs shared among various combinations of conditions tested. The bar graph to the left indicates the quantity of sRNAs with a TPM >300 under the conditions shown. The connected nodes indicate shared conditions giving rise to the number of sRNAs expressed as indicated by the bar graph at the top.

Verification of select *B. bacilliformis* sRNAs

Of the 160 putative sRNAs discovered in *B. bacilliformis*, we assigned name designations to 22 of them based on appraisal of general relevance, including six widely-conserved sRNAs, such as BbtmRNA (***B***. ***b****aciliformis* **tmRNA**), and 15 novel *Bartonella*-specific sRNAs. We also gave a name designation to the ***B***. ***b****acilliformis* **g**rou**p I** intron (BbgpI), a group I intron with related elements previously identified, but not characterized, in other alphaproteobacteria [211]. To corroborate RNA-Seq expression results and verify sRNA expression, we chose eight novel sRNAs at random and conducted Northern blot analyses. Results of the Northern blots confirmed the expression of all eight sRNAs (**Figure 4.2**).

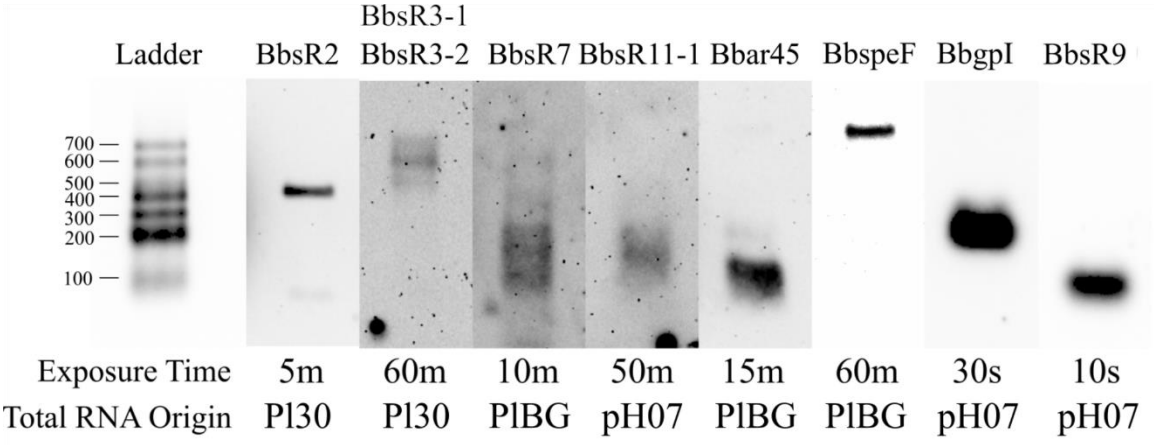


Figure 4.2: Northern blot analyses confirm expression of eight putative *B. bacilliformis* sRNAs. Eight separate Northern blots were run under identical experimental conditions (see Chapter 4: Materials and Methods). RNA ladders from the respective blots were aligned with each other for presentation of the resolved total RNA. sRNA designations are shown above each blot. Exposure times (minutes, m; seconds, s) and origin of the RNA are indicated below each blot.

We analyzed the differential expression of *B. bacilliformis* genes across the ten tested conditions by performing relevant pairwise comparisons (**Table 4.2**) using the DESeq2 package in R version 3.4.4 [193]. For this analysis, transcriptomes from all ten conditions were compared simultaneously, while specific, relevant pairwise comparisons were made (**Table 4.2**). Results showed the greatest number of significant differentially expressed sRNAs by comparing solid-to-liquid media and host cell types. Other sRNA candidates were also found to be differentially regulated by these comparisons (see **Table S4.3**). We then utilized quantitative reverse transcription PCR (qRT-PCR) to validate the DESeq2 results. In doing so, we confirmed eleven sRNAs to be significantly differentially expressed under the relevant conditions (**Figure S4.3**).

Table 4.2: DESeq2 comparisons made.

Comparison	**Controlled Conditions**	**No. sRNA DEGs**
Pl30[a] *vs.* Pl25	Temperature	0
Pl30[a] *vs.* Pl37	Temperature	2
Pl25[a] *vs.* Pl37	Temperature	0
Pl30[a] *vs.* pH07	Solid/liquid media	12
pH07[a] *vs.* pH08	pH level	0
pH07[a] *vs.* pH06	pH level	0
pH06[a] *vs.* pH08	pH level	7
Pl37[a] *vs.* PlBG	Blood gas	1
PlBG[a] *vs.* HBBG	Solid/liquid media, human/sheep blood	6
PlBG[a] *vs.* HUVEC	Solid/liquid media, cell type	18
pH07[a] *vs.* HUVEC	Temperature, cell type	6
pH07[a] *vs.* HB37	Temperature, human/sheep blood	3
HBBG[a] *vs.* HUVEC	Cell type	17

DEGs, significant differentially expressed genes.

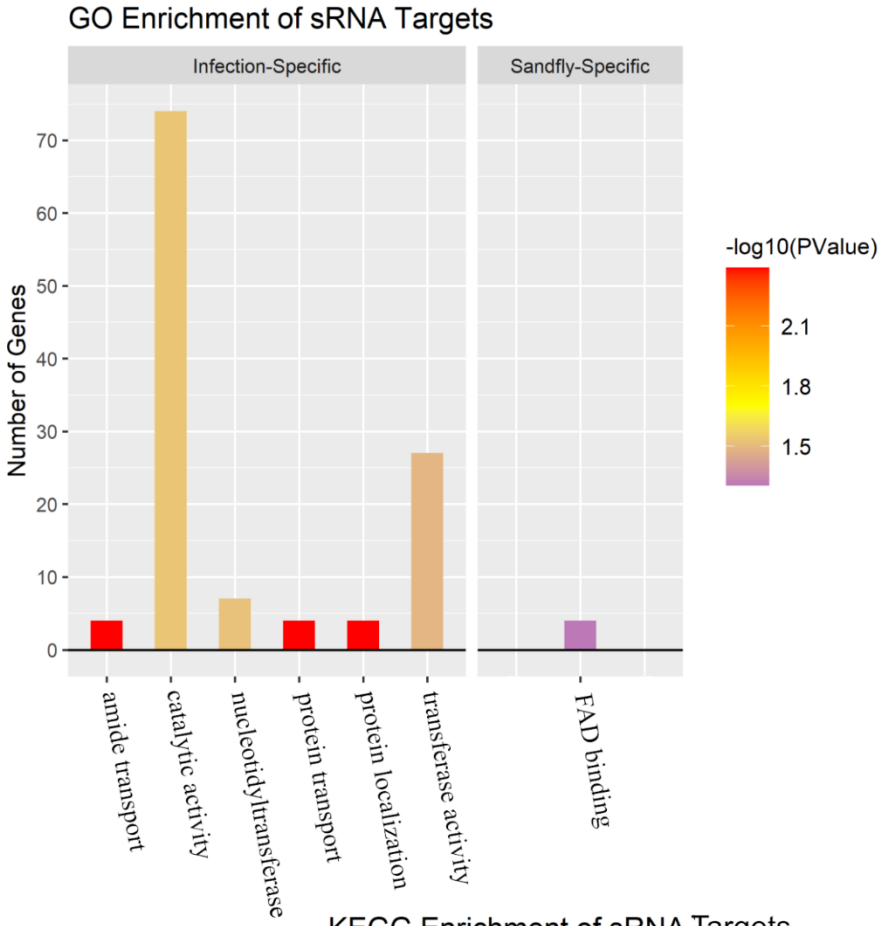
[a]Reference dataset

Condition-specific sRNAs target mRNAs enriched in specific pathways

We grouped several sRNAs based on their expression patterns across the ten conditions tested by using heatmap comparisons combined with data from the DESeq2 analysis. For example, multiple sRNAs were significantly and strictly expressed under conditions used to simulate or actually infect human cells (i.e., PlBG, HUVEC, HB37, and HBBG; see **Table 4.1**). These sRNAs were classified as human "infection-specific" based on their restricted upregulation (defined by TPM greater than the mean TPM plus one standard deviation) in at least two of these four conditions. Based on this definition, we identified 24 infection-specific sRNAs (see **Table S4.4**).

We were also curious to determine if the predicted mRNA targets of the infection-specific sRNAs significantly corresponded to particular gene classifications to provide clues regarding their upregulation under infection-specific conditions. Using gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses and the IntaRNA 2.0 sRNA target prediction program [43], we determined that the predicted mRNA targets of these sRNAs (**Table S4.5**) were enriched for several GO terms, including protein/amide transport and nucleotidyltransferase activity (**Figure 4.3A**). The pool of mRNA targets was also enriched for the glycerophospholipid/glycerolipid metabolism and nucleotide excision repair KEGG pathways (**Figure 4.3B**).

A


GO Enrichment of sRNA Targets
Infection-Specific
Sandfly-Specific
Number of Genes
0
10
20
30
40
50
60
70
amide transport
catalytic activity
nucleotidyltransferase
protein transport
protein localization
transferase activity
FAD binding
-log10(PValue)
2.1
1.8
1.5

B

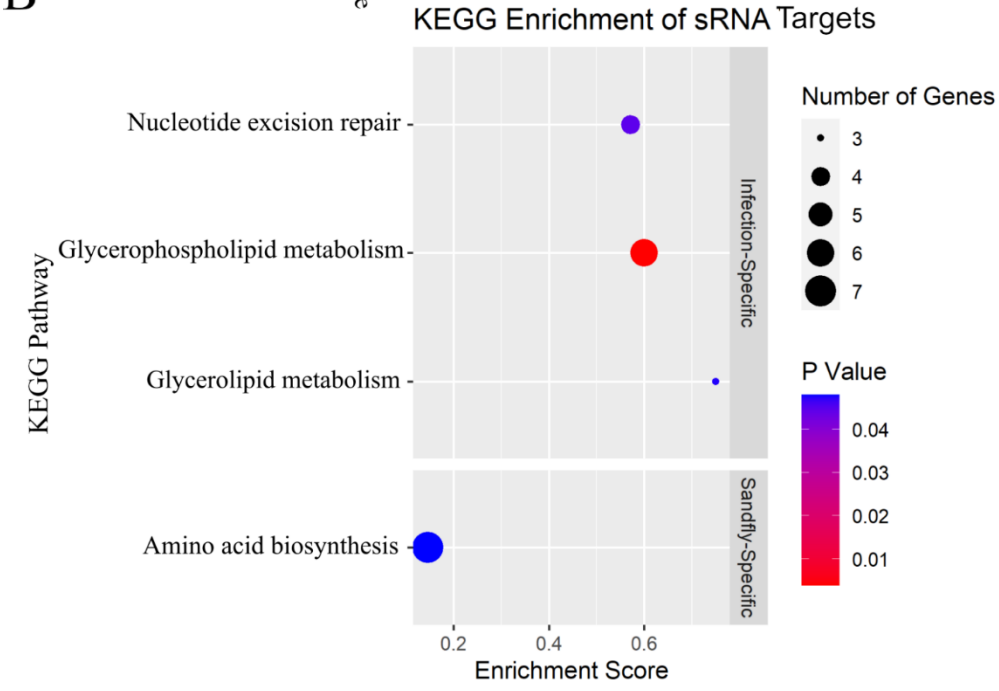

KEGG Enrichment of sRNA Targets
KEGG Pathway
Nucleotide excision repair
Glycerophospholipid metabolism
Glycerolipid metabolism
Amino acid biosynthesis
Infection-Specific
Sandfly-Specific
0.2
0.4
0.6
Enrichment Score
Number of Genes
3
4
5
6
7
P Value
0.04
0.03
0.02
0.01

Figure 4.3: Condition-specific sRNA targets are enriched in several GO terms and KEGG pathways. A) Faceted bar graph of GO enrichment terms for infection-specific and sand fly-specific sRNA targets. Height of the bars indicates the number of sRNA targets containing that GO term, while the color displays the significance of enrichment. **B)** Faceted dot plot of KEGG enrichment terms for infection-specific and sand fly-specific sRNA targets. The enrichment score refers to the ratio of the number of gene targets corresponding to a particular pathway to the total number of genes in that pathway. Dot colors represent significance (p-value) of enrichment for that particular KEGG pathway.

We also determined that eight sRNAs were only expressed under conditions simulating the sand fly vector (i.e., Pl25, pH06, pH07, and pH08; see **Table 4.1**). Despite the fact that Pl25 clustered separately from pH06, pH07, and pH08 on the heatmap (**Figure S4.2**), we included it in the sand fly-specific conditions due to upregulation of several sRNAs under both pH06/pH08 and Pl25 conditions (BB026-1, BB103-2, BB103-3, and BB124; **Table S4.4**). These and other sRNAs were classified as "sand fly-specific sRNAs" based on their restricted upregulation in at least two of these four conditions. The predicted mRNA targets of the sand fly-specific sRNAs (**Table S4.6**) were enriched for the flavin adenine dinucleotide (FAD) binding GO term (**Figure 4.3A**) and the amino acid biosynthesis KEGG pathway (**Figure 4.3B**).

BbgpI is a group I intron that splices *in vivo* and self-splices *in vitro*

BB009 was initially identified as a sRNA of interest based on its high expression across multiple conditions (**Table S4.4**). A BLAST search of the BB009 gene sequence

(hereafter referred to as BbgpI) showed that it was highly homologous to a group I intron conserved in several other alphaproteobacteria and encoded in host $tRNA_{CCU}^{Arg}$ genes [211]. Since *B. bacilliformis* has no annotated $tRNA_{CCU}^{Arg}$ gene, we initially assumed that BbgpI was encoded in an IGR. However, such a location would be novel for group I introns, which are selfish genetic elements found in tRNA, rRNA, and rarely, protein-coding genes (reviewed in [74]).

To address this discrepancy, 5' RACE was utilized to determine the 5' end of the putative spliced-out RNA segment. From these results, we determined that BbgpI was flanked by CCT DRs, identical to those produced by the $tRNA_{CCU}^{Arg}$ alphaproteobacterial group I intron (**Figure S4.4A**) [211]. Also, the predicted secondary structure of BbgpI possessed conserved group I intron stem structures (**Figure S4.4B**). Finally, we scanned the locus and flanking sequences with the tRNAscan-SE 2.0 web server and identified a $tRNA_{CCU}^{Arg}$ gene (**Figure S4.4C**), but only when a sequence with BbgpI spliced out was used [212]. Taken together, these results suggest that BbgpI is a member of a conserved, alphaproteobacterial group I intron family and disrupts an unannotated $tRNA_{CCU}^{Arg}$ gene (locus: c42404-42711) of *B. bacilliformis*.

Although homology and structural results suggested that BbgpI was a group I intron, it was unclear whether BbgpI was able to self-splice or whether a protein cofactor was required [211]. To address this question, we examined BbgpI's ribozyme activity *in vitro*. Following *in vitro* transcription, cDNA synthesis, and PCR analysis with the primers shown in **Figure S4.4A**, we determined that BbgpI self-splices *in vitro* and is spliced *in vivo* (**Figure 4.4**), in keeping with other group I introns.

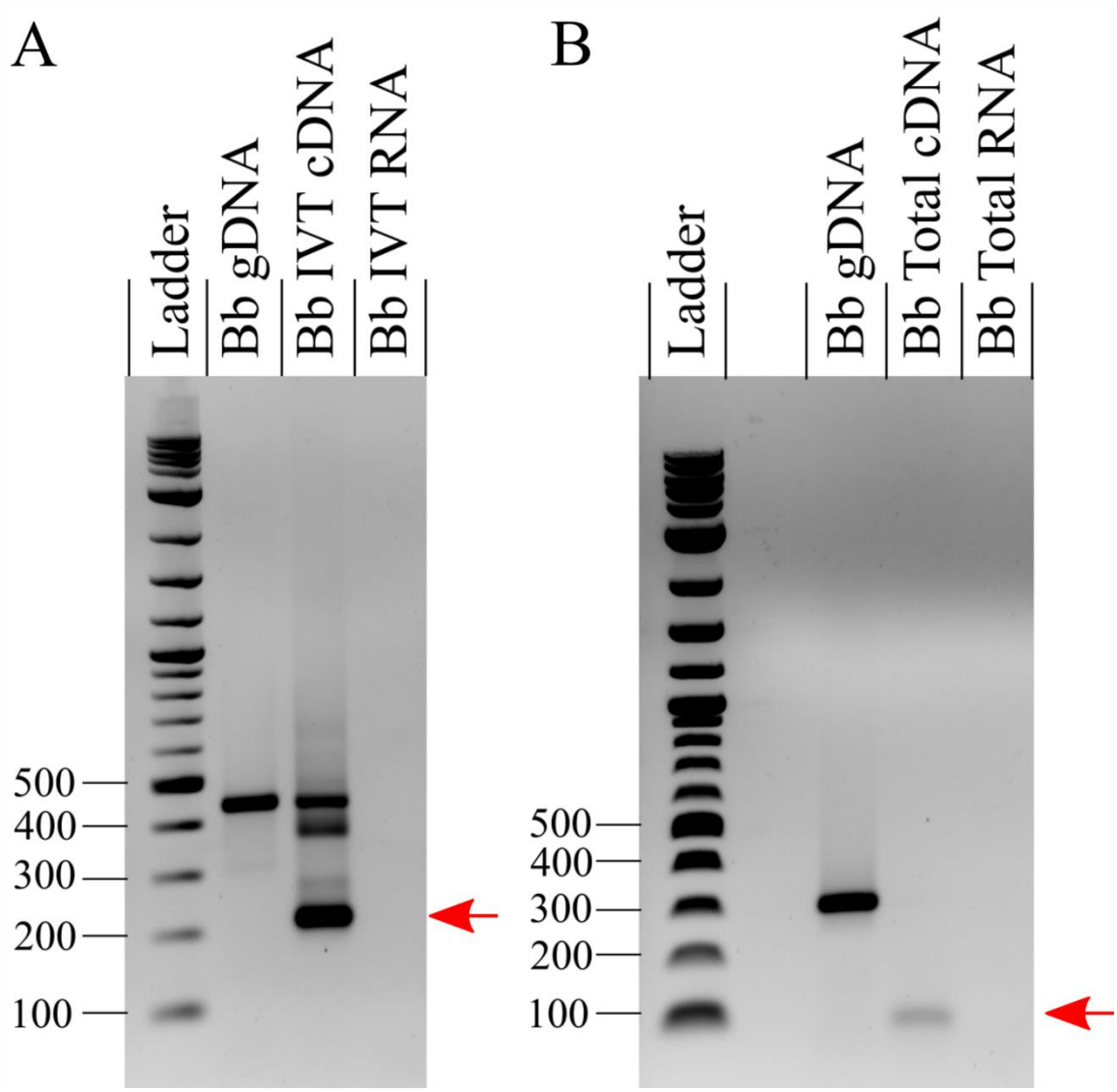


Figure 4.4: BbgpI self-splices *in vitro* and is spliced *in vivo*.

A) PCR analysis of an *in vitro*-transcribed (IVT) region of *B. bacilliformis* (Bb) genomic DNA (gDNA) containing BbgpI. Ethidium bromide-stained agarose (1%) gels are shown. PCR on the resulting cDNA using "Nested Primers" (see **Figure S4.4A)** produced unspliced DNA (450-bp band), a partially-spliced product of ~390 bp, plus a 218-bp band corresponding to the BbgpI flanking region where BbgpI self-spliced out (indicated by the red arrow). PCR on a Bb IVT RNA negative control did not produce product. **B)** As in A) but utilizing cDNA synthesized from *B. bacilliformis* total RNA using "Splice

Flank Primers" (see **Figure S4.4A)**. A 308-bp amplicon was produced from gDNA, whereas a 76-bp band, corresponding to the BbgpI flanking region with BbgpI spliced out, was produced from cDNA generated from total RNA (indicated by the red arrow). PCR on a Bb total RNA negative control did not produce product.

BbsR9 is a sand fly-specific sRNA

BB092 (hereafter referred to as BbsR9) was initially identified as a sRNA of interest due to its restricted high-level expression under conditions that simulated the sand fly vector (**Table 4.1, Table S4.4**). In addition, BbsR9 was found to have well-defined, predicted sigma-70 promoter and Rho-independent terminator regions, and it was conserved among several other *Bartonella* spp. (**Table S4.3**). To elucidate the BbsR9 gene, the TSS was determined by 5' RACE. These results showed two possible sites with equal representation among the six clones sequenced (**Figure S4.5A**). We therefore wished to determine if two distinct transcripts of BbsR9 were made or if there was a single, dominant transcript for the sRNA. In addition, we wanted to confirm BbsR9 expression across the conditions examined. To this end, we set out to determine if BbsR9 would remain highly expressed in sheep blood shifted to 37^0 C or if we would see a downregulation of the sRNA, as in *B. bacilliformis* shifted to human blood at 37^0 C (HB37/HBBG; see **Table S4.4**). Northern blot analyses showed BbsR9 transcript and, together with the intensity of the signal, indicated a single, dominant TSS (bolded underlined in **Figure S4.5A**). Interestingly, we saw a distinct downregulation of BbsR9 when *B. bacilliformis* was shifted to sheep blood at 37^0 C compared to 30^0 C (**Figure S4.5B**). This decrease in RNA suggests that BbsR9 is primarily expressed under

conditions that simulate the sand fly vector and not the human host. Taken as a whole, results of the Northern blots and RNA-Seq suggest that both a liquid medium (Pl30 vs. pH07; see **Table S4.4**) and a temperature below 37^0 C upregulate BbsR9.

BbsR9 targets transcripts of *ftsH*, *nuoF*, and *gcvT in vitro*

Since BbsR9 expression was restricted to sand fly-like conditions, we were interested in characterizing its mRNA targets to shed light on the sRNA's role in regulation. To that end, we first utilized the TargetRNA2 [42], IntaRNA [43] and CopraRNA [44] algorithms to determine potential mRNA targets (**Table 4.3**). From these results, we selected transcripts of *ftsH*, *nuoF*, *gcvT*, *trmD*, *hflK*, and a predicted DNA response regulator (RS02100) as potential targets for characterization based on shared predictions between algorithms and the strength of predicted binding events.

Table 4.3: mRNA targets for BbsR9, as predicted by the indicated algorithms.

Rank	**TargetRNA2**	**IntaRNA**	**CopraRNA**
1	***nuoF* (0.0001)**	***gcvT* (0.0033)**	RS06660 (0.0013)
2	RS01360 (0.012)	RS01025 (0.0043)	***nuoF*** (0.0103)
3	*czrB* (0.004)	***trmD* (0.0060)**	RS02895 (0.0113)
4	*ftsE* (0.010)	***ftsH* (0.0071)**	***gcvT*** (0.0153)
5	RS05725 (0.012)	**RS02100 (0.0090)**	RS02955 (0.0232)
6	*rplX* (0.033)	DUF475 (0.0125)	*efp* (0.0286)
7	*flgC* (0.043)	Pseudogene (0.0132)	***ftsH*** (0.0288)
8	*aroP* (0.044)	*tonB* (0.0136)	***trmD*** (0.0477)
9		***hflK* (0.0148)**	
10		***nuoF* (0.0276)**	

p-values < 0.05 are indicated in parentheses; bolded gene targets were chosen for further study.

To demonstrate physical interactions between BbsR9 and the chosen mRNA candidates, RNA-RNA EMSAs were done using *in vitro*-transcribed BbsR9 and segments of the target mRNAs of interest with their predicted sRNA target regions. Results of the EMSAs showed that BbsR9 bound mRNAs of *ftsH*, *nuoF*, and *gcvT in vitro*, as judged by the novel hybrid RNA species showing markedly slower migration during gel electrophoresis (**Figure 4.5**). Hybrid RNAs were not observed for the other three candidate mRNAs, suggesting that sRNA binding did not occur.

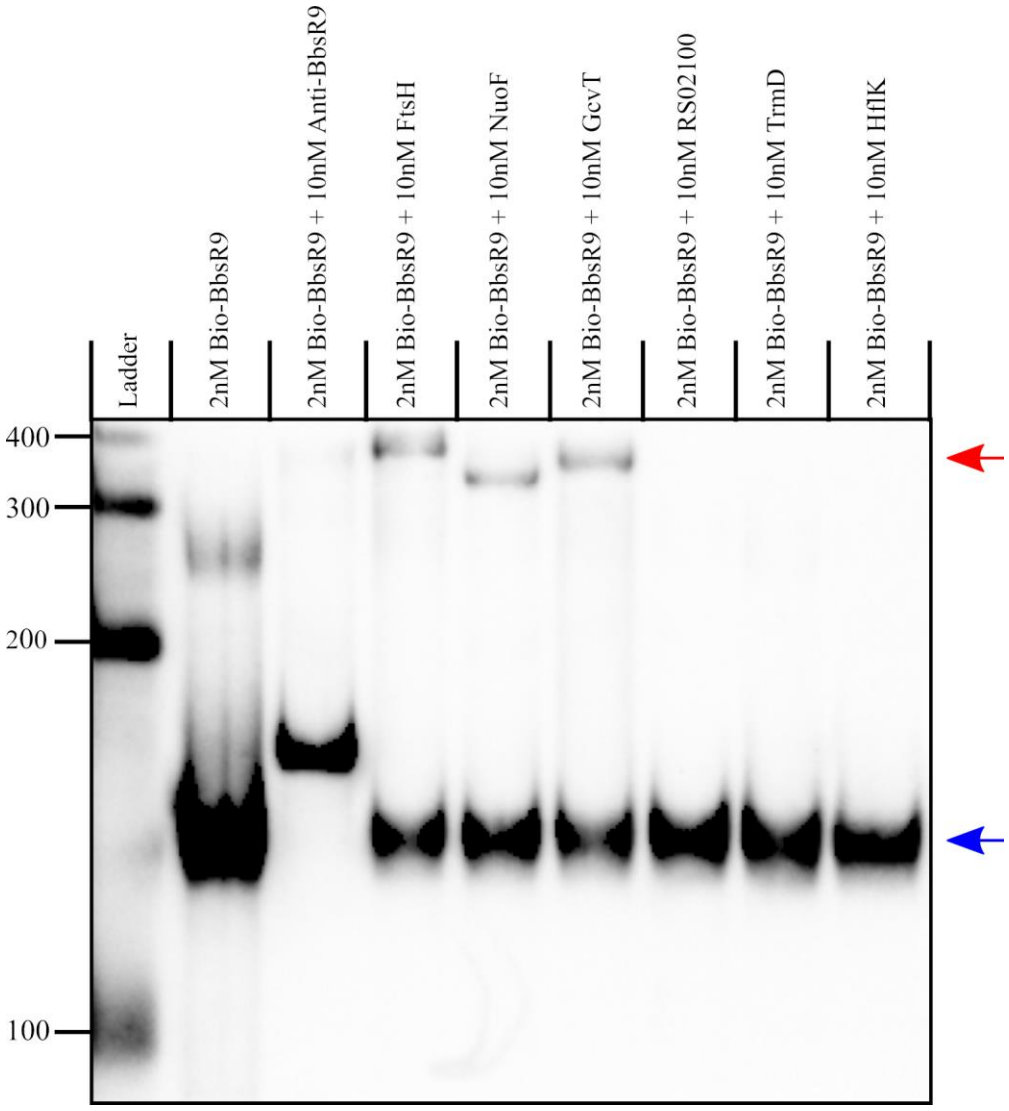


Figure 4.5: BbsR9 targets transcripts of *ftsH*, *nuoF* and *gcvT in vitro*.

RNA-RNA EMSA of biotin-labeled *in vitro*-transcribed BbsR9 binding to *in vitro*-transcribed mRNA segments of the *ftsH*, *nuoF*, *gcvT*, *trmD*, BARBAKC583_RS02100 and *hflK* genes. Red and blue arrows indicate bands corresponding to BbsR9 bound and unbound to target RNAs, respectively. Base values of the RNA size standard (ladder) are shown on the left.

We further characterized BbsR9-mRNA interactions by mutagenizing the predicted sRNA-binding regions of *ftsH*, *nuoF*, and *gcvT* (**Figure 4.6**). The predicted *ftsH*-binding region was extensive, so we created two distinct mutants for this target as well as a double-mutant (**Figure 4.6**). RNA-RNA EMSAs conducted with the mutagenized target mRNAs showed complete elimination of BbsR9 binding to all three targets *in vitro* regardless of increasing target quantity present in the hybridization reaction (**Figure 4.7**). As expected, wild-type targets showed dose-dependent hybridization and signal intensity. Interestingly, abrogation of *ftsH* transcript binding by BbsR9 was only observed with mutation 1 (Mut1) alone or in combination with mutation 2 (Mut2), whereas Mut2 alone did not prevent BbsR9 binding to the RNA (**Figure 4.7**). In consideration of the RNA secondary structure predictions and the EMSA results, we conclude that BbsR9 primarily targets mRNA transcripts via multiple GC-rich regions of a large, predicted stem-loop structure (see **Figure 4.6**) [145].

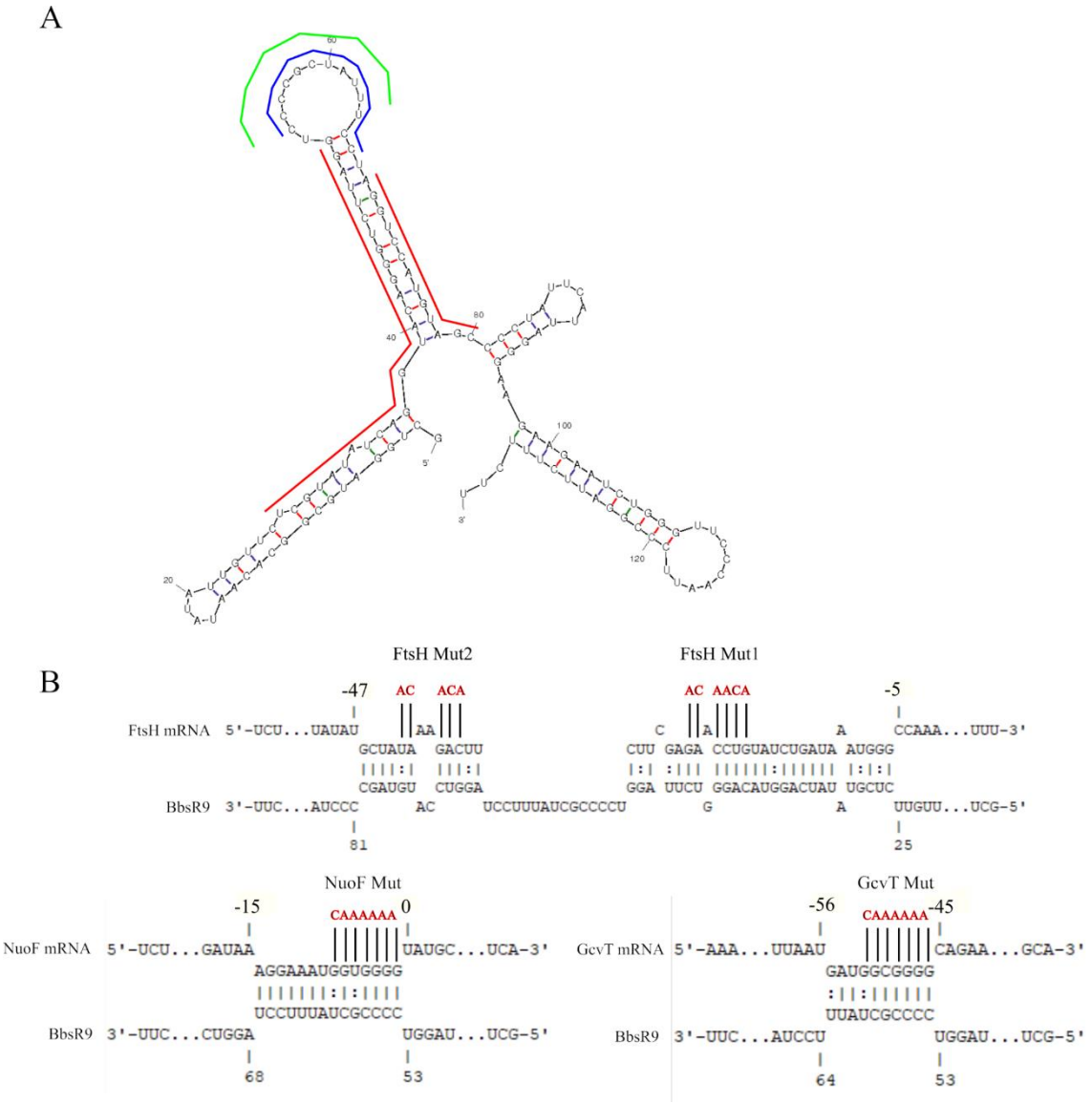


Figure 4.6: BbsR9 binds its targets through several GC-rich predicted seed regions. **A)** Mfold secondary structure prediction of BbsR9 (ΔG = -46.9 J mol^{-1}) with predicted seed regions for *ftsH*, *nuoF*, and *gcvT* transcript binding indicated by red, blue, and green lines, respectively. **B)** Predicted IntaRNA BbsR9 target seed regions of the indicated transcripts. For the mRNA targets, nucleotide position +1 represents the first nucleotide of the respective start codon. Mutagenized bases of each mRNA are indicated in red.

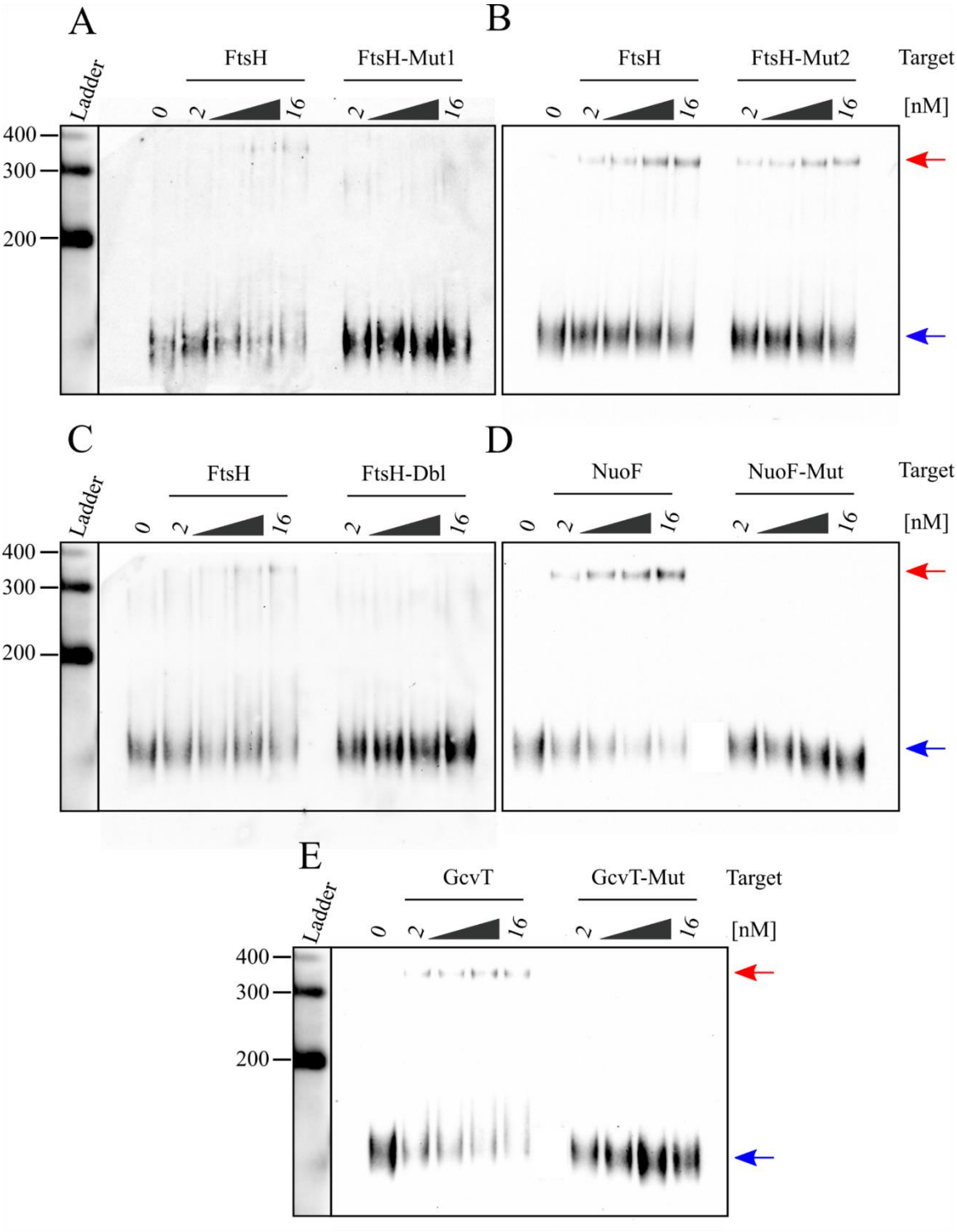


Figure 4.7: BbsR9 binds to *ftsH*, *nuoF*, and *gcvT* transcripts via specific GC-rich seed regions.

RNA-RNA EMSAs showing dose-dependency of biotin-labeled BbsR9 binding to wild-type but not mutated, *in vitro*-transcribed segments of **A-C)** *ftsH*, **D)** *nuoF* and **E)** *gcvT*. Mutated regions correspond to those shown in **Figure 4.6B**, and "Dbl" specifies the double *ftsH* mutant. Red and blue arrows indicate bands corresponding to BbsR9 bound and unbound to target RNAs, respectively. All lanes contained 2 nM biotin-labeled BbsR9 in addition to increasing amounts of the indicated target (2 nM, 4 nM, 8 nM, and 16 nM).

Discussion

In this study, we performed an extensive transcriptomic analysis of *B. bacilliformis* grown *in vitro* then shifted to one of 10 distinct conditions that mimic environments encountered by the bacterium during its natural life cycle. We chose these conditions in order to control for a variety of environmental factors that may directly influence expression of certain sRNAs. For example, temperature (25^0C, 30^0C, 37^0C), pH levels (pH 6, pH 7.4, pH 8.2), solid/liquid substrates, and presence of a blood-gas atmosphere (5% CO_2, 2.5% O_2, and 92.5% N_2 at 100% humidity) were all examined. In addition, we included RNA-Seq experiments from experimental infections of low-passage human vascular endothelial cells (HUVE) and fresh human blood samples (HB37 and HBBG). From these experiments, we discovered 160 sRNAs expressed by *B. bacilliformis* in at least one of the conditions tested.

Although we initially approached sRNA discovery using an automated approach, some clear-cut sRNAs were missed during the process. This issue led us to manually curate the 10 stranded RNA-Seq alignments, scanning each annotated gene, leader

region, and IGR for aligned reads forming peaks that could represent novel sRNAs. These peaks were required to surpass a pre-determined read coverage threshold determined independently for each condition based on reads aligned to the *rpoD* (locus tag: BARBAKC583_RS04670) gene, which was consistently expressed across all 10 conditions (TPM ~300; **Table S4.4**). We remained consistent by using *rpoD* as the housekeeping gene in qRT-PCR analyses (**Figure S4.3**) and employing a 300 TPM threshold for the purpose of the UpSet plot (**Figure 4.1**).

The putative sRNAs identified were organized into three categories (IGR, *cis*-anti, or leader) depending on location of the corresponding sRNA locus. Each sRNA category has implications for its potential function. For example, IGR sRNAs are likely *trans*-acting with small seed regions that often bind multiple mRNAs. *Cis*-anti sRNAs most likely target the gene to which they are antisense, so target identification via algorithms such as IntaRNA would not be useful. Putative leader sRNAs are peaks that were identified sense to and in the 5' UTRs of protein-coding genes. Although the identified peaks appeared distinct from those within the actual coding sequence, the possibility remains that these peaks are not *trans*-acting sRNAs. More likely, these leader RNAs may serve as *cis*-acting regulatory components, like riboswitches, which are co-transcribed with the downstream protein-coding gene and harbor regulatory stem-loops that influence translation of the respective transcript [213]. Determining whether the identified leader sRNAs are *cis*- and/or *trans*-acting elements would require further experiments such as Northern blots and 5' RACE experiments to see if there is read-through into the downstream gene.

We performed Northern blot analyses on the putative leader sRNAs, BbspeF and BbsR7, and found that these are likely *cis*-acting leader RNAs, since the RNA sizes suggest read-through into the downstream gene (**Figure 4.2**). Northern blot analysis also verified the existence of six other sRNAs, although some of the results raise additional questions. For example, the presence of BbsR2 was detected, but the apparent band of ~450 bases is considerably larger than its predicted 284-base band (see **Figure 4.2**, **Table S4.3**). Although we identified a putative promoter element for BbsR2, we did not identify a Rho-independent terminator, so it is possible that the sRNA extends further downstream than predicted. It was also unclear whether BbsR3-1 / BbsR3-2 represented two distinct sRNAs. However, Northern blot analysis utilizing a probe against BbsR3-1 confirmed that there was a single transcript produced whose length (~600 bp) was equal to the sum of the predicted sizes of BbsR3-1 and BbsR3-2, indicating that this locus probably produces a single sRNA species (**Figure 4.2**). The BbsR7 blot also requires explanation. Here, several bands were identified, including smaller bands of ~200 bases and a larger band of ~600 bases. Since BbsR7 is predicted to be a leader sRNA, it is possible that the smaller bands represent the sRNA being independently expressed, while the larger band may represent BbsR7 being co-transcribed with the downstream gene (BARBAKC583_RS01695), which is 225 bp long (**Figure 4.2**). We also probed in the BbsR11-1 / BbsR11-2 region to determine if the two corresponding RNA-Seq peaks represented two distinct sRNAs. In this case, the Northern blot showed a single band that corresponded only to the predicted size of BbsR11-1 (**Figure 4.2**), suggesting that the locus harbors two distinct sRNAs. Northern blots for BbgpI and BbsR9 produced single

bands (**Figure 4.2**) that corresponded well to the estimated sizes of their respective peaks by RNA-Seq.

Among the sRNAs analyzed by Northern blot, Bbar45 and BbspeF are intriguing, non-coding RNA elements worthy of further characterization. Bbar45 belongs to the αr45 sRNA family first described in *Sinorhizobium meliloti*, but it is widely conserved in other Rhizobiales [214]. Functional characterization of sRNAs in the αr45 family has not been performed, although the *S. meliloti* αr45 can be co-immunoprecipitated with Hfq [215]. Since Hfq is an RNA chaperone that facilitates sRNA-mRNA interactions, we hypothesize that the *S. meliloti* αr45 sRNA may be *trans*-acting [16]. Here, we confirmed that Bbar45 is independently expressed from BbspeF, which lies immediately downstream (**Figure 4.2**). While this observation was previously observed in *S. meliloti*, it was unclear whether it was the case for other alphaproteobacteria [214]. Based on Northern blot results showing a transcript >700 bases (**Figure 4.2**), BbspeF is likely a leader RNA that is not independently expressed from its downstream gene. The *speF* leader RNA was initially discovered during a search for alphaproteobacterial riboswitches and was named for its upstream location relative to the *Bacillus subtilus speF* ortholog, which codes for an ornithine decarboxylase protein involved in polyamine biosynthesis [216]. However, metabolites of the polyamine biosynthesis pathway of *B. subtilus* were not shown to bind to the *speF* leader *in vitro*, leaving the element's function unclear [216]. More experiments are needed to determine the regulatory role of the BbspeF leader as well as the function of the Bbar45 sRNA in *B. bacilliformis*.

An RNA secondary structure prediction of BbsR14 showed two stem-loops with nearly identical sequences of TTCCTCCTAA. Remarkably, these are anti-SD motifs

most often found in 16S rRNA, where they function in translational initiation. The presence of SD sequences outside of a RBS is rare, as they are selected against in the context of mRNAs, since they can cause ribosome stalling due to hybridization with 16S rRNA [217, 218]. One way in which sRNAs regulate translation is to bind directly to the RBS to occlude the ribosome and inhibit translational initiation [1]. In most cases, this is accomplished via a seed region that overlaps the SD sequence and extends up and/or downstream [6]. The predicted BbsR14 secondary structure displays unique potential seed regions solely comprised of anti-SD sequences. We speculate that this arrangement could provide opportunities for indiscriminate translational repression by the BbsR14 sRNA.

We also analyzed each of the identified *B. bacilliformis* sRNAs and discovered that BB019, BB113, and BB125-2 possessed a single anti-SD sequence (CCTCCT). Interestingly, of the four sRNAs that contain anti-SD sequences, BbsR14 and BB113 were significantly upregulated at pH08 relative to pH06 (see **Table S4.3**). Conditions of pH08 and pH06 were designed to simulate the initial and late stages of the sand fly after feeding, respectively. Thus, downregulation of translation may be advantageous for bacterial survival during initial stages within the sand fly's midgut. As *B. bacilliformis* persists in the sand fly and infection proceeds, "gearing up" for a subsequent mammalian infection may occur as the insect prepares for another blood meal. Supporting this notion, we also identified 6S RNA as a sand fly-specific sRNA that was upregulated at pH08 vs. pH06, although not significantly (see **Table S4.4**). 6S RNAs function by binding to and sequestering the RNA polymerase holoenzyme [11]. The resulting global repression of transcription during the initial stages of sand fly infection and, to a lesser extent,

throughout a sand fly infection, could conceivably promote persistence of *B. bacilliformis* in the insect.

The mRNA target enrichment analyses for potential sand fly and infection-specific sRNAs provided insight into the regulation of pathways necessary for bacterial survival in these disparate environments. For example, targets of sand fly-specific sRNAs were significantly enriched for genes involved in the FAD-binding GO term and the biosynthesis of amino acids KEGG pathway (**Figure 4.3**). FAD-binding proteins include a wide array of proteins that participate in numerous biological processes. Enrichment of these genes may reflect a relatively low availability of FAD during residence in the sand fly. *B. bacilliformis* encodes a bifunctional riboflavin kinase/FAD synthetase (BARBAKC583_RS05700), and although this gene is relatively lowly expressed in all conditions tested, there is a downregulation of its expression under sand fly-like conditions (pH07, average TPM = 49.04) compared to human blood infections (HBBG, average TPM = 84.94). Enrichment of genes involved in the biosynthesis of amino acids is possibly explained by the likely downregulation of transcription and translation under sand fly-like conditions, where *B. bacilliformis* enters into a stationary phase that may promote persistence.

The human infection-specific sRNA targets were enriched in multiple GO terms associated with transferase activities, transporters, and the phospholipid biosynthetic process and KEGG pathways associated with glycerolipid/glycerophospholipid metabolism and nucleotide excision repair (**Figure 4.3**). Among these, there is a clear regulation of cell wall constituents during human infection conditions that would presumably be associated with morphological changes to the bacterium in the human host

or perhaps as a means of expressing outer membrane proteins/transporters that aid in bacterial growth and replication during infection. This may very well also be in response to stressors encountered under these conditions, since nucleotide excision repair also seems to be significantly regulated by infection-specific sRNAs.

When analyzing mRNA targets of the infection-specific sRNAs, it was clear that numerous sRNAs were predicted to target the same mRNA in several cases (see **Table S4.5**). For example, of the 19 presumed *trans*-acting, infection-specific sRNAs, three independently target BARBAKC583_RS04310 transcripts, coding for lysylphosphatidylglycerol synthetase; an enzyme previously shown to augment a pathogen's defense against host cationic antimicrobial immune peptides [219]. Additionally, four of the 19 predicted *trans*-acting, infection-specific sRNAs target BARBAKC583_RS00395 transcripts, coding for cobaltochelatase subunit CobT, which is involved in the synthesis of cobalamin (vitamin B_{12}), an essential coenzyme for many biological reactions [220]. It is difficult to ascribe roles to these mechanisms without knowing whether the sRNA-mediated regulation is positive or negative, although it is worth noting that redundant targeting is not a result of sRNA duplication, and each predicted binding site on these transcripts is unique. We hypothesize that redundant regulation of particular mRNAs may serve to "hyper-regulate" protein production in response to subtle differences in environmental cues. This kind of redundant regulation of mRNAs from multiple "sibling sRNAs" has been described in other pathogens, so further research into the function of sibling sRNAs of *B. bacilliformis* could be fruitful [221].

We found conservation of some sRNAs among other alphaproteobacteria species using discontinuous megaBLAST analysis (**Table S4.3**). Unfortunately, we were only

able to analyze IGR and leader sRNAs, since *cis*-anti sRNAs showed broad sequence conservation due to their close linkage to protein-coding genes. The majority of analyzed sRNAs was unique to *B. bacilliformis*, while the BB036 sRNA group was unique to the KC583 strain of *B. bacilliformis*. Five other sRNAs were widespread in *Bartonella* spp., including BbsR9 which was characterized in this study. Conservation of BbsR9 in other *Bartonella* spp. further highlights its potential importance. Since *Bartonella* spp. are typically transmitted to mammals by various arthropods (ticks, sand flies, fleas, lice, etc.), it is possible that BbsR9 plays a role in persistence in many vectors. Seven more sRNAs were found in additional alphaproteobacteria, including ubiquitous sRNAs like 6S RNA and tmRNA, conserved alphaproteobacteria sRNAs like Bbar45, and the $tRNA_{Arg}^{CCU}$ group I intron.

BbgpI is a member of a $tRNA_{CCU}^{Arg}$ group I intron family first identified in *Agrobacterium tumefaciens* and later found in other alphaproteobacteria [222, 211]. Group I introns are selfish genetic elements that insert into tRNAs, rRNAs, and protein-coding genes. Although group I introns are ribozymes and RNA splicing is auto-catalytic, they sometimes require protein co-factors for self-splicing *in vitro*, and it is presumed that all group I introns require protein co-factors to some extent for splicing *in vivo* [74]. Here, we have demonstrated that BbgpI self-splices *in vitro* and is spliced *in vivo*. Furthermore, we have shown that BbgpI is not located in an IGR as presumed, but rather within an unannotated $tRNA_{CCU}^{Arg}$ gene. Since the flanking $tRNA_{CCU}^{Arg}$ gene retains all necessary tRNA domains (see **Figure S4.4C**), we predict that the tRNA is functional following intron splicing. This novel tRNA gene might have implications for future analyses of *B. bacilliformis* involving codon bias, conservation of tRNA genes, amino

acid scavenging, etc. Furthermore, this discovery suggests further optimization may be required for current tRNA scanning algorithms.

We also characterized the targeting and molecular interactions of BbsR9, as the sRNA was only appreciably expressed under pH06, pH07, and pH08 conditions (**Table S4.4**). For reference, these conditions reflect a liquid blood / serum environment at 30^0 C (**Table 4.1**) and simulate the sand fly's midgut following a blood meal. It is interesting to note that the Pl30 condition is identical to pH07 except that Pl30 represents a solid medium. Furthermore, Northern blot analyses indicated that, in addition to the liquid medium requirements, BbsR9 expression was restricted to temperatures < 37^0 C (**Figure S4.5B**). The regulatory mechanisms that facilitate such an expression pattern warrant further investigation.

We verified several mRNA targets of BbsR9 using RNA-RNA EMSAs. Among the targets were transcripts of the *ftsH*, *nuoF*, and *gcvT* genes. First, *ftsH* codes for the FtsH zinc metalloprotease; a membrane-anchored, universal protease with various functions. FtsH has been extensively studied in *E. coli*, where it is the only protease essential for survival [reviewed in 223]. FtsH has also been described as required for regulation of optimal ratios of phospholipids and LPS in the outer membrane [223]. Whether BbsR9 regulation of *ftsH* transcripts is involved in bacterial protein turnover and/or modulation of membrane architecture in the context of the sand fly is unknown, but would be interesting to investigate. Second, the *nuoF* gene codes for the NADH-quinone oxidoreductase subunit F, a component of the type I NADH dehydrogenase enzyme and the initial step in the electron transport chain. NuoF is a component of the peripheral fragment of the NADH dehydrogenase complex and plays a role in oxidation

of NADH to generate a proton motive force [reviewed in 224]. Regulation of *nuoF* transcripts could conceivably play a role in helping to establish the stationary phase as *B. bacilliformis* persists in the sand fly. Finally, we determined that BbsR9 targets transcripts of *gcvT*, which codes for the glycine cleavage system aminomethyltransferase, GcvT. The glycine cleavage system responds to high concentrations of glycine, breaking the amino acid down to CO_2, ammonia, and NADH [225]. In addition to redox reactions, NADH can be used to produce energy through cellular respiration. Of note is the potential interplay between sRNA targeting of *nuoF* and *gcvT* transcripts in this regard. Interestingly, the glycine cleavage system has been implicated in contributing to bacterial persistence in animal and plant hosts [226]. In fact, *gcvT* is essential for persistence of a closely-related pathogen, *Brucella abortus*, in its animal host [227]. However, to our knowledge, the role of a glycine cleavage system in pathogen persistence in its arthropod vector has not been explored, to date. It is conceivable that *B. bacilliformis* utilizes regulation of *nuoF* and *gcvT* to fine-tune levels of NAD+/NADH, thereby contributing to regulation of metabolism and persistence of the bacterium in the sand fly.

This study has provided further insight into the regulation of numerous processes by *B. bacilliformis* in response to conditions encountered in the context of its sand fly vector and human host. We believe the results provide a strong foundation for future studies examining sRNA-mediated regulation in *B. bacilliformis* and the regulatory mechanisms required for vector-host transmission.

Acknowledgments

The authors wish to thank Patty Langasek and Auguste Dutcher for technical assistance. This work was supported by NIH grant R21AI128575 (to MFM). RR was supported by NIH grants AI133023 and DE028409.

Supplementary Material

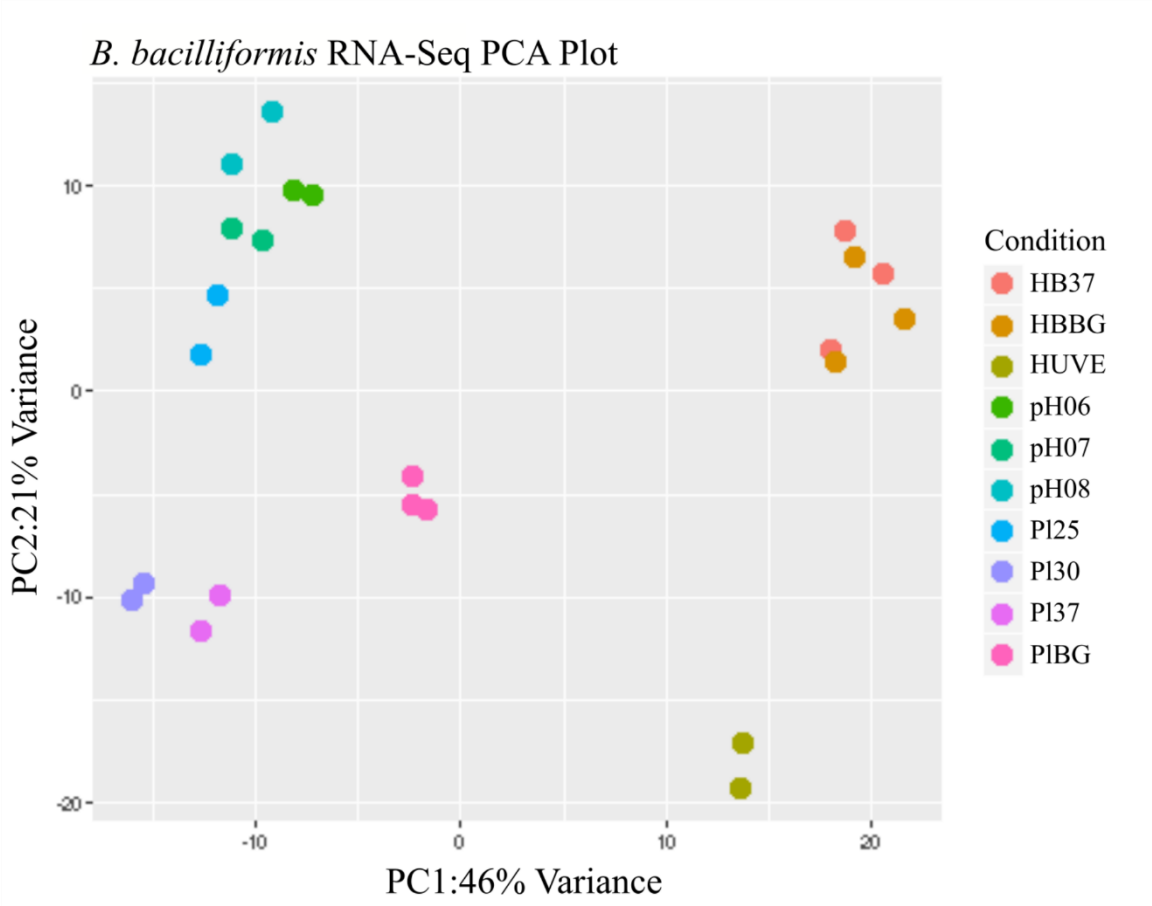


Figure S4.1: *B. bacilliformis* RNA-Seq PCA plot. Axes indicate the percentage of total variance that can be accounted for by two principle components. Colored dots indicate the retained biological replicates of the RNA-Seq analyses, and their distance apart is representative of overall relatedness in gene expression profiles. Experimental conditions are shown on the right.

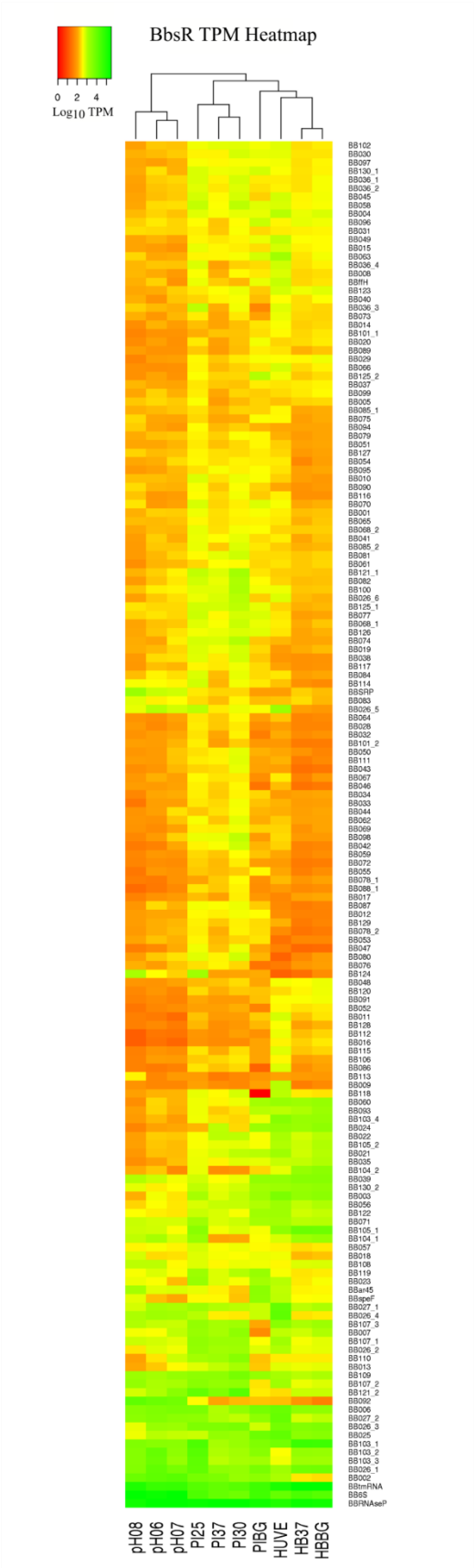

BbsR TPM Heatmap
0 2 4
Log10 TPM
pH08
pH06
pH07
PI25
PI37
PI30
PIBG
HUVE
HB37
HBBG

Figure S4.2: *B. bacilliformis* sRNAs group into specific expression patterns. Heatmap of *B. bacilliformis* sRNA TPMs across the tested conditions (shown at the bottom). sRNAs group vertically based on similarity in expression patterns. Conditions group horizontally based on similarity in overall expression patterns. The $\log_{10}$ of the TPM value for each sRNA is indicated by a color gradient.

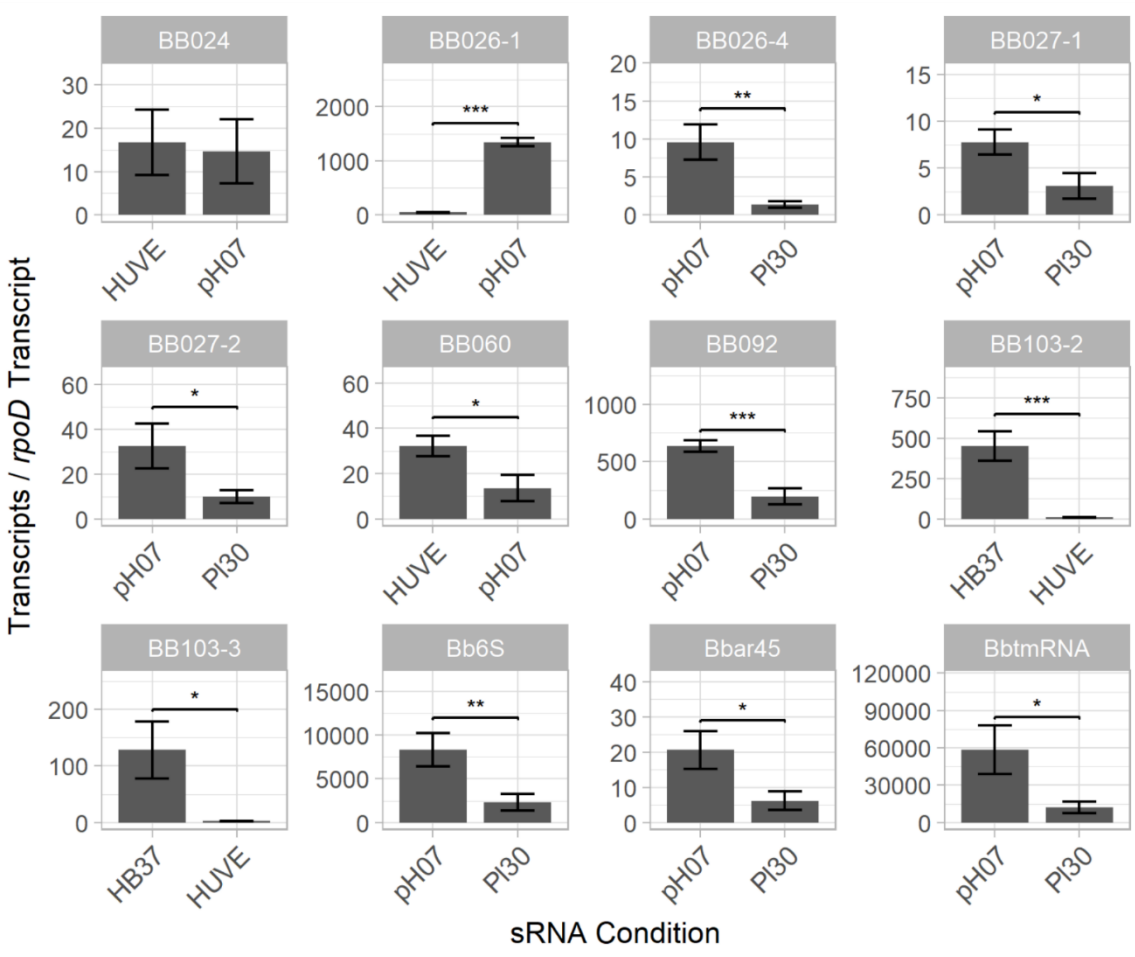


Figure S4.3: qRT-PCR confirmation of differential expression of several identified sRNAs. Faceted bar graph displaying the number of sRNA transcripts / *rpoD* transcript for select, differentially-expressed sRNAs and BB024, which was not shown to be differentially expressed. The condition / source of the total RNA is noted on the x-axis.

Significance was determined by students t-test (N = 9; * = p<0.05, ** = p<0.01, *** = p<0.001).

Figure S4.4: BbgpI is a group I intron inserted into an unannotated $tRNA_{CCU}^{Arg}$ gene of *B. bacilliformis*. A) Nucleotide sequence of BbgpI (bolded and underlined) and flanking chromosomal regions. Primer binding sites used for *in vitro* transcription (IVT) and PCR assays designed to show splicing of BbgpI *in vitro* and *in vivo* are indicated. **B)** Sequence of BbgpI outlining the conserved, characteristic stem structures (P1 to P9) with putative base pairings highlighted in green and yellow. Nucleotides predicted to

participate in base pairing are bolded and underlined. **C)** Sequence coding for the $tRNA_{CCU}^{Arg}$ immediately flanking BbgpI. The two bolded underlined nucleotides represent the ends of the spliced out BbgpI. Conserved tRNA features are also outlined.

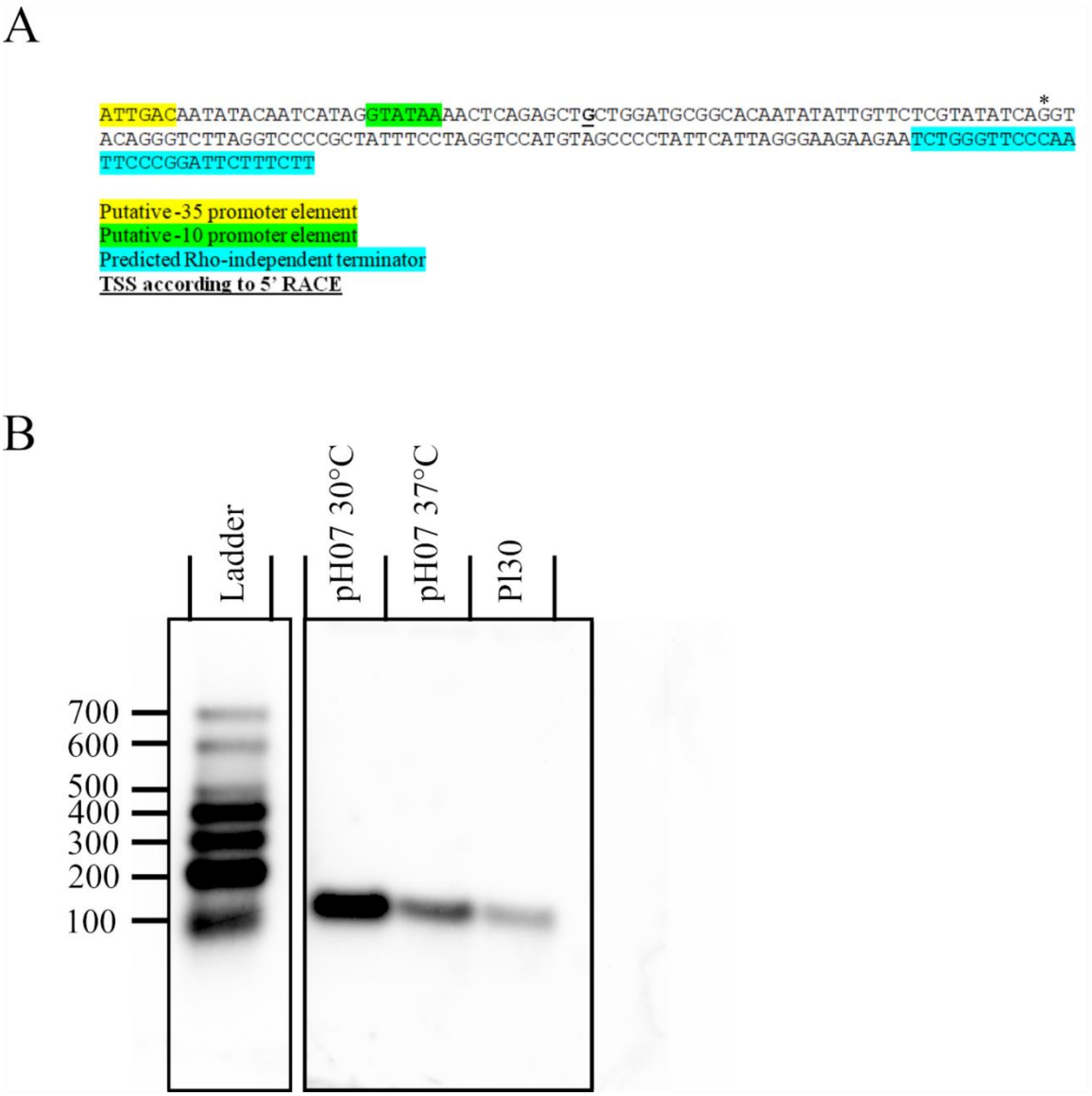


Figure S4.5: BbsR9 is a sand fly-specific sRNA. A) Nucleotide sequence of the *bbsR9* gene with predicted promoter elements and Rho-independent terminator plus experimentally-determined TSS's, highlighted in various colors or underlined,

respectively. An asterisk indicates the alternative TSS found by 5' RACE analysis. **B)** Northern blot analysis of BbsR9 expression under the indicated conditions. The RNA ladder (2 min exposure) and resolved total RNA samples (30s exposure) were from the same blot but imaged using different exposure times.

Table S4.1: Bacterial strains, primers, and plasmids used in the study.

Purpose	Name	Sequence	Tm	Length	Reference
Northern blots	Bb_gpI_Probe_F	TAATACGACTCACTATAGGGGGACTGTCTCTTC	60.6	33	This Study
	Bb_gpI_Probe_R	GGGGAAAGCTTCACAGGAAACTATATGCC	60.8	29	This Study
	Bb_sRNA9_Probe_F	TAATACGACTCACTATAGGGGAATTGGGAACC	60.2	32	This Study
	Bb_sRNA9_Probe_R	CAATATATTGTTCTCGTATATCAGGTACAGGGTC	57.7	34	This Study
	Bb_sRNA2_Probe_F	TAATACGACTCACTATAGGGCCCCAACAAACTCTTG	62.7	36	This Study
	Bb_sRNA2_Probe_R	GGCTGTCTTCGTATCAAAGATAATGATTTATGGG	58.8	34	This Study
	Bb_sRNA7_Probe_F	TAATACGACTCACTATAGGGCCCGATGAAAAAATAC	60.3	36	This Study
	Bb_sRNA7_Probe_R	GGTATACGGGATAAAAATTCTACTAAAATGTTTATTTTGGAGG	59.3	43	This Study
	Bb_ar45_Probe_F	TAATACGACTCACTATAGGGCTATAGCTGAAAGAC	59.4	35	This Study
	Bb_ar45_Probe_R	GTCGCACGTGCCTGTGG	59.9	17	This Study
	Bb_SpeF_Probe_F	TAATACGACTCACTATAGGGGGAATAACGGTATTG	59.6	35	This Study
	Bb_SpeF_Probe_R	GAGATAACTAAAGAGGCACACCTCATTG	57.1	28	This Study
	Bb_sRNA3-1_Probe_F	TAATACGACTCACTATAGGGCTAAAATACAATATGGG	58.6	37	This Study
	Bb_sRNA3-1_Probe_R	CTGATTAGGATTTTTCAAAAATTAAAGAATATAGCGGCC	58.5	39	This Study
	Bb_sRNA11-2_Probe_F	TAATACGACTCACTATAGGGGCGGCACAATATATTG	61.4	36	This Study
	Bb_sRNA11-2_Probe_R	GAAAGAATCCGGGAATTGGGAACCCAGATTC	62.2	31	This Study
5' RACE	Bbgpl_GSP1	GCCAGCCTATTGGTTTAAG	51.1	19	This Study
	Bbgpl_GSP2	CTGGACTGTCTCTTCACCTTGAAGCTTTATAC	59.8	32	This Study
	BbsR9_GSP1	GGGAACCCAGATTCTTC	49.5	17	This Study
	BbsR9_GSP2	GAATAGGGGCTACATGGACCTAGGAAATAG	59.3	30	This Study
qRT_PCR	BB_rpoD_qRT_F	CGATCCGGTGCGTATGTATTTG	63	22	This Study
	BB_rpoD_qRT_R	GTTTCACGTCCTGCCTCAATG	63	21	This Study
	BB_6S_qRT_F	CTGCTCTTCTCGACAGGAAATAC	63	23	This Study
	BB_6S_qRT_R	ATTCCGGGAACCTACACAAG	62	20	This Study
	BB_tmRNA_qRT_F	GGCGAAATAGGATCGACAAGAG	62	22	This Study
	BB_tmRNA_qRT_R	GACGTGCTTCCGCATAGTTG	63	20	This Study
	BB_ar45_qRT_F	TCCGCTGGTTCCCAAATG	62	18	This Study
	BB_ar45_qRT_R	ACCGTCGTTGCTTCAAGATG	63	20	This Study
	BB_BB026-4_qRT_F	GGTTAGATTCGTAATATCAATGAGGAG	61	27	This Study
	BB_BB026-4_qRT_R	GATGCTGGTGTGCCATTTAC	61	20	This Study
	BB_BB027-1_qRT_F	AATAGGAGATTTGATATGGATCCTG	60	25	This Study
	BB_BB027-1_qRT_R	TAGCAGCAGTAGTAGGCTTTAG	61	22	This Study
	BB_BB027-2_qRT_F	ATAACGGCTGCAGTAGTAGGG	63	21	This Study
	BB_BB027-2_qRT_R	ACTGCCACCATTCCTATGAAATTAG	63	25	This Study
	BB_BB092_qRT_F	CGTATATCAGGTACAGGGTCTTAG	62	24	This Study
	BB_BB092_qRT_R	CCAGATTCTTCTTCCCTAATGAATAG	62	26	This Study
	BB_BB026-1_qRT_F	GGCTGTCTTCGTATCAAAGATAATG	62	25	This Study
	BB_BB026-1_qRT_R	CCCAACAAACTCTTGAACCAAAG	63	23	This Study
	BB_BB103-2_qRT_F	CATCTCAATTCTTCAACGAGGAGAG	63	25	This Study
	BB_BB103-2_qRT_R	ACCAATTTGGGCAATATACTGGAAG	63	25	This Study
	BB_BB103-3_qRT_F	CTGATACGAAACGATATGAGTAGAAG	61	26	This Study
	BB_BB103-3_qRT_R	ACGCGATTATTCATGGATCAGTG	63	23	This Study
	BB_BB024_qRT_F	AGGTGTAGCGTCTTCTCCTAC	63	24	This Study
	BB_BB024_qRT_R	CCAAACCACATATCCTGCAATC	62	22	This Study
	BB_BB060_qRT_F	AAACTGTTGGGTAGTAGCAAGG	62	22	This Study
	BB_BB060_qRT_R	CCCTACCTCCAAGTCAAAGATG	62	22	This Study
Group I intron Splicing	Bb_Intron_Splice_NEW_F	GTGTCTGCGTAGCTCAGTAGGATAG	58.1	25	This Study
	Bb_Intron_Splice_NEW_R	GGTGTCTACGGCAGGATTTGAAC	58	23	This Study
	Bb_Intron_F+T7	TAATACGACTCACTATAGGGCATTTATCGCAAAAGATG	60.5	38	This Study
	Bb_Intron_R_Long	GGATTTGGGTCTTCTGGGACTTCAATG	59.4	27	This Study
	Bb_Intron_Nested_F	CACATTGATTCATACTTATCTAAATTGATTCACC	55	34	This Study
	Bb_Intron_Nested_R	CCTAAAATGAGTAAAACTTTTTTCATGATATTTTCC	55	36	This Study
Q5 mutagenesis / EMSAs	BbsR9_F+T7	TAATACGACTCACTATAGGGCTGGATGCGGCAC	64.6	33	This Study
	BbsR9_R	CAAGAAAGAATCCGGGAATTGGGAACCCAG	62.5	30	This Study
	Bb_FtsH_F+T7	TAATACGACTCACTATAGGAAAGATTAAGAAACCGC	58.9	36	This Study
	Bb_FtsH_R	CACCATTGCTGGCGCGTTG	60.5	19	This Study
	Bb_NuoF_F+T7	TAATACGACTCACTATAGGGAGTTCAGAAAAATGCTAG	59.7	38	This Study
	Bb_NuoF_R	CAATAATCCAATCACGGCCTTTTTCG	57	26	This Study
	Bb_GcvT_F+T7	TAATACGACTCACTATAGGAATGATATAATTGGGGG	58.3	36	This Study
	Bb_GcvT_R	CCTGCAAAAGCACCAAATTTTGCC	58.5	24	This Study
	Bb_0612RR_F+T7	TAATACGACTCACTATAGGTGTTGTTGGCATTGGGAC	62.8	37	This Study
	Bb_0612RR_R	CCCCATCACAAGCAATATCAGCTGC	60.3	25	This Study
	Bb_TrmD_F+T7	TAATACGACTCACTATAGGAAAGAGGACATGCTTTG	60.1	36	This Study
	Bb_TrmD_R	GACCATATTCCCCGCTCTAAAGCG	59.5	24	This Study
	Bb_HflK_F+T7	TAATACGACTCACTATAGGAGACTCTATAAGTAGTTG	57.3	37	This Study
	Bb_HflK_R	CCGGAACCAAAAGGATTCTTAGGTG	57.8	25	This Study
	BbsR9_Anti_F+T7	TAATACGACTCACTATAGGCAAGAAAGAATCCGGG	61.4	35	This Study
	BbsR9_Anti_R	GGTACAGGGTCTTAGGTCCCC	58.7	21	This Study
	Q5_FtsH_Mut1_F	ACATATCTGATAAATGGGCCAAATATG	61	27	This Study
	Q5_FtsH_Mut1_R	TTGTTCGAAGAAGTCTTTATAGC	56	23	This Study
	Q5_FtsH_Mut2_F	ACATTCTTCGAGAACCTGTATC	58	22	This Study
	Q5_FtsH_Mut2_R	TTGTTAGCATATAATCTTGTCTCAG	57	25	This Study
	Q5_NuoF_Mut_F	AAATATGCTAGCTGATAAAGATC	57	23	This Study
	Q5_NuoF_Mut_R	TTTGATTTCCTTTATCACCGAC	56	22	This Study
	Q5_GcvT_Mut_F	AAACAGAAGTTGTCGCACTATG	60	22	This Study
	Q5_GcvT_Mut_R	TTTGATCATTAAGTTTGAAAAACTTTAAATAAC	57	33	This Study
	Q5_FtsH_DBL_R	TTGTTCGAAGAATGTTTGTTAGC	57	23	This Study

Table S4.2: Quality control results for *B. bacilliformis* RNA-Seq analyses.

Treatment	Raw reads	Reads after Filter, Trim, and Pairing	% Reads Discarded	Reads with Bowtie Alignments	Bowtie Ambiguous Reads	Bowtie Unambiguous Reads	featureCounts Assigned Reads	% Ambiguous Reads	% Reads with alignments
PI25-1	27456299	19381531	29.40952821	19259963	15456384	3803579	5005237	80.25136912	99.37276369
PI25-2	26751499	20410930	23.7017335	20242734	6106703	14136031	21216644	30.16738253	99.17595132
PI25-3	29674598	22015946	25.80878096	21868359	12367018	9501341	N/A	56.552108	99.32963589
PI30-1	32477931	24510162	24.53287126	24286427	7555333	16731094	N/A	31.10928174	99.08717454
PI30-2	24793860	17701522	28.6052192	17613689	14915011	2698678	3318580	84.67851908	99.50381103
PI30-3	30826768	23509074	23.73811617	23272863	4084481	19188382	22756550	17.55040194	98.99523478
PI37-1	28604496	21510901	24.79888127	21334528	7177523	14157005	N/A	33.64275507	99.18007619
PI37-2	28000000	21457560	23.36585714	21137084	6885979	14251105	18471790	32.57771507	98.50646579
PI37-3	31215787	23553447	24.54636175	23218990	12797482	10421508	11913139	55.11644563	98.58000827
pH06-1	25786111	20047573	22.25437562	19059991	1100188	17959803	19954630	5.772237773	95.07380769
pH06-2	24396811	18378162	24.66981853	17540083	700167	16839916	19016429	3.991811213	95.43981058
pH06-3	23771077	17572118	26.07773724	3693991	1639030	2054961	N/A	44.37016766	21.0218882
pH07-1	24411875	18695408	23.41674697	17406219	1200794	16205425	13930221	6.89864927	93.10424785
pH07-2	26051988	20215780	22.40215987	19223158	600622	18622536	14354650	3.124471016	95.08986544
pH07-3	25747656	19883785	22.7743877	18096033	450605	17645428	N/A	2.49007614	91.00899552
pH08-1	24800450	17562342	29.18538978	2177927	1321785	856142	N/A	60.69005068	12.4011194
pH08-2	27032587	20701159	23.42146536	19330064	772789	18557275	16932480	3.997860535	93.3767235
pH08-3	20319866	15693461	22.76789128	13822106	794086	13028020	13152053	5.745043483	88.07557492
HUVE-1	47751481	20940177	56.14758629	232722	181151	51571	73518	77.84008388	1.111365964
HUVE-2	37532653	17616747	53.06287834	268246	201862	66384	98155	75.25256667	1.522676122
HUVE-3	40748520	18015913	55.7875648	381259	310618	70641	101436	81.47165051	2.116234687
PIBG-1	16404919	7117417	56.61412897	5654883	217769	5437114	8478003	3.850990374	79.45133747
PIBG-2	17299526	7265107	58.00401121	5635336	198919	5436417	8461078	3.529851636	77.56714388
PIBG-3	18410720	8451627	54.09398981	6497556	237713	6259843	9391536	3.658498672	76.87935116
HB37-1	41282856	20609226	50.07800332	11071031	10268079	802952	801918	92.74726988	53.71881021
HB37-2	41671146	21428176	48.57790568	11149925	10401491	748434	734112	93.28754229	52.03394353
HB37-3	50959552	25411079	50.13480692	13882484	12993736	888748	743190	93.59806213	54.63161954
HBBG-1	46693159	22819698	51.12839121	10077985	8780305	1297680	1407121	87.12361648	44.16353363
HBBG-2	16812873	8000076	52.4169605	3336897	3025888	311009	356387	90.67969434	41.71081625
HBBG-3	38931496	18430339	52.65956643	10488831	9477791	1011040	983573	90.36079426	56.91067864
Replicates removed from global analysis									

Table S4.3: Putative sRNAs identified in *B. bacilliformis* by RNA-Seq analyses.

Name	Start	End	Length	Strand	RFAM Designation	Conserved in α?	Given Name	Location	Putative Promoter Sequence**	Predicted Rho-Independent Terminator****	Anti-SD Sequence?	Peak TPM	Peak TPM Condition	DESeq2 Differential Expression
BB001	1565	1990	426	R				Cis-anti	CTTGAAG-N17-CAAAAT	N/A	No	496.2816733	PI25	
BB002	4992	5346	355	F				Cis-anti	ATTGCC-N10-CATTTT	N/A	No	17276.28927	pH07	↑pH07 vs PI30
BB003	23471	23637	167	F		No		Leader*	CTTATC-N17-CTCTAA	N/A	No	5292.6659	HUVE	↓pH08 vs pH06, ↑HB37 vs pH07, ↑HUVE vs pH07, ↑HBBG vs PIBG
BB004	32276	32563	288	R				Cis-anti	Cannot Decipher	N/A	No	1058.11888	HUVE	↑HUVE vs PIBG
BB005	35430	35620	191	F		No		IGR	GAAGAG-N13-GTAAAA	N/A	No	416.2446332	PI25	
BB006	42480	42726	247	R		Yes	**Bbgpl**	Group 1 intron	GATTCATAC-N18-CTTTAT***	N/A	No	9939.711374	HUVE	
BB007	50218	50430	213	R		No		IGR	GTGCTA-N19-CAAAAT	N/A	No	2922.443742	PI30	↑HUVE vs PIBG
BB008	68255	68455	201	R				Cis-anti	AAAGAA-N17-CAAAAA	N/A	No	647.3358684	HUVE	
RF01793	76870	76922	53	F	ffH leader	Yes		Leader*	CTGGGAC-N11-GTATAA	CACCCGACGACGCGTTGCGTTCTCGGGTGTTATTTTGGTTT	No	850.8311306	HUVE	
BB009	86533	86877	345	F				Cis-anti	CTGTTAG-N19-GAAAAA	N/A	No	1065.599064	HUVE	↑HUVE vs HBBG, ↑HUVE vs PIBG
BB010	97732	98160	429	R				Cis-anti	CTGTC-N14-CAAAAT	N/A	No	650.427422	PI25	
BB011	101286	101502	217	F		No		IGR	GATCGTC-N16-GTATAA	N/A	No	707.1929022	HUVE	
BB012	105538	105915	378	R				Cis-anti	CTCTAG-N18-CTTTTT	N/A	No	461.4989028	PI30	↓HBBG vs PIBG
BB013	112234	112614	381	F				Cis-anti	GTTGAT-N13-GTTAAA	N/A	No	1333.821864	PI30	
BB014	133192	133383	192	F		No		Leader*	GTTCTT-N20-CAGATAT	N/A	No	425.4313761	HUVE	
BB015	137410	137664	255	F		No		IGR	TTTGAA-N18-CAATTT	N/A	No	692.8676713	HUVE	
BB016	152703	152910	208	F		No		Leader*	Cannot Decipher	N/A	No	345.6360544	HUVE	
BB017	165406	165713	308	F				Cis-anti	GTTTGCC-N17-GTTTAA	N/A	No	320.9188772	PI30	
BB018	174210	174396	187	F		No		IGR	CACGTT-N17-CTATAA	N/A	No	713.5680009	pH08	↑pH08 vs pH06
BB019	177094	177355	262	R		No		IGR	GTGCTT-N16-GTATAA	N/A	Yes; 1 (**CCTCCT**)	711.1817861	PI30	
BB020	181780	182124	345	F		No		IGR	TATGAG-N14-CAAAAAT	N/A	No	464.9245512	PIBG	
BB021	201988	202232	245	F		No		IGR	CAAGAA-N17-GATATA	N/A	No	1795.004731	HUVE	↓pH07 vs PI30
BB022	202233	202467	235	F		No		Leader*	Cannot Decipher	N/A	No	1287.866693	HUVE	
RF00013	206709	206871	163	R	6S RNA	Yes	**Bb6S**	6S RNA	CTAGTC-N17-CTATAAT	N/A	No	402121.5657	pH08	↓HUVE vs pH07, ↑pH07 vs PI30
BB023	208306	208580	275	R		No		Leader*	CTTCTT-N15-GATTTT	N/A	No	2172.245172	PI25	
BB024	214513	214777	265	F		No	**BbsR1**	Leader*	CTTGCAC-N19-TATTT	N/A	No	2906.471248	HBBG	↑PI37 vs PI30
BB025	218599	218718	120	F		No		Leader*	GGCAC-N15-CATTAT	N/A	No	13148.5228	HUVE	
BB026_1	220801	221084	284	F		No	**BbsR2**	IGR	CATGTA-N18-GATATT	N/A	No	12820.65011	PI25	↑pH07 vs PI30, ↓HUVE vs HBBG, ↓HUVE vs pH07
BB026_2	221246	221514	269	R		No	**BbsR3-1**	IGR	GATTTA-N17-CAAATA	N/A	No	2419.190792	PI25	
BB026_3	221512	221820	309	R		No	**BbsR3-2**	IGR	GATTTG-N18-GATATA	N/A	No	6879.351935	HUVE	↑HUVE vs PIBG
BB026_4	221947	222234	288	F		No	**BbsR4**	IGR	CTATGTG-N10-CATTAT	N/A	No	13303.01973	HUVE	↑HUVE vs HBBG, ↑pH07 vs PI30, ↑HUVE vs PIBG
BB026_5	222235	222425	191	R		No		IGR	CAACAA-N18-CAAATA	N/A	No	1991.765384	HUVE	↑pH07 vs PI30, ↑HUVE vs HBBG, ↑HUVE vs PIBG
BB026_6	222725	223048	324	R		No		IGR	GTTCTT-N22-CTTACT	N/A	No	1058.003412	PI30	↑HUVE vs PIBG
BB027_1	223964	224233	270	R		No	**BbsR5**	IGR	Cannot Decipher	N/A	No	11853.57524	HUVE	↑HUVE vs HBBG, ↑pH07 vs PI30, ↑HUVE vs PIBG
BB027_2	224237	224539	303	R		No	**BbsR6**	IGR	CAAACTG-N19-CATAAT	N/A	No	11219.10742	HUVE	↑HUVE vs HBBG, ↑pH07 vs PI30, ↑HUVE vs PIBG
BB028	262399	262562	164	R				Cis-anti	CTTGAA-N16-CTAAAA	N/A	No	307.2329244	PI25	
BB029	263407	263730	324	R		No		IGR	CATCAT-N17-CATTAA	N/A	No	487.6770919	HBBG	
BB030	279639	279909	271	F		No		IGR	GCTGAC-N16-GTTATA	N/A	No	698.5395197	PI30	
BB031	280558	280767	210	F		No		Leader*	Cannot Decipher	N/A	No	417.5108195	PI30	
BB032	284831	285238	408	R				Cis-anti	CAATTT-N16-CTATACT	N/A	No	340.8093002	PI30	
BB033	286460	286750	291	R				Cis-anti	GATTCAG-N15-CATTAT	N/A	No	454.4660937	PI30	
BB034	296552	296829	278	R				Cis-anti	CATGTA-N19-CTAATT	N/A	No	342.3876178	PI30	
BB035	300016	300280	265	F		No		IGR	GTTATG-N17-GATATAAT	N/A	No	1171.63052	HBBG	
BB036_1	303123	303376	254	R		No; Only KC583		IGR	CTGCAC-N12-CTTTTT	N/A	No	710.178005	PI30	
BB036_2	303097	303405	309	F		No; Only KC583		IGR	CTTTCC-N18-GATAAT	N/A	No	616.6129156	PI30	
BB036_3	303381	303550	170	R		No; Only KC583		IGR	CTTGGG-N16-CAAAAGT	N/A	No	960.8786332	PI25	
BB036_4	303790	304035	246	F		No; Only KC583		IGR	GAAGTA-N12-CTTTTT	N/A	No	771.0831325	HUVE	↑HUVE vs PIBG
BB037	320278	320469	192	F				Cis-anti	GTTTGAAG-N14-CTTTAAT	N/A	No	351.982407	HBBG	
BB038	345872	346134	263	R		No		IGR	CAATGTA-N16-CAATAA	N/A	No	735.9567972	PI30	
BB039	348864	349306	443	F		No	**BbsR7**	Leader*	CTTGCA-N21-CATATAT	N/A	No	3708.411037	HBBG	↓HUVE vs HBBG
BB040	364947	365233	287	R				Cis-anti	CTTTAT-N17-GTAAAA	N/A	No	433.1188295	PI30	
BB041	371607	372062	456	R				Cis-anti	CATGTG-N17-CTTTTT	N/A	No	538.6591709	PI30	
BB042	385701	386036	336	R				Cis-anti	CTAAC-N14-CAATAT	N/A	No	774.3039086	PI30	
BB043	390191	390526	336	R				Cis-anti	CCTAAC-N16-GTAAAA	N/A	No	526.7062237	PI30	
BB044	417847	418086	240	F				Cis-anti	CAAATA-N17-CAATAT	N/A	No	434.8789108	PI30	
BB045	422851	423150	300	R		No		IGR	CAAAAG-N15-CTATAT	N/A	No	735.9480016	HUVE	
BB046	428891	429229	339	R		No		IGR	Cannot Decipher	N/A	No	365.7986215	PI30	
BB047	435331	435585	255	R				Cis-anti	GAAAGCC-N14-CAAAAA	N/A	No	485.7051311	PI30	
BB048	452671	452952	282	F		No		IGR	CATGAA-N18-CTCTTA	N/A	No	503.7204688	HBBG	
BB049	456023	456304	282	R				Cis-anti	GTGAGAC-N17-GTAAAT	N/A	No	943.3427482	HUVE	
BB050	489865	490191	327	R				Cis-anti	CTGCAC-N16-GATAAT	N/A	No	560.6879339	PI30	
BB051	490903	491111	209	F				Cis-anti	GTAAAC-N14-GAAAAA	AGCTTACCTGTTTCTTTTAAATCGGGTAAGGAGAGATGAGCTATTTTCTT	No	360.3237013	PIBG	
BB052	498847	499083	237	F				Cis-anti	CAAAAAC-N14-CAATAT	N/A	No	372.0046788	HUVE	↑HBBG vs PIBG
BB053	500909	501262	354	R				Cis-anti	ATTAAG-N16-CAAATT	N/A	No	355.9714491	PI25	
BB054	511310	511467	158	F				Cis-anti	GGTAA-N14-GTAATT	N/A	No	368.8290748	PI25	
BB055	531995	532227	233	F		No		IGR	CATCAA-N14-GATAT	TCAGCCTAATCACTAAGGTTGAATTTCATTT	No	379.6425804	PIBG	
BB056	560531	560858	328	R				Cis-anti	GTTGAG-N15-GTTTAA	N/A	No	1801.852905	HUVE	
BB057	563437	563763	327	F		No		IGR	GATTATCG-N18-GTTATT	N/A	No	716.3879653	PI25	
BB058	568738	568983	246	R				Cis-anti	CTATAC-N17-GTAAAA	N/A	No	979.7847614	PI30	
BB059	570985	571338	354	R				Cis-anti	GTTTCAA-N17-CTTTAC	N/A	No	399.325032	PI30	
BB060	587515	587806	292	F		Yes; Only *Bartonella* spp.	**BbsR8**	Leader*	TTTCAG-N19-CATTTT	N/A	No	2353.470966	HBBG	↓pH08 vs pH06, ↑HB37 vs pH07, ↑HUVE vs pH07, ↑HBBG vs PIBG
BB061	594424	594678	255	F				Cis-anti	CTAGCC-N14-CATAAT	N/A	No	400.9741484	PI37	
BB062	601997	602269	273	R				Cis-anti	CTTTCC-N11-CAATAA	N/A	No	574.1521929	PI30	↑HUVE vs PIBG
BB063	614479	614769	291	F				Cis-anti	CTTGAT-N16-GTAATT	N/A	No	1664.856056	HUVE	
BB064	617574	617939	366	R				Cis-anti	GAATG-N20-GTAAAT	N/A	No	375.459916	PI30	
BB065	636778	637032	255	F				Cis-anti	GAAGTG-N18-CAATAT	N/A	No	366.4170991	PI25	
BB066	644476	644755	280	F		No		IGR	CTTAAT-N18-GATTAA	N/A	No	714.1882733	HUVE	
BB067	655501	655800	300	R				Cis-anti	CTTTAC-N15-CATTTA	N/A	No	381.0416561	PI25	
BB068_1	721160	721467	308	R				Cis-anti	CAACAC-N18-CTAATAT	N/A	No	824.707146	PI30	
BB068_2	721465	722043	579	R				Cis-anti	GCTGTA-N19-CTATTA	N/A	No	532.5312284	PI25	
BB069	739764	740082	319	F				Cis-anti	CTTCAG-N14-GAAATA	N/A	No	477.6685311	PI30	
BB070	786325	786705	381	F				Cis-anti	GAATTG-N21-GATTAA	N/A	No	678.7599738	PI25	↑pH08 vs pH06
BB071	789268	789466	199	F		Yes; Only *Bartonella* spp.		Leader*	TTTGTG-N18-GTTTAT	N/A	No	1627.280536	PI25	
BB072	790468	790830	363	R				Cis-anti	Cannot Decipher	N/A	No	316.0681082	PI30	
BB073	791491	791700	210	R				Cis-anti	CTACAC-N18-GTTTAA	N/A	No	519.9037998	HUVE	
BB074	804697	804969	273	R				Cis-anti	CAATTC-N18-CATAAA	N/A	No	1098.905034	PI30	
BB075	819616	819870	255	R		Yes; Only *Bartonella* spp.		IGR	AAAGTC-N16-CTTAAT	TGGTTCGGATATTCGGACCATTTT	No	420.7543702	PIBG	
BB076	834748	835083	336	R				Cis-anti	GCAGAC-N16-CTATAA	N/A	No	350.5645079	PI37	
BB077	836431	836757	327	F				Cis-anti	CATTG-N16-GTATAT	N/A	No	800.0383303	PI25	
BB078_1	844892	845173	282	F				Cis-anti	Cannot Decipher	N/A	No	384.5918016	PI30	
BB078_2	845180	845497	318	F				Cis-anti	CCTGTA-N19-GATATT	N/A	No	339.6999519	PI30	
BB079	855197	855424	228	R				Cis-anti	ATGGA-N16-CTATAT	N/A	No	413.5966927	PI30	
BB080	867684	867937	254	R		No		IGR	CTTCAA-N14-CTATA	N/A	No	769.9230033	PI30	
BB081	893155	893571	417	R				Cis-anti	CTTGAT-N19-CTATAT	N/A	No	781.0749478	PI30	
BB082	897559	897876	318	F				Cis-anti	CAATCAG-N13-GAAATT	N/A	No	1183.09352	PI30	
BB083	912647	912921	275	F		No		IGR	Cannot Decipher	CCTCTTCGAAGTGGAGGTTATTTTTTT	No	611.6681933	pH08	
BB084	916424	916831	408	R				Cis-anti	AAGGAG-N12-CATATA	N/A	No	650.8805063	PI25	
BB085_1	921473	921817	345	R				Cis-anti	GTTGAT-N17-GATATA	N/A	No	361.150714	HUVE	
BB085_2	921977	922303	327	R				Cis-anti	CAAGTT-N15-CAAAAT	N/A	No	633.245647	PI30	↓PI37 vs PI30
BB086	933217	933624	408	F		No		IGR	TATGAT-N18-CTATT	N/A	No	637.0525186	HUVE	↑HUVE vs HBBG, ↓PIBG vs PI37, ↑HUVE vs PIBG
BB087	941224	941847	624	R				Cis-anti	GCAGAA-N17-CTTTTA	N/A	No	499.0692369	PI30	
RF01849	955017	955372	356	F	tmRNA	Yes	**BbtmRNA**	tmRNA	Cannot Decipher	N/A	No	314430.1099	HB37	↓HUVE vs HBBG, ↑pH07 vs PI30
RF00010	988347	988732	386	R	Rnase P RNA	Yes		Rnase P	Cannot Decipher	N/A	No	457521.8377	PI37	↓HUVE vs HBBG
BB088	990118	990253	136	F		No		IGR	TTTTAC-N14-CTATAA	N/A	No	335.1273341	PI30	
BB089	992341	992595	255	R		No		IGR	GAGAG-N19-GATATT	N/A	No	316.3931867	PIBG	
BB090	1034595	1034803	209	R				Cis-anti	CTAGCC-N11-CATTTT	N/A	No	442.0277833	PI25	
BB091	1045088	1045342	255	R				Cis-anti	GAAGAA-N19-CATAAA	N/A	No	463.1306877	HBBG	↑HBBG vs PIBG
BB092	1051920	1052052	133	R		Yes; Only *Bartonella* spp.	**BbsR9**	IGR	ATTGAC-N15-GTATAA***	TCTGGGTTCCCAATTCCCGGATTCTTTCTT	No	13936.21138	pH08	↑pH07 vs PI30, pH07 vs HUVE
BB093	1055361	1055559	199	R		No		Leader*	TTTTGTA-N13-CATTAAT	N/A	No	1911.894896	HBBG	
BB094	1057607	1057843	237	R				Cis-anti	CTAAAA-N19-CATATT	N/A	No	390.3329281	PI25	
BB095	1071748	1072200	453	R				Cis-anti	CTTCAA-N18-GTTTAT	N/A	No	434.2029275	PI37	
BB096	1079900	1080116	217	F				Cis-anti	CGACAT-N21-GTAAAA	N/A	No	847.52094	PIBG	
BB097	1083853	1084188	336	R				Cis-anti	GATGTAC-N15-GTTATT	N/A	No	472.2336635	HUVE	
BB098	1103684	1103956	273	R				Cis-anti	GTACTA-N14-CAATAT	N/A	No	948.8994593	PI30	
BB099	1116571	1116807	237	R				Cis-anti	GTAAATC-N16-GAATAA	N/A	No	376.046376	HBBG	
RF00518	1119924	1120074	151	R	speF leader	Yes; Only *Bartonella* spp.	**BbspeF**	Leader	Cannot Decipher	N/A	No	3120.285529	PIBG	↑pH08 vs pH06
RF02347	1120082	1120229	148	R	ar45	Yes	**Bbar45**	IGR	GATTGC-N17-CTTATA	N/A	No	3236.169339	PIBG	↑pH07 vs PI30
BB100	1120921	1121265	345	F				Cis-anti	CTTGAT-N18-GAAATT	N/A	No	1207.950247	PI30	
BB101_1	1124085	1124530	446	R				Cis-anti	CATTG-N18-CTTTTT	N/A	No	384.3275378	HUVE	
BB101_2	1124568	1124689	122	F		No		IGR	GTTAAA-N17-CATAAA	AGCTCACACAATATGTGAGCTTTTTT	No	324.1249053	PI25	
BB102	1129498	1129887	390	R				Cis-anti	Cannot Decipher	N/A	No	596.2911073	PI30	
BB103_1	1131141	1131342	202	F		No	**BbsR10**	IGR	ATTGAC-N20-GTTATT	TACCATGATTGGTATTTTTTT	No	43933.97802	HB37	↓HUVE vs HBBG
BB103_2	1131369	1131526	158	F		No	**BbsR11-1**	IGR	CATGGTG-N14-CAATTT	N/A	No	16854.32834	PI25	↓HUVE vs HBBG
BB103_3	1131528	1131747	220	F		No	**BbsR11-2**	IGR	CTTTTC-N15-CAAATT	CCCTGCCTATTTGGGGTTTTT	No	13789.28927	PI25	↓HUVE vs HBBG
BB103_4	1132079	1132331	253	R		No		IGR	GATAAAAC-N14-CAATAT	N/A	No	3763.101715	HUVE	↑HB37 vs pH07, ↑HUVE vs pH07, ↑HUVE vs PIBG
BB104_1	1133539	1133739	201	R		No		IGR	GTTCTA-N14-CTTTAA	N/A	No	2329.960124	HBBG	↑pH07 vs PI30
BB104_2	1133800	1134198	399	R		No		IGR	AATGGG-N11-CATTTT	N/A	No	3223.724178	HBBG	↓HUVE vs HBBG
BB105_1	1134984	1135170	187	R		No		IGR	CATAAA-N17-CAAATA	N/A	No	9354.669137	HB37	↓HUVE vs HBBG, ↑HBBG vs PIBG
BB105_2	1135171	1135386	216	R		No		IGR	CTTGAA-N17-GTATTT	N/A	No	1169.877316	HBBG	
BB106	1142717	1143115	399	R				Cis-anti	GATTA-N16-CTAAAA	N/A	No	363.8787504	HUVE	
BB107_1	1154245	1154411	167	R		No		IGR	TAATTG-N12-CTATAT	N/A	No	2724.520796	PI25	
BB107_2	1154407	1154540	134	R		No		IGR	CAATAT-N20-CTTTT	N/A	No	3357.102372	PI25	↑HUVE vs PIBG
BB107_3	1154533	1154694	162	R		No		IGR	CTTGCG-N12-CTTTTT	N/A	No	3482.169956	HUVE	↑HUVE vs PIBG
BB108	1159243	1159542	300	F				Cis-anti	Cannot Decipher	N/A	No	983.7782236	pH08	↓HUVE vs HBBG
BB109	1201930	1202214	285	F		No	**BbsR12**	Leader*	GATCAG-N19-CATTTT	N/A	No	6402.064419	PI30	
BB110	1204027	1204317	291	F				Cis-anti	CATGAC-N15-CAATAT	N/A	No	3573.184729	PI30	
BB111	1206802	1207035	234	R				Cis-anti	GAAAT-N15-CAAAAA	N/A	No	585.7390241	PI30	
BB112	1208356	1208691	336	R		No		IGR	CTTTAC-N19-CAAATT	N/A	No	516.0034285	HUVE	↑HBBG vs PIBG, ↑HUVE vs PIBG
BB113	1228413	1228502	90	R		No		IGR	CAACTA-N18-CAATTA	N/A	Yes; 1 (**CCTCCT**)	347.5906739	pH08	↑pH08 vs pH06
BB114	1231513	1231794	282	R				Cis-anti	CTTAAA-N18-GATATT	N/A	No	749.7778248	PI25	
BB115	1232243	1232549	307	R		No		IGR	GGACAG-N10-CTAATAT	AGTGCAGCTAAGAGTCTTCATGCTGCACTATTATATTTT	No	603.5837774	HUVE	
BB116	1233880	1234161	282	F				Cis-anti	GAAATT-N14-CATTAA	N/A	No	463.5127007	PI30	↑pH08 vs pH06
BB117	1249516	1249932	417	F				Cis-anti	CAATC-N16-GTTAAA	N/A	No	476.5439599	PI30	
BB118	1257316	1257687	372	F				Cis-anti	CTTCTT-N13-CTAATT	N/A	No	1008.298517	HUVE	↑HUVE vs HBBG, ↑HUVE vs pH07, ↑HUVE vs PIBG
BB119	1261442	1261755	314	R		No		IGR	GTTTTA-N17-CTTTTT	N/A	No	2207.422201	PIBG	
BB120	1263272	1263490	219	R				Cis-anti	CAGGAG-N17-CTAATA	N/A	No	486.1326461	HBBG	
BB121_1	1277701	1278072	372	F				Cis-anti	CTGGAT-N14-GAATAA	N/A	No	1511.296872	PI30	
BB121_2	1278196	1278377	182	F			**BbsR13**	Cis-anti	Cannot Decipher	N/A	No	18412.30814	PI30	
RF00169	1281192	1281290	99	R	SRP sRNA	Yes		SRP sRNA	Cannot Decipher	N/A	No	2006.872267	pH08	
BB122	1302253	1302426	174	R		No		IGR	CATTGAC-N20-CTATGT	N/A	No	1770.798921	HUVE	
BB123	1350043	1350441	399	R				Cis-anti	Cannot Decipher	N/A	No	719.0327519	HUVE	
BB124	1360532	1360645	114	R		No	**BbsR14**	IGR	CTAGTC-N14-GTATAA	AGGCTGCATTATAATTTGTTTATGCAGCCTTTTTTTCTT	Yes; 2 (TT**CCTCCT**AA)	1635.901098	PI25	↑pH08 vs pH06
BB125_1	1384358	1384612	255	R				Cis-anti	GATCAC-N12-GTTTAA	N/A	No	963.5677666	PI30	
BB125_2	1384921	1385123	203	R		No		IGR	GAAGTA-N14-CAATAA	N/A	Yes; 1 (**CCTCCT**)	1108.332252	PIBG	
BB126	1385968	1386222	255	R				Cis-anti	Cannot Decipher	N/A	No	642.7871349	PI30	
BB127	1410250	1410486	237	R				Cis-anti	GAGATA-N17-CTTTTT	N/A	No	351.9288207	PI25	
BB128	1422463	1422762	300	R				Cis-anti	GTGTCA-N20-CATTTT	N/A	No	420.8572183	HUVE	
BB129	1429268	1429522	255	R				Cis-anti	CAAGTC-N12-CTTTAT	N/A	No	329.4885998	PI25	
BB130_1	1438680	1438878	199	R		No		Leader*	GTTTTG-N15-GAAATCTAT	N/A	No	727.2795699	PIBG	
BB130_2	1438979	1439214	236	R		No		IGR	GTTTCC-N17-CTATAA	N/A	No	2770.117379	HBBG	

BB025 duplicated in adjacent genomic areas several times with no transcription
BB026-2/BB026-3 are duplicated at BB027-1/BB027-2 with slight changes.
BB049 is duplicated at three other locations in the genome with no transcription
BB071 is duplicated in many locations (13) in the genome with varying homology. Among these duplications is a portion of BB098
BB084 is duplicated in many locations (11) in the genome with varying homology. None of these are transcribed.
BB104-1 is duplicated within BB104-2
A portion of BB116 is duplicated 4000bp downstream
A portion of BB118 is duplicated 800bp downstream

A Cis-anti sRNA is characterized by all of or a portion of the sRNA sequence running antisense to an annotated gene sequence
An intergenic region (IGR) sRNA is characterized by the sRNA sequence being wholly contained within an intergenic region
A potential leader sRNA is characterized by being immediately upstream and sense to an annotated gene sequence while having an identifiable peak separate from the peaks generated by the mRNA

*Leader sRNAs need to be verified as such.
**Consensus aphaprotobacterial promoter sequence: CTTGAC-N_{17}-CTATAT
***Verified by 5' RACE analysis
****Rho-independent terminators predicted by the ARNold web tool

Table S4.4: Average TPMs of identified *B. bacilliformis* sRNAs.

Name	PI25	PI30	PI37	pH06	pH07	pH08	PIBG	HUVE	HB37	HBBG	AVG	St Dev	Cut-Off	Specific
BB001	496.281673	355.0882	261.7784	214.4198	203.5735	136.2476	436.7377	326.6814	168.5673	179.7341	277.91	121.44	399.35	
BB002	6780.15889	2252.933	5268.244	14597.06	17276.29	9683.019	3215.727	1281.349	293.2723	255.9822	6090.4	6012.1	12103	S
BB003	1336.93564	929.6449	607.1764	463.9976	279.8368	134.6836	1831.387	5292.666	2633.749	3345.881	1685.6	1647.1	3332.7	I
BB004	604.971755	551.7564	428.0488	221.896	175.7485	155.7173	540.7636	1058.119	532.0631	667.8628	493.69	271.18	764.88	
BB005	416.244633	158.835	91.33097	237.0055	199.4454	191.0331	217.4196	173.8117	222.1369	238.4801	214.57	83.368	297.94	
BB006	5226.38713	2724.366	1161.087	4826.729	5530.814	3564.889	6155.516	9939.711	3037.116	2912.592	4507.9	2446.4	6954.3	
BB007	2221.09914	2922.444	1356.977	419.9187	938.3363	455.1365	53.03346	946.4493	783.363	1200.609	1129.7	867.3	1997	
BB008	420.472325	381.7612	136.8421	127.0299	89.12428	153.627	302.3291	647.3359	226.5039	256.0073	274.1	171.56	445.67	
BBffH	511.694974	152.5887	201.0233	218.9877	72.9172	135.0689	522.7546	850.8311	232.1132	277.8666	317.58	239.61	557.19	
BB009	102.076297	50.95892	57.87471	48.0806	47.83368	70.01687	101.8461	1065.599	37.67142	52.89733	163.49	317.75	481.24	
BB010	650.427422	506.7517	272.6527	169.8482	151.5391	169.0627	415.7373	210.6175	94.98924	78.73111	272.04	190.6	462.64	
BB011	176.443376	92.18907	86.55054	33.82523	22.26634	40.55508	203.0318	707.1929	176.2003	283.013	182.13	203.09	385.21	
BB012	383.073643	461.4989	328.183	117.7198	115.005	93.85092	394.7877	57.75204	42.35532	41.0943	203.53	167.25	370.78	
BB013	604.125006	1333.822	924.0228	188.7438	377.0009	122.7182	190.8603	589.4512	496.7412	559.3282	538.68	370.04	908.72	
BB014	167.432504	170.2782	111.0845	144.1497	91.69551	59.87598	281.2	425.4314	207.9048	239.3835	189.84	106.63	296.47	
BB015	523.596022	320.5067	366.6048	93.15389	70.26132	111.1741	253.1459	692.8677	338.7034	385.9669	315.6	196.94	512.54	
BB016	71.7684769	55.80893	57.06662	26.55306	19.25249	15.26721	110.8978	345.6361	218.7939	301.1687	122.22	121.92	244.14	I
BB017	163.262924	320.9189	102.982	68.5975	87.02759	42.2084	80.2017	49.33542	107.5717	124.6905	114.68	80.816	195.5	
BB018	554.527535	405.3938	447.8946	245.9185	268.2929	713.568	515.1939	481.5326	151.9881	176.8074	396.11	181.57	577.68	
BB019	632.180909	711.1818	481.6992	156.1502	231.8251	132.4912	236.2538	87.8217	129.5953	159.0255	295.82	226.97	522.8	
BB020	298.008697	172.179	98.92058	101.7274	65.37053	56.65238	464.9246	229.1876	243.2976	347.2185	207.75	134.54	342.29	I
BB021	491.341899	966.3944	657.9737	170.6068	163.185	78.04279	1040.212	1795.005	765.9353	1082.971	721.17	530.5	1251.7	
BB022	426.462993	995.9694	1017.476	203.1249	188.3647	94.10134	429.8425	1287.867	717.4385	1087.647	644.83	431.75	1076.6	I
BB6S	25043.3746	8439.385	11618.14	220820.8	335928.5	402121.6	36686.4	11180.84	27009.04	18460.69	109731	151393	261123	S
BB023	2172.24517	436.5188	632.8404	414.7754	164.9013	511.5875	1505.105	1053.713	297.739	426.8075	761.62	633.89	1395.5	
BB024	148.076796	236.9635	961.3708	104.5203	123.7008	47.12176	1954.901	2378.895	2030.043	2906.471	1089.2	1116.4	2205.6	I
BB025	818.264627	2134.598	2551.78	1032.291	907.2288	495.1546	8951.143	13148.52	3247.903	3704.811	3699.2	4140.8	7839.9	I
BB026_1	12820.6501	2448.968	2900.803	12724.51	6223.749	4174.382	5797.677	1616.631	7375.695	7737.729	6382.1	3944.7	10327	S
BB026_2	2419.19079	1724.189	1608.817	619.6049	474.9493	283.0937	620.3166	1172.613	516.745	565.6484	1000.5	702.35	1702.9	
BB026_3	4916.71227	5971.758	5097.68	2209.464	1991.992	537.6892	759.3877	6879.352	4333.144	4949.106	3764.6	2221.3	5986	
BB026_4	1061.48397	298.3998	322.9151	1719.788	1543.905	1228.067	830.6596	13303.02	274.8566	304.8648	2088.8	3977.3	6066.1	
BB026_5	1012.29667	511.6615	305.6435	1071.646	737.442	536.3487	424.5746	1991.765	95.29877	78.1965	676.49	572.14	1248.6	
BB026_6	510.588768	1058.003	658.4091	191.0989	263.423	95.61526	204.7118	636.6194	159.519	141.2912	391.93	313.61	705.54	
BB027_1	1854.05251	1018.427	1146.852	1379.768	1241.823	718.0987	826.3864	11853.58	983.9935	1202.708	2222.6	3398.7	5621.2	
BB027_2	3046.59559	948.8716	1180.915	4320.998	3147.575	3232.133	2828.439	11219.11	717.0406	958.5495	3160	3092	6252	
BB028	307.232924	302.4516	192.9082	53.1616	88.00946	81.2821	63.48193	91.12551	45.91663	45.72519	127.13	102.83	229.96	
BB029	367.994486	299.5016	191.0167	93.41232	103.1092	47.2494	432.7158	407.2171	374.8025	487.6771	280.47	159.14	439.61	
BB030	350.135266	698.5395	349.6149	159.8816	197.9999	138.6489	554.5447	561.7907	283.3865	278.3033	357.28	188.9	546.18	
BB031	325.17987	417.5108	193.5004	240.3349	229.219	258.1843	236.0756	307.8706	328.2474	357.3796	289.35	69.288	358.64	
BB032	274.683035	340.8093	115.0336	90.12971	122.1232	67.41984	37.38817	128.6255	56.76688	68.02435	130.1	99.414	229.51	
BB033	296.702866	454.4661	240.2845	94.89685	94.57767	31.34044	175.4968	144.5604	90.13126	94.26719	171.67	127.03	298.7	
BB034	319.673295	342.3876	196.3561	117.6612	131.1013	48.45703	233.7548	88.64983	104.6191	117.0662	169.97	99.777	269.75	
BB035	516.799307	605.5225	344.4244	94.60037	173.1977	60.82428	549.8415	851.0611	911.3474	1171.631	527.92	370.87	898.79	I
BB036_1	660.541698	710.178	334.8262	202.4536	235.1499	104.7914	306.203	315.5596	266.5505	437.9428	357.42	194	551.42	
BB036_2	590.525059	616.6129	291.697	179.9272	203.364	95.07	270.6425	339.3522	266.7184	436.4141	329.03	171.32	500.36	
BB036_3	960.878633	474.0072	133.5273	157.8306	137.8946	210.9479	47.9791	610.1676	258.0122	228.6069	321.99	281.05	603.04	
BB036_4	642.650509	214.9473	104.6245	184.3745	179.5666	149.6257	241.683	771.0831	260.8984	339.856	308.93	221.34	530.28	
BB037	288.188925	175.2369	189.1463	124.1646	109.7646	151.6657	313.0461	282.5354	320.3999	351.9824	230.61	89.785	320.4	I
BB038	598.51536	735.9568	310.019	252.0444	270.1891	81.96199	191.1948	75.06511	63.64589	67.17829	264.58	233.32	497.9	
BB039	740.919339	528.6875	396.6356	810.5112	472.2024	1128.306	2645.049	2548.956	3522.978	3708.411	1650.3	1315	2965.2	I
BB040	224.107301	433.1188	346.3147	72.28228	184.9189	136.6577	163.2562	401.5428	258.1302	403.9554	262.43	126.87	389.3	
BB041	510.944396	538.6592	394.3467	206.2553	133.939	90.72333	249.6925	324.1868	115.2657	174.6566	273.87	162.15	436.02	
BB042	344.554468	774.3039	429.0012	41.75375	124.027	34.98344	147.0514	215.7558	43.85225	49.66122	220.49	237.36	457.85	
BB043	223.142756	526.7062	249.5801	74.7058	154.064	59.60163	106.7515	99.47767	25.55547	30.14216	154.97	150.94	305.92	
BB044	186.708545	434.8789	269.5569	99.42693	266.9863	72.75748	145.1509	132.9191	99.82227	117.9773	182.62	111.67	294.29	
BB045	718.71061	690.5309	379.5445	168.9944	198.0634	108.8854	214.7779	735.948	351.9168	452.5732	401.99	239.79	641.79	
BB046	329.54722	365.7986	209.84	86.0165	104.7747	93.65417	28.53721	166.3197	27.31351	33.67072	144.55	122.48	267.03	
BB047	480.974279	485.7051	176.2431	54.15996	127.6259	37.61825	110.4182	41.9719	20.132	22.72946	155.76	179.88	335.64	
BB048	157.734861	250.5587	176.2653	70.9104	108.2427	75.31724	117.8684	444.2793	397.9973	503.7205	230.29	161.32	391.61	I
BB049	357.655523	357.8118	294.852	142.4805	108.6153	113.7462	514.4467	943.3427	409.7021	389.3256	363.2	245.58	608.78	
BB050	261.814354	560.6879	329.8038	88.66957	183.8473	73.81209	138.5793	119.2558	73.21432	47.164	187.68	159.01	346.69	
BB051	303.810337	240.2345	173.5363	118.0903	147.9681	116.7088	360.3237	205.6907	106.0328	117.4115	188.98	87.953	276.93	
BB052	102.76717	133.9297	101.0641	48.24658	60.19751	30.46652	43.50422	372.0047	272.2816	309.8956	147.44	124.13	271.56	I
BB053	355.971449	225.915	156.1479	176.8417	139.892	113.7646	280.7633	26.82198	51.71602	50.89614	157.87	106.19	264.07	
BB054	368.829075	350.2621	269.9674	133.913	251.83	125.2511	324.0297	311.7748	50.54563	96.16062	228.26	116.25	344.51	
BB055	189.612668	197.1108	162.0038	80.39799	66.94351	31.93417	379.6426	114.2548	40.54335	58.07848	132.05	105.85	237.9	
BB056	656.905509	869.9735	855.7954	408.5809	296.2914	230.1841	1135.258	1801.853	925.9694	1179.522	836.03	474.12	1310.2	
BB057	716.387965	492.6844	420.6755	429.868	344.2598	353.9569	382.8162	349.5526	257.2441	303.031	405.05	128.07	533.11	
BB058	901.029707	979.7848	403.5959	249.6212	260.5086	124.7752	607.806	664.0311	365.2419	426.619	498.3	283.47	781.77	
BB059	237.061931	399.325	183.6403	77.56931	72.38496	48.3676	181.0029	119.7821	45.33751	58.02973	142.25	112.13	254.38	
BB060	489.911111	494.0511	355.1127	232.7186	147.4718	59.38701	1475.846	2214.144	1891.576	2353.471	971.37	909.91	1881.3	I
BB061	231.830413	379.1072	400.9741	147.3797	156.6597	62.10264	232.3166	277.3936	192.7724	234.4331	231.5	103.01	334.51	
BB062	433.256896	574.1522	185.4082	69.11379	178.9611	74.28101	129.2165	197.9015	104.7574	109.6332	205.67	166.64	372.31	
BB063	354.946475	432.1115	365.282	124.0033	131.3285	230.7329	706.3074	1664.856	282.746	460.0475	475.24	451.63	926.86	
BB064	290.537767	375.4599	197.7904	60.06465	108.4338	67.61872	97.8107	185.0351	46.72835	54.68406	148.42	112.12	260.54	
BB065	366.417099	330.3586	254.2579	156.575	175.416	163.2833	347.8948	292.6381	141.1781	187.1022	241.51	87.126	328.64	
BB066	352.476646	210.6201	155.9131	63.21188	64.10958	63.57965	283.8488	714.1883	243.2809	342.099	249.33	197.27	446.6	
BB067	381.041656	304.2419	282.9994	137.0866	101.3921	80.3013	63.23738	318.677	41.58271	84.78748	179.53	127.15	306.69	
BB068_1	598.461769	824.7071	429.242	170.4434	191.7078	111.5518	521.3269	322.4612	128.6248	186.8065	348.53	238.8	587.34	
BB068_2	532.531228	443.0651	268.7777	162.4738	151.8778	125.825	408.1832	220.797	204.3874	208.0139	272.59	139.35	411.95	
BB069	327.144906	477.6685	312.0723	55.16504	108.2412	71.08251	225.0267	230.5119	69.47676	110.7217	198.71	140.79	339.5	
BB070	678.759974	326.4783	230.0698	83.50372	105.3729	279.6675	628.2335	436.1977	121.1444	135.4579	302.49	215.9	518.39	
BB071	1627.28054	793.3775	665.3765	732.7659	715.7462	845.0456	1382.095	1444.228	1013.123	1050.844	1027	343.7	1370.7	
BB072	213.597306	316.0681	156.3719	61.78344	56.86545	46.30738	91.29691	86.16907	33.83925	38.70862	110.1	92.106	202.21	
BB073	313.477678	294.7395	113.3382	79.78813	152.9438	136.4413	106.1369	519.9038	231.9374	221.6717	217.04	133.25	350.29	
BB074	690.565003	1098.905	666.3118	110.0169	320.0579	130.9471	190.5273	149.7323	110.9173	114.6807	358.27	343.31	701.58	
BB075	231.511706	127.348	119.3441	115.7535	108.9518	236.0343	420.7544	377.4143	97.87516	126.7315	196.17	118	314.17	I
BB076	262.584561	234.9488	350.5645	93.63779	185.514	135.5438	28.10665	35.70313	80.65548	94.45938	150.17	105.71	255.89	
BB077	800.03833	713.2503	369.8534	283.3159	233.8144	228.3348	668.6046	293.7326	97.69315	109.3272	379.8	254.95	634.75	
BB078_1	106.401429	384.5918	199.1735	35.19846	67.99452	38.28889	64.48811	158.0028	74.28925	107.9322	123.64	105.23	228.87	
BB078_2	318.4134	339.7	133.7285	132.4709	164.7989	85.04491	256.1764	57.05093	28.61479	34.11559	155.01	114.11	269.12	
BB079	341.284745	413.5967	214.7857	170.9629	190.54	127.3889	363.9222	136.2971	88.21985	109.8781	215.69	115.99	331.68	
BB080	585.832779	769.923	346.3523	195.4594	272.788	91.03496	145.0927	17.45049	51.96482	92.61719	256.85	247.15	504	
BB081	378.85474	781.0749	559.4355	152.8784	228.8427	86.99601	244.7029	245.4552	189.5073	228.6644	309.64	210.67	520.31	
BB082	747.130919	1183.094	524.0637	179.8843	297.0433	88.77553	387.3977	222.0479	165.4853	171.5766	396.65	340.58	737.23	
BB083	243.620532	353.3191	164.0545	330.9215	452.4502	611.6682	168.1633	179.0767	257.5662	155.0259	291.59	149.05	440.64	S
BB084	650.880506	532.7412	162.3119	342.9731	261.7926	166.1786	244.013	296.2548	96.07459	119.6164	287.28	180.47	467.75	
BB085_1	261.35661	224.4993	143.3102	157.8852	183.6809	135.2653	219.35	361.1507	124.5502	129.8162	194.09	74.849	268.94	
BB085_2	529.190118	633.2456	193.7919	189.4115	184.223	89.22076	296.6596	203.1408	138.8989	149.0771	260.69	178.76	439.44	
BB086	279.416263	239.0659	176.3913	34.20663	42.66434	41.99699	17.13943	637.0525	62.09834	101.4947	163.15	190.09	353.24	
BB087	453.347933	499.0692	415.7385	246.0219	141.9824	74.13605	201.7042	53.87698	36.53076	44.66469	216.71	179.79	396.49	
BBtmRNA	19442.4032	17322.65	12334.52	49432.05	50553.26	110030.4	39388.58	128807.8	314430.1	294803.5	103655	112749	216404	I
BBRNAseP	240813.909	428996.7	457521.8	308500.5	336893.2	185428.4	188274.5	109688.5	224791.5	188524.3	266943	113222	380165	
BB088_1	273.198869	335.1273	190.9538	33.37076	53.57955	21.70031	59.0506	81.69049	54.01417	61.2507	116.39	110.16	226.56	
BB089	189.143969	162.6434	91.97164	67.46679	67.94362	92.30516	316.3932	225.4663	138.8458	178.6638	153.08	78.998	232.08	
BB090	442.027783	416.8685	187.1979	147.7688	140.0678	247.8772	229.3037	118.1707	68.42478	84.99072	208.27	129.87	338.14	
BB091	122.043331	87.45106	83.21218	55.09316	44.48606	44.19333	173.0611	404.3914	418.7804	463.1307	189.58	170.09	359.67	I
BB092	379.443361	158.6806	112.4885	9123.282	7221.958	13936.21	181.4977	167.9381	75.54182	39.03012	3139.6	5069.4	8209	S
BB093	495.083703	266.1492	196.1989	222.1314	136.1442	94.43841	1061.172	987.3418	1281.689	1911.895	665.22	615.09	1280.3	I
BB094	390.332928	213.9006	158.4259	112.2414	83.44362	200.7864	142.8467	121.2921	113.1242	108.9979	164.54	89.537	254.08	
BB095	267.271777	327.7481	434.2029	104.0519	156.1351	84.53462	405.9492	276.4481	98.77019	121.6737	227.68	132.14	359.81	
BB096	385.701359	390.6854	183.6712	297.8884	229.8164	211.9839	847.5209	353.9396	545.1214	475.4792	392.18	197.31	589.49	
BB097	426.323702	437.3399	371.4102	95.54965	137.3572	123.0601	433.6923	472.2337	260.4838	357.7289	311.52	145.52	457.04	
BB098	514.018604	948.8995	530.1956	85.33739	215.824	50.86943	85.8162	243.106	81.67664	83.52095	283.93	293.67	577.59	
BB099	312.144434	228.3959	123.7823	215.0763	136.1833	118.6554	204.4773	243.458	298.0347	376.0464	225.63	85.67	311.3	
BBspeF	371.647908	178.4116	371.404	167.675	114.5098	575.6127	3120.286	979.8654	394.0825	413.1213	668.66	896.39	1565.1	
BBar45	555.441642	195.0332	327.4399	587.9792	649.9371	1453.576	3236.169	947.8318	347.7166	361.9515	866.31	910.01	1776.3	
BB100	597.579464	1207.95	596.6917	321.1579	354.6874	175.2911	465.6404	496.2974	183.0318	194.9301	459.33	308.54	767.87	
BB101_1	131.932162	176.779	178.7252	69.28139	71.66459	46.00189	256.3726	384.3275	231.5393	290.7084	183.73	108.85	292.58	
BB101_2	324.124905	201.442	98.3094	90.7163	148.0383	95.86954	78.18405	79.58947	35.00911	35.95488	118.72	87.336	206.06	
BB102	477.082624	596.2911	579.7693	190.9088	177.2367	99.1084	457.4175	592.2828	254.6559	313.7954	373.85	189.34	563.2	
BB103_1	24534.225	2150.582	1351.537	3681.434	3372.439	4771.812	4876.949	6963.81	43933.98	42319.41	13796	16820	30616	I
BB103_2	16854.3283	1550.926	1179.037	11196.35	2982.026	4067.151	3508.035	407.0011	2431.215	2253.889	4643	5233.9	9876.9	S
BB103_3	13789.2893	1318.755	1001.653	9408.019	2354.007	3997.431	1547.845	478.9296	1725.089	1658.47	3727.9	4375.4	8103.3	S
BB103_4	482.761923	226.4913	262.2218	157.7326	90.18855	85.24615	518.4092	3763.102	2178.283	2359.409	1012.4	1286.1	2298.4	I
BB104_1	2157.92981	122.0493	105.6033	1334.557	702.142	951.6676	408.4418	686.1369	1793.063	2329.96	1059.2	811.43	1870.6	
BB104_2	532.204368	72.3747	57.81754	188.3075	54.27911	129.972	226.0384	556.4679	2272.04	3223.724	731.32	1101.2	1832.5	I
BB105_1	3383.89401	648.5271	461.4739	745.745	416.5371	795.3154	447.2481	2877.825	9354.669	8704.32	2783.6	3459.6	6243.2	I
BB105_2	423.00542	569.6361	544.7256	185.1116	161.5011	100.1391	436.3821	1127.41	807.6754	1169.877	552.55	380.04	932.59	I
BB106	199.842177	299.0306	233.9579	92.6139	86.15237	42.76914	108.9409	363.8788	95.22931	120.2618	164.27	105.35	269.62	
BB107_1	2724.5208	1991.73	1957.317	735.6687	591.4861	819.6458	405.3686	1622.051	449.5969	575.7949	1187.3	817.57	2004.9	
BB107_2	3357.10237	2699.52	2443.829	1231.982	1359.229	2417.168	281.7174	1614.161	511.5469	635.8276	1655.2	1037.7	2692.9	
BB107_3	2442.60441	1685.996	1369.376	885.6213	829.3546	1153.534	122.8438	3482.17	1240.386	1873.785	1508.6	938.05	2446.6	
BB108	612.931082	511.3048	366.5154	978.8257	376.9907	983.7782	475.2998	364.2498	639.2356	671.1033	598.02	231.12	829.15	S
BB109	2977.43093	6402.064	3043.709	1453.525	2355.855	1984.529	2440.322	1395.491	1041.63	1238.558	2433.3	1564.5	3997.8	
BB110	1445.06128	3573.185	1898.523	283.9721	806.8761	101.7905	175.0421	305.3503	310.6251	305.489	920.59	1107.6	2028.2	
BB111	312.432795	585.739	346.6421	87.40798	180.0605	90.42414	111.5499	141.5414	37.041	52.38694	194.52	172.34	366.86	
BB112	76.1275193	94.02042	58.14484	30.7334	27.76421	15.00712	64.51041	516.0034	168.4895	227.2601	127.81	151.68	279.49	
BB113	88.7740386	50.48516	34.7599	47.65324	81.69832	347.5907	75.53579	100.3604	85.48432	77.64176	98.998	89.759	188.76	
BB114	749.777825	467.3767	145.7038	405.2286	350.9563	446.856	302.4716	176.7304	65.2524	53.7115	316.41	215.42	531.82	
BB115	283.23527	255.1752	127.6819	72.88804	60.17154	40.41213	114.1625	603.5838	182.9434	226.9835	196.72	165.84	362.57	
BB116	386.860684	463.5127	243.8813	79.4913	93.09005	181.7094	197.565	330.4772	65.23436	65.89707	210.77	143.21	353.98	
BB117	354.25929	476.544	298.6725	167.834	145.4245	77.8761	241.56	107.1798	65.40299	70.69686	200.55	139.09	339.63	
BB118	537.653253	687.8962	379.7209	144.1069	145.3259	102.9878	1	1008.299	308.6478	275.0549	359.07	308.62	667.69	
BB119	1805.89812	1109.787	638.1871	648.5879	419.8632	437.4673	2207.422	439.663	283.7755	315.136	830.58	669.64	1500.2	
BB120	246.505926	194.2599	135.0042	74.92066	48.83287	51.80266	179.4514	245.3647	355.49	486.1326	201.78	139.59	341.37	I
BB121_1	920.792856	1511.297	636.4139	210.2122	341.356	142.6457	279.7619	203.7383	195.2214	181.4373	462.29	443.68	905.97	
BB121_2	11669.8659	18412.31	4809.886	1185.018	3377.111	733.89	316.4279	255.3965	833.0573	858.2561	4245.1	6083.4	10328	
BBSRP	300.271822	234.3135	225.6228	860.9071	768.3485	2006.872	92.8433	91.03039	212.8621	166.312	495.94	594.92	1090.9	
BB122	816.675959	439.3797	660.8004	322.2178	287.6851	407.6271	1617.527	1770.799	761.6592	890.7983	797.52	517.99	1315.5	I
BB123	303.839729	519.6717	407.9949	130.9959	209.8089	125.5209	197.5667	719.0328	495.4105	610.5364	372.04	209.73	581.77	I
BB124	1635.9011	112.0891	99.01927	395.3434	219.4397	1270.394	126.5486	16.23152	24.17041	55.54666	395.47	574.74	970.2	S
BB125_1	738.31857	963.5678	735.2077	236.6555	240.5886	283.0279	556.2148	220.8494	137.5356	151.8695	426.38	296.31	722.69	
BB125_2	410.697517	230.9142	220.4415	113.7877	67.18871	66.46507	1108.332	474.5114	191.0529	249.5645	313.3	309.71	623	
BB126	466.407064	642.7871	520.5039	183.3066	273.1611	122.6094	593.7753	287.2139	113.0735	139.5737	334.24	204.14	538.38	
BB127	351.928821	250.3317	261.0218	162.0284	176.9269	224.7689	306.5443	232.7855	82.16717	107.4522	215.6	84.572	300.17	
BB128	151.470901	80.84457	62.41229	52.83837	36.23647	40.46633	153.1045	420.8572	100.7948	142.9293	124.2	113.54	237.74	
BB129	329.4886	327.1539	187.9458	134.3244	148.619	92.32172	249.8832	139.7903	34.81202	46.25113	169.06	104.93	273.99	
BB130_1	657.554405	523.09	492.2728	268.1249	141.0131	128.048	727.2796	414.2803	274.2309	349.6073	397.55	203.86	601.41	
BB130_2	951.774683	743.4078	545.7157	319.3034	398.5611	566.6704	1522.55	1874.871	2341.665	2770.117	1203.5	872.11	2075.6	I
rpoD	348.921354	245.5633	298.6134	260.3189	255.3644	254.9965	312.7889	371.0605	269.5809	258.6272	287.58	43.792	331.38	

I: Upregulated in ONLY two of 4 of PIBG, HUVE, HB37, and HBBG (1 STDEV higher than mean)

S: Upregulated in ONLY two of four of PI25, pH06, pH07, pH08 (1STDEV higher than mean)

Table S4.5: Predicted IntaRNA targets of *B. bacilliformis* infection-specific sRNAs.

An "X" indicates transcripts of the indicated gene to which the sRNA is predicted to bind ($p < 0.01$). Targets with a FDR < 0.05 are indicated with a red "X".

Gene Name	Product	BB003	BB016	BB020	BB022	BB024	BB025	BB035	BB039	BB048	BB060	BB075	BB093	BB103-1	BB103-4	BB104-2	BB105-1	BB105-2	BB122	BB130-2
BARBAKC583_RS05120	DUF1561	X	X																	X
BARBAKC583_RS05100	glutathione S-transferase	X	X	X					X											
BARBAKC583_RS06570	cysteine biosynthesis protein	X				X				X									X	X
BARBAKC583_RS04740	Potassium transporter	X																		
BARBAKC583_RS05285	MFS transporter	X						X												
BARBAKC583_RS00785	ABC transporter	X																		
BARBAKC583_RS05700	bifunctional riboflavin kinase/FAD synthetase		X																	
BARBAKC583_RS02940	NAD kinase		X																	
BARBAKC583_RS04335	DUF1561		X																	
BARBAKC583_RS00225	1-acyl-sn-glycerol-3-phosphate acyltransferase		X																	
BARBAKC583_RS00920	DUF1376		X												X					
BARBAKC583_RS00500	YdcF family protein		X								X									
BARBAKC583_RS04305	protein translocase subunit SecDF		X		X				X											
BARBAKC583_RS05805	DUF1561		X			X														
BARBAKC583_RS02365	replicative DNA helicase		X						X											
BARBAKC583_RS04885	metallophosphatase		X							X										
BARBAKC583_RS00875	dUTP diphosphatase			X		X														
BARBAKC583_RS06040	tRNA threonylcarbamoyltransferase TsaB			X																X
BARBAKC583_RS01605	lytic murein transglycosylase			X																
BARBAKC583_RS00560	ATP synthase subunit alpha			X																
BARBAKC583_RS04685	Na/Pi cotransporter family protein			X																
BARBAKC583_RS00520	YggS family pyridoxal phosphate-dependent enzyme			X			X													
BARBAKC583_RS06070	apolipoprotein N-acyltransferase			X																
BARBAKC583_RS03885	ABC transporter ATP-binding protein			X																
BARBAKC583_RS02890	glutamate--tRNA ligase			X		X					X									
BARBAKC583_RS03690	DUF1561			X																
BARBAKC583_RS03935	tyrosine--tRNA ligase			X																
BARBAKC583_RS00740	Holliday junction branch migration DNA helicase RuvB			X				X												X
BARBAKC583_RS06260	heme exporter protein CcmC			X																
BARBAKC583_RS02000	DUF721			X																
BARBAKC583_RS00090	septation protein A			X															X	
BARBAKC583_RS05960	cell wall hydrolase			X																
BARBAKC583_RS02725	transcription termination/antitermination protein NusG				X							X								
BARBAKC583_RS03670	bifunctional N-acetylglucosamine...				X															
BARBAKC583_RS03850	S49 family peptidase				X							X								
BARBAKC583_RS04485	MFS transporter				X					X										
BARBAKC583_RS06645	PTS IIA-like nitrogen-regulatory protein				X															
BARBAKC583_RS00985	5-formyltetrahydrofolate cyclo-ligase				X															
BARBAKC583_RS04990	ABC transporter permease				X	X														
BARBAKC583_RS04665	DUF541				X				X											
BARBAKC583_RS06305	FxsA cytoplasmic membrane protein				X															
BARBAKC583_RS01780	Na+/H+ antiporter NhaA					X														
BARBAKC583_RS02820	phosphatidate cytidylyltransferase					X														
BARBAKC583_RS00285	diaminopimelate decarboxylase						X													
BARBAKC583_RS06155	type I pantothenate kinase						X													
BARBAKC583_RS01055	DNA gyrase inhibitor YacG							X												
BARBAKC583_RS05860	DUF2093							X					X			X				
BARBAKC583_RS05965	phosphate transport system regulatory protein PhoU							X												
BARBAKC583_RS06225	DNA replication and repair protein RecF							X												
BARBAKC583_RS04370	membrane protein							X												
BARBAKC583_RS06580	bifunctional folylpolyglutamate synthase/dihydrofolate synthase							X	X											
BARBAKC583_RS06045	ribosomal-protein-alanine N-acetyltransferase							X												
BARBAKC583_RS01225	sensor histidine kinase							X												
BARBAKC583_RS05905	nucleoside deaminase								X											
BARBAKC583_RS02115	aminopeptidase N								X		X									
BARBAKC583_RS00630	metalloprotease								X											
BARBAKC583_RS06285	ornithine carbamoyltransferase								X				X							
BARBAKC583_RS04845	photosystem reaction center subunit H								X											
BARBAKC583_RS01815	ATP synthase subunit b 2								X											
BARBAKC583_RS00600	monofunctional biosynthetic PG transglycosylase								X											
BARBAKC583_RS02025	S9 family peptidase								X											
BARBAKC583_RS02980	NADPH:quinone oxidoreductase								X						X					
BARBAKC583_RS05130	type II toxin-antitoxin system HicB family antitoxin								X											
BARBAKC583_RS05275	transcriptional repressor								X								X			
BARBAKC583_RS00200	shikimate kinase									X										
BARBAKC583_RS06280	acetylornithine transaminase									X										
BARBAKC583_RS05870	3-deoxy-D-manno-octulosonic acid transferase									X										
BARBAKC583_RS02490	CDP-diacylglycerol--glycerol-3-phosphate									X					X					
BARBAKC583_RS00375	bifunctional...									X										
BARBAKC583_RS06430	tRNA pseudouridine synthase TruA									X										
BARBAKC583_RS02210	hemin ABC transporter substrate-binding protein									X										
BARBAKC583_RS03255	30S ribosomal protein S3									X										
BARBAKC583_RS05845	alpha-hydroxy-acid oxidizing enzyme IldD									X										
BARBAKC583_RS02435	molecular chaperone SurA									X										
BARBAKC583_RS05900	rRNA pseudouridine synthase										X									
BARBAKC583_RS01230	two-component system response regulator										X									
BARBAKC583_RS01520	ribosomal RNA large subunit methyltransferase E										X									
BARBAKC583_RS02605	PAS domain-containing sensor histidine kinase										X									
BARBAKC583_RS04760	SH3 domain-containing protein										X									
BARBAKC583_RS04340	DUF1561										X									
BARBAKC583_RS06030	YbaB/EbfC family nucleoid-associated protein										X									
BARBAKC583_RS01425	glutamine synthetase											X								
BARBAKC583_RS02335	CDP-diacylglycerol--serine O-phosphatidyltransferase											X								
BARBAKC583_RS04110	glycerol-3-phosphate acyltransferase											X								
BARBAKC583_RS05445	flagellar hook protein FlgL											X								
BARBAKC583_RS02635	ATP-dependent Clp protease ATP-binding subunit											X								
BARBAKC583_RS02540	triose-phosphate isomerase											X								
BARBAKC583_RS02245	DNA polymerase											X								
BARBAKC583_RS04505	ATP-dependent DNA helicase											X								
BARBAKC583_RS06685	RNA methyltransferase											X								
BARBAKC583_RS06585	complex I NDUFA9 subunit family protein											X								
BARBAKC583_RS05420	flagellar biosynthesis protein FliQ											X								
BARBAKC583_RS04725	pyridoxal phosphate-dependent aminotransferase											X	X							
BARBAKC583_RS04530	DNA ligase LigA											X								
BARBAKC583_RS02340	oxidoreductase											X								
BARBAKC583_RS01690	ABC transporter ATP-binding protein												X							
BARBAKC583_RS03490	preprotein translocase subunit SecY												X							
BARBAKC583_RS03925	peptide chain release factor 2												X							
BARBAKC583_RS04840	redox-regulated ATPase YchF												X							
BARBAKC583_RS04090	Holliday junction resolvase RuvX												X							
BARBAKC583_RS06680	class I SAM-dependent rRNA methyltransferase												X					X		
BARBAKC583_RS04310	lysylphosphatidylglycerol synthetase													X	X		X			
BARBAKC583_RS04695	chorismate lyase													X						
BARBAKC583_RS01715	peptide chain release factor N(5)-glutamine methyltransferase													X						
BARBAKC583_RS00395	cobaltochelatase subunit CobT													X		X	X			X
BARBAKC583_RS04765	excinuclease ABC subunit UvrB														X					
BARBAKC583_RS01990	site-specific DNA methyltransferase														X					
BARBAKC583_RS03875	proline--tRNA ligase														X					
BARBAKC583_RS04200	adenylosuccinate lyase														X					
BARBAKC583_RS03650	deoxyguanosinetriphosphate triphosphohydrolase															X				
BARBAKC583_RS00855	bipolymer transporter ExbB															X				
BARBAKC583_RS00665	cell division protein ZapA															X				
BARBAKC583_RS00415	tyrosine recombinase XerC															X				
BARBAKC583_RS02425	ribosomal RNA small subunit methyltransferase A															X				
BARBAKC583_RS06465	molecular chaperone DnaK																X			
BARBAKC583_RS03270	30S ribosomal protein S17																X			
BARBAKC583_RS02125	hydroxymethylpyrimidine/phosphomethylpyrimidine kinase																X			
BARBAKC583_RS03080	phosphate acyltransferase																X			
BARBAKC583_RS04260	ribonuclease D																X			
BARBAKC583_RS05995	molybdopterin-synthase adenylyltransferase																X			
BARBAKC583_RS03480	50S ribosomal protein L30																	X		
BARBAKC583_RS05335	ATP-dependent protease																	X		
BARBAKC583_RS02560	enolase																	X		
BARBAKC583_RS01730	protein translocase subunit SecA																	X		
BARBAKC583_RS04940	histidine phosphotransferase																	X		
BARBAKC583_RS06220	3-dehydroquinate dehydratase																	X		
BARBAKC583_RS06140	HslU--HslV peptidase ATPase subunit																		X	
BARBAKC583_RS06125	polyribonucleotide nucleotidyltransferase																		X	
BARBAKC583_RS00255	DNA polymerase III subunit delta																		X	
BARBAKC583_RS06600	double-strand break repair protein AddB																		X	
BARBAKC583_RS01570	excinuclease ABC subunit UvrA																		X	
BARBAKC583_RS04815	polyphenol oxidase																		X	
BARBAKC583_RS00615	ferredoxin family protein																		X	
BARBAKC583_RS03870	MBL fold metallo-hydrolase																		X	
BARBAKC583_RS00735	tol-pal system-associated acyl-CoA thioesterase																		X	
BARBAKC583_RS02380	30S ribosomal protein S18																			X
BARBAKC583_RS00095	signal recognition particle-docking protein FtsY																			X
BARBAKC583_RS02870	amino acid ABC transporter permease																			X

Table S4.6: Predicted IntaRNA targets of *B. bacilliformis* sand fly-specific sRNAs.

An "X" indicates transcripts of the indicated gene to which the sRNA is predicted to bind ($p < 0.01$). Targets with a FDR < 0.05 are indicated with a red "X".

Gene Name	Product	BB026-1	BB083	BB092	BB103-2	BB103-3	BB124
BARBAKC583_RS04075	cold shock protein	X					
BARBAKC583_RS01990	site-specific DNA-methyltransferase	X					
BARBAKC583_RS00630	metalloprotease	X					
BARBAKC583_RS05275	transcriptional repressor	X					
BARBAKC583_RS06065	HlyC/CorC family transporter	X					
BARBAKC583_RS02255	bifunctional (p)ppGpp synthetase/guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase	X					
BARBAKC583_RS05190	exonuclease	X					
BARBAKC583_RS05865	tetraacyldisaccharide 4'-kinase	X					
BARBAKC583_RS02830	outer membrane protein assembly factor BamA		X				
BARBAKC583_RS00345	signal recognition particle protein		X				
BARBAKC583_RS04685	Na/Pi cotransporter family protein		X				
BARBAKC583_RS02620	Trk system potassium transporter TrkA		X				
BARBAKC583_RS01875	FAD-binding oxidoreductase		X				
BARBAKC583_RS00690	phosphoglycerate kinase		X				
BARBAKC583_RS00660	DUF1036 domain-containing protein		X				
BARBAKC583_RS04320	BolA family transcriptional regulator		X				
BARBAKC583_RS00065	tRNA uridine-5-carboxymethylaminomethyl(34) MnmG		X				
BARBAKC583_RS06555	2-polyprenylphenol 6-hydroxylase		X				
BARBAKC583_RS02125	hydroxymethylpyrimidine/phosphomethylpyrimidine kinase		X				
BARBAKC583_RS06655	LPS export ABC transporter ATP-binding protein		X				
BARBAKC583_RS00200	shikimate kinase		X				
BARBAKC583_RS05325	glycine cleavage system aminomethyltransferase GcvT			X			
BARBAKC583_RS00430	tRNA methyltransferase TrmD			X			
BARBAKC583_RS02100	DNA-binding response regulator			X			
BARBAKC583_RS02225	energy transducer TonB			X			
BARBAKC583_RS03820	NADH-quinone oxidoreductase subunit NuoF			X			
BARBAKC583_RS00835	ATP-dependent zinc metalloprotease FtsH			X			
BARBAKC583_RS04900	tRNA 2-thiouridine(34) synthase MnmA				X		
BARBAKC583_RS02220	TonB-dependent hemoglobin.transferrin/lactoferrin receptor protein				X		
BARBAKC583_RS00455	succinate dehydrogenase hydrophobic membrane anchor protein				X		
BARBAKC583_RS04560	cell division protein FtsA				X		
BARBAKC583_RS01535	glutamate-5-semialdehyde dehydrogenase				X		
BARBAKC583_RS05065	16S rRNA (cytidine(1402)-2'-O)-methyltransferase				X		
BARBAKC583_RS03490	preprotein translocase subunit SecY					X	
BARBAKC583_RS01680	metal ABC transporter substrate-binding protein					X	
BARBAKC583_RS05755	peroxiredoxin					X	
BARBAKC583_RS00070	ribosomal RNA small subunit methyltransferase G					X	
BARBAKC583_RS01630	phospholipase					X	
BARBAKC583_RS06440	2345-tetrahydropyridine-26-dicarboxylate N-succinyltransferase					X	
BARBAKC583_RS02150	phosphomethylpyrimidine synthase					X	
BARBAKC583_RS00380	SURF1 family protein					X	
BARBAKC583_RS04465	protease modulator HflC					X	
BARBAKC583_RS05815	amino acid permease						X
BARBAKC583_RS04535	DNA repair protein RecN						X
BARBAKC583_RS04720	cold shock protein						X
BARBAKC583_RS03380	30S ribosomal protein S10						X
BARBAKC583_RS03220	30S ribosomal protein S10						X
BARBAKC583_RS06280	acetylornithine transaminase						X
BARBAKC583_RS02910	DUF2059 domain-containing protein						X

Chapter 5: Conclusions and Future Directions

QMITEs as a source for sRNAs and highly basic proteins

We have shown that QMITE1 and QMITE2 copies can serve as sources for novel sRNAs. Specifically, the sRNAs CbsR3 and CbsR13 were derived from QMITE1 loci, whereas CbsR16 was derived from a QMITE2 locus. Although QMITE1 and QMITE2 exist in multiple copies throughout *C. burnetii* genomes, RNA-Seq analysis has shown that certain loci have unambiguous reads mapping to them, indicating that these loci are transcriptionally active. Indeed, we have shown that predicted promoter elements exist within the confines of QMITE copies (see **Figure S2.11**). When transcribed, CbsR13 and CbsR16 produce large, stable stem-loop structures (see **Figures 2.1B, 2.4B**) that may serve as substrates for RNase III degradation. The roles for these sRNAs remain unclear, although their high ambiguous expression (see **Figure S2.3**) suggests that they may serve some adaptive role.

In addition to being the source for sRNAs, some QMITE1 copies contain a short ORF coding for highly basic proteins (average pI ~12.4). These uncharacterized proteins may confer some adaptive advantage for *C. burnetii* within the host, where it must survive low pH conditions within acidified phagolysosomes. Indeed, while the extracellular pH in this niche is ~ 4.5-5, *C. burnetii* maintains an intracellular pH of 5.1 to 6.95 [228]. Determining the roles of these proteins within the intracellular niche could provide further insights into how *C. burnetii* survives such an extreme environment.

enhC and QMITEs as a timeline for *C. burnetii* strain divergence

We determined that QMITE copies show inter-strain sequence and linkage conservation (see **Table 2.1**). While this is useful for establishing a timeline for IS1111 insertion and determining recent horizontal gene acquisition, it also allows for tracing individual QMITE insertions in order to gauge their effect on strain phenotype. For example, it is very interesting to note a QMITE2 insertion in the 3' end of the *C. burnetii* Dugway strain's *enhC* gene. This insertion effectively provides a C-terminal extension for the resulting EnhC protein, but it is very possible that the strong secondary structure of the QMITE2 insertion may cause ribosome stalling, leading to degradation of the transcript via tmRNA [229, 14]. Since EnhC has been implicated in the virulence of *L. pneumophila*, a close pathogenic relative of *C. burnetii*, it is conceivable that this QMITE2 insertion in the Dugway strain *enhC* gene renders it avirulent [161, 162, 139]. The avirulent nature of the Dugway strain remains a mystery, so this could provide valuable information on the factors necessary for successful host infection.

CbsR12 is a *trans*-acting sRNA that also binds CsrA

In this study, we carried out a comprehensive characterization of a highly expressed, infection-specific sRNA of *C. burnetii*, named CbsR12. This sRNA was found to be necessary for CCV expansion during early infection of a human monocyte-derived alveolar macrophage cell line (THP-1s) (see **Figure 3.4**). Growth rate also correlated to CbsR12 expression, both *in vitro* and during THP-1 infection (see **Figures 3.2, 3.3**). We also determined that the regulation of CbsR12 was dependent on the genomic context, as a transposon-based genetic complement of a *cbsR12* mutant targeted towards a different genomic locus resulted in the dysregulation of CbsR12 expression (see **Figures 3.2, 3.3**). As a result, we conclude that there may be an unknown transcriptional regulator of

CbsR12 expression. Indeed, there is a predicted PmrA binding site within the *cbsR12* promoter, although this element was also present in the *cbsR12* complementation cassette. As a result, it does not fully explain the noted dysregulation of CbsR12 expression. Determining the mechanisms for CbsR12 regulation would help in understanding its expression pattern during infection.

We determined that CbsR12 is a *trans*-acting sRNA that engages in the post-transcriptional regulation of *carA* and *metK*. We also determined that CbsR12 binds to transcripts of *cvpD* and *ahcY in vivo*. Additionally, CbsR12 binds to CsrA-2, but not CsrA-1, *in vitro*. These data, in addition to the *E. coli* biofilm induction data, indicate that in addition to its in-*trans* activities, CbsR12 also serves as an RsmY/Z sRNA of *C. burnetii*. These sRNAs act by binding CsrA, sequestering it away from its regulatory activities [29]. We have also determined that CbsR1, another highly expressed infection-specific sRNA, also harbors multiple CsrA-binding sites. Furthermore, CbsR1 seems to contain a classical GacA/LetA-binding site that is common amongst RsmY/Z sRNAs [27]. Since the CsrA regulon of *C. burnetii* is wholly unknown, we hope that these conclusions will lead to more research into this important regulatory network. For example, determining the regulatory mechanisms of CbsR1 and CbsR12 expression would be important, as would determining the repertoire of mRNAs to which CsrA-1 and CsrA-2 bind. The regulation of CsrA-1 and CsrA-2 production would also be a fruitful area of research. Despite all that is unknown, we have developed a model of what we know about the *C. burnetii* CsrA regulon, to date (**Figure 5.1**).

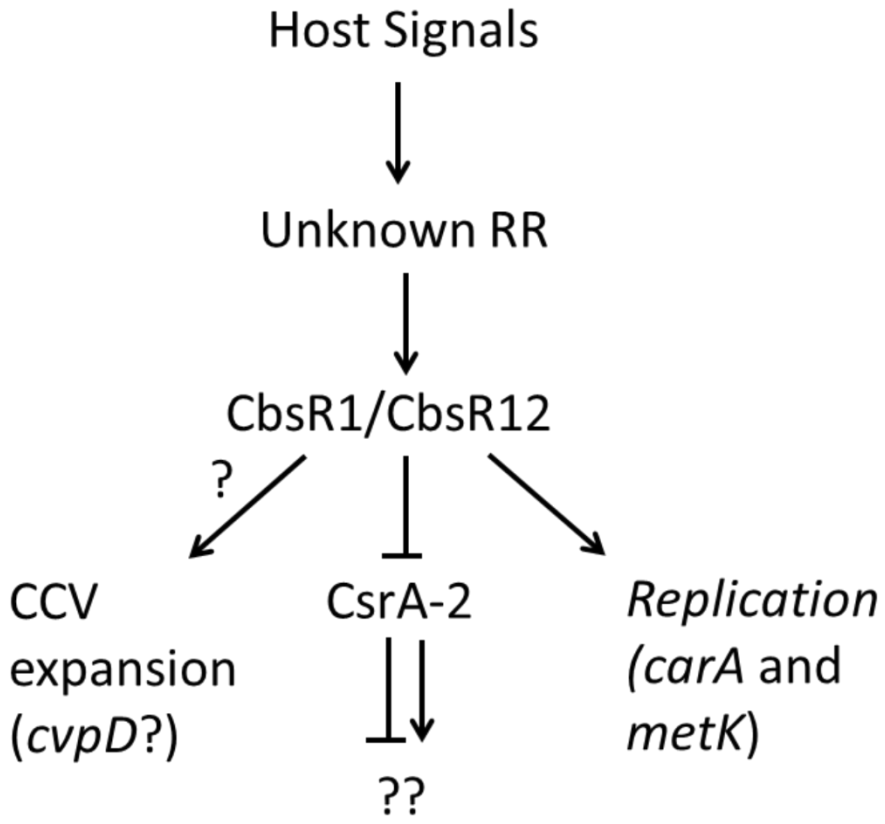


Figure 5.1: What is known of the CsrA regulatory cascade/regulon in *C. burnetii*. An uncharacterized host signal triggers a putative sensor kinase(s), leading to regulation of CbsR1/CbsR12 by an unknown response regulator. In turn, CbsR12 may promote CCV expansion via regulation of the CvpD T4BSS effector. CbsR12 also promotes replication through in-*trans* regulation of CarA and MetK production. CbsR1/CbsR12 may also bind CsrA-1/CsrA-2, leading to sequestration and indirect regulation of the CsrA-1/CsrA-2 targetome.

Determining the role of the methionine cycle in *C. burnetii*

Our research into CbsR12 has lead to many additional questions regarding the methionine cycle in *C. burnetii*. Since *C. burnetii* is a semi-auxotroph for methionine, it is presumed that the bacterium can scavenge it from the host [182]. Indeed, an ABC transporter for methionine has been predicted in *C. burnetii* [182]. We determined that CbsR12 negatively regulates MetK production. MetK is responsible for converting methionine to SAM, which is an essential methyl group donor in bacterial cells [166]. Although the regulation of MetK production in the context of retaining scavenged methionine is intuitive, *C. burnetii* would seemingly also require a mechanism for producing or scavenging SAM. Indeed, SAM scavenging has been reported in another obligate intracellular pathogen, *Rickettsia prowazekii* [185]. Furthermore, we have determined, through position-specific iterated BLAST (PSI-BLAST) homology and *in silico* protein folding analysis [230], that CBU_0636 may be a SAM transporter homolog (**Figure 5.2**). If so, it may provide a compensatory mechanism for the down-regulation of MetK production by CbsR12. Also, determining the nature of *ahcY* regulation by CbsR12 may also provide insights into the role of the methionine cycle during infection.

A B

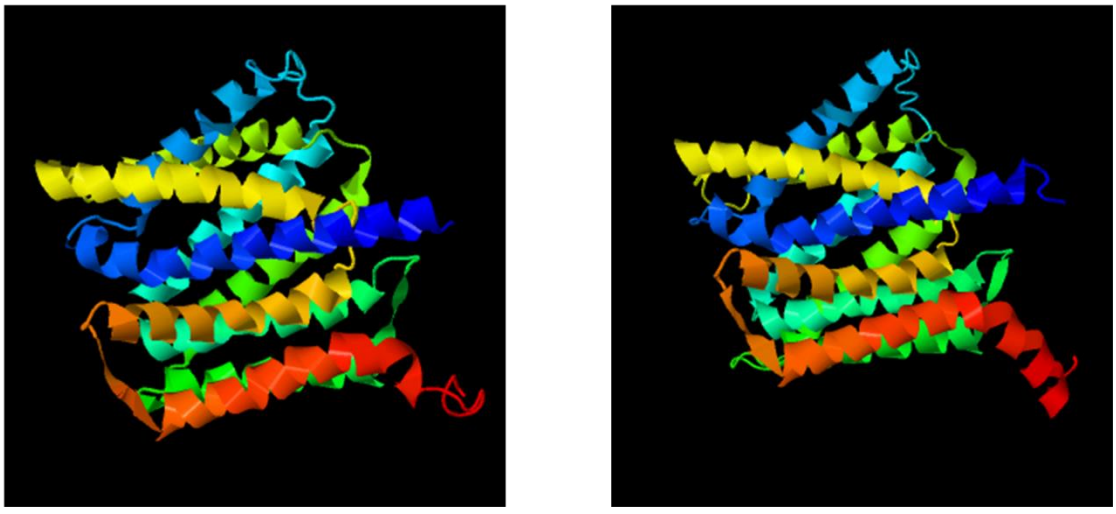


Figure 5.2: CBU_0636 may be a SAM transporter homolog. I-TASSER protein structure comparisons of the *R. prowazekii* Madrid E SAM transporter (*sam* gene) (**A**) and the *C. burnetii* RSA493 CBU_0636 gene product (**B**).

BbsRs as a means of rapid regulation in rapidly changing environments

We performed total RNA-Seq analysis on *B. bacilliformis* grown *in vitro* then shifted to one of ten conditions designed to mimick the various environments encountered by the pathogen during its life cycle. In doing so, we discovered 160 novel sRNAs, some of which were found to be differentially expressed under certain conditions (see **Table S4.3**). We hope that the discovery of these sRNAs leads to many future characterization studies. Additionally, since we performed total RNA-Seq, global gene expression analyses can be carried out. Alternatively, the results may be used as reference datasets in future studies.

Determining the *B. bacilliformis* Hfq targetome would be useful in further characterizing the identified sRNAs. For example, by performing an *in vivo* crosslinking analysis such as CLASH [48], one could simultaneously determine the repertoire of Hfq-binding sRNAs along with the mRNAs they target. It is worth noting, though, that the majority (149 out of 160) of the sRNAs we identified do not have a predicted Rho-independent terminator, which is thought to be essential for binding Hfq (see **Table S4.3**) [20]. However, this doesn't rule out the possibility of some unknown sRNA chaperone being involved.

Among the *B. bacilliformis* sRNAs we identified, the BB103 grouping (BB103-1, BB103-2, BB103-3, and BB103-4; see **Table S4.3**) is ideal for future characterization.

While each sRNA shows high expression under certain conditions (see **Table S4.4**), we determined that BB103-1 and BB103-4 are infection-specific, while BB103-2 and BB103-3 are sand fly-specific. Furthermore, BB103-1 and BB103-4 are both predicted to target transcripts of the RS04310 gene, coding for lysylphosphatidylglycerol synthetase, which is involved in the defense against host antimicrobial peptides by other pathogens (see **Table S4.5**) [219]. Determining the targetomes of these sRNAs and how they are regulated may help in understanding how *B. bacilliformis* adapts to the mammalian host and sand fly vector.

BbsR9 is a *Bartonella*-specific sRNA uniquely expressed in the arthropod vector

We determined that BbsR9 is a highly expressed, sand fly-specific sRNA that targets transcripts of the *ftsH*, *nuoF*, and *gcvT* genes, *in vitro* (see **Figures 4.6, 4.7**). We hypothesize that the regulation of these transcripts aids *B. bacilliformis* in persistence in the sandfly vector. Additionally, BbsR9 has a predicted Rho-independent terminator and is also conserved in some other *Bartonella* spp. Since pathogenic *Bartonella* spp. are vector-borne, we predict that BbsR9 is involved in persistence in a wide array of arthropod vectors. Experimental infections of *L. verrucarum* with a *bbsR9* mutant strain could be done to help determine its role in persistence. Additionally, since BbsR9 may be bound by Hfq, CLASH [48] could be useful in determining its repertoire of mRNA targets. Finally, since BbsR9 is appreciably expressed under very limited conditions (pH06, pH07, and pH08; see **Tables 4.1, S4.4**), it would be prudent to identify the specific regulators involved, as they may also be generically involved in the transition of *B. bacilliformis* from vector to host and back again.

References

1. Carrier MC, Lalaouna D, Masse E. Broadening the definition of bacterial small RNAs: Characteristics and mechanisms of action. Annu Rev Microbiol. 2018;72:141-61.
2. Vanderpool CK, Balasubramanian D, Lloyd CR. Dual-function RNA regulators in bacteria. Biochimie. 2011;93(11):1943-9.
3. Mizuno T, Chou MY, Inouye M. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). Proc Natl Acad Sci U S A. 1984;81(7):1966-70.
4. Novick RP, Ross HF, Projan SJ, Kornblum J, Kreiswirth B, Moghazeh S. Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. EMBO J. 1993;12(10):3967-75.
5. Wassarman KM, Storz G. 6S RNA regulates *E. coli* RNA Polymerase activity. Cell. 2000;101(6):613-23.
6. Masse E and Gottesman S. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. Proc Natl Acad Sci U S A. 2002;99(7):4620-5.
7. Keiler KC, Waller PRH, Sauer RT. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. Science. 1996;271(5251):990-3.
8. Sledjeski DD, Gupta A, Gottesman S. The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in *Escherichia coli*. EMBO J. 1996;15(15):3993-4000.

9. Majdalani N, Chen S, Murrow J, St John K, Gottesman S. Regulation of RpoS by a novel small RNA: the characterization of RprA. Mol Microbiol. 2004;39(5):1382-94.
10. Wehner S, Damm K, Hartmann RK, Marz M. Dissemination of 6S RNA among bacteria. RNA Biol. 2014;11(11):1467-78.
11. Wassarman KM. 6S RNA, a global regulator of transcription. Microbiol Spectr. 2018;6(3):10.1128/microbiolspec.RWR-0019-2018.
12. Hindley J. Fractionation of 32P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. J Mol Biol. 1967;30(1):125-36.
13. Trotochaud AE, Wassarman KM. 6S RNA function enhances long-term cell survival. J Bacteriol. 2004;186(15):4978-85.
14. Hayes CS, Keiler KC. Beyond ribosome rescue: tmRNA and co-translational processes. FEBS Lett. 2010;584(2):413-419.
15. Ray BK, Apirion D. Characterization of 10S RNA: A new stable RNA molecule from *Escherichia coli*. Mol Gen Genet. 1979;174(1):25-32.
16. Updergrove TB, Zhang A, Storz G. Hfq: the flexible RNA matchmaker. Curr Opin Microbiol. 2016;30:133-138.
17. Vogel J, Luisi BF. Hfq and its constellation of RNA. Nat Rev Microbiol. 2011;9(8):578-89.
18. Fernandez MTF, Eoyang L, August JT. Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. Nature. 1968;219(5154):588-90.
19. Sobrero P, Valverde C. The bacterial protein Hfq: Much more than a mere RNA-binding factor. Crit Rev Microbiol. 2012;38(4):276-99.

20. Otaka H, Ishikawa H, Morita T, Aiba H. PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. Proc Natl Acad Sci U S A. 2011;108(32):13059-64.
21. Robinson KE, Orans J, Kovach AR, Link TM, Brennan RG. Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. Nucleic Acids Res. 2014;42(4):2736-49.
22. Sauer E, Schmidt S, Weichenrieder O. Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. Proc Natl Acad Sci U S A. 2012;109(24):9396-401.
23. Vinent HA, Henderson CA, Ragan TJ, Garza-Garcia A, Cary PD, Gowers DM, *et al*. Characterization of *Vibrio cholera* Hfq provides novel insights into the role of the Hfq C-terminal region. J Mol Biol. 2012;420(1-2):56-69.
24. Bohn C, Rigoulay C, Bouloc P. No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*. BMC Microbiol. 2007;7:10.
25. Romeo T, Gong M, Liu MY, Brun-Zinkernagel AM. Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. J Bacteriol. 1993;175(15):4744-4755.
26. Jackson DW, Suzuki K, Oakford L, Simecka JW, Hart ME, Romeo T. Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*. J Bacteriol. 2002;184(1):290-301.

27. Rasis M, Segal G. The LetA-RsmYZ-CsrA regulatory cascade, together with RpoS and PmrA, post-transcriptionally regulates stationary phase activation of *Legionella pneumophila* Icm/Dot effectors. Mol Microbiol. 2009;72(4):995-1010.
28. Mercante J, Edwards AN, Dubey AK, Babitzke P, Romeo T. Molecular geometry of CsrA (RsmA) binding to RNA and its implications for regulated expression. J Mol Biol. 2009;392(2):511-28.
29. Liu MY, Gui G, Wei B, Preston III JF, Oakford L, Yuksel U, Giedroc DP, Romeo T. The RNA molecule CsrB binds to the global regulatory protein CsrA and anatagonizes its activity in *Escherichia coli*. J Biol Chem. 1997;272(28):17502-17510.
30. Heeb S, Blumer C, Haas D. Regulatory RNA as mediator in GacA/RsmA-dependent global control of exoproduct formation in *Pseudomonas fluorescens* CHA0. J Bacteriol. 2002;184(4):1046-1056.
31. Sahr T, Bruggemann H, Jules M, Lomma M, Albert-Weissenberger C, Cazalet C, Buchrieser C. Two small ncRNAs jointly govern virulence and transmission in *Legionella pneumophila*. Mol Microbiol. 2009;72(3):741-762.
32. Molofsky AB, Swanson MS. *Legionella pneumophila* CsrA is a pivotal repressor of transmission traits and activator of replication. Mol Microbiol. 2003;50(2):445-461.
33. Nevo O, Zusman T, Rasis M, Lifshitz Z, Segal G. Identification of *Legionella pneumophila* effectors regulated by the LetAS-RsmYZ-CsrA regulatory cascade, many of which modulate vesicular trafficking. J Bacteriol. 2014;196(3):681-692.
34. Chakravarty S, Massé E. RNA-dependent regulation of virulence in pathogenic bacteria. Front Cell Infect Microbiol. 2019;9:337.

35. Janzon L, Arvidson S. The role of the delta-lysin gene (*hld*) in the regulation of virulence genes by the accessory gene regulator (*agr*) in *Staphylococcus aureus*. EMBO J. 1990;9(5):1391-99.
36. Chevalier C, Boisset S, Romilly C, Masquida B, Fechter P, Geissmann T, *et al*. *Staphylococcus aureus* RNAIII binds to two distant regions of *coa* mRNA to arrest translation and promote mRNA degradation. PLoS Pathog. 2010;6(3):e1000809.
37. Boisset S, Geissmann T, Huntzinger E, Fechter P, Bendridi N, Possedko M. *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. Genes Dev. 2007;21(11):1353-66.
38. Morfeldt E, Taylor D, von Gabain A, Arvidson S. Activation of alpha-toxin translation in *Staphylococcus aureus* by the *trans*-encoded antisense RNA, RNAIII. EMBO J. 1995;14(18):4569-77.
39. Khandige S, Kronborg T, Uhlin BE, Møller-Jensen J. sRNA-mediated regulation of P-fimbriae phase variation in uropathogenic *Escherichia coli*. PLoS Pathog. 2015;11(8):e1005109.
40. Lane MC, Mobley HLT. Role of P-fimbrial-mediated adherence in pyelonephritis and persistence of uropathogenic *Escherichia coli* (UPEC) in the mammalian kidney. Kidney Int. 2007;72(1):19-25.
41. Pain A, Ott A, Amine H, Rochat T, Bouloc P, Gautheret D. An assessment of bacterial small RNA target prediction programs. RNA Biol. 2015;12(5):509-13.
42. Kery MB, Feldman M, Livny J, Tjaden B. TargetRNA2: identifying targets of small regulatory RNAs in bacteria. Nucleic Acids Res. 2014;42(Web Server issue):W124-9.

43. Mann M, Wright PR, Backofen R. IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. Nucleic Acids Res. 2017;45(W1):W435-W439.
44. Wright PR, Georg J, Mann M, Sorescu DA, Richter AS, Lott S, *et al*. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. Nucleic Acids Res. 2014;42(Web Server issue):W119-W123.
45. Bak G, Han K, Kim K, Lee Y. Electrophoretic mobility shift assay of RNA-RNA complexes. Methods Mol Biol. 2015;1240:153-63.
46. Schoenfelder SMK, Lange C, Prakash SA, Marincola G, Lerch MF, Wencker FDR. The small non-coding RNA RsaE influences extracellular matrix composition in *Staphylococcus epidermidis* biofilm communities. PLoS Pathog. 2019;15(3):e1007618.
47. Kudla G, Granneman S, Hahn D, Beggs JD, Tollervey D. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. Proc Natl Acad Sci U S A. 2011;108(24):10010-5.
48. Iosub IA, van Nues RW, McKellar SW, Nieken KJ, Marchioretto M, Sy B, *et al*. Hfq CLASH uncovers sRNA-target interaction networks linked to nutrient availability adaptation. eLife. 2020;9:e54655.
49. Werren JH, Nur U, Wu CI. Selfish genetic elements. Trends Evol Evol. 1988;3(11):297-302.
50. Gershenson S. A new sex-ratio abnormality in *Drosophila obscura*. Genetics. 1928;13(6):488-507.
51. McClintock B. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A. 1950;36(6):344-55.

52. Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. Nature. 1980;284(5757):604-7.

53. Newton ILG, Bordenstein SR. Correlations between bacterial ecology and mobile DNA. Curr Microbiol. 2011;62(1):198-208.

54. Blanc G, Ogata H, Robert C, Audic S, Suhre K, Vestris G. Reductive genome evolution from the mother of *Rickettsia*. PLoS Genet. 2007;3(1):e14.

55. Werren JH. Selfish genetic elements, genetic conflict, and evolutionary innovation. Proc Natl Acad Sci U S A. 2011;108 Suppl 2(Suppl 2):10863-70.

56. Beare PA, Unsworth N, Andoh M, Voth DE, Omsland A, Gilk SD, *et al*. Comparative genomics reveal extensive transposon-mediated genomic plasticity and diversity among potential effector proteins within the genus *Coxiella*. Infect Immun. 2009;77(2):642-56.

57. Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev. 2014;38(5):865-91.

58. Babakhani S, Oloomi M. Transposons: the agents of antibiotic resistance in bacteria. J Basic Microbiol. 2018;58(11):905-17.

59. Touchon M, Rocha EPC. Causes of insertion sequences abundance in prokaryotic genomes. Mol Biol Evol. 2007;24(4):969-81.

60. Mira A, Ochman H, Moran NA. Deletional bias and the evolution of bacterial genomes. Trends Genet. 2001;17(10):589-96.

61. Delihas N. Impact of small repeat sequences on bacterial genome evolution. Genome Biol Evol. 2011;3:959-73.

62. Chen Y, Zhou F, Li G, Xu Y. A recently active miniature inverted-repeat transposable element, Chunjie, inserted into an operon without disturbing the operon structure in *Geobacter uraniireducens* Rf4. Genetics. 2008;179(4):2291-7.

63. Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bally P, Bourras S, *et al*. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by repeat-induced point mutations. Nat Commun. 2011;doi:10.1038/ncomms1189.

64. Kang S, Lebrun MH, Farrall L, Valent B. Gain of virulence caused by insertion of a Pot3 transposon in a *Magnaporthe grisea* avirulence gene. Mol Plant Microbe Interact. 2001;14(5):671–674.

65. Black CG, Fyfe JA, Davies JK. A promoter associated with the *neisserial* repeat can be used to transcribe the *uvrB* gene from *Neisseria gonorrhoeae*. J Bacteriol. 1995;177(8): 1952–1958.

66. Buisine N, Tang CM, Chalmers R. Transposon-like Correia elements: Structure, distribution and genetic exchange between pathogenic *Neisseria* sp. FEBS Lett. 2002;522(1-3):52-8.

67. Delihas N. Small mobile sequences in bacteria display diverse structure/function motifs. Mol Microbiol. 2008;67(3):475-81.

68. Snyder LA, Shafer WM, Saunders NJ. Divergence and transcriptional analysis of the division cell wall (*dcw*) gene cluster in *Neisseria* spp. Mol Microbiol. 2003;47(2):431-42.

69. Chen SL, Shapiro L. Identification of long intergenic repeat sequences associated with DNA methylation sites in *Caulobacter crescentus* and other alpha-proteobacteria. J Bacteriol. 2003;185(16):4997-5002.

70. Ogata H, Audic S, Abergel C, Fournier PE, Claverie JM. Protein coding palindromes are a unique but recurrent feature in *Rickettsia*. Genome Res. 2002;12(5):808–816.

71. Ogata H, Audic S, Barbe V, Artiguenave F, Fournier PE, Raoult D, Claverie JM. Selfish DNA in protein-coding genes of *Rickettsia*. Science. 2000;290(5490):347-50.

72. Mazzone M, De Gregorio E, Lavitola A, Pagliarulo C, Alifano P, Di Nocera PP. Whole-genome organization and functional properties of miniature DNA insertion sequences conserved in pathogenic *Neisseriae*. Gene. 2001;278(1-2):211-22.

73. De Gregorio E1, Silvestro G, Petrillo M, Carlomagno MS, Di Nocera PP. Enterobacterial repetitive intergenic consensus sequence repeats in yersiniae: genomic organization and functional properties. J Bacteriol. 2005;187(23):7945-54.

74. Hausner G, Hafez M, Edgell DR. Bacterial group I introns: mobile RNA catalysts. Mob DNA. 2014;5:8.

75. Wilson GW, Edgell DR. Phage T4 *mobE* promotes *trans* homing of the defunct homing endonuclease I-TevIII. Nucleic Acids Res. 2009;37(21):7110-23.

76. Angelakis E, Raoult D. Q fever. Vet Microbiol. 2010;140(3-4):297-309.

77. Minnick MF, Raghavan R. Genetics of *Coxiella burnetii*: on the path of specialization. Future Microbiol. 2011;6(11):1297-1314.

78. Romano PS, Gutierrez MG, Berón W, Rabinovitch M, Colombo MI. The autophagic pathway is actively modulated by phase II *Coxiella burnetii* to efficiently replicate in the host cell. Cell Microbiol. 2007;9(4):891-909.

79. Gutierrez MG, Vazquez CL, Munafo DB, Zoppino FCM, Beron W, Rabinovitch M, Colombo MI. Autophagy induction favours the generation and maturation of the *Coxiella*-replicative vacuoles. Cell Microbiol. 2005;7(7):981-993
80. Howe D, Melnicakova J, Barak I, Heinzen RA. Maturation of the *Coxiella burnetii* parasitophorous vacuole required bacterial protein synthesis but not replication. Cell Microbiol. 2003;5(7):469-480.
81. Zamboni DS, McGrath S, Rabinovitch M, Roy CR. *Coxiella burnetii* express type IV secretion system proteins that function similarly to components of the *Legionella pneumophila* Dot/Icm system. Mol Microbiol. 2003;49(4):965-76.
82. Chen C, Banga S, Mertens K, Weber MM, Gorbaslieva I, Tan Y, Luo ZQ, Samuel JE. Large-scale identification and translocation of type IV secretion substrates by *Coxiella burnetii*. Proc Natl Acad Sci U S A. 2010;107(50):21755-21760.
83. Cunha LD, Ribeiro JM, Fernandes TD, Massis LM, Khoo CA, Moffatt JH, Newton HJ, Roy CR, Zamboni DS. Inhibition of inflammasome activation by *Coxiella burnetii* type IV secretion system effector IcaA. Nat Commun. 2015;6:10205.
84. Larson CL, Beare PA, Voth DE, Howe D, Cockrell DC, Bastidas RJ, Valdivia RH, Heinzen RA. *Coxiella burnetii* effector proteins that localize to the parasitophorous vacuole membrane promote intracellular replication. Infect Immun. 2015;83(2):661-670.
85. Martinez E, Allombert J, Cantet F, Lakhani A, Yandrapalli N, Neyret A, Norville IH, Favard C, Muriaux D, Bonazzi M. *Coxiella burnetii* effector CvpB modulates phophoinositide metabolism for optimal vacuole development. Proc Natl Acad Sci U S A. 2016;113(23):E3260-E3269.

86. Martinez E, Cantet F, Fava L, Norville I, Bonazzi M. Identification of OmpA, a *Coxiella burnetii* protein involved in host cell invasion, by multi-phenotypic high-content screening. PLoS Pathog. 2014;10(3):e1004013.

87. Newton HJ, Kohler LJ, McDonough JA, Temoche-Diaz M, Crabill E, Hartland EL, Roy CR. A screen of *Coxiella burnetii* mutants reveals important roles for Dot/Icm effectors and host autophagy in vacuole biogenesis. PLoS Pathog. 2014;10(7):e1004286.

88. Weber MM, Chen C, Rowin K, Mertens K, Galvan G, Zhi H, Dealing CM, Roman VA, Banga S, Tan Y, Luo ZQ, Samuel JE. Identification of *Coxiella burnetii* Type IV secretion substrates required for intracellular replication and *Coxiella*-containing vacuole formation. J Bacteriol. 2013;195(17):3914-3924.

89. Weber MM, Faris R, McLachlan J, Tellez A, Wright WU, Galvan G, Luo ZQ, Samuel JE. Modulation of the host transcriptome by *Coxiella burnetii* nuclear effector Cbu1314. Microbes Infect. 2016;18(5):336-345.

90. Mahapatra S, Gallaher B, Smith SC, Graham JG, Voth DE, Shaw EI. *Coxiella burnetii* employs the Dot/Icm Type IV secretion system to modulate host NF-κB/RelA activation. Front Cell Infect Microbiol. 2016;6:188.

91. Beare PA, Sandoz KM, Larson CL, Howe D, Kronmiller B, Heinzen RA. Essential role for the response regulator PmrA in *Coxiella burnetii* Type 4B secretion and colonization of mammalian host cells. J Bacteriol. 2014;196(11):1925-40.

92. Hackstadt T. The role of lipopolysaccharides in the virulence of *Coxiella burnetii*. Ann N Y Acad Sci. 1990;590:27-32.

93. Hackstadt T, Peacock MG, Hitchcock PJ, Cole RL. Lipopolysaccharide variation in *Coxiella burnetii*: intrastraine heterogeneity in structure and antigenicity. Infect Immun. 1985;48(2):359-65.

94. Fiset P. Phase variation of *Rickettsia* (*Coxiella*) *burneti*; a study of the antibody response in guinea pigs and rabbits. Can J Microbiol. 1957;3(3):435-45.

95. Hackstadt T. Steric hindrance of antibody binding to surface proteins of *Coxiella burneti* by phase I lipopolysaccharide. Infect Immun. 1988;56(4):802-7.

96. Beare PA, Jeffrey BM, Long CM, Martens CM, Heinzen RA. Genetic mechanisms of *Coxiella burnetii* lipopolysaccharide phase variation. PLoS Pathog. 2018;14(3):e1006922.

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

98. Omsland A, Beare PA, Hill J, Cockrell DC, Howe D, Hansen B, Samuel JE, Heinzen RA. Isolation from animal tissue and genetic transformation of *Coxiella burnetii* are facilitated by an improved axenic growth medium. Appl Environ Microbiol. 2011;77(11):3720-3725.

99. Seshadri R, Paulsen IT, Eisen JA, Read TD, Nelson KE, Nelson WC, *et al*. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. Proc Natl Acad Sci U S A. 2003;100(9):5455-60.

100. Raghavan R, Hicks LD, Minnick MF. Toxic introns and parasitic intein in *Coxiella burnetii*: Legacies of a promiscuous past. J. Bacteriol. 2008;190(17):5934-5943.

101. Hicks LD, Warrier I, Raghavan R, Minnick MF. Ribozyme stability, exon skipping, and a potential role for RNA helicase in Group I intron splicing by *Coxiella burnetii*. J Bacteriol. 2011;193(19):5292–5299.

102. Warrier I, Walter MC, Frangoulidis D, Raghavan R, Hicks LD, Minnick MF. The intervening sequence of *Coxiella burnetii*: Characterization and evolution. Front Cell Infect Microbiol. 2016;6(83):10.3389/fcimb.2016.00083.

103. Hoover TA, Vodkin MH, Williams JC. A *Coxiella burnetti* repeated DNA element resembling a bacterial insertion sequence. J Bacteriol. 1992;174(17):5540–5548.

104. Partridge SR, Hall RM. The IS1111 family members IS4321 and IS5075 have subterminal inverted repeats and target the terminal inverted repeats of Tn21 family transposons. J Bacteriol. 2003;185(21):6371–6384.

105. Nocera PPD, Gregorio ED, Rocco F. GTAG- and CGTC-tagged palindromic DNA repeats in prokaryotes. BMC Genomics. 2013;14:522.

106. Schultz MG. A history of bartonellosis (Carrión's disease). Am J Trop Med Hyg. 1968;17(4):503-15.

107. Gomes C, Pons MJ, Del Valle Mendoza J, Ruiz J. Carrion's disease: an eradicable illness? Infect Dis Poverty. 2016;5(1):105.

108. Gray GC, Johnson AA, Thornton SA, Smith WA, Knobloch J, Kelley PW, *et al*. An epidemic of Oroya fever in the Peruvian Andes. Am J Trop Med Hyg. 1990;42(3):215-21.

109. Hertig M. Carrión's disease. V. Studies on Phlebotomus as the possible vector. Proc Soc Exp Biol Med. 1937;37(3):598-600.

110. Battisti JM, Lawyer PG, Minnick MF. Colonization of *Lutzomyia verrucarum* and *Lutzomyia longipalpis* sand flies (Diptera: Psychodidae) by *Bartonella bacilliformis*, the etiologic agent of Carrión's disease. PLoS Negl Trop Dis. 2015;9(10):e0004128.

111. Maguiña C, Gotuzzo E. Bartonellosis. New and old. Infect Dis Clin North Am. 2000;14(1):1-22.

112. Minnick MF, Anderson BE, Lima A, Battisti JM, Lawyer PG, Birtles RJ. Oroya fever and verruga peruana: bartonelloses unique to South America. PLoS Negl Trop Dis. 2014;8(7):e2919.

113. Maguiña C, Garcia PJ, Gotuzzo E, Cordero L, Spach DH. Bartonellosis (Carrión's disease) in the modern era. Clin Infect Dis. 2001;33(6):772-9.

114. Reynafarje C, Ramos J. The hemolytic anemia of human bartonellosis. Blood. 1961;17:562-78.

115. Maguiña C, Guerra H, Ventosilla P. Bartonellosis. Clin Dermatol. 2009;27(3):271-80.

116. Amano Y, Rumbea J, Knobloch J, Olson J, Kron M. Bartonellosis in Ecuador: serosurvey and current status of cutaneous verrucous disease. Am J Trop Med Hyg. 1997;57(2):174-9.

117. Blazes DL, Mullins K, Smoak BL, Jiang J, Canal E, Solorzano N, *et al*. Novel *Bartonella* agent as cause of verruga peruana. Emerg Infect Dis. 2013;19(7):1111-4.

118. Eremeeva ME, Gerns HL, Lydy SL, Goo JS, Ryan ET, Mathew SS, *et al*. Bacteremia, fever, and splenomegaly caused by a newly recognized *Bartonella* species. N Engl J Med. 2007;356(23):2381-7.

119. Ulloa GM, Vàsquez-Achaya F, Gomes C, Del Valle LJ, Ruiz J, Pons MJ, *et al*. Molecular detection of *Bartonella bacilliformis* in *Lutzomyia maranonensis* in Cajamarca, Peru: a new potential vector of Carrion's disease in Peru? Am J Trop Med Hyg. 2018;99(5):1229-1233.

120. Birtles RJ, Canales J, Ventosilla P, Alvarez E, Guerra H, Llanos-Cuentas A, *et al*. Survey of *Bartonella* species infecting intradomicillary animals in the Huayllacallàn Valley, Ancash, Peru, a region endemic for human bartonellosis. Am J Trop Med Hyg. 1999;60(5):799-805.

121. Hertig M. Phlebotomus and Carrión's disease III. Field studies on Phlebotomus. Am J Trop Med Hyg. 1942;S1-22(4):23-60.

122. Bentzel DE, Espinosa BJ, Canal E, Blazes DL, Hall ER. Susceptibility of owl monkeys (*Aotus nancymaae*) to experimental infection with *Bartonella bacilliformis*. Comp Med. 2008;58(1):76-80.

123. Benson LA, Kar S, McLaughlin G, Ihler GM. Entry of *Bartonella bacilliformis* into erythrocytes. Infect Immun. 1986;54(2):347-53.

124. Scherer DC, DeBuron-Connors I, Minnick MF. Characterization of *Bartonella bacilliformis* flagella and effect of antiflagellin antibodies on invasion of human erythrocytes. Infect Immun. 1993;61(12):4962-71.

125. Mitchell SJ, Minnick MF. Characterization of a two-gene locus from *Bartonella bacilliformis* associated with the ability to invade human erythrocytes. Infect Immun. 1995;63(4):1552-62.

126. Coleman SA, Minnick MF. Establishing a direct role for the *Bartonella bacilliformis* invasion-associated locus B (IalB) protein in human erythrocyte parasitism. Infect Immun. 2001;69(7):4373-81.

127. Hendrix LR. Contact-dependent hemolytic activity distinct from deforming activity of *Bartonella bacilliformis*. FEMS Microbiol Lett. 2000;182(1):119-24.

128. Verma A, Ihler GM. Activation of Rac, Cdc42 and other downstream signaling molecules by *Bartonella bacilliformis* during entry into human endothelial cells. Cell Microbiol. 2002;4(9):557-69.

129. Garcia FU, Wojta J, Broadley KN, Davidson JM, Hoover RL. *Bartonella bacilliformis* stimulates endothelial cells in vitro and is angiogenic *in vivo*. Am J Pathol. 1990;136(5):1125-35.

130. Minnick MF, Smitherman LS, Samuels DS. Mitogenic effect of *Bartonella bacilliformis* on human vascular endothelial cells and involvement of GroEL. Infect Immun. 2003;71(12):6933-42.

131. Hicks LD, Minnick MF. Human vascular endothelial cells express epithelial growth factor in response to infection by *Bartonella bacilliformis*. PLoS Negl Trop Dis. 2020;14(4):e0008236.

132. Caceres AG. Geographic distribution of *Lutzomyia verrucarum* (Townsend, 1913) (Diptera, Psychodidae, Phlebotominae), vector of human bartonellosis in Peru. Rev Inst Med Trop Sao Paulo. 1993;35(6):485-90.

133. Garreaud RD. The Andes climate and weather. Adv. Geosci. 2009;22:3-11.

134. Santos VC, Araujo RN, Machado LA, Pereira MH, Gontijo NF. The physiology of the midgut of *Lutzomyia longipalpis* (Lutz and Neiva 1912): pH in different

physiological conditions and mechanisms involved in its control. 2008;211(Pt 17):2792-8.

135. Gontijo NF, Almeida-Silva S, Costa FF, Mares-Guia ML, Williams P, Melo MN. *Lutzomyia longipalpis*: pH in the gut, digestive glycosidases, and some speculations upon Leishmania development. Exp Parasitol. 1998;90(3):212-9.

136. McCaul TF, Williams JC. Developmental cycle of *Coxiella burnetii*: structure and morphogenesis of vegetative and sporogenic differentiations. J Bacteriol. 1981;147(3):1063-76.

137. Moos A, Hackstadt T. Comparative virulence of intra- and interstrain lipopolysaccharide variants of *Coxiella burnetii* in the guinea pig model. Infect Immun. 1987;55(5):1144-50.

138. Williams JC, Peacock MG, McCaul TF. Immunological and biological characterization of *Coxiella burnetii*, phases I and II, separated from host components. Infect Immun. 1981;32(2):840-51.

139. Stoenner HG, Lackman DB. The biologic properties of *Coxiella burnetii* isolated from rodents collected in Utah. Am J Hyg. 1960;71(1):45-51.

140. Beare PA, Jeffrey BM, Martens CA, Heinzen RA. Draft genome sequences of the avirulent *Coxiella burnetii* Dugway 7D77-80 and Dugway 7E65-68 strains isolated from rodents in Dugway, Utah. Genome Announc. 2017;5(39):e00984-17.

141. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, *et al*. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647-1649.

142. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology. 2007; 56(1): 564-577.

143. Price MN, Dehal PS, Arkin AP. FastTree 2 -- Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 2010; 5(3):e9490.

144. Ge R, Mai G, Zhang R, Wu X, Wu Q, Zhou F. MUSTv2: An improved de novo detection program for recently active miniature inverted repeat transposable elements (MITEs). J Integr Bioinform. 2017;10;14(3).

145. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31(13):3406-15.

146. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, *et al*. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639-45.

147. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics. 2012;28(4):464-9.

148. Sewitz S, Crellin P, Chalmers R. The positive and negative regulation of Tn10 transposition by IHF is mediated by structurally asymmetric transposon arms. Nucleic Acids Res. 2003;31(20):5868-76.

149. Friedman DI. Integration host factor: a protein for all reasons. Cell. 1988;55:545-554.

150. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, *et al*. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research. 2016;44(D1):D279-85.

151. Shen J, Liu J, Xie K, Xing F, Xiong F, Xiao J, *et al*. Translational repression by a miniature inverted-repeat transposable element in the 3′ untranslated region. Nature Communications. 2017;8:14651.

152. Kacharia FR, Millar JA, Raghavan R. Emergence of new sRNAs in enteric bacteria is associated with low expression and rapid evolution. J Mol Evol. 2017;84(4):204-13.

153. Kaessmann H. Origins, evolution, and phenotypic impact of new genes. Genome Res. 2010;20(10):1313-26.

154. Voth DE, Beare PA, Howe D, Sharma UM, Samoilis G, Cockrell DC, *et al*. The *Coxiella burnetii* cryptic plasmid is enriched in genes encoding type IV secretion system substrates. J Bacteriol. 2011;193(7):1493–1503.

155. Moses AS, Millar JA, Bonazzi M, Beare PA, Raghavan R. Horizontally acquired biosynthesis genes boost *Coxiella burnetii*'s physiology. Front Cell Infect Microbiol. 2017;7:174.

156. Willems H, Thiele D, Frolich-Ritter R, Krauss H. Detection of *Coxiella burnetii* in cow's milk using the polymerase chain reaction (PCR). Zentralbl Veterinarmed B. 1994;41(9):580-7.

157. Duron O. The IS1111 insertion sequence used for detection of *Coxiella burnetii* is widespread in *Coxiella*-like endosymbionts of ticks. FEMS Microbiol Lett. 2015;362(17):fnv132.

158. Pasternak C, Dulermo R, Ton-Hoang B, Debuchy R, Siguier P, Coste G, *et al*. ISDra2 transposition in *Deinococcus radiodurans* is downregulated by TnpB. Mol. Microbiol. 2013;88(2):443-455.

159. Roberts SB, Spencer-Smith R, Shah M, Nebel JC, Cook RT, Snyder LAS. Correia repeat enclosed elements and non-coding RNAs in the *Neisseria* species. Microorganisms. 2016;4(31):doi:10.3390/microorganisms4030031.

160. Siddique A, Buisine N, Chalmers R. The transposon-like Correia elements encode numerous strong promoters and provide a potential new mechanism for phase variation in the meningococcus. PLoS Genetics. 2011;7(1):e1001277.

161. Liu M, Haenssler E, Uehara T, Losick VP, Park JT, Isberg RR. The *Legionella pneumophila* EnhC protein interferes with immunestimulatory muramyl peptide production to evade innate immunity. Cell Host Microbe. 2012;12(2):166–176.

162. Liu M, Conover GM, Isberg RR. *Legionella pneumophila* EnhC is required for efficient replication in tumor necrosis factor α-stimulated macrophages. Cell Micrbiol. 2008;10(9):1906–1923.

163. Sandoz KM, Popham DL, Beare PA, Sturdevant DE, Hansen B, Nair V, *et al*. Transcriptional profiling of *Coxiella burnetii* reveals extensive cell wall remodeling in the small cell variant developmental form. PLoS One. 2016;11(2):e0149957.

164. Wachter S, Raghavan R, Wachter J, Minnick MF. Identification of novel MITEs (miniature inverted-repeat transposable elements) in *Coxiella burnetii*: implications for protein and small RNA evolution. BMC Genomics. 2018;19(1):247.

165. Piette J, Nyunoya H, Lusty CJ, Cunin R, Weyens G, Crabeel M. DNA sequence of the *carA* gene and the control region of *carAB*: tandem promoters, respectively

controlled by arginine and the pyrimidines, regulate the synthesis of carbamoyl-phosphate synthetase in *Escherichia coli* K-12. Proc Natl Acad Sci U S A**.** 1984;81(13):4134-4138.

166. Urig S, Gowher H, Hermann A, Beck C, Fatemi M, Humeny A, *et al*. The *Escherichia coli* Dam DNA methyltransferase modifies DNA in a highly processive reaction. J Mol Biol. 2002;319(5):1085-1096.

167. Churchill ME, Chen L. Structural basis of acyl-homoserine lactone-dependent signaling. Chem Rev. 2011;111(1):68-85.

168. Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics. 2010;26(4):493-500.

169. Yakhnin AV, Yakhnin H, Babitzke P. Gel Mobility Shift Assays to Detect Protein-RNA Interactions. Methods Mol Biol. 2012;905:201-211.

170. Battisti JM, Hicks LD, Minnick MF. A unique *Coxiella burnetii* lipoprotein involved in metal binding (LimB). Microbiology. 2011;157(Pt 4):966-976.

171. Heinzen RA, Scidmore MA, Rockey DD, Hackstadt T. Differential interaction with endocytic and exocytic pathways distinguish parasitophorous vacuoles of *Coxiella burnetii* and *Chlamydia trachomatis*. Infect Immun. 1996;64(3):796-809.

172. Smith TA, Driscoll T, Gillespie JJ, Raghavan R. A *Coxiella*-like endosymbiont is a potential vitamin source for the Lone Star tick. Genome Biol Evol. 2015;7(3):831-838.

173. Schubert M, Lapouge K, Duss O, Oberstrass FC, Jelesarov I, Haas D, *et al*. Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. Nat Struct Mol Biol. 2017;14(9):807-813.

174. Abbott ZD, Yakhnin H, Babitzke P, Swanson MS. CsrR, a paralog and direct target of CsrA, promotes *Legionella pneumophila* resilience in water. mBio. 2015;6(3):e00595.

175. Marden JN, Diaz MR, Walton WG, Gode CJ, Betts L, Urbanowski ML, *et al*. An unusual CsrA family member operates in series with RsmA to amplify posttranscriptional responses in *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A. 2013;110(37):15055-15060.

176. Lou YC, Weng TH, Li YC, Kao YF, Lin WF, Peng HL, *et al*. Structure and dynamics of polymyxin-resistance-associated response regulator PmrA in complex with promoter DNA. Nat Commun. 2015;6:8838.

177. Parshin A, Shiver AL, Lee J, Ozerova M, Schneidman-Duhovny D, Gross CA. DksA regulates RNA polymerase in *Escherichia coli* through a network of interactions in the secondary channel that includes Sequence Insertion 1. Proc Natl Acad Sci U S A. 2015;112(50):E6862-71.

178. Zusman T, Aloni G, Halperin E, Kotzer H, Degtyar E, Feldman M, *et al*. The response regulator PmrA is a major regulator of the *icm*/*dot* type IV secretion system in *Legionella pneumophila* and *Coxiella burnetii*. Mol Microbiol. 2007;63(5):1508-1523.

179. Kay E, Humair B, Denervaud V, Riedel K, Spahr S, Eberl L, *et al*. Two GacA-sependent small RNAs modulate the quorum-sensing response in *Pseudomonas aeruginosa*. J Bacteriol. 2006;188(16):6026-6033.

180. Shmike RT, Berlin CM, Sweeney EW, Carroll WR. The generation of energy by the arginine dihydrolase pathway in *Mycoplasma hominis* 07. J Biol Chem. 1966;241(10):2225-2236.

181. Ferla MP, Patrick WM. Bacterial methionine biosynthesis. Microbiology. 2014;160(Pt 8):1571-1584.

182. Sandoz KM, Beare PA, Cockrell DC, Heinzen RA. Complementation of arginine auxotrophy for genetic transformation of *Coxiella burnetii* by use of a defined axenic medium. Appl Environ Microbiol. 2016;82(10):3042-3051.

183. Sanchez-Romero MA, Cota I, Casadesus J. DNA methylation in bacteria: from the methyl group to the methylome. Curr Opin Microbiol. 2015;25:9-16.

184. Garcia-Del Portillo F, Pucciarelli MG, Casadesus J. DNA adenine methylase mutants of *Salmonella typhimurium* show defects in protein secretion, cell invasion, and M cell cytotoxicity. Proc Natl Acad Sci U S A. 1999;96(20)::11578-11583.

185. Tucker AM, Winkler HH, Driskell LO, Wood DO. S-adenosylmethionine Transport in *Rickettsia prowazekii*. J Bacteriol. 2003;185(10):3031-3035.

186. Kuley R, Kujit E, Smits MA, Roest HIJ, Smith HE, Bossers A. Genome plasticity and polymorphisms in critical genes correlate with increased virulence of Dutch outbreak-related *Coxiella burnetii* strains. Front Microbiol. 2017;8:1526.

187. Glazunova O, Roux V, Freylikman O, Sekeyova Z, Fournous G, Tyczka J, *et al*. *Coxiella burnetii* genotyping. Emerg Infect Dis. 2005;11(8):1211-1217.

188. Muller P, Gimpel M, Wildenhain T, Brantl S. A new role for CsrA: promotion of complex formation between an sRNA and its mRNA target in *Bacillus subtilus*. RNA Biol. 2019;16(7):972-987.

189. Romeo T, Babitzke P. Global regulation by CsrA and its RNA antagonists. Microbiol Spectr. 2018; 6(2):10.1128/microbiolspec.RWR-0009-2017.

190. Coleman SA, Fischer ER, Howe D, Mead DJ, Heinzen RA. Temporal analysis of *Coxiella burnetii* morphological differentiation. J Bacteriol. 2004;186(21):7344-7352.

191. Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics. 2009;25(15):1974-1975.

192. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-930.

193. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

194. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, *et al*. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676-682.

195. Kamentsky L, Jones TR, Fraser A, Bray MA, Logan DJ, Madden KL, *et al*. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. Bioinformatics. 2011;27(8):1179-1180.

196. Lustig Y, Wachtel C, Safro M, Liu L, Michaeli S. 'RNA walk' a novel approach to study RNA-RNA interactions between a small RNA and its target. Nucleic Acids Res. 2010;38(1):e5.

197. Martinez E, Cantet F, Bonazzi M. Generation and multi-phenotypic high-content screening of *Coxiella burnetiii* transposon mutants. J Vis Exp. 2015;99:e52851.

198. Beare PA, Larson CL, Gilk SD, Heinzen RA. Two systems for targeted gene deletion in *Coxiella burnetii*. Appl Environ Microbiol. 2012;78(13):4580-4589.

199. Choi KH, Gaynor JB, White KG, Lopez C, Bosio CM, Karkhoff-Schweizer RR, *et al*. A Tn7-based broad-range bacterial cloning and expression system. Nat Methods. 2005;2(6):443-448.

200. Wang L, Yang G, Qi L, Li X, Jia L, Xie J, *et al*. A novel small RNA regulates tolerance and virulence in *Shigella flexneri* by responding to acidic environmental changes. Front Cell Infect Microbiol. 2016;6:24.

201. Tu N, Carroll RK, Weiss A, Shaw LN, Nicolas G, Thomas S, *et al*. A family of genus-specific RNAs in tandem with DNA-binding proteins control expression of the *badA* major virulence factor gene in *Bartonella henselae*. Microbiologyopen. 2017;6(2):e00420.

202. Wachter S, Bonazzi M, Shifflett K, Moses AS, Raghavan R, Minnick MF. A CsrA-binding, *trans*-acting sRNA of *Coxiella burnetii* is necessary for optimal intracellular growth and vacuole formation during early infection of host cells. J Bacteriol. 2019;201(22):e00524-19.

203. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al*. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

204. MacLellan SR, MacLean AM, Finan TM. Promoter prediction in the rhizobia. Microbiology. 2006;152(Pt 6):1751-63.

205. Strimmer K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics. 2008;24(12):1461-62.

206. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, *et al*. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36(10):3420-35.

207. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57.

208. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33(18):2938-2940.

209. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 1st ed. New York: Springer-Verlag; 2016.

210. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of intersecting sets. IEEE Trans Vis Comput Graph. 2014;20(12):1983-92.

211. Paquin B, Heinfling A, Shub DA. Sporadic distribution of tRNA(Arg)CCU introns among alpha-purple bacteria: evidence for horizontal transmission and transposition of a group I intron. J Bacteriol. 1999;181(3):1049-53.

212. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol. 2019;1962:1-14.

213. Pavlova N, Kaloudas D, Penchovsky R. Riboswitch distribution, structure, and function in bacteria. Gene. 2019;708:38-48.

214. del Val C, Romero-Zaliz R, Torres-Quesada O, Peregrina A, Toro N, Jiménez-Zurdo JI. A survey of sRNA families in α-proteobacteria. RNA Biol. 2012;9(2):119-129.

215. Torres-Quesada O, Oruezabal RI, Peregrina A, Jofré E, Lloret Javier, Rivilla R, *et al*. The *Sinorhizobium meliloti* RNA chaperone Hfq influences central carbon metabolism and the symbiotic interaction with alfalfa. BMC Microbiol. 2010;10:71.

216. Corbino KA, Barrick JE, Lim J, Welz R, Tucker BJ, Puskarz I, *et al.* Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. Genome Biol. 2005;6(8):R70.

217. Li G-W, Oh E, Weissman JS. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature. 2012;484(7395):538-41.

218. Diwan GD, Agashe D. The frequency of internal Shine-Dalgarno-like motifs in prokaryotes. Genome Biol Evol. 2016;8(6):1722-33.

219. Ernst CM, Staubitz P, Mishra NN, Yang S, Hornig G, Kalbacher H, *et al.* The bacterial defensing resistance protein MprF consists of separable domains for lipid lysinylation and antimicrobial peptide repulsion. PLoS Pathog. 2009;5(11):e1000660.

220. Banerjee R, Ragsdale SW. The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes. Annu Rev Biochem. 2003;72:209-47.

221. Caswell CC, Oglesby-Sherrouse AG, Murphy ER. Sibling rivalry: related bacterial small RNAs and their redundant and non-redundant roles. Front Cell Infect Microbiol. 2014;4:151.

222. Reinhold-Hurek B, Shub DA. Self-splicing introns in tRNA genes of widely divergent bacteria. Nature. 1992;357(6374):173-6.

223. Bittner LM, Arends J, Narberhaus F. When, how, and why? Regulated proteolysis by the essential FtsH protease in *Escherichia coli*. Biol Chem. 2017;398(5-6):625-635.

224. Falk-Krzesinski HJ, Wolfe AJ. Genetic analysis of the *nuo* locus, which encodes the proton-translocating NADH dehydrogenase in *Escherichia coli*. J Bacteriol. 1998;180(5):1174-84.

225. Kikuchi G, Motokawa Y, Yoshida T, Hiraga K. Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. Proc Jpn Acad Ser B Phys Biol Sci. 2008;84(7):246-63.

226. Vereecke D, Cornelis K, Temmerman W, Holsters M, Goethals K. Versatile persistence pathways for pathogens of animals and plants. Trends Microbiol. 2002;10(11):485-8.

227. Hong PC, Tsolis RM, Ficht TA. Identification of genes required for chronic persistence of *Brucella abortus* in mice. Infect Immun. 2000;68(7):4102-7.

228. Hackstadt T. Estimation of the cytoplasmic pH of *Coxiella burnetii* and effect of substrate oxidation on proton motive force. J Bacteriol. 1983;154(2):591-7.

229. Del Campo C, Barthomomäus A, Fedyunin I, Ignatova Z. Secondary structure across the bacterial transriptome reveals versatile roles in mRNA regulation and function. PLoS Genet. 2015;11(10):e1005613.

230. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. 2015;43(W1):W174-81.