Summer 2020

Helicobacter pylori gene regulation by virulence region located sRNAs

Veronica Janette Albrecht

Helicobacter pylori gene regulation by virulence region located sRNAs

---

A Thesis

Presented To

Eastern Washington University

Cheney, Washington

In Partial Fulfillment of the Requirements

For the degree

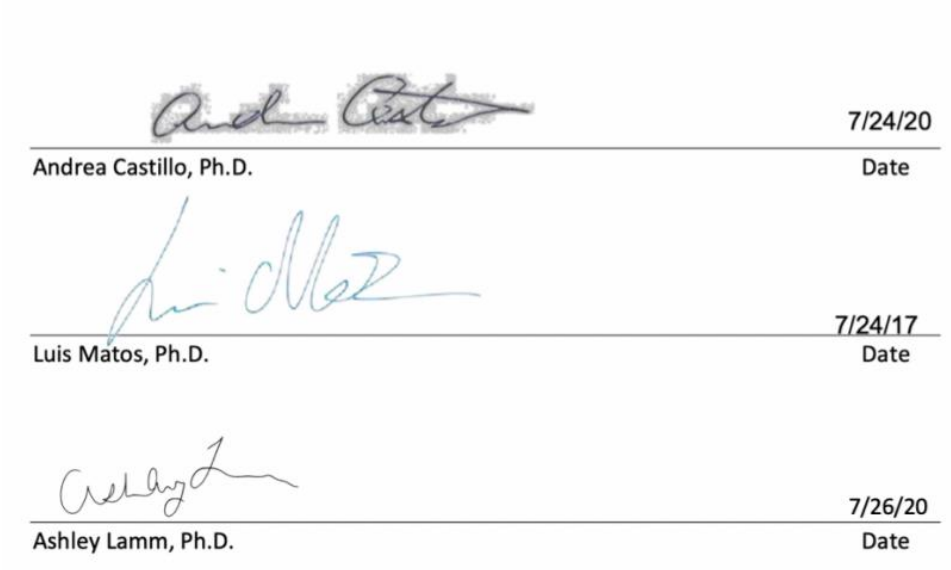
Master of Science in Biology

By

Veronica Janette Albrecht

Spring 2020

THESIS OF Veronica Janette Albrecht APPROVED BY



7/24/20

Andrea Castillo, Ph.D. Date

7/24/17

Luis Matos, Ph.D. Date

7/26/20

Ashley Lamm, Ph.D. Date

TABLE OF CONTENTS

ABSTRACT

Helicobacter pylori gene regulation by virulence region located sRNAs

By

Veronica Janette Albrecht

Spring 2020

Approximately 50% of the human population is infected with *Helicobacter pylori*, which can lead to gastrointestinal diseases such as ulcers and gastric adenocarcinoma. *Helicobacter pylori* strains are genetically variable, and some contain a DNA region called the cytotoxin associated gene pathogenicity island (cagPAI) that encodes virulence factors. Gastrointestinal disease associated with *H. pylori* are more likely to occur in infections with cagPAI positive strains. *Helicobacter pylori* has few known transcriptional regulators, but still must regulate expression to survive a constantly changing environment. A mechanism to facilitate this regulation was revealed in a transcriptome analysis conducted by Sharma *et al.* (2010) that identified 60 previously unknown small RNAs (sRNA) and suggested their role in gene regulation may be significant. Small RNAs are short non-coding transcripts that bind to target mRNAs through complementary base-pairing and regulate gene expression. Several sRNAs were identified in the cagPAI, and, to date, only one has been characterized. To learn more about cagPAI sRNAs and the genes they regulate, I characterized transcriptional regulatory sequences of two cagPAI sRNAs*, HPnc2620* and *HPnc2665*, and used a bioinformatic approach to predict their target mRNAs. The results indicate that *HPnc2620* promoter is TGTCCA- 23 nucleotides (nt) -TAAAAT and is controlled with two terminators, a Rho-dependent terminator, and a Rho-independent terminator. *HPnc2665* has the promoter consensus sequences GTCAAA- 26 nt -TTGCAA and a transcriptional Rho-independent terminator. Both sRNAs were highly conserved in *H. pylori,* but not in non-*pylori Helicobacter* and were predicted to regulate various virulence factors including chemotaxis and flagellar genes, vacuolating cytotoxin A, cagPAI genes, and urease gene *ureB*.

ACKNOWLEDGEMENTS

I thank Drs. Andrea Castillo, Luis Matos, and Ashley Lamm for serving on my thesis committee, advice and editing of this document. I thank Dr. O'Quinn for aid in editing this document and Dr. Magori for assistance with data analysis. This project was supported by a Biology Graduate Mini-grant Award to Veronica Albrecht and a EWU Foundation Start Something Big grant to Andrea Castillo.

CHAPTER I: INTRODUCTION

Helicobacter pylori characteristics

Helicobacter is a diverse genus of bacteria containing at least 32 species[1]. Many members of this genus infect the stomachs of animals including cheetahs (*H. acinonychis*), dogs (*H. bizzozeronii*), and humans (*H. pylori)*[1]. *Helicobacter* are gram negative and spiral-to-curved in shape. These bacteria are motile with multiple polar or bipolar flagella, that produce a corkscrew-like motion and allow them to colonize and persist in the stomachs of animals. *Helicobacter* are microaerophilic and neutrophilic, they require oxygen concentrations of 5-10% and a more neutral pH, making the gastric mucosa (~6.1) an ideal home for them[2].

Helicobacter pylori is highly prevalent among humans

Evidence suggests that *H. pylori* has infected humans for thousands of years, but it has only been linked to human disease for about 40 years[3,4]. In 2019, Maixner *et al.* found DNA from a virulent strain of *H. pylori* in a 5300-year-old human mummy from the Italian Alpine glacier[3]. *Helicobacter pylori* was discovered in the stomach in 1906, but was thought to be a contaminant from the mouth[5]. It was not until 1987 that Robin Warren and Barry Marshall linked *H. pylori* infection to duodenal ulcer disease with their famous experiment where Marshall consumed *H. pylori* and developed gastritis[4]. *Helicobacter pylori* has since been linked to other gastric diseases such as, gastric adenocarcinoma and mucosa-associated lymphoid tissue (MALT) lymphoma[6,7]. The high

prevalence of *H. pylori* infection and link to gastric cancer led it to be classified as a group 1 carcinogen to humans in 1994[8].

Fifty percent of humans worldwide are infected with *H. pylori* but infection distribution is not homogeneous (Figure 1)[9]. The huge variation in prevalence is attributed to the varying levels of urbanization, sanitation and access to clean water[9–13]. Infection is typically sustained throughout life, unless antimicrobial intervention occurs[11–13]. Although the details of how *H. pylori* is transmitted are not entirely known, transmission is hypothesized to occur via oral-to-oral and fecal-to-oral routes, where stomach contents are transmitted through saliva or through fecal contamination of food and fomites[10,12,13].

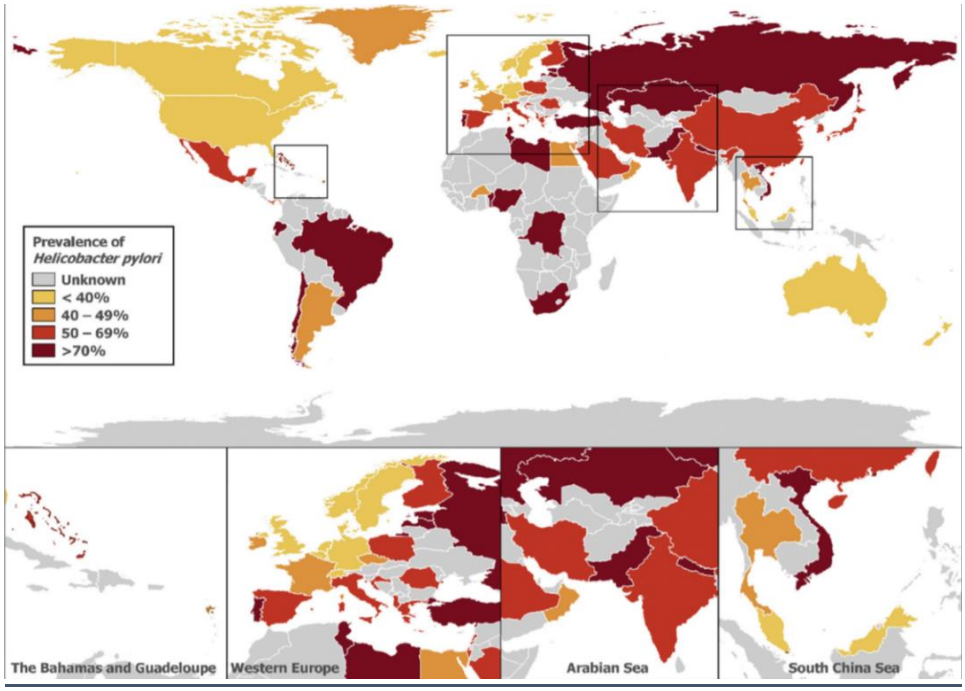


Figure 1. Map depicting the prevalence of Helicobacter pylori infection (Hooi, J. et al., 2017)

Transmission of *Helicobacter pylori* likely fecal-to-oral

Research into *H. pylori* transmission has been ongoing, but it has been difficult to find a causative correlation because the evidence for one route of transmission can be evidence for another. For example, high intrafamilial clustering of infection could point to person-person transmission or contaminated food and water[12]. The major hypothesis for transmission is fecal-contaminated food or water; evidence supporting this hypothesis includes the fact that infection rates are higher in lower economic regions where access to clean water is difficult, families tend to consume the same items, and *H. pylori* has been detected in water and food[10,12–15]. However, several studies failed to detect *H. pylori* in food and water sources, so other transmission routes such as, person-to-person or vector borne transmission (via house fly excrement contamination of food and water) also have been proposed[12,13,16,17].

***Helicobacter pylori* has evolved many virulence factors for survival**

Helicobacter pylori has evolved many virulence factors to colonize, survive and thrive within the human stomach. Virulence factors are microbe-produced molecules that enhance their ability to colonize a host and evade the host's immune response[18]. The stomach is an inhospitable environment for many bacteria, with a lumen pH of ~1.4 and a constant efflux of material into the intestine[2]. To survive the low pH in the gastric lumen, *H. pylori* produces urease, an enzyme that breaks down host-produced urea into ammonia and carbon dioxide to create a buffer against low pH of the lumen as it makes

its way to the more neutral gastric mucosa[19,20]. Urease-negative *H. pylori* is unable to colonize the stomach[20]. *Helicobacter pylori* has urease both within its cytoplasm and on its outer membrane and releases urease into the gastric mucosa by undergoing autolysis, or by secretion of outer membrane vesicles containing urease[21]. The free urease disrupts gastric epithelial tight junctions and can be internalized by gastric epithelial cells where the subunit urease A (UreA) localizes to the nuclei and alters gastric cell morphology and induces inflammation[20–23].

Helicobacter pylori uses flagellar-based motility and chemotaxis to navigate through the gastric lumen to the gastric mucosa and to avoid being expelled into the small intestine[24]. Chemotaxis is the ability for *H. pylori* to sense environmental cues and move either away or towards a different environmental niche[25]. Some factors are chemorepellent and direct *H. pylori* away from acidic pH, bile, and reactive oxygen species (ROS); other factors can be chemoattractants, such as arginine, and the cells will move up a concentration gradient[25]. Not only does chemotaxis direct motility, it promotes colonization, and modulates host immune responses[25]. Mutations of the chemotaxis proteins and chemoreceptors leads to alterations in swimming patterns and reduction or complete loss in the ability to colonize the stomach[25,26]. One study observed that chemoreceptor (tlpD, senses pH and ROS) deficient *H. pylori* mutants colonized mice gastric glands at significantly lower levels than wild type *H. pylori*; however, in mutant mice unable to produce hydrogen peroxide tlpD deficient *H. pylori* colonized the gastric glands at levels equivalent to the wild type *H. pylori*[26]. Chemotaxis has also been shown to influence the host's immune response. In a study conducted by

Rolig *et al.* in 2011, mice infected with chemotaxis protein (che-) deficient *H. pylori* had significantly lower levels of CD4+ T cells, IL17 and T regulatory cells than mice infected with wild type (che+) *H. pylori*[27].

Once *H. pylori* reaches the gastric epithelium, it uses adhesin proteins on its surface to attach to host-cell receptors and prevent it from being expelled into the small intestine and to modulate the immune system. *Helicobacter pylori* strains may possess any combination of several adhesin proteins; one example, sialic acid-binding adhesin (SabA) binds to sialyl-Lewis A antigens sLe^x and sLe^a on gastric epithelial cells[28,29]. Unemo, *et al.* (2005) found that SabA is critical for nonopsonic activation of neutrophils, meaning *H. pylori* can bind to and interact with neutrophils without being phagocytized[29]. Additionally, when *H. pylori* strains with SabA are present, gastritis increases in patients because neutrophils invade the gastric mucosa and cause damage to the gastric epithelial cells through oxidative bursts[29]. Another example of an adhesin common in *H. pylori* strains is blood group antigen binding adhesin (BabA) which allows the bacterium to adhere to fucosylated Lewis B blood-group antigen on host gastric epithelial cells[30]. Interestingly, BabA is active during early, acute infection, but is then intentionally silenced by phase variation or gene conversion[31]. In rhesus macaques, *H. pylori* shuts down BabA expression by recombination between *babA* and *babB* causing a duplication of *babB* in *babA;* this phase variation alters the outer membrane proteins masking the bacterium from immune cells[30,32]. Similarly, during gene conversion, BabA expression is lost because of slipped strand mispairing of a CT repeat region in the 5' region leading to a frame shift in the *babA* open reading frame[30,32,33]. Both mechanisms

of suppression have been seen in multiple animal models and human clinical isolates, indicating modulating BabA is important in *H. pylori*-host interactions[30–34].

Helicobacter pylori infections can persist, often life-long, due to virulence factors that alter the host immune response. Sialic acid-binding adhesin is an example of a virulence factor that changes the immune response and *H. pylori* has many others that allow it to evade and circumvent the host's defenses. Vacuolating toxin A (VacA) is an exotoxin secreted by *H. pylori* and is taken up by the epithelial cells by endocytosis[35]. When VacA is taken up by host cells, it triggers vacuolar degeneration (the formation of cytoplasmic vesicles); this is thought to be responsible for ulcer formation, because the gastric epithelium is compromised[36]. Vacuolating toxin A also inhibits activation of immune cells, specifically T cells, by blocking transcription factors that are essential for their activation[36]. *Helicobacter pylori* uses a cag-type 4 secretion system (cag-T4SS), composed of a pilus, a needle-like structure, that spans the *H. pylori* membrane and facilitates transfer of the cytotoxin associated gene A protein (CagA) from the bacterium's cytoplasm directly to the host cell cytoplasm[37,38]. The CagA protein is a cytotoxin that modulates immune cell signaling and maturation[39]. When CagA enters the hosts cytoplasm it alters dendritic cell maturation through activation of transcription factors that lead to increased levels of interleukin-10 (IL-10)[39]. Interleukin-10 is a cytokine that immune cells recognize and is critical to maintaining dendritic cells in an immature state and triggers T regulatory cell (Immune cells essential for turning down/off immune responses) differentiation[39]. Usually, IL-10 is used by the human immune system to ensure the immune response is not activated when it does not need

to be, but *H. pylori* highjacks this safeguard to prevent itself from being attacked[39]. There are also changes CagA induces in the gastric epithelial cells, such as cytoskeletal rearrangements, suppression of the mitochondrial apoptosis pathway, and disruption of cell-to-cell junctions[40]. While research is ongoing into the mechanisms behind virulence and understanding the complex host-microbe interactions, it is equally important to investigate how *H. pylori* controls expression of its genes.

Helicobacter pylori **has few regulatory proteins but many small RNAs**

We know that *H. pylori* changes its expression profile in response to different environmental conditions. For instance, when *H. pylori* is starved for iron, at least 183 genes have altered expression[41]. In the study conducted by Merrell and colleagues in 2003, they found that many genes were significantly regulated at the transcriptional level during iron starvation[41]. A microarray analysis showed that many genes (*cagA, vacA, tlpB,* etc.) were upregulated, in low iron conditions[41]. A separate study conducted by Merrell *et al.* (2003) on transcriptional regulation in acidic pH conditions found similar results with about 118 genes having altered expression. Perhaps unsurprising, several of the genes shown to have increased expression were associated with urease which protects *H. pylori* from the low pH in the gastric lumen[42]. While it is apparent that *H. pylori* regulates its genes, it has significantly fewer global regulatory proteins and two-component regulator systems compared to *Escherichia coli* and *Haemophilus influenzae*[43]. Only four proteins in *H. pylori* have a helix-turn-helix motif common to

transcription factors, whereas in *E. coli* there are at least 148 proteins and in *H. influenzae* 34 proteins.[43,44]. *Haemophilus influenzae* and *H. pylori* have similar numbers of two-component regulator systems, but *E. coli* has three times as many[43,44]. A clue about how *H. pylori* may regulate gene expression came from a transcriptome analysis conducted by Sharma *et al.* (2010); they found that small RNAs (sRNAs) were abundant (~200 sRNA identified)[45,46].

Basics of small RNAs

Many organisms, including bacteria, use sRNAs to respond to changes in environmental conditions[47]. Small RNAs are 50-450 nucleotides in length and found within coding and noncoding DNA regions[47,48]. There are four main classes of sRNAs based on the mode of regulation: RNAs that alter protein activity, clustered regularly interspaced short palindromic repeats (CRISPRs), *trans*-encoded base-pairing RNAs, and *cis*-encoded base-pairing RNAs[49]. Protein activity can be modulated by RNAs through mechanisms such as protein sequestration[50]. RNA that act to target foreign DNA and trigger degradation are part of CRISPR-Cas systems. The last two RNA groups that alter expression do so by base pairing with target mRNAs and differ based on their location in the genome compared to their target. *Cis*-encoded sRNAs (*cis*-sRNAs) reside within DNA regions that encode their target mRNAs, but on the opposite DNA strand (Figure 2A)[47,48]. *Trans*-encoded sRNAs (*trans*-sRNAs) reside at a genomic location that is distinct from their respective target mRNAs (Figure 2B)[47]. *Cis*-sRNAs and *trans*-sRNAs were

discovered in the 1980s and have been found to regulate a multitude of processes including, metabolism, metal regulation, quorum sensing, biofilm formation, environmental stress response, and pathogenesis[49,51–55].

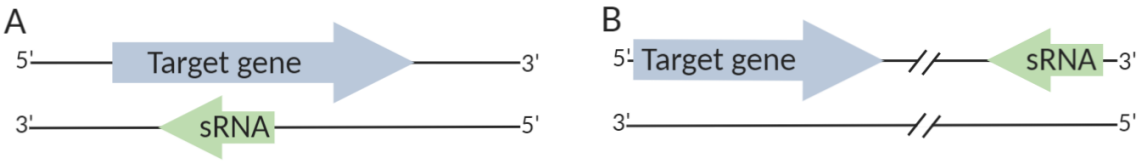


Figure 2. sRNA location on the chromosome in comparison to its target. ***2A.*** *cis-sRNA. The sRNA is in the same region as the target but on the opposite strand.* ***2B.*** *trans-sRNA. The sRNA genomic location is distinct from its target*

Cis-sRNAs and *trans*-sRNAs use several mechanisms to alter expression; in most cases, they trigger post-transcriptional inhibition, but some lead to activation. The mechanisms for down regulation include transcription attenuation, translational inhibition, or direct mRNA degradation. During transcription attenuation the mRNA target has two conformations, one in the presence of the sRNA and one without. When the sRNA is present, it base pairs to the 5' untranslated region of the mRNA inducing a stem-loop that halts transcription prematurely[56,57]. Small RNAs can inhibit translation in several ways, including direct blockage of the Ribosome Binding Site (RBS), blockage of the ribosome standby site, and structural changes downstream of the RBS. Blocking the RBS is the most common mechanism; the sRNA base pairs to the RBS of the target mRNA and prevents the ribosome from accessing the RBS (Figure 3A) [56,58–60]. A sRNA can also block the ribosome standby site or translation enhancer elements preventing/decreasing translation of the mRNA (Figure 3B)[57,58,61]. The least common mechanism for translational inhibition is the induction of structural changes

downstream of the RBS. The sRNA binds to the mRNA and triggers a secondary structure that blocks the RBS and inhibits translation of the mRNA (Figure 3C)[56,58]. Promoting mRNA degradation is another sRNA expression regulation mechanism[56,58]. During this process the sRNA base pairs with the mRNA and the complex recruits ribonucleases (RNases) that degrades the mRNA and leaves the sRNA intact (Figure 3D)[56,58].

Small RNAs can also increase target expression through several methods including stabilization, activation of translation initiation, activation by translation coupling, and regulation of transcription antitermination. A sRNA can stabilize mRNAs by recruiting RNases to cut a bicistronic transcript within an untranslated region and leaving both mRNAs intact and stable (Figure 3E)[54,58]. Messenger RNAs may have an intrinsic secondary structures that occludes the RBS, and sRNA base pairing with the mRNA changes this structure to reveal the RBS and promote translation (Figure 3F)[62]. During translation coupled activation, the sRNA base pairs with one mRNA of a bicistronic transcript which prevents a secondary structure from forming and allows the ribosome to associate with the RBS of the second mRNA[63]. Lastly, a sRNA can prevent termination of a transcript by targeting RNA binding of the Rho protein. The sRNA binds to the docking point for the Rho protein, which is required for Rho-dependent

termination, preventing Rho from terminating transcription and up regulating expression.

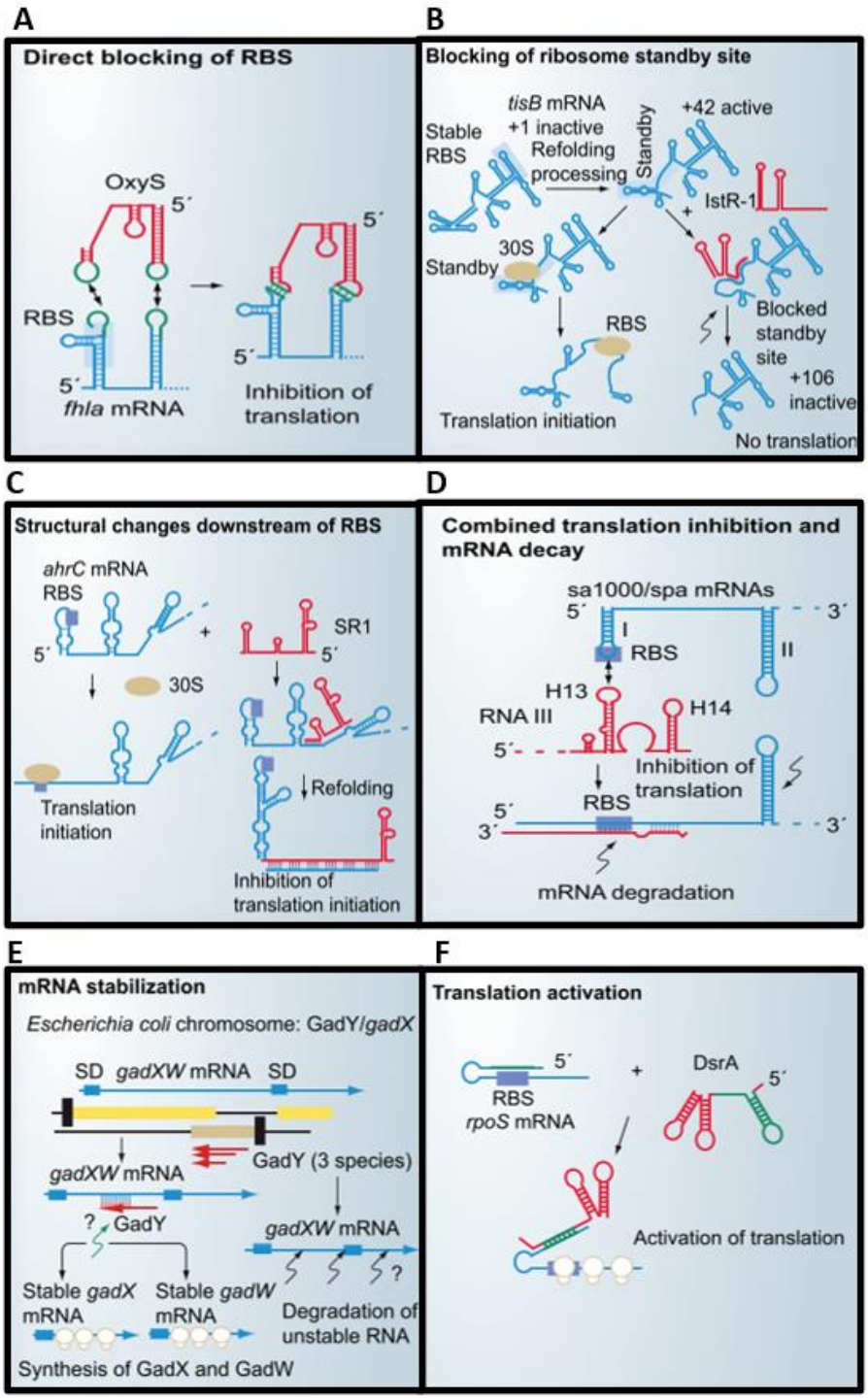


Figure 3.

Overview of regulatory mechanisms employed by small RNAs. Small RNAs are drawn in red, mRNAs in blue. Regions complementary between sRNA and mRNA in green and ribosome binding sites in light blue. Black arrows denote RNase III action. Yellow symbols indicate ribosomes. ? hypothesized. RBS: Ribosome-binding site. Adapted from Brantl, et al. (2009) ***3A.*** *Direct blocking of RBS. OxyS base pairs to the RBS of fhla and inhibits the ribosome from associating with the mRNA*[64,65]*.* ***3B****. Blocking the ribosome standby site. When IstR-1 is absent tisB mRNA forms a stem loop that allows translation. When IstR-1 is present it base pairs with tisB causing a conformational change and closing the ribosome standby site from ribosome access*[66]*.* ***3C****. Structural changes downstream of RBS. When SR1 is present it base pairs to ahrC downstream of the RBS causing a conformational change of ahrC causing the RBS to be blocked, turning off translation*[51]*.* ***3D****. Combined translational inhibition and mRNA decay. When sa1000/spa and RNA III combine, it not only blocks the RBS but also recruits RNases to the mRNA for degradation*[67]*.* ***3E.*** *mRNA stabilization. The gadY sRNA base pairs between the gadX and gadW mRNA and triggers the two mRNAs to be cut and separated making two stable mRNAs*[54]*.* ***3F.*** *Translation activation. rpoS has a secondary structure that prevents translation from occurring, but when dsrA base pairs upstream of RBS and causes a conformational change allowing the RBS to be accessible*[68]*.*

Only four small RNAs have be characterized in *Helicobacter pylori*

Small RNAs have been most extensively studied in the model organism *E. coli,* but research is lacking in non-model bacteria, such as *H. pylori*[69]*.* Four studies have contributed to identifying sRNAs in *H. pylori*. Six sRNAs were identified in 2009 Xiao *et al.* using a bioinformatics approach to identify transcriptional promoters and terminators in intergenic regions of the *H. pylori* genome[70]. Wen *et al.* (2011) found a sRNA, now named *5'ureB-sRNA*, while characterizing *ureAB* in the urease gene cluster[69]. Ta *et al.* (2012) identified three sRNAs while characterizing the operon structure of the cytotoxin associated gene pathogenicity island (cagPAI) and functional promoter assay[71]. In 2010, Sharma and colleagues performed a transcriptome analysis that identified at least 60 sRNAs[45]. Most recently, in April 2020, Du *et al.* found about 160 sRNAs within *H. pylori* using an Illumina Hiseq2000 to construct an sRNA library[46] Even

with the large amount of potential sRNAs identified, only four sRNAs (Table 1) have been fully characterized in *H. pylori*[45,69,72,73].

We are interested in how sRNAs contribute to *H. pylori* gene expression regulation in general, but more specifically how they regulate expression of virulence factors. Two studies, Sharma *et al*. (2010) and Ta *et al*. (2012), identified sRNAs within the clinically important cytotoxin associated gene pathogenicity island (cagPAI) and, to date, only 1 sRNA (*CncR1*) has been fully characterized[73]. Another sRNA (*HPnc4160*) located outside of the cagPAI was found to regulate the *cagA* gene, located within the cagPAI (acting in trans)[74]. Small RNAs encoded in the cagPAI may regulate *H. pylori* virulence factors because of their location in this clinically important region[71].

Table 1. Characterized sRNAs in H. pylori

sRNA Name	Year characterized	Length in nucleotides	Target gene	Function	Citation
5'ureB-sRNA	2012	292	*ureAB*	Down regulates urease production in nearly neutral acidic conditions	Wen *et al.,* 2012
RepG	2013	87	*tlpB*	Down regulates a chemotaxis receptor	Peritzsch *et al.,* 2013
CncR1	2016	213	*cagP*	Down regulates a fimbrial assembly protein for the T4SS.	Vannini *et al.,* 2016
HPnc4160	2020	10	*cagA horB hopE omp14 hofC hpaA*	Down regulates various genes	Kinoshita-Daitoku *et al.,* 2020

Research objectives

As mentioned previously, *H. pylori* strains containing the cagPAI have been linked to a higher incidence of disease[40,75]. The cagPAI is a large genomic region in *H. pylori* (approximately 35-40 kilobases and 2.5% of its entire genome) that encodes two essential virulence factors, the cag-T4SS and CagA[43,76]. The objective of this work was to characterize a putative sRNA, called *HPnc2665,* found within the cagPAI during a functional promoter assay performed by Ta *et al.* in 2012 and later partially sequenced by the Castillo lab (Castillo-Garcia and Castillo, unpublished)[71]. I intended to define the 5' and 3' ends by identifying the promoter consensus sequences and the transcriptional terminator, predict a potential secondary structure, and identify likely targets. Originally, this project was going to include lab work to functionally confirm the promoters, terminators, and targets; however, the COVID-19 pandemic prevented me from including these components. Therefore, I used bioinformatic analyses and another sRNA (*HPnc2620*, also found within the cagPAI) to test my methods. Promoters and terminators fall within a certain distance to the 5' and 3' ends but the ends of *HPnc2665* had not been defined. The transcriptional ends of *HPnc2620* were defined by Sharma *et al.* (2010) during RNA sequencing and 454 pyrosequencing[45]. I identified the promoters, terminators, level of conservation, and targets for *HPnc2620* and *HPnc2665*.

CHAPTER II: METHODS

Alignments of *HPnc2620* and *HPnc2665*

HPnc2620 and HPnc2665, listed in Table 2, were aligned using the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST)[77]. The BLAST algorithm parameters were altered in the following ways: by the organism, either *Helicobacter* excluding *H. pylori* or *H. pylori* only; the program selected was BLASTN rather than megaBLAST; the maximum target sequences was increased to 500 from the standard 100. These parameters were selected to get a unique view of sRNA sequence conservation in the genus *Helicobacter* (without the conservation within *H. pylori* skewing the results) and conversely, to determine whether the sRNAs had high conservation with the *pylori* species. The program BLASTN rather than megaBLAST was selected to ensure that all similar sequences were identified as opposed to only highly similar sequences. Lastly, the target sequences were increased so that all available genomes within each specified group was represented and not only the top 100.

Table 2. Sequences from H. pylori strain G27 used for analyses.

Name	Sequence	Location in G27 genome
HPnc2620	AATATCTGTTGTGTGAAAATTTCAGAGCAGTCATAATTCAAAG AGCAAAAAACTATTTTTTAACCATAAAATATTGTTTCAATCAC TCTTATATCTATTTTTCAAAACCTACAAAAAACGCTTTCATAA ATAGCCCTAAAAACCGATTTAAAAAAGTTTTAATATTA	533269-533435
HPnc2665	GCGCGACAAGCCCATTAGGATCATTGTGGTCTTTCCCGAAAGC ATTAATAAGTTGAGTGATAGGATAATCTAGGTTCATATCTCCT GTGATAAGGTTGGTTACTGCCGCTGCAAGCGTGTTAGAATCTG CTAGGCTAAAAGAGATGCTGTTGCCATTTTCATCTTTTTCATC GCTTTTGGTCGCTAGGTTTTTCACAAGCTCTTTGACAACAGAA ATAGCTGTTTGTTTTTGCTCC	544206-544441
P_*HPnc2620*	TGTTTCTCTTTAGATTGTTCAAATCGTAAAGTTTTATATTAAAAT TATAC	533219-533268
P_*HPnc2665*	GATCTGTTGCTTTATTGTCAAAAAGCCATTGAAATTCACCATT GGTTGATTTGCAAAAAGGCGCTAATCGCGCGACAAGCCCATTA GGATC	544137-544227
T_*HPnc2620*	ACGCTTTCATAAATAGCCCTAAAAACCGATTTAAAAAAGTTTT AATATTA	533386-533435
TI_*HPnc2665*	AAGAGATGCTGTTGCCATTTTCATCTTTTTCATCGCTTTTGGT CGCTAGG	544344-544393
TII_*HPnc2665*	CTTTTTCATCGCTTTTGGTCGCTAGGTTTTTCACAAGCTCTTT GACAACA	544368-544417
TIII_*HPnc2665*	GGTTTTTCACAAGCTCTTTGACAACAGAAATAGCTGTTTGTTT TTGCTCC	544392-544441
TIV_*HPnc2665*	CAGAAATAGCTGTTTGTTTTTGCTCCATTGTTGCATTTGTTTT TTGCACA	544416-544465
TV_*HPnc2665*	CCATTGTTGCATTTGTTTTTTGCACACAAGCCGCCCAAGCAAAAGGATTT	544440-544489
TVI_*HPnc2665*	ACAAGCCGCCCAAGCAAAAGGATTTAATCCTGTATCTGTCCCTAGCTCAA	544465-544514
TVII_*HPnc2665*	TAATCCTGTATCTGTCCCTAGCTCAATCTTGACATACTCCCCACCCATTG	544489-544538
TVIII_*HPnc2665*	ATCTTGACATACTCCCCACCCATTGCGACAATATTCCCATAAGCGCCATA	544514-544563
TIX_*HPnc2665*	GCGACAATATTCCCATAAGCGCCATAATCTTTATCCATATAGACCATAGT	544538-544587

Promoter Identification

Because the transcription start site (TSS) of *HPnc2620* had been identified by Sharma *et al.* (2010), I analyzed the sequence 50 nucleotides (nt) upstream of the TSS for *H. pylori* promoter consensus sequences (Table 2). A 91 nt DNA region identified as a functional promoter by Ta *et al*. (2012) is located upstream of *HPnc2665* and overlaps its putative TSS (SMARTer RACE, Castillo-Garcia and Castillo, unpublished). I analyzed sequence from the *HPnc2665* putative TSS to the 5' end of functional promoter for promoter consensus sequences (Table 2)[71]. The indicated sequences were analyzed using the promoter prediction program Virtual Footprint (http://www.prodoric.de/vfp/vfp_promoter.php) and previously identified *H. pylori* promoter consensus sequences (Table 3)[45,78–85]. Sequences with multiple matches were considered likely promoters (*i.e*. identified with both methods or sharing homology with multiple promoter sequences).

Table 3 Previously identified H. pylori consensus sequences

-10 consensus sequence	-35 consensus sequence	Citation
TGATAA	GTGAGC	Spohn *et al.,* 1997
TAAAAT	TACCCA	Spohn *et al*., 1997
TATAAT	-	Spohn *et al.,* 1997
TATaaT	-	Forsyth *et al.,* 1999
TATAAT	TTAAGC	Vanet *et al.,* 2000
TaAA	cCGAT	Josenhans *et al.,* 2002
tttGCtT	Ggaa	Niehus *et al*., 2004
tTTGCTT	TGGAA	Pereira *et al*., 2006
tttGCtT	GGaA	Sharma *et al*., 2010

Terminator prediction

In bacteria, termination of transcription is directed by two mechanisms, Rho-independent termination, and Rho-dependent termination. *Helicobacter pylori* uses both mechanisms to terminate transcription and, in some cases, uses both a Rho-independent terminator (RIT) and a Rho-dependent terminator (RDT) to terminate transcription for a single gene[86–88]. *HPnc2620* and *HPnc2665* were analyzed for consensus sequences consistent with both terminator types[86]. Rho-independent terminators are characterized by an intrinsic secondary hairpin structure with a minimum free energy (ΔG) <-3.0 kcal/mol, a loop size of 3-10 nt, a stem between 4-15 nt and a thymine-rich region should be within 20 nt of the hairpin structure [86,87]. Rho-dependent terminators can be identified in *H. pylori* by four cytosines followed by a 12 nt spacer and a four to ten thymine stretch at the 3' [86,87,89].

Bacterial transcriptional terminators are typically located within approximately 50 nt of the 3' end of the transcript[87,89]. For *HPnc2620*, 50 nt upstream of the 3' end was analyzed because the ends were previously defined by Sharma *et al.* (2010)[45]. The putative 3' end of *HPnc2620* was identified by SMARTer RACE (Garcia-Castillo and Castillo, unpublished); I analyzed the 50 nt upstream of the SMARTer RACE identified 3' end (Table 2) and staggered 50 nt regions incrementally following the identified 3'end (Figure 4).

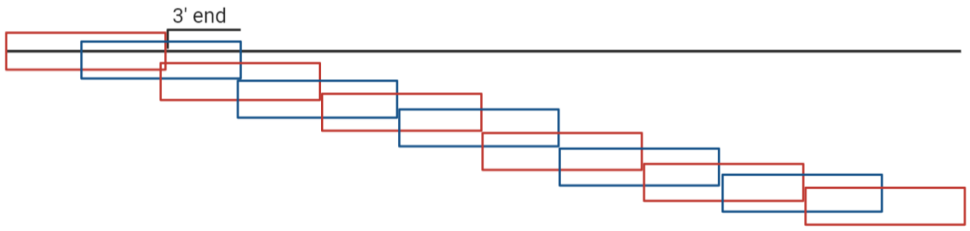


Figure 4. Depiction of HPnc2665 staggered terminator regions. The sequence is represented as a black line and the RACE determined 3' end is labeled. The 50 nt upstream (in the first red box) of 3'end was analyzed and following sequences (blue and red boxes) were staggered to ensure maximum coverage. The sequences overlapped by 25 nt.

Sequences were analyzed for RITs using criteria adapted from Lesnik *et al* (2003), as done previously by Castillo *et al.* (2008)[86,87]. Lesnik and their colleagues in 2003 analyzed over 130 RITs in *E. coli* and found that they shared similar stem-loop structure characteristics such as an 11 nt adenosine-rich region followed by, a variable-length hairpin, a variable-length spacer, and a 12 nt thymine-rich region[87]. Then, in 2008, Castillo *et al.* identified terminators in *H. pylori* using the above mentioned criteria and tested them using a functional terminator assay[86]. Based on their results, Castillo *et al.* (2008) concluded that *H. pylori* RITs vary somewhat from *E. coli* RITs. The characteristics for RITs in *H. pylori* are as follows; the secondary structure has a minimum free energy

(ΔG) <-3.0 kcal/mol, the loop is about 3-10 nt, the stem is between 4-15 nt and a thymine-rich region should be within 20 nt of the hairpin structure[86,87].

The program RNAfold (http://rna.tbi.univie.ac.at//cgi-bin/RNAWebSuite/RNAfold.cgi) was used to predict secondary hairpin structures (represented in a bracket notation) based on thermostability (ΔG); a lower ΔG correlates with a more stable secondary structure[90]. The predicted structures were then compared to the RIT characteristics mentioned above[86,90]. I designed the graphic representation of the secondary structures based on the RNAfold provided bracket notation and Biorender (https://biorender.com/).

Additionally, Castillo *et al.* (2008) determined that *H. pylori* uses both RIT and RDT to halt transcription, so the designated sequences (listed in Table 2) also were evaluated for RDTs using criteria from Castillo *et al.* (2008) and Petersen and Krogh (2003)[86,88]. The criteria used for RDT are as follows: a four cytosine tract, a spacer region of about 12 nt, and a 4-10 thymine tract.

Target prediction

Targets were predicted using the sRNA target prediction program called TargetRNA2 (http://cs.wellesley.edu/~btjaden/TargetRNA2/)[91]. All searches were done using the full sequences of each sRNA (Table 2) and within the *H. pylori* G27 chromosome[92]. TargetRNA2 uses the conservation and accessibility of the sRNA along with the accessibility of the mRNA and the energy of hybridization for the sRNA-mRNA

complex to form to identify targets[91]. TargetRNA2 orders the predicted targets based on the p value calculated and, in order to reduce false positives, targets with a p value less than or equal to 0.02 were reported[48,91,93].

Statistical Analysis

A Mann-Whitney U nonparametric test was conducted in RStudio, on the E-values collected during BLAST searches. P values of less than 0.05 were interpreted as the tested groups being significantly different from one another.

Experimental methods halted due to COVID-19 pandemic

A. Bacterial strains and growth conditions

The bacteria used in this study include *H. pylori* strain G27 and *E. coli* DH5α. All *E. coli* was grown on Luria Burtani (LB) media with 1.5% agar, LB + ampicillin (amp, 100 μg/ml), LB + chloramphenicol (cm, 20 μg/ml), or LB + ampicillin and chloramphenicol. *Escherichia coli* cultures were grown in broth agitated at 220 rpm for 18-24 hours at 37°C or on plates grown at 37 °C. All *H. pylori* was grown on Columbia blood agar + betacyclodextrin and, where needed, supplemented with kanamycin (kan, 15ug/ml) or cm (20ug/ml, see Appendix I). Cultures of *H. pylori* were grown for 48 hrs at 37°C under microaerophilic conditions with a gas mixture of 5-10% O_2, 10% CO_2, and 80-95% N_2.

B. Oligonucleotides

All oligonucleotides used in this study are listed in Table 4.

Table 4. Oligonucleotides

Name	Sequence
HPnc2620-F	GGATAACAAGATACTGAGCACATAGA ATATCTGTTGTGTGAAAATTTC
*HPnc2620*g27-R	CCTTTCGTTTTATTTGATGCTAATATTAAAA CTTTTTTAAATCGGTTTTTAG
HPnc2620-26695-R	CCTTTCGTTTTATTTGATGCTTCTTACAA CTTATCTTGCTTTAAC
pJV300-F	GCATCAAATAAAACGAAAGGCTC
pJV300-R	CTATGTGCTCAGTATCTTGTTATCC
pXG10sf-F	CTCGCTAGCAAAGGAGAAGAAC
pXG10sf-R	CAATGCATGTGCTCAGTATCTC
pXG FlgA-F	GAGATACTGAGCACATGCATTGGTTTTAG GCGTAGAAAAAG
pXG FlgA-R	GTTCTTCTCCTTTGCTAGCGAGTAAGATTT GCGCTTTAGAG
pXGVacA-F	GAGATACTGAGCACATGCATTTGACTATATATTTA TAGCCTTAATCGTAAATG
pXG VacA-R	AGTTCTTCTCCTTTGCTAGCGAAACTATAC CTCATTCCTAAATTG
CatSeqSt	GAAGTATTATGAGGAGGGCG
TnpRbk75	TCAGTAAAGATGCGATTTGC

C. Plasmids used in this study

All plasmids used in this study are listed in Table 5. Plasmid maps are in Appendix II. Plasmids were digested using standard protocols from New England BioLabs (Appendix III); a five µl aliquot of the 20 µl digestion reaction was checked by gel electrophoresis to ensure the plasmid was completely digested. Completely digested samples were then resolved by gel electrophoresis and visualized by ultraviolet exposure. Desired bands were excised using a razor

blade. DNA was isolated from the gel using Nucleospin Gel and PCR Clean-up Kit (Appendix VI).

The inserts for cloning were prepared by polymerase chain reaction (PCR); for a detailed protocol see Appendix V. Following PCR, five µl of insert amplicon were verified by gel electrophoresis. If the insert was the correct size, the remaining sample was purified using Monarch PCR and DNA Cleanup kit (Appendix VI).

The purified linearized plasmid and insert were ligated together using a NEBuilder HiFi Assembly Master Mix protocol (Appendix VII). The complete plasmid was transformed into *E. coli* DH5α and stored at -80°C in 25-40% glycerol solution for future use.

Plasmids were extracted from *E. coli* using QIAprep Miniprep kit (per manufacturer's protocol). *Escherichia coli* was taken from frozen storage and plated on LB media with appropriate antibiotics and grown for 24 hours in standard conditions (listed above). The cells were scrapped and added to the kit-provided P1 buffer. After lysing, washing, and centrifuging the plasmid was ready for use.

Table 5. Plasmids for this study.

Plasmid name	Plasmid backbone	marker	Reference
pJV300	pZE12-luc	Amp	Sittka *et al*., 2007
pXG-10sf	pXG10	Cm	Corcoran *et al*., 2012
pJV2620	pJV300	Amp	This study
pXGVacA	pXG-10sf	Cm	This study
pXGFlgA	pXG-10sf	Cm	This study
pCmut-tnpR1	pNR9589	Amp	Castillo *et al.,* 2008
pCT-PHPnc2620-tnpR1	pCT-tnpR1	Amp	Castillo *et al.,* 2008

D. Promoter assay

Helicobacter pylori was grown on standard media and in standard conditions as above. *Helicobacter pylori* is naturally competent, so cells were transformed by plating cultures on new media, adding 5 µl of pCmut-tnpR1, swirling the cells and plasmid together, and then allowing them to grow for 2 hours in appropriate conditions[71,94]. The bacteria were grown on Cm + CBA plates for 48 hrs for transformant selection. They were then replica plated to media plates with Kan to test the promoters; if the promoters are functional, transformants will not grow on Kan. Total Cm resistant transformants and Kan resistant transformants were counted and used to determine promoter efficiency (Appendix IX).

E. Terminator assay

The terminator assay was also performed in *H. pylori* and all cultures were grown on standard media and conditions and transformed as mentioned above. Once *H. pylori* cells were transformed with pCT-tnpR1, the bacteria were

grown on Cm selective media for 48 hrs in standard conditions. Chloramphenicol resistant transformants were counted and then replica plated onto Kan + CBA to test terminator functionality. Cultures were grown for 48 hrs in standard conditions and transformants were counted. The terminator efficiency was calculated by (total Kan transformants/total Cm transformants) * 100. See Appendix X for more details.

F. GFP plasmid based expression system

The GFP plasmid based expression system was used by Urban and Vogel (2007) and Corcoran *et al.* (2012) to test sRNAs and predicted targets[95,96]. *Escherichia coli* is electrically transformed (detailed protocol in Appendix XI) with pJV2620 and grown on LB + amp agar. Then, *E. coli* + pJV2620 is co-transformed with a pXG containing the appropriate insert. The co-transformed *E. coli* was grown in LB broth + amp and cm overnight and 1 ml of culture was transferred to 50 ml of sterile LB broth with antibiotics and grown until the cell density had an optical density of 0.5. 50 µl of culture was plated, in triplicate, and fluorescence was imaged, and fluorescence intensity was quantified using imageJ.

CHAPTER III: RESULTS

Characterizing *HPnc2665* and *HPnc2620* sRNA transcripts.

I used computer algorithms to identify the promoter and transcriptional terminator consensus sequences to define the 5' and 3' ends for the two sRNAs called *HPnc2620* and *HPnc2665*. *HPnc2665* was found by Ta *et al.* (2012) during a functional promoter assay and later sequenced via SMARTer RACE (Castillo-Garcia and Castillo, unpublished); however, we were unsure as to whether the actual 5' and 3' ends were identified. For this reason, I tested my methods for identifying promoters and terminators on another sRNA, *HPnc2620*; *HPnc2620* was originally identified by Sharma *et al.* (2010) by RNA sequencing and they established the 5' and 3' ends, but not the transcriptional regulatory sequences. I also determined the level of conservation for both sRNAs against sequenced *H. pylori* strains and *Helicobacter*-non-*pylori* strains.

Identification of the *HPnc2665* gene

A functional promoter assay was performed by Ta *et al.* (2012) to determine the transcriptional organization of the cagPAI and found a previously unidentified promoter that was antisense and intergenic to *cagE*. The authors confirmed this result by RT-PCR expression of a transcript downstream of the functional promoter and hypothesized that the transcript was a sRNA because of its location to *cagE* and its small size[71]. The Castillo lab sought to determine the 5' and 3' ends of the sRNA (hereby known as *HPnc2665*, based on the naming convention established by Sharma *et al.* (2010)) by

SMARTer RACE. Shown below in Figure 5 is a depiction of the region identified by Ta *et al.* (2012) and sequenced by the Castillo lab.

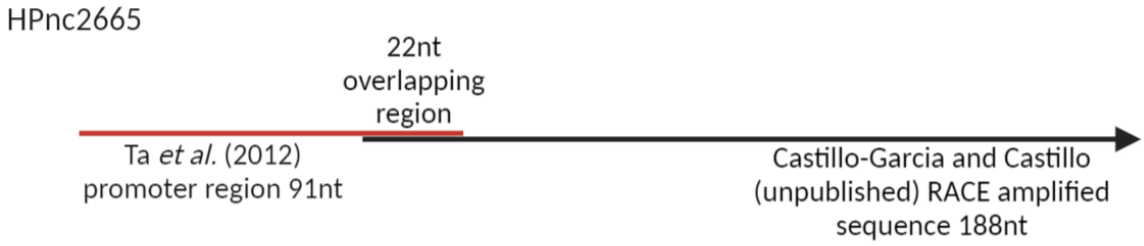


Figure 5. Regions previously identified for HPnc2665. The red line represents the region shown to have a functional promoter in Ta et al (2012). The black arrow represents the SMARTer RACE amplified sequence identified by our research group.

In silico identification of *HPnc2620* and *HPnc2665* promoters

The process of transcription begins when the sigma (σ) factor subunit of RNA polymerase complexes with the DNA upstream of a gene at the regulatory promoter sequences[97]. Generally, the promoter consists of a -10 consensus sequence and the -35 consensus sequence and they are so named because of their distance upstream of the transcription start site (TSS or 5' end). To predict promoter consensus sequences the promoter prediction software Virtual Footprint (http://www.prodoric.de/vfp/vfp_promoter.php) was used in conjunction with previously identified and tested promoter consensus sequences listed in Table 3[45,78,80–84]. For *HPnc2620* and *HPnc2665*, I analyzed sequences P_*HPnc2620* and P_*HPnc2665* (listed in Table 2) based on their location with respect to their known TSS or putative TSS. For *HPnc2620*, the sequence TAAAAT located eight nt upstream of the TSS (determined by counting from the center of the sequence to the TSS) was predicted to be a promoter consensus sequence by Virtual Footprint. Additionally, it shares strong

homology, which I define here as having a four to six nt match, to at least 26 known -10 promoter consensus sequences. As shown in Table 6, the predicted -10 sequence shares 100% homology to at least two *H. pylori* promoters. The distance from the TSS along with the homology to known promoters show that the predicted sequence is a likely -10 consensus sequence for *HPnc2620*. A -35 consensus sequence was not predicted by Virtual Footprint. However, 32 nt from the TSS is the sequence TGTTCA that has a four of six nt match (shown in Table 7) to two other -35 consensus sequences identified in *H. pylori.* Lastly, I used BLAST along with MAFFT (a multiple sequence alignment tool, https://mafft.cbrc.jp/alignment/server/) and Weblogo (http://weblogo.threeplusone.com/) to analyze the level of conservation of the promoter region (TSS to -35) for *HPnc2620*. Figure 6 shows a graphical representation of the multiple sequence alignment (called a sequence logo). The P_*HPnc2620* is highly conserved among *H. pylori*.

For P_*HPnc2665*, the -10 consensus sequence is predicted to be TTGCAA which is located 16 nt upstream of the putative TSS. The -10 consensus sequence was predicted by Virtual Footprint and shares partial homology with seven other -10 promoter consensus sequences in *H. pylori* (shown in Table 8). A -35 consensus sequence was not predicted with Virtual Footprint, but I identified a potential -35 consensus sequence based on distance from the TSS and homology to another *H. pylori* -35 consensus sequence (shown in Table 9). I analyzed the level of conservation of P_*HPnc2665* in *H. pylori* and found that it is highly conserved (Figure 7).

Table 6. Comparison of the -10 consensus sequence for HPnc2620 against other known H. pylori -10 promoter consensus sequences. Bolded letters match HPnc2620.

Associated gene	Sequence	bp matches to -10 consensus	Reference
HPnc2620	TAAAAT	-	This Study
-	**TA**T**AAT**	5	Vanet *et al.*, 2000
flaA	**TT AAA**A	4	Josenhans *et al.*, 2002
fliA	**TT AAA**C	4	Josenhans *et al.*, 2002
HP1051	**TT AAA**A	4	Josenhans *et al.*, 2002
HP0472	**TT AAA**A	4	Josenhans et al., 2002
P1 cagA	**TA**T**AAT**	5	Spohn *et al.,* 1997
cagB P2	**TAAAAT**	**6**	Spohn *et al.,* 1997
vacA	**TAAAA**G	5	Forsyth and Cover, 1999
katA	A**A**T**AAT**	4	Forsyth and Cover, 1999
cheY	**TA**TT**AT**	4	Forsyth and Cover, 1999
ureA	**TA**C**AAT**	5	Forsyth and Cover, 1999
cagA	**TA**T**AAT**	5	Forsyth and Cover, 1999
hspA	**TA**T**AGT**	4	Forsyth and Cover, 1999
hpaA	**TAA**C**AT**	5	Forsyth and Cover, 1999
sodB	**TA**C**AAT**	5	Forsyth and Cover, 1999
MT 54(2)	**T**G**AAAT**	5	McGowan *et al.,* 2003
hypo 97(2)	**TA**T**AAT**	5	McGowan *et al*., 2003
bisC	**TA**G**AAT**	5	McGowan *et al.,* 2003
spaB	**TA**T**AAT**	5	McGowan *et al.,* 2003
hpaA 797(3)	**TAAAAT**	**6**	McGowan *et al.,* 2003
hypo 878(1)	**TA**T**AA**A	4	McGowan *et al*., 2003
hopC	G**AAAAT**	5	McGowan *et al*., 2003
ppk	**TA**T**AAT**	5	McGowan *et al.,* 2003
soj	**TA**G**AAT**	5	McGowan *et al.*, 2003
nuoA	**T**TT**AAT**	4	McGowan *et al*., 2003
murE	**TAAA**T**T**	5	McGowan *et al.,* 2003

Table 7. Comparison of the -35 consensus sequence for HPnc2620 against other known H. pylori -35 promoter consensus sequences. Bolded sequences match HPnc2620.

Associated gene	Sequence	bp matches to -35 consensus	Reference
HPnc2620	TGTCCA	-	This study
cagB P2	T**A**C**CCA**	4	Mobley et al., 2001
sodB	T**TA**C**CA**	4	Mobley et al., 2001

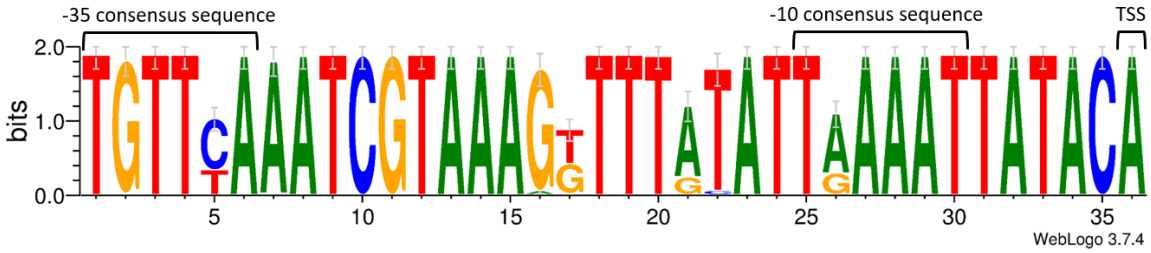


Figure 6. P_HPnc2620 is highly conserved among *H. pylori. The frequency of bases at each position is shown as the height of the letters. The level of conservation is represented by the total height of stacked letters in bits. 215 sequences were used to build this figure.*

Table 8. Comparison of the -10 consensus sequence for HPnc2665 against other known H. pylori -10 promoter consensus sequences. Bolded letters match HPnc2665

Associated gene	Sequence	bp matches to -10 consensus	Reference
HPnc2665	TTGCAA	-	This Study
HP0115	T**TTGC**TT	4	Neihus *et al.,* 2004
HP0367	T**TTGC**TT	4	Neihus *et al.,* 2004
HP0870	T**TTGC**TT	4	Neihus *et al.,* 2004
HP1076	**TTGC**GT	4	Neihus *et al.,* 2004
HP1154	T**TTGC**TT	4	Neihus *et al.*, 2004
HP1120	T**TTGC**TT	4	Neihus *et al*., 2004
HP1233	T**TTGC**TT	4	Neihus *et al.,* 2004

Table 9. Comparison of the -35 consensus sequence for HPnc2665 against other known H. pylori -35 promoter consensus sequences. Bolded letters match Hpnc2665

Associated gene	Sequence	bp matches to -35 *HPnc2665*	Reference
HPnc2665	GTCAAA	-	This Study
hypo 97 (1)	**GTCAAA**	**6**	McGowan *et al.,* 2003

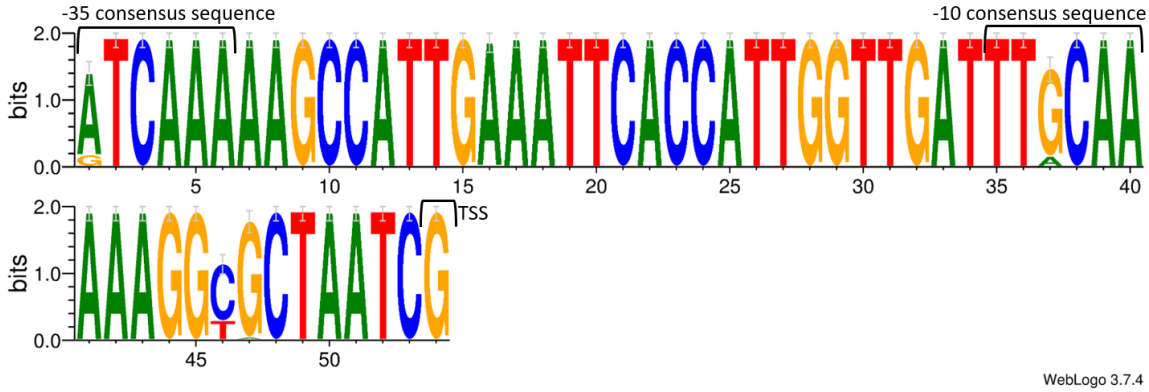


Figure 7. P_HPnc2665 is highly conserved among H. pylori. the frequency of bases at each position is shown as the height of the letters. The level of conservation is represented by the total height of stacked letters in bits. 215 sequences were used in the creation of this figure.

In silico **identification of *HPnc2620* and *HPnc2665* transcriptional terminators**

Transcriptional terminators are typically located within 50 nt of the transcripts 3' end, so I sought to identify the *HPnc2665* terminator to help locate the 3' end. To test my method for identifying transcriptional terminators, I again used *HPnc2620*, for which the 3' end is known. As shown by Castillo *et al* (2008), *H. pylori* uses either RIT or RDT and, in some cases, both RIT and RDT to terminate transcription of a gene; both sRNAs were analyzed for consensus sequences consistent with both terminator types[86]. Briefly, RITs are intrinsically formed secondary hairpin structures followed by a short stretch of thymine nucleotides, that disrupt RNA polymerase; therefore, I used these characteristics to identify potential RITs[86,87]. More specifically, *H. pylori* RITs consist of a secondary structure with a minimum free energy (ΔG) <-3.0 kcal/mol, a loop size of 3-10 nt, a stem between 4-15 nt and a thymine-rich region within 20 nt of the hairpin

structure[86,87]. RDTs rely on the Rho protein to terminate transcription and can be identified in *H. pylori* by four cytosines followed by a 12 nt spacer and a four to ten thymine stretch[86,87,89].

Analysis of the 50 nt upstream of the 3' end for *HPnc2620* revealed a RIT seven nt from the 3' end (from the end of the thymine stretch, figure 8). The calculated ΔG was -2.00 kcal/mol which is slightly higher than the identified average of -3.00 kcal/mol for RIT stem-loops reported in previous literature, but the secondary structure matched the other criteria for a *H. pylori* RIT (stem-loop and ΔG determined using http://rna.tbi.univie.ac.at//cgi-bin/RNAWebSuite/RNAfold.cgi) [86,87]. The characteristics of the hairpin are as follows; the loop size is 14 nt, the stem is 10 base pairs (bp) with a four nt bulge, and there is a stretch of four thymines directly following the hairpin. I used visual inspection to search for RDT characteristics. I observed an imperfect RDT match, three cytosines followed by three thymines, ten nucleotides downstream (Figure 8).

Analysis of the sequence windows described above (Table 2, Figure 4) revealed a RIT within the sequence T_*HPnc2665*_III (Table 2) for *HPnc2665* that is located 67 nt downstream from the cloned 3' end; the dot-bracket representation is shown in Figure 9A and the stem-loop structure in Figure 9B. RIT_*HPnc2665* has a ΔG= -8.40 kcal/mol, loop size of 11 nt, a stem length of 15 bp with a three bulges (two of which are seven nt and the other is one nt) and a six thymine stretch directly following the hairpin structure. No RDT were identified for *HPnc2665*.

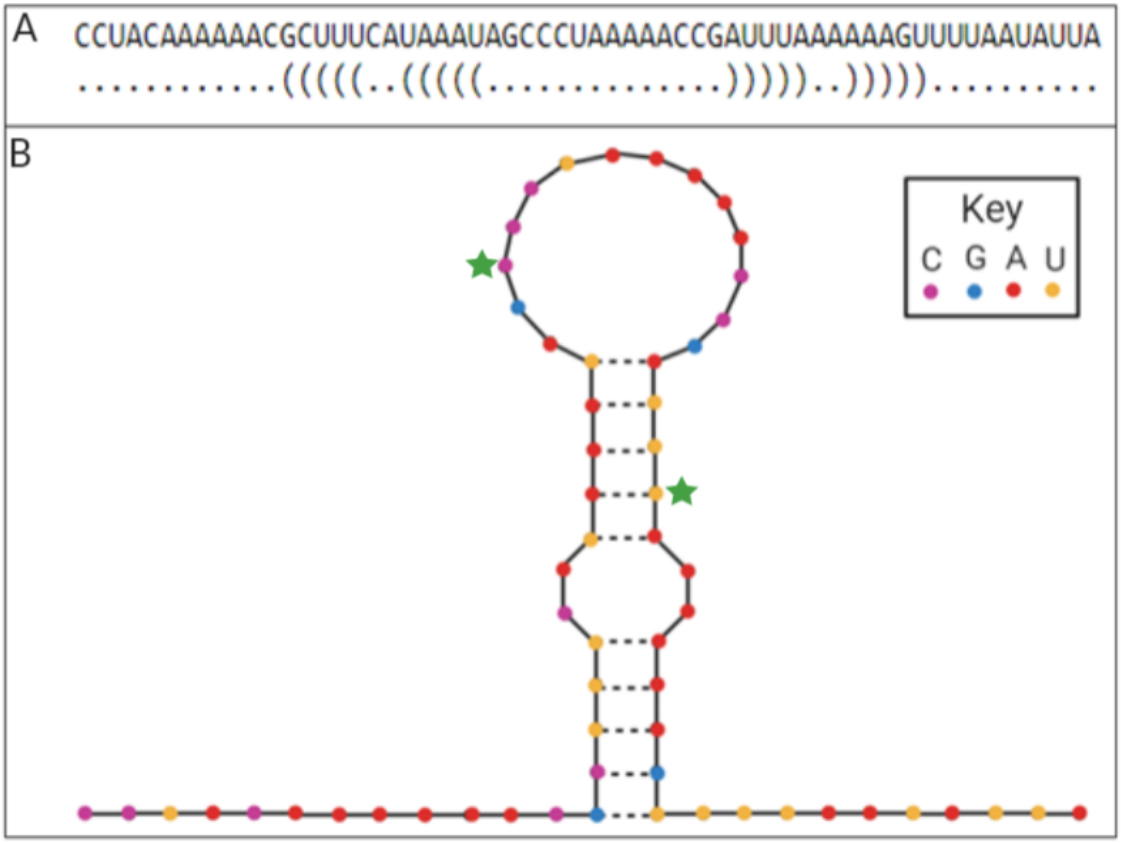


Figure 8. HPnc2620 transcriptional terminator. **8A.** Dot-bracket representation of RIT_*HPnc2620*. Brackets correspond to base pairs while dots represent unpaired nucleotides. **8A**. Graphic representation of RIT structure. Green stars show the RDT. Dots represent nucleotides, dashed lines represent hydrogen bonding, and solid lines represent phosphodiester bonds.

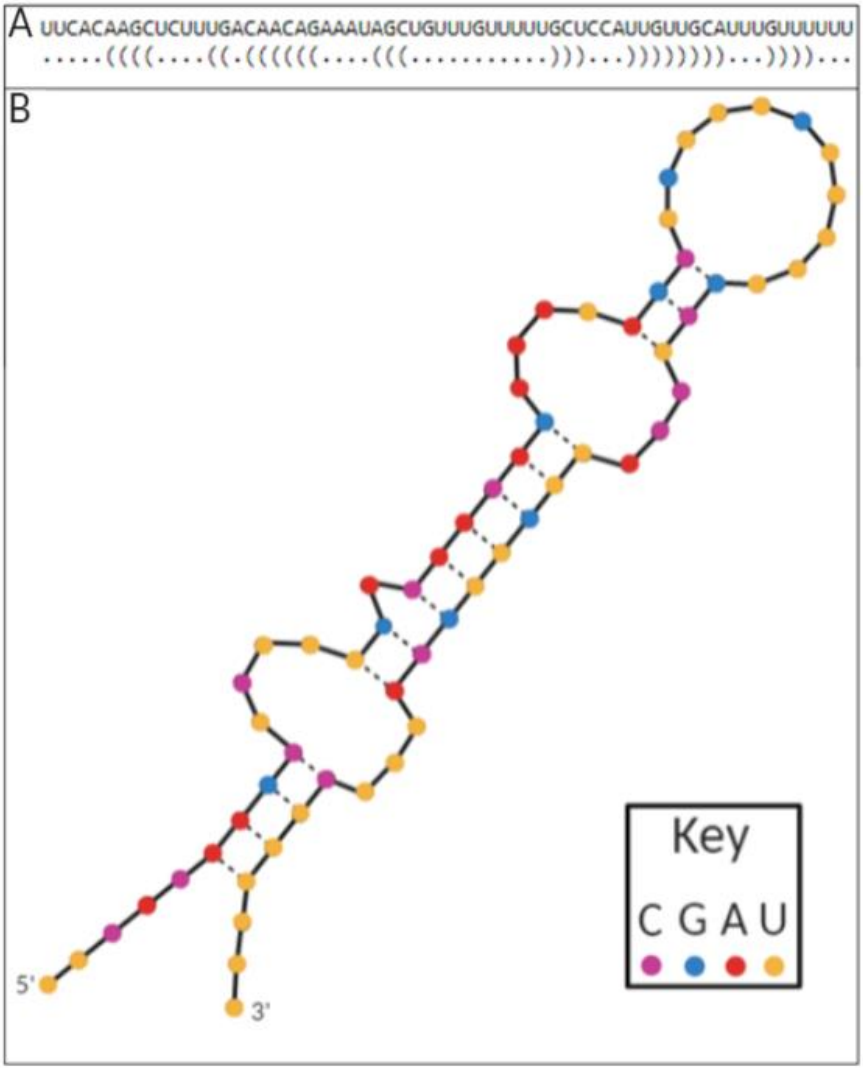


Figure 9. HPnc2665 transcriptional terminator. ***9A****. Dot-Bracket representation of T_HPnc2665 structure. Brackets correspond to base pairs while dots represent unpaired nucleotides.* ***9B****. Graphic representation of T_HPnc2665 structure. Dots represent nucleotides, dashed lines represent hydrogen bonding, and solid lines represent phosphodiester bonds.*

HPnc2620 **and** ***HPnc2665*** **are highly conserved among** ***Helicobacter pylori***

To learn about the potential significance of the sRNAs in *H. pylori* and further delineate the 5’ and 3’ ends of *HPnc2665*, I analyzed the identified gene sequences and a 100 bp non-gene sequence by BLAST against available *H. pylori* sequences

(https://blast.ncbi.nlm.nih.gov/Blast.cgi). I hypothesized that the gene sequences would exhibit higher conservation than non-gene sequences.

Alignment of *HPnc2620* and *HPnc2665* to *H. pylori* strains showed that they were highly conserved. For *HPnc2620*, the median values for query cover (proportion of sRNA sequence that matches the sequence), percent identity (proportion of an aligned sequence that matches the sRNA sequence), and E-values (statistical likelihood that the match was random chance) were 1.00, 1.00, and 9e-70, respectively. *HPnc2665* had query cover, percent identity, and E-value median values of 0.98, 0.97, and 1.1e-84, respectively. A Mann-Whitney U nonparametric test was performed on the E-values for each sRNA against the E-values of a 100 bp non-gene sequence control region to determine whether the sRNAs have significant conservation. For *HPnc2620* and *HPnc2665*, the p values were less than 2.2e-16. There was a significant difference in the E-values between each sequence, *HPnc2620*, *HPnc2665* and the 100 bp non-gene sequence control. The control had E-values closer to zero than both sRNAs and *HPnc2665* had lower E-values than *HPnc2620.* as shown in figure 10. These results indicate that both sRNAs are more conserved than a non-gene sequence and HPnc2665 is more highly conserved than *HPnc2620*.

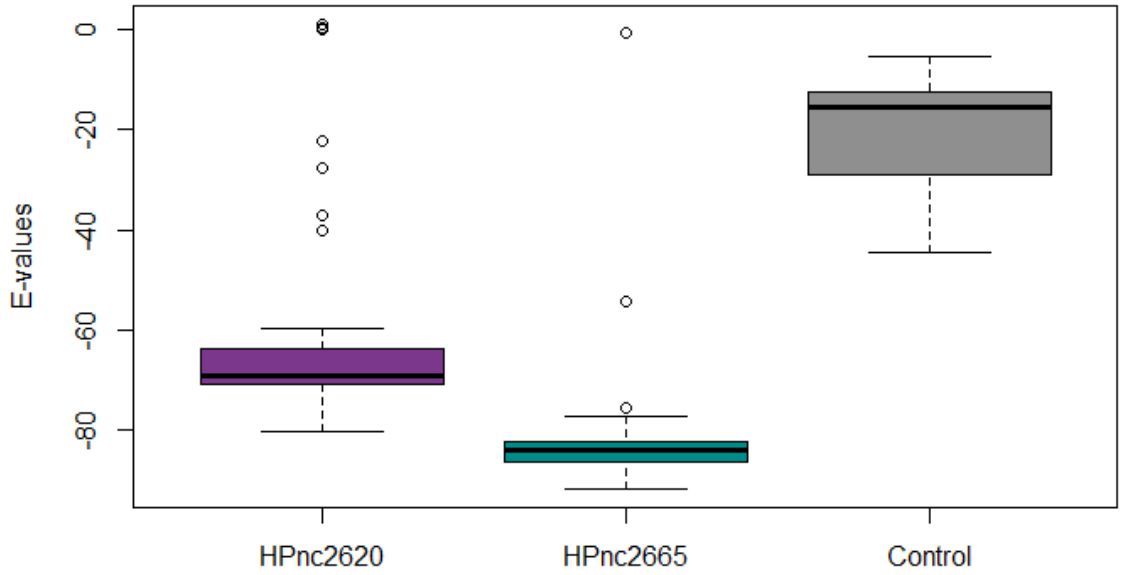


Figure 10. HPnc2620 and HPnc2665 are highly conserved among Helicobacter pylori. Across the X axis is the sequence names. Across the Y axis are the log10 transformed E-values. The open circles represent outliers.

***HPnc2620* and *HPnc2665* are not conserved among other *Helicobacter* species**

The high level of conservation of both sRNAs in *H. pylori* brought up the question: are *HPnc2620* and *HPnc2665* also highly conserved in other *Helicobacter* species? Alignment of *HPnc2620* and *HPnc2665* to non-*pylori Helicobacter* species showed that they were not well conserved. For *HPnc2620*, the median values for query cover was 0.35, percent identity was 0.92, and E-value was 1.5. *HPnc2665* had median values of 0.3 for query cover, 0.92 for percent identity, and 1.7 for E-value.

A Mann-Whitney U nonparametric test was performed on the E-values for each sRNA against the E-values of a 100 bp non-gene sequence control region to determine whether the sRNAs have significant conservation. For *HPnc2620* and *HPnc2665*, the p values were 0.76 and 0.46, respectively. Both the p values and the boxplot (figure 11) indicates that *HPnc2620* and *HPnc2665* are not conserved in non-*pylori Helicobacter*.

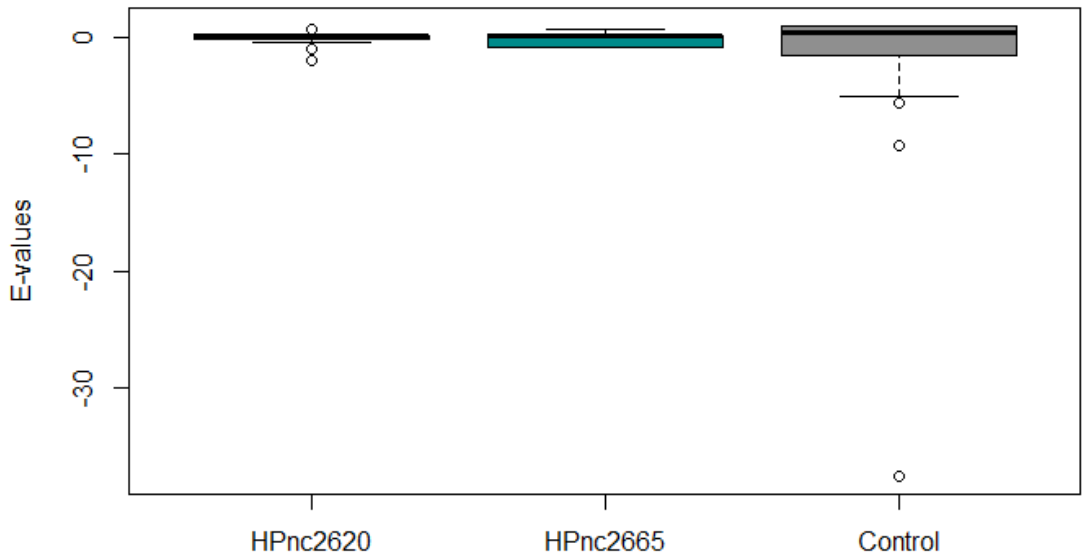


Figure 11 HPnc2620 and HPnc2665 are not highly conserve among non-pylori Helicobacter. Across the X axis is the sequence names. Across the Y axis are the log10 transformed E-values. The open circles represent outliers.

Target prediction and testing for *HPnc2620* and *HPnc2665*

With the 5' and 3'ends characterized by way of promoter and terminator consensus sequence identification, I moved forward to identify target genes regulated by *HPnc2620* and *HPnc2665*. Targets for *HPnc2620* and *HPnc2665* were predicted by analyzing their gene sequences (*HPnc2620* and *HPnc2665* in Table 2) with the computer program TargetRNA2; this program uses the sRNA conservation, accessibility of the sRNA against a chosen organism's mRNA, and the energy of hybridization to predict mRNA targets[91]. TargetRNA2 provides predicted targets with a p value of less than 0.05; here I only report targets with a p value of less than or equal to 0.02.

Of the ten mRNA targets reported for *HPnc2620* in Table 5, two are known virulence factors, a flagellar biosynthesis gene (*flgA*) and *vacA* (bolded in Table 10). The p values and ΔG for the likelihood and stability of the mRNA and sRNA complementary

base pairing for the *flgA* transcript are 0.014 and -10.91 kcal/mol and for *vacA* are 0.02 and -10.06 kcal/mol. TargetRNA2 also predicts where the sRNA and mRNA will complex and the *flgA* transcript is predicted to base pair with *HPnc2620* at nucleotides 112-98 (Figure 12), the region between stem loops two and three, while the *vacA* transcript base pairs at nucleotides 15-4 within the first stem loop shown in Figure 12.

Table 10. Potential targets for HPnc2620. Virulence targets are bolded.

mRNA	mRNA function	Energy of hybridization (kcal/mol)	p value	sRNA Binding location
Type 1 restriction enzyme S	endonuclease	-13.6	0.002	132-119
Fumarate hydratase	catalyzes fumarate into to L-malate[98]	-12.14	0.007	97-84
Hypothetical protein *HPG27_900*	Unknown function	-12.08	0.007	40-26
30S ribosomal protein S19	One of the proteins that make up the 30S ribosomal subunit[99]	-11.28	0.012	138-129
Flagellar basal body P-ring biosynthesis protein f*lgA*	periplasmic flagellar protein that chaperones P-ring formation[100]	-10.91	0.014	112-98
pH-dependent sodium/proton antiporter nhaA	aids in maintaining homeostasis in high salinity, lithium toxicity, and alkaline pH environments[101]	-10.52	0.018	33-19
Signal peptidase I *lepB*	cleaves the hydrophobic N-terminal sequence from their natural substrates[102]	-10.4	0.019	25-10
Vacuolating cytotoxin A	binds host cells and inserts into the membrane forming anion-selective channels to influx Cl-ions[103]	-10.06	0.02	15-4
50S ribosomal protein L2 *rplB*	50S ribosomal subunit[104]	-10.05	0.02	126-116
Proline peptidase	cleaves amino acids N-terminal to a proline residue[105]	-10.01	0.02	128-116

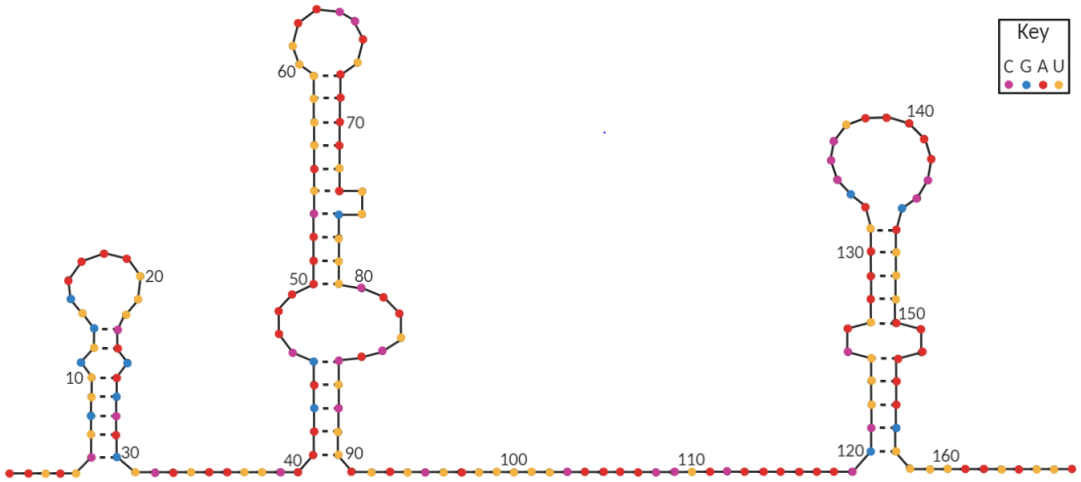


Figure 12. Predicted secondary structure for HPnc2620. The numbers show the nucleotide location in the sRNA. The colors represent the nucleotides, purple for cytosine, blue for guanine, red for adenosine, and yellow for uracil.

TargetRNA2 predicted 53 mRNA targets for *HPnc2665*; listed in Table 11 are 33 mRNA targets with a p value of less than or equal to 0.02. Five of the reported targets bolded in Table 11 are known virulence factors. Methyl-accepting chemotaxis protein (TlpB) has a ΔG of -13.74 kcal/mol, p value is 0.002, and complementary base pairs to *HPnc2665* between the first two stem loops (nts 42-29) shown in Figure 13. Penicillin-binding protein 1A has a ΔG and p value of -13.71 kcal/mol and 0.002, respectively, and binds to *HPnc2665* at the fifth stem loop (Figure 13). Urease subunit beta (*ureB*) transcript complexes with *HPnc2665* within the loop of the fourth stem loop structure (Figure 13), has a p value of 0.009, and a ΔG= -11.63 kcal/mol. *CagF* has a p value of 0.009 and a ΔG= -10.97 kcal/mol and complementary base pairs to *HPnc2665* within the second stem loop (Figure 13) at nucleotides 102 to 93. Lastly, Flagellar basal body protein (*fliL*) complexes with *HPnc2665* between stem loops one and two (nt 41-26) and has a p value and ΔG of 0.02 and -10.16 kcal/mol, respectively.

Table 11. Potential targets for HPnc2665. Virulence factors are bolded.

mRNA	target function	Energy of hybridization (kcal/mol)	P value	sRNA Binding location
Cell division inhibitor (*minD*)	adenosine 5' triphosphate-dependent membrane-binding protein that controls the position of the cell division septum[106]	-14.36	0.001	161-144
coproporphyrinogen III oxidase (*HPG27_627*)	catalyzes the oxidative decarboxylation of coprogen to protogen[107]	-13.97	0.002	180-163
methyl-accepting chemotaxis protein (*tlpB*)	membrane spanning chemoreceptor protein[24]	-13.74	0.002	42-29
penicillin-binding protein 1A (*HPG27_557*)	Component of beta-lactamase	-13.71	0.002	207-191
cytochrome c biogenesis protein (*HPG27_244*)	polysaccharide production[108]	-13.64	0.002	123-110
ExsB trans-regulatory protein (*HPG27_600*)	An outer membrane lipoprotein that allows attachment to bacterial membranes[109]	-13.22	0.002	125-112
outer membrane protein (*hopZ*)	adhesion to gastric epithelial cells[110]	-13.12	0.003	60-48
hydrogenase expression/formation protein (*hypD*)	synthesis of cofactor NiFe(CN)2CO[111]	-12.72	0.003	173-161
outer membrane protein (*horC*)	transporter of the resistance-nodulation-cell division superfamily[112]	-12.62	0.005	89-78
Hypothetical protein (HPG27_720)	Unknown function	-12.59	0.005	175-155

Hypothetical protein (*HPG27_817*)	Unknown function	-12.42	0.005	102-93
succinyl-CoA-transferase subunit A (*HPG27_647*)	forms acetoacetyl CoA from succinyl CoA and acetoacetate[113]	-12.21	0.006	170-151
Hypothetical protein (*HPG27_737*)	Unknown function	-12.19	0.007	90-76
ferrochelatase (*hemH*)	Iron metabolism[114]	-12.16	0.007	176-163
hypothetical protein (*HPG27_15*)	Unknown function	-12.11	0.007	153-139
Hypothetical protein (*HPG27_972*)	Unknown function	-11.74	0.007	71-57
CDP-diacylglycerol pyrophosphatase (*HPG27_825*)	regulation of phospholipid metabolism by inositol[115]	-11.74	0.009	191-176
Urease subunit beta (*ureB*)	catalytic subunit for urease[69]	-11.63	0.009	169-155
hypothetical protein (*HPG27_583*)	unknown function	-11.57	0.009	181-167
adenylate kinase (*adk*)	catalyzes the formation of two ADP molecules from ATP and AMP[116]	-11.54	0.01	119-101
hypothetical protein (*HPG27_773*)	Unknown function	-11.34	0.01	195-177
hypothetical protein (*HPG27_341*)	Unknown function	-11.18	0.011	228-218
ABC transporter permease (*HPG27_572*)	Unknown function	-11	0.012	200-182
cag pathogenicity island protein F (*cagF*)	accessory protein that interacts with CagA[117]	-10.97	0.014	102-93
30S ribosomal protein S11 (*rpsK*)	Ribosomal protein[118]	-10.92	0.014	72-62
glutamine synthetase (*glnA*)	enzyme for nitrogen assimilation[119]	-10.92	0.014	87-75

cytochrome c-type biogensis protein (*HPG27_1019*)	polysaccharide production[108]	-10.82	0.015	102-88
typeI R-M system specificity subunit (*HPG27_1455*)	subunit of restriction-modification system enzymes[120]	-10.81	0.015	209-199
thioredoxin (*HPG27_1381*)	thio-oxireductase enzyme that controls redox homeostasis[121]	-10.67	0.016	77-57
Flagellar basal body-associated protein (*FliL*)	Single transmembrane protein with a large periplasmic region and associates with the flagellar basal body[122]	-10.16	0.02	41-26
Hypothetical protein (*HPG27_334*)	Unknown function	-10.14	0.02	170-155
elongation factor Ts (*tsf*)	protein synthesis	-10.09	0.02	173-156
GMP synthase (*guaA*)	catalyzes glutamine of GMP[123]	-10.05	0.02	177-167

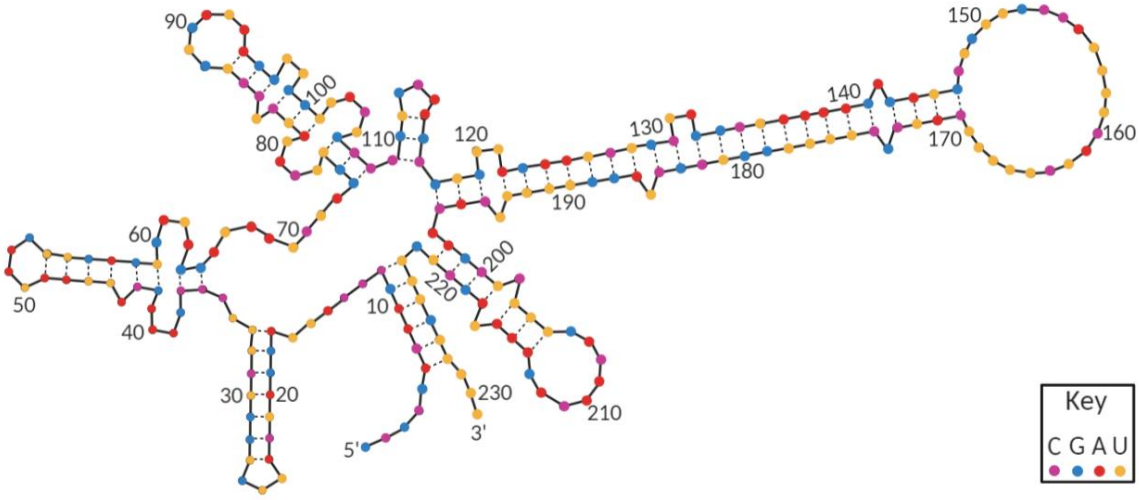


Figure 13. predicted secondary structure for HPnc2665. The numbers show the nucleotide location in the sRNA. The colors represent the nucleotides, purple for cytosine, blue for guanine, red for adenosine, and yellow for uracil.

CHAPTER IV: DISCUSSION

In silico analysis for promoter consensus sequences successful for *H. pylori* sRNAs

HPnc2620 was identified by Sharma *et al.* (2010) by RNA sequencing which allowed them to define the 5' and 3' ends[45]. I used *HPnc2620* as a test to confirm that predictions for regulatory elements were positionally consistent with respect to the 5' and 3' ends of the transcript. The rationale for this was that promoters and terminators are found within a reliable nucleotide distance from the 5' and 3' ends of transcripts, respectively. Promoter consensus sequences are approximately 10 and 35 nt upstream of the TSS, the first nucleotide at the 5' end of a transcript, and transcriptional terminators are about 50 nt upstream of the 3' end.

The promoter consensus sequence for *HPnc2620* is TGTCCA- 23 nt -TAAAAT. The -10 consensus sequence, TAAAAT, was predicted by Virtual Footprint and shared perfect homology to at least two promoters and strong homology to 25 other *H. pylori* promoters (shown in Table 6). My predicted -10 sequence is 8 nt from the *HPnc2620* TSS, increasing the likelihood that it is the promoter sequence and is consistent with Sharma and colleagues (2010) identification of the 5' end[61]. The predicted -35 consensus sequence (TGTCCA) was not predicted by Virtual Footprint; it shares a four of six bp match to two *H. pylori* -35 promoter consensus sequences and is an appropriate distance (31 nt) from the TSS. The low level of conservation observed in the predicted -35 sequence is not uncommon; in *Mycobacterium,* Bashyam *et al.* (1996) found no apparent conserved -35 sequence when they aligned the 24 mycobacterial promoters[124]. Further, Bashyam *et al.* tested the impact of losing the -35 consensus

sequence and found that it did not eliminate transcription but when the -10 consensus sequence was removed the gene was not transcribed[124]. This could be similar in *H. pylori,* with the -10 consensus sequence being integral to transcription and the -35 playing a lesser role. Previous research to identify -35 consensus sequences in *H. pylori* have been unsuccessful[78–80]. Spohn and colleagues were characterizing regulatory elements for *cagA* and found that it contained a -10 sequence like the σ^{70} *E. coli* recognized promoter (TATAAT) but could not identify a -35 consensus sequence[80]. Similarly, when Forsyth *et al.* (1999) analyzed 11 different genes in *H. pylori* they were able to find a -10 sequence among all of them but could not find any conservation among them in the -35 region[78]. McGowan's research group found similar results when they evaluated 28 genes for promoters; they were unable to find any apparent conservation in the -35 region[79]. These results could indicate that, like *Mycobacterium,* the -35 consensus sequence in *H. pylori* is not vital for transcription and so conservation is lacking. I propose the above-mentioned sequence (TGTCCA- 23 nt -TAAAAT) is the promoter for *HPnc2620*. The promoter could be experimentally tested by first determining the efficiency of the promoter as is, then mutating it separately in two separate strains and testing its functionality. This experiment would test the prediction that the -35 consensus sequence is not necessary for *HPnc2620.*

HPnc2665 was originally identified as a putative sRNA within the cagPAI in 2012 by Ta *et al.* Since then, the Castillo lab worked to define the 5' and 3' ends using SMARTer RACE. My work predicts the promoter for *HPnc2665* is GTCAAA- 27 nt -TTGCAA. The -10 consensus sequence (TTGCAA) is within 15 nt of the 5' end identified

by SMARTer RACE, it shared partial conservation with at least seven other -10 consensus sequences in *H. pylori* and was predicted by Virtual Footprint, making it very likely the *HPnc2665* promoter. The -35 sequence (GTCAAA) shares complete homology to one *H. pylori* -35 sequence and is 45 nt from the TSS but was not predicted by Virtual Footprint. The fact that the predicted promoter is also within in the functional promoter sequence identified by Ta et al (2012) strongly supports it as the *HPnc2665* promoter. Before the COVID-19 pandemic I planned to test predicted promoters using the promoter assay established in Ta *et al.* (2012). Essentially, the predicted promoter is cloned into a plasmid upstream of a promoterless gene that, when expressed, converts *H. pylori* cells from kanamycin resistant to kanamycin sensitive.

In silico analysis for transcriptional terminators successful for *H. pylori* sRNAs

Transcriptional terminators are located withing 50 nt of the transcript 3' end. I used previously identified characteristics for RIT and RDT to identify terminators (listed in methods). I began by trying to identify the *HPnc2620* transcriptional terminator to test this method of identification. Sharma *et al.* (2010) were able to define the 3' end for *HPnc2620*, so I used it to test if predicted terminators were within the correct distance to the end.

I identified a potential RIT and RDT for *HPnc2620*; however, these fell short of the above-mentioned criteria indicating that they may be weak terminators of transcription. The RIT (Figure 8B) has a ΔG of -2.00 kcal/mol, a loop size of 14 nt, a stem

length of 10 bp with a 4 nt bulge, and a 4 thymine stretch within 20 nt of the stem. The ΔG and loop do not fall within the established characteristics which led me to question if the 3' end had been correctly identified. During reverse transcription, the cDNA may become degraded for several reasons, such as poor RNA quality, low RNA abundance, low sample purity, and when the RNA is sequenced the 3' end would be farther upstream than the true 3' end[125]. To check whether the true 3' end was farther downstream, I analyzed 50 nt regions up to 200 nt downstream to address the chance that the 3' end was incorrectly defined and was unable to identify any terminators. Therefore, I predict that the RIT mentioned above is a weak terminator. Similarly, I identified a RDT overlapping the RIT; however, it was also a weak terminator (Figure 8B). There were three cytosines followed by three thymines ten nucleotides downstream leading me to the conclusion that it may be a weak RDT. The presence of weak RIT and RDT indicates *HPnc2620* transcription is terminated with both terminators and this is supported by previous research.

In 2008, a study conducted by Castillo and colleagues showed that *H. pylori* uses both forms of termination for a single gene and both terminators worked independently of each other[86]. It is obvious that sometimes *H. pylori* relies on RDTs and RITs to terminate transcription for a gene. This could be the case for *HPnc2620*, if the RIT is unable to form (due to its weak pairing) then the RDT takes over and terminates transcription. Two weak terminators working in concert to halt transcription has been observed in *Bacillus subtilis.* Many transcripts in *B. subtilis* have weak RITs and when NusA (a termination cofactor for RDT) was knocked out, the weak RIT alone was

ineffective in terminating transcription[126]. *Helicobacter pylori* may have similar methods for terminating expression to *B. subtilis* and require both RDT and RIT to effectively terminate transcription. It is possible that because both terminators for *HPnc2620* overlap they work interchangeably instead of in concert. One way to identify how the terminators work to halt expression would be to mutate the stem of the RIT and test the termination efficiency of the RDT and conversely mutate the RDT and test the RIT[86]. If they work in concert, then when one terminator is broken termination efficiency would be low or fail completely.

The predicted terminator for *HPnc2665* is shown in Figure 9 within the TIII_*HPnc2665* sequence (Table 2) and is downstream of the experimentally identified 3' end. Our lab previously predicted *HPnc2665* to be about 188 nt, a typical length for sRNAs; however, I predict that *HPnc2665* is 236 nt based on the location of the RIT. The discrepancy in length from the SMARTer RACE reaction to my current terminator prediction may be due to the reverse transcription reaction. Cloned transcripts may be shorter at the 3'end because a 5'/3' RACE approach does not guarantee to clone the full length of the 3' transcript[127,128]. There are other reasons the RACE reaction may have produced a short transcript, for instance if the RNA was of low quality. Testing the functionality of the RIT would be an ideal way to determine the 3'end. If the RIT terminates transcription, then the previously predicted 3' end was incorrect.

HPnc2620 and _HPnc2665_ are highly conserved among _H. pylori,_ but not conserved among non-_pylori Helicobacter_ species.

With the ends defined based on the regulatory sequences, next I analyzed whether *HPnc2620* and *HPnc2665* were conserved among *H. pylori* and non-*pylori Helicobacter.* I was interested to determine whether the level of conservation would be different between the sRNAs because *HPnc2620* is intergenic to *cag13* and *cag14,* while *HPnc2665* is antisense to *cag23* (Figure 14). I predicted *HPnc2665* would have higher conservation than *HPnc2620* because of its location within a gene. To address these questions, I used BLAST to check conservation and a Mann-Whitney U nonparametric test to determine statistically significance among groups. My groups included both sRNAs and a noncoding 100 bp sequence as a control.

In *H. pylori* all groups were significantly different (Figure 10), with *HPnc2665* being the most conserved (median query cover= 1.1e-84), *HPnc2620* being highly conserved (median query cover= 9e-70), but less than *HPnc2665*, and the noncoding sequence had the lowest conservation (median query cover= 0.62). Groups were not significantly different (p values > 0.05) when conservation was analyzed in non-*pylori Helicobacter* (Figure 11). The median E-values for *HPnc2620*, *HPnc2665*, and non-gene control were 1.5, 1.7, and 2.7, respectively. This low level of conservation is not overly surprising when one keeps in mind that only *H. pylori* contains the cagPAI which is where all three sequences were located. Based on a PubMed search performed on July 4, 2020, only one other species of *Helicobacter* encodes a secretion system, but it is a type VI secretion rather than the cag-T4SS[129]. Additionally, to determine whether the

sRNAs had any conservation in other species, I performed a BLAST search among all genomes except *H. pylori* and BLAST was unable to identify significant similarities.

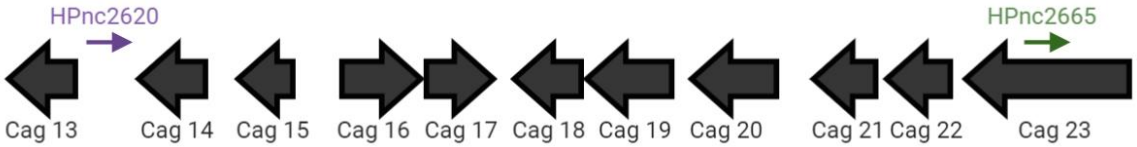


Figure 14. Schematic of HPnc2620 and HPnc2665 locations in the cagPAI. Arrows denote genes.

HPnc2620 **may control three virulence genes**

Next, I used TargetRNA2 to predict targets *HPnc2620* regulate. All targets with a p value less than or equal to 0.02 are listed in Table 10. The first target is a restriction enzyme which has yet to be characterized in *H. pylori*; however, in general, a restriction enzyme recognizes a specific nucleotide sequence and cuts it[130]. The main function for restriction enzymes is to defend the bacterium from bacteriophage invasion[130]. Fumarate hydratase, *lepB*, and proline peptidase are all involved with making or breaking proteins within the cell[98]. Two targets are subunits of the ribosome, indicating *HPnc2620* may play a role in controlling protein synthesis. The five above mentioned targets could indicate that *HPnc2620* plays a role in maintaining homeostasis of the cell. One predicted target has not been characterized and its function is unknown. Testing the regulation of a hypothetical gene could be interesting and complex because the first task would be to confirm that the gene encodes a transcript and a protein. Once the gene had a confirmed function, as opposed to a degenerate gene, regulation by *HPnc2620* would need to be tested. Lastly, two targets are virulence factors; *vacA* is

associated with disease causing strains and plays a role in chronic inflammation of the gastric lining and *flgA* is a flagellar protein that chaperones the p-ring formation[100,103]. Both targets are important in human health, without motility *H. pylori* would have greatly reduced chances for colonization and *vacA* is strongly associated with severe disease phenotypes. The results here indicate that *HPnc2620* likely controls various cellular functions from maintaining cell homeostasis to key virulence factors and, in all cases, acts as a *trans*-sRNA. Interestingly, a cagPAI gene, *Cag11*, was predicted as a target with a p value of 0.03 and thus was not included in Table 10.

HPnc2665 predicted to control five virulence associated genes

HPnc2665 had 33 estimated targets that had a p value less than or equal to 0.02. The specific function of each target can be found in Table 11 and I will discuss the most intriguing ones here. Five virulence factors were predicted as targets; two (*tlpB* and *fliL*) encode components for chemotaxis and motility both of which are essential for *H. pylori* to colonize the stomach[25,26]. One of the virulence factors is the *hopZ* an adhesin allowing *H. pylori* to attach to gastric epithelial cells and it exhibits phase variation, meaning it may play a role in evading immune attack[110,131]. Additionally, *HPnc2665* targets a subunit of urease called UreB. Urease is an essential virulence factor that enables *H. pylori* to survive within the gastric lumen by breaking down urea into ammonia and carbon dioxide and raising the pH around itself[20]. Without urease to protect the cells, *H. pylori* is unable to colonize the stomach[20,132]. Not only does urease play an essential role

in persistence but it also plays a role in triggering more severe disease phenotypes. One study done by Olivera-Severo *et al.* (2017) found that urease produced by *H. pylori* is internalized by the gastric epithelial cells and induces an angiogenic response[22]. This is significant because angiogenesis (the formation of new blood vessels from pre-existing vasculature) is essential for tumor growth, and metastatic dissemination[22]. Additionally, *HPnc2665* is not the first sRNA to be implicated in regulating *ureB; 5'-ureB-sRNA* down regulates UreB expression in neutral pH conditions. 5'-*ureB-sRNA* is a *cis*-sRNA to *ureB* and shares complementarity to the 5' coding region of *ureB,* while *HPnc2665* is a *trans*-sRNA to *ureB* and is complementary to the 5' untranslated region. It may be that sRNAs play a significant role in regulating urease genes. The cagPAI gene *cagF* (*cag22*) is another target of *HPnc2665*. This result was unexpected because I had anticipated *cagE,* the gene trans to *HPnc2665,* to be a target of the sRNA rather than a gene downstream. The targets of *HPnc2665* indicate that it is a *trans*-sRNA.

Future directions: Experimentally testing the promoters, terminators, and targets for *HPnc2620* and *HPnc2665*

This bioinformatics approach to characterize sRNAs was an ideal experiment and was useful to understand how *HPnc2620* and *HPnc2665* are regulated, to develop a list of potential targets, and tentatively classify these sRNAs as *trans*-sRNAs. This project provides a strong foundation for future research into both sRNAs. If not for the COVID-19 pandemic, this study would have included functional assays to test the regulatory

elements and targets predicted in this study. I had begun developing plasmids to test promoters, terminators, and targets (Appendix X for plasmid maps) but was unable to progress farther. The goal of this work was to take the predicted regulatory regions and confirm their function and efficiency. With the promoter and terminator confirmed it would also define the 5’ and 3’ ends of *HPnc2665*, which had not been defined previously. With the 5’ and 3’ ends defined it would allow for TargetRNA2 to predict targets more accurately. Lastly, and possibly most importantly for future directions is testing the targets. TargetRNA2 does not indicate how the target is regulated, so without a functional assay one cannot predict if targets are up regulated or down-regulated.

REFERENCES

1. Haesebrouck, F. *et al.* Gastric helicobacters in domestic animals and nonhuman primates and their significance for human health. *Clin. Microbiol. Rev.* **22**, 202–223 (2009).

2. Schreiber, S., Thanh Hoa Nguyen, Stuben, M. & Scheid, P. Demonstration of a pH gradient in the gastric gland of the acid-secreting guinea pig mucosa. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **279**, 597–604 (2000).

3. Maixner, F. *et al.* Helicobacter pylori in ancient human remains. *world J. Gastroenterol.* **25**, 6289–6298 (2019).

4. Coghlan, J. G. *et al.* Campylobacter pylori and the recurrence of duodenal ulcers-a 12-month follow-up study. *Lancet* **330**, 1109–1111 (1987).

5. Buckley, M. J. M. & O'Morain, C. A. Helicobacter biology - Discovery. *Br. Med. Bull.* **54**, 7–16 (1998).

6. Wang, C., Yuan, Y. & Hunt, R. H. The association between Helicobacter pylori infection and early gastric cancer: a meta-analysis. *Am. J. Gastroenterol.* **102**, 1789–1798 (2007).

7. Graham, D. Y. History of Helicobacter pylori, duodenal ulcer, gastric ulcer and gastric cancer. *World J. Gastroenterol.* **20**, 5191–5204 (2014).

8. Lyon, F. R. Schistosomes, Liver flukes, and Helicobacter pylori. in *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, No. 61. IARC Working Group on the Evaluation of Carcinogenic Risk to Humans* 218–220 (1994).

9. Hooi, J. K. Y. *et al.* Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. *Gastroenterology* **153**, 420–429 (2017).

10. Aziz, R. K., Khalifa, M. M. & Sharaf, R. R. Contaminated water as a source of Helicobacter pylori infection: A review. *J. Adv. Res.* **6**, 539–547 (2015).

11. Nagy, P., Johansson, S. & Molloy-Bland, M. Systematic review of time trends in the prevalence of Helicobacter pylori infection in China and the USA. *Gut Pathog.* **8**, (2016).

12. Brown, L. M. Helicobacter pylori : Epidemiology and Routes of Transmission. *Epidemiol. Rev.* **22**, (2000).

13. Quaglia, N. C. & Dambrosio, A. Helicobacter pylori: A foodborne pathogen? *World J. Gastroenterol.* **24**, 3472–3487 (2018).

14. Vesga, F., Moreno, Y., Ferrus, M. A., Campos, C. & Trespalacios, A. A. Detection of Helicobacter pylori in drinking water treatment plants in Bogota, Colombia, using

cultural and molecular techniques. *Int. J. Hyg. Environ. Health* **221**, 595–601 (2018).

15. Fujimura, S., Kato, S. & Kawamura, T. Helicobacter pylori in Japanese river water and its prevalence in Japansese children. *Lett. Appl. Microbiol.* **38**, 517–521 (2004).

16. Janzon, A. *et al.* Failure to detect Helicobacter pylori DNA in drinking and environmental water in Dhaka, Bangladesh, using highly sensitive real-time PCR assays. *Appl. Environ. Microbiol.* **75**, 3039–3044 (2009).

17. Grübel, P. *et al.* Vector potential of houseflies (Musca domestica) for Helicobacter pylori. *J. Clin. Microbiol.* **35**, 1300–1303 (1997).

18. Cross, A. S. What is a virulence factor? *Crit. Care* **12**, 196 (2008).

19. Celli, J. P. *et al.* Helicobacter pylori moves through mucus by reducing mucin viscoelasticity. *Proc. Natl. Acad. Sci.* **106**, 14321–14326 (2009).

20. Scott, D. R. *et al.* The role of internal urease in acid resistance of Helicobacter pylori. *Gastroenterology* **114**, 58–70 (1998).

21. Phadnis, S. H. *et al.* Surface localization of Helicobacter pylori urease and a heat shock protein homolog requires bacterial autolysis. *Infect. Immun.* **64**, 905–912 (1996).

22. Olivera-Severo, D. *et al.* A new role for Helicobacter pylori urease: Contributions to angiogenesis. *Front. Microbiol.* **8**, 1–11 (2017).

23. Lee, J. H., Jun, S. H., Kim, J. M., Baik, S. C. & Lee, J. C. Morphological changes in human gastric epithelial cells induced by nuclear targeting of Helicobacter pylori urease subunit A. *J. Microbiol.* **53**, 406–414 (2015).

24. Croxen, M. A., Sisson, G., Melano, R. & Hoffman, P. S. The Helicobacter pylori chemotaxis receptor tlpB (HP0103) is required for pH taxis and for colonization of the gastric mucosa. *J. Bacteriol.* **188**, 2656–2665 (2006).

25. Johnson, K. S. & Ottemann, K. M. Colonization, localization, and inflammation: The roles of H. pylori chemotaxis in vivo. *Curr. Opin. Mircobiology* **41**, 51–57 (2018).

26. Collins, K. D., Hu, S., Grasberger, H., Kao, J. Y. & Ottemann, K. M. Chemotaxis allows bacteria to overcome host-generated reactive oxygen species that constrain gland colonization. *Infect. Immun.* **86**, 1–12 (2018).

27. Rolig, A. S., Carter, J. E. & Ottemann, K. M. Bacterial chemotaxis modulates host cell apoptosis to establish a T-helper cell, type 17 (Th17)-dominant immune response in Helicobacter pylori infection. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 19749–19754 (2011).

28. Mahdavi, J. *et al.* Helicobacter pylori SabA Adhesin in Persistent Infection and Chronic Inflammation. *Science (80-.).* **297**, 573–578 (2002).

29. Unemo, M. *et al.* The sialic acid binding SabA adhesin of Helicobacter pylori is essential for nonopsonic activation of human neutrophils. *J. Biol. Chem.* **280**, 15390–15397 (2005).

30. Solnick, J. V., Hansen, L. M., Salama, N. R., Boonjakuakul, J. K. & Syvanen, M. Modification of Helicobacter pylori outer membrane protein expression during experimental infection of rhesus macaques. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2106–2111 (2004).

31. Kable, M. E. *et al.* Host Determinants of Expression of the Helicobacter pylori BabA Adhesin. *Sci. Rep.* **7**, 1–12 (2017).

32. Styer, C. M. *et al.* Expression of the BabA adhesin during experimental infection with Helicobacter pylori. *Infect. Immun.* **78**, 1593–1600 (2010).

33. Ohno, T. *et al.* Effects of blood group antigen-binding adhesin expression during Helicobacter pylori infection of Mongolian gerbils. *J. Infect. Dis.* **203**, 726–735 (2011).

34. Colbeck, J. C., Hansen, L. M., Fong, J. M. & Solnick, J. V. Genotypic profile of the outer membrane proteins BabA and BabB in clinical isolates of Helicobacter pylori. *Infect. Immun.* **74**, 4375–4378 (2006).

35. Palframan, S. L., Kwok, T., Gabriel, K. & Cover, T. Vacuolating cytotoxin A (VacA), a key toxin for Helicobacter pylori pathogenesis. **2**, 1–9 (2012).

36. Boncristiano, M. *et al.* The Helicobacter pylori Vacuolating Toxin Inhibits T Cell Activation by Two Independent Mechanisms. *J. Exp. Med.* **198**, 1887–1897 (2003).

37. Shariq, M. *et al.* Biochemical analysis of CagE: A VirB4 homologue of helicobacter pylori cag-T4SS. *PLoS One* **10**, 1–23 (2015).

38. Haas, R., Gerhard, M. & Naumann, M. The Helicobacter pylori type IV secretion system encoded by the cag Pathogenicity island: Architecture, Function, and Signaling. in *Tyoe IV Secretion in Gram-Negative and Gram-Positive Bacteria* 187–220 (2018).

39. Kaebisch, R., Mejías-Luque, R., Prinz, C. & Gerhard, M. Helicobacter pylori Cytotoxin-Associated Gene A Impairs Human Dendritic Cell Maturation and Function through IL-10–Mediated Activation of STAT3 . *J. Immunol.* **192**, 316–323 (2014).

40. Alzahrani, S. *et al.* Effect of Helicobacter pylori on gastric epithelial cells. *World J. Gastroenterol.* **20**, 12767–12780 (2014).

41. Merrell, D. S. *et al.* Growth Phase-Dependent Response of Helicobacter pylori to Iron Starvation. *Infect. Immun.* **71**, 6510–6525 (2003).

42. Merrell, D. S., Goodrich, M. L., Otto, G., Tompkins, L. S. & Falkow, S. pH-regulated gene expression of the gastric pathogen Helicobacter pylori. *Infect. Immun.* **71**, 3529–3539 (2003).

43. Tomb, J. *et al.* The complete genome sequence of the gastric pathogen Helicobacter pylori. **389**, 1211–1212 (1997).

44. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science (80-.).* **269**, 496–512 (1995).

45. Sharma, C. M. *et al.* The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464**, 250–255 (2010).

46. Du, J., Zhang, W., Li, X. & Li, Y. Bioinformatics analysis of small RNAs in Helicobacter pylori and the role of NAT-67 under tinidazole treatment. *Mol. Med. Rep.* **22**, 1227–1234 (2020).

47. Waters, L. S. & Storz, G. Regulatory RNAs in Bacteria. *Cell* **136**, 615–628 (2009).

48. Harris, J. F., Micheva-Viteva, S., Li, N. & Hong-Geller, E. Small RNA-mediated regulation of host-pathogen interactions. *Virulence* (2013). doi:10.4161/viru.26119

49. Michaux, C., Verneuil, N., Hartke, A. & Giard, J. Physiological roles of small RNA molecules. *Microbiology* **160**, 1007–1019 (2014).

50. Babitzke, P. & Romeo, T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr. Opin. Mircobiology* **10**, 156–163 (2007).

51. Heidrich, N., Moll, I. & Brantl, S. In vitro analysis of the interaction between the small RNA SR1 and its primary target ahrC mRNA. *Nucleic Acids Res.* **35**, 4331–4346 (2007).

52. Masse, E. & Gottesman, S. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 4620–4625 (2002).

53. Bardill, J. P., Zhao, X. & Hammer, B. K. The Vibrio cholerae quorum sensing response is mediated by Hfq-dependent sRNA/mRNA base-pairing interactions. *Mol. Microbiol.* **80**, 1381–1394 (2011).

54. Opdyke, J. A., Fozo, E. M., Hemm, M. R. & Storz, G. GadY, a small-RNA regulator of acid response genes in Escherichia coli. *J. Bacteriol.* **186**, 6698–6705 (2004).

55. Kreikemeyer, B., Boyle, M. D., Buttaro, B. A., Heinemann, M. & Podbielski, A. Group A streptococccal growth phase-associated virulence factor reguation by a novel operon (Fas) with homologies to two component-type regulators requires a small RNA molecule. *Mol. Microbiol.* **39**, 392–406 (2001).

56. Brantl, S. Bacterial chromosome-encoded small regulatory RNAs. *Future*

Microbiol. **4**, 85–103 (2009).

57. Brantl, S., Birch-Hirschfeld, E. & Behnke, D. RepR Protein Expression on Plasmid pIP501 Is Controlled by an Antisense RNA-mediated Transcription Attenuation Mechanism. *J. Bacteriol.* **175**, 4052–4061 (1993).

58. Dutta, T. & Srivastava, S. Small RNA-mediated regulation in bacteria : A growing palette of diverse mechanisms. *Gene* **656**, 60–72 (2018).

59. Smirnov, A., Wang, C., Drewry, L. L. & Vogel, J. Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA. *Eur. Mol. Biol. Organ. J.* **36**, 1029–1045 (2017).

60. Bos, J. *et al.* The sRNA RyhB Regulates the Synthesis of the Escherichia coli Methionine Sulfoxide Reductase MsrB but Not MsrA. *PLoS One* **8**, (2013).

61. Sharma, C. M. *et al.* Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol. Microbiol.* **81**, 1144–1165 (2011).

62. 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* **14**, 4569–4577 (1995).

63. Sonnleitner, E. *et al.* The small RNA PhrS stimulates synthesis of the Pseudomonas aeruginosa quinolone signal. *Mol. Microbiol.* **80**, 868–885 (2011).

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

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

66. Darfeuille, F., Unoson, C., Vogel, J. & Wagner, E. G. H. An Antisense RNA Inhibits Translation by Competing with Standby Ribosomes. *Mol. Cell* **26**, 381–392 (2007).

67. Huntzinger, E. *et al.* Staphylococcus aureus RNAIII and the endoribonuclease III coordinately regulate spa gene expression. *EMBO J.* **24**, 824–835 (2005).

68. Sledjeski, D. D., 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.* **15**, 3993–4000 (1996).

69. Wen, Y., Feng, J., Scott, D. R., Marcus, E. A. & Sachs, G. A cis -Encoded Antisense Small RNA Regulated by the HP0165-HP0166 Two-Component System Controls Expression of ureB in Helicobacter pylori . **193**, 40–51 (2011).

70. Li, X. X. *et al.* Bacterial microbiota profiling in gastritis without Helicobacter pylori

infection or non-steroidal anti-inflammatory drug use. *PLoS One* **4**, 1–9 (2009).

71. Ta, L. H. *et al.* Conserved Transcriptional Unit Organization of the Cag Pathogenicity Island among Helicobacter pylori Strains. *Front. Cell. Infect. Microbiol.* (2012). doi:10.3389/fcimb.2012.00046

72. Pernitzsch, S. R., Tirier, S. M., Beier, D. & Sharma, C. M. A variable homopolymeric G-repeat defines small RNA-mediated posttranscriptional regulation of a chemotaxis receptor in Helicobacter pylori. *Proc. Natl. Acad. Sci.* (2014). doi:10.1073/pnas.1315152111

73. Vannini, A., Roncarati, D. & Danielli, A. The cag-pathogenicity island encoded CncR1 sRNA oppositely modulates Helicobacter pylori motility and adhesion to host cells. *Cell. Mol. Life Sci.* (2016). doi:10.1007/s00018-016-2151-z

74. Kinoshita-Daitoku, R. *et al.* Helicobacter small RNA regulates host adaptation and carcinogenesis. *bioRxiv* **Preprint**, 1–9 (2020).

75. Brandt, S., Kwok, T., Hartig, R., König, W. & Backert, S. NF-κB activation and potentiation of proinflammatory responses by the Helicobacter pylori CagA protein. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 9300–9305 (2005).

76. Kao, C. Y., Sheu, B. S. & Wu, J. J. Helicobacter pylori infection: An overview of bacterial virulence factors and pathogenesis. *Biomed. J.* **39**, 14–23 (2016).

77. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

78. Forsyth, M. H. & Cover, T. L. Mutational analysis of the vacA promoter provides insight into gene transcription in Helicobacter priori. *J. Bacteriol.* **181**, 2261–2266 (1999).

79. McGowan, C. C., Necheva, A. S., Forsyth, M. H., Cover, T. L. & Blaser, M. J. Promoter analysis of Helicobacter pylori genes with enhanced expression at low pH. *Mol. Microbiol.* **48**, 1225–1239 (2003).

80. Spohn, G., Beier, D., Rappuoli, R. & Scarlato, V. Transcriptional analysis of the divergent cagAB genes encoded by the pathogenicity island of Helicobacter pylori . *Mol. Microbiol.* **26**, 361–372 (1997).

81. Pereira, C. V., Pereira, L. J., Gonçalves, R. B. & Höfling, J. F. In vitro bacterial plaque suppression and recolonization by S. mutans and S. sobrinus. *Brazilian J. Microbiol.* **37**, 20–25 (2006).

82. Vanet, A., Marsan, L., Labigne, A. & Sagot, M. F. Inferring regulatory elements from a whole genome. An analysis of Helicobacter pylori σ80 family of promoter signals. *J. Mol. Biol.* **297**, 335–353 (2000).

83. Josenhans, C. *et al.* Functional characterization of the antagonistic flagellar late regulators FliA and FlgM of Helicobacter pylori and their effects on the H. pylori

transcriptome. *Mol. Microbiol.* **43**, 307–322 (2002).

84. Niehus, E. *et al.* Genome-wide analysis of transcriptional hierarchy and feedback regulation in the flagellar system of Helicobacter pylori. *Mol. Microbiol.* **52**, 947–961 (2004).

85. Münch, R. *et al.* Virtual Footprint and PRODORIC: An integrative framework for regulon prediction in prokaryotes. *Bioinformatics* **21**, 4187–4189 (2005).

86. Castillo, A. R., Arevalo, S. S., Woodruff, A. J. & Ottemann, K. M. Experimental analysis of Helicobacter pylori transcriptional terminators suggests this microbe uses both intrinsic and factor-dependent termination. *Mol. Microbiol.* **67**, 155–170 (2008).

87. Lesnik, E. A. *et al.* Prediction of rho-independent transcriptional terminators in Escherichia coli. *Nucleic Acids Res.* **29**, 3583–3594 (2001).

88. Petersen, L. & Krogh, A. Modeling of Rho Dependent Transcription Termination Sites in the Bacterium Helicobacter pylori. *ECCB 2003 Eur. Conf. Comput. Biol.* (2003).

89. d'Aubenton Carafa, Y., Brody, E. & Thermes, C. Prediction of rho-independent Escherichia coli transcription terminators. A statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.* **216**, 835–858 (1990).

90. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).

91. Kery, M. B., Feldman, M., Livny, J. & Tjaden, B. TargetRNA2: Identifying targets of small regulatory RNAs in bacteria. *Nucleic Acids Res.* **42**, 124–129 (2014).

92. Baltrus, D. A. *et al.* The complete genome sequence of Helicobacter pylori strain G27. *J. Bacteriol.* **91**, 447–448 (2009).

93. Tjaden, B. TargetRNA: a tool for predicting targets of small RNA action in bacteria. *Nucleic Acids Res.* (2008). doi:10.1093/nar/gkn264

94. Castillo, A. R., Woodruff, A. J., Connolly, L. E., Sause, W. E. & Ottemann, K. M. Recombination-Based In Vivo Expression Technology Identifies Helicobacter pylori Genes Important for Host Colonization □ †. **76**, 5632–5644 (2008).

95. Urban, J. H. & Vogel, J. Translational control and target recognition by Escherichia coli small RNAs in vivo. *Nucleic Acids Res.* (2007). doi:10.1093/nar/gkl1040

96. Corcoran, C. P. *et al.* Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol. Microbiol.* (2012). doi:10.1111/j.1365-2958.2012.08031.x

97. Davis, M. C., Kesthely, C. A., Franklin, E. A. & MacLellan, S. R. The essential activities of the bacterial sigma factor. *Can. J. Microbiol.* **63**, 89–99 (2017).

98. de Pádua, R. A. P., Kia, A. M., Costa-Filho, A. J., Wilkinson, S. R. & Nonato, M. C.

Characterisation of the fumarate hydratase repertoire in Trypanosoma cruzi. *Int. J. Biol. Macromol.* **102**, 42–51 (2017).

99. Lövgren, J. M. *et al.* The PRC-barrel domain of the ribosome maturation protein RimM mediates binding to ribosomal protein S19 in the 30S ribosomal subunits. *Rna* **10**, 1798–1812 (2004).

100. Matsunami, H., Yoon, Y. H., Meshcheryakov, V. A., Namba, K. & Samatey, F. A. Structural flexibility of the periplasmic protein, FlgA, regulates flagellar P-ring assembly in Salmonella enterica. *Sci. Rep.* **6**, 1–3 (2016).

101. Padan, E. Functional and structural dynamics of NhaA, a prototype for Na+ and H+ antiporters, which are responsible for Na+ and H+ homeostasis in cells. *Biochim. Biophys. Acta - Bioenerg.* **1837**, 1047–1062 (2014).

102. De Rosa, M. *et al.* Design, synthesis and in vitro biological evaluation of oligopeptides targeting E. coli type I signal peptidase (LepB). *Bioorganic Med. Chem.* **25**, 897–911 (2017).

103. Kern, B. *et al.* Characterization of Helicobacter pyloriVacA-containing vacuoles (VCVs), VacA intracellular trafficking and interference with calcium signalling in T lymphocytes. *Cell. Microbiol.* **17**, 1811–1832 (2015).

104. Chandler, J. R. *et al.* Bactobolin resistance is conferred by mutations in the L2 ribosomal protein. *MBio* **3**, 1–8 (2012).

105. Weaver, J., Watts, T., Li, P. & Rye, H. S. Structural basis of substrate selectivity of e. coli prolidase. *PLoS One* **9**, (2014).

106. Cai, M. *et al.* Probing transient excited states of the bacterial cell division regulator MinE by relaxation dispersion NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 25446–25455 (2019).

107. Pratibha, P., Singh, S. K., Srinivasan, R., Bhat, S. R. & Sreenivasulu, Y. Gametophyte development needs mitochondrial coproporphyrinogen III oxidase function. *Plant Physiol.* **174**, 258–275 (2017).

108. Chen, L. *et al.* XC-0531 encodes a c-type cytochrome biogenesis protein and is required for pathogenesis in Xanthomonas campestris pv. campestris. *BMC Microbiol.* **17**, 1–13 (2017).

109. Perdu, C. *et al.* ExsB is required for correct assembly of the Pseudomonas aeruginosa Type III secretion apparatus in the bacterial membrane and full virulence in vivo. *Infect. Immun.* **83**, 1789–1798 (2015).

110. Kennemann, L. *et al.* In vivo sequence variation in HopZ, a phase-variable outer membrane protein of helicobacter pylori. *Infect. Immun.* **80**, 4364–4373 (2012).

111. Nutschan, K., Golbik, R. P. & Sawers, R. G. The iron–sulfur-containing HypC-HypD scaffold complex of the [NiFe]-hydrogenase maturation machinery is an ATPase.

FEBS Open Bio **9**, 2072–2079 (2019).

112. Iijima, K., Suzuki, K., Asano, S., Ogata, T. & Kitagawa, Y. HorC, a hop-resistance related protein, presumably functions in homodimer form. *Biosci. Biotechnol. Biochem.* **73**, 1880–1882 (2009).

113. Corthésy-Theulaz, I. E. *et al.* Cloning and characterization of helicobacter pylori succinyl CoA:Acetoacetate CoA-transferase, a novel prokaryotic member of the CoA- transferase family. *J. Biol. Chem.* **272**, 25659–25667 (1997).

114. Peng, X., Dong, H. & Wu, Q. A newcis-encoded sRNA, BsrH, regulating the expression of hemH gene in Brucella abortus 2308. *FEMS Microbiol. Lett.* **362**, 1–7 (2015).

115. Oshiro, J. *et al.* Regulation of the DPP1-encoded diacylglycerol pyrophosphate (DGPP) phosphatase by inositol and growth phase: Inhibition of DGPP phosphatase activity by CDP-diacylglycerol and activation of phosphatidylserine synthase activity by DGPP. *J. Biol. Chem.* **275**, 40887–40896 (2000).

116. Thach, T. T. & Lee, S. New crystal structures of adenylate kinase from Streptococcus pneumoniae D39 in two conformations. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **70**, 1468–1471 (2014).

117. Bonsor, D. A. *et al.* Characterization of the translocation-competent complex between the Helicobacter pylori oncogenic protein CagA and the accessory protein CagF. *J. Biol. Chem.* **288**, 32897–32909 (2013).

118. Archambault, L., Borchert, J. S., Bergeron, J., Snow, S. & Schlax, P. J. Measurements of mRNA degradation in borrelia burgdorferi. *J. Bacteriol.* **195**, 4879–4887 (2013).

119. Aurass, P. *et al.* glnA truncation in Salmonella enterica results in a small colony variant phenotype, attenuated host cell entry, and reduced expression of flagellin and SPI-1- associated effector genes. *Appl. Environ. Microbiol.* **84**, 1–15 (2018).

120. Liu, Y. P. *et al.* Structural basis underlying complex assembly and conformational transition of the type I R-M system. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 11151–11156 (2017).

121. Léveillard, T. & Aït-Ali, N. Cell Signaling with Extracellular Thioredoxin and Thioredoxin-Like Proteins: Insight into Their Mechanisms of Action. *Oxid. Med. Cell. Longev.* **2017**, (2017).

122. Takekawa, N. *et al.* Structure of vibrio flil, a new stomatin-like protein that assists the bacterial flagellar motor function. *MBio* **10**, 1–13 (2019).

123. Tiedeman, A. A., Smith, J. M. & Zalkin, H. Nucleotide sequence of the guaA gene encoding GMP synthetase of Escherichia coli K12. *J. Biol. Chem.* **260**, 8676–8679 (1985).

124. Bashyam, M. D., Kaushal, D., Dasgupta, S. K. & Tyagi, A. K. A study of the mycobacterial transcriptional apparatus: Identification of novel features in promoter elements. *J. Bacteriol.* **178**, 4847–4853 (1996).

125. ThermoFisher. Reverse transcription troubleshooting guide. Available at: https://www.thermofisher.com/us/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/rt-education/reverse-transcription-troubleshooting.html. (Accessed: 3rd July 2020)

126. Mondal, S., Yakhnin, A. V., Sebastian, A., Albert, I. & Babitzke, P. NusA-dependent transcription termination prevents misregulation of global gene expression. *Nat. Microbiol.* 1570–1573 (2018). doi:10.1038/s41395-018-0061-4.

127. Akhter, S. *et al.* Integrative analysis of three RNA sequencing methods identifies mutually exclusive exons of MADS-box isoforms during early bud development in picea abies. *Front. Plant Sci.* **871**, 1–18 (2018).

128. TakaraBio. *SMARTer RACE 5'3' Kit User Manual_052617*. **634859**,

129. Sirianni, A., Kaakoush, N. O., Raftery, M. J. & Mitchell, H. M. The pathogenic potential of helicobacter pullorum: Possible role for the type VI secretion system. *Helicobacter* **18**, 102–111 (2013).

130. Wilkowska, K., Targonska, M., Smarz, A. & Sektas, M. Periplasmic expression of a restriction endonuclease in Escherichia coli and its effect on the antiviral activity of the host. *Acta Biochim. Pol.* **66**, 83–89 (2019).

131. Peck, B., Ortkamp, M., Diehl, K. D., Hundt, E. & Knapp, B. Conservation, localization and expression of HopZ, a protein involved in adhesion of Helicobacter pylori. *Nucleic Acids Res.* **27**, 3325–3333 (1999).

132. Eaton, K. A., Brooks, C. L., Morgan, D. R. & Krakowka, S. Essential role of urease in pathogenesis of gastritis induced by Helicobacter pylori in gnotobiotic piglets. *Infect. Immun.* **59**, 2470–2475 (1991).

APPENDIX

I. Media for this study

a. Sheep's blood agar with betacyclodextrin

i. Dissolve correct mass of Columbia Agar (check container) in 500 ml distilled water.

ii. Cover with tinfoil and autoclave tape and autoclave.

iii. WHILE AUTOCLAVING: Mix fresh ß Cyclodextrin by adding 1 g betacyclodextrin to 5 ml DMSO. (1 ml per 100 ml agar)

1. Filter sterilize the mix using a syringe and filter

iv. Cool media to about 55˚C (still hot to touch but tolerable)

v. Add 5 ml blood per 100 ml agar and betacyclodextrin.

vi. Swirl gently and pour

b. Sheep's blood agar with betacyclodextrin and chloramphenicol

i. Make sheep's blood agar with betacyclodextrin.

ii. Once cooled enough to pour, add 50 µl of chloramphenicol (20 mg/ml) per 100 ml agar.

iii. Swirl and pour.

c. Sheep's blood agar with betacyclodextrin and kanamycin

i. Make sheep's blood agar with betacyclodextrin.

ii. Once cooled enough to pour, add 75 µl of kanamycin (30 mg/ml) per 100ml agar.

iii. Swirl and pour.

d. Sheep's blood agar with betacyclodextrin, chloramphenicol, and kanamycin

i. Make sheep's blood agar with betacyclodextrin.

ii. Once cooled enough to pour, add 75 µl of kanamycin (30 mg/ml) and 50 µl of chloramphenicol (20 mg/ml) per 100 ml agar.

iii. Swirl and pour.

II. Plasmid maps

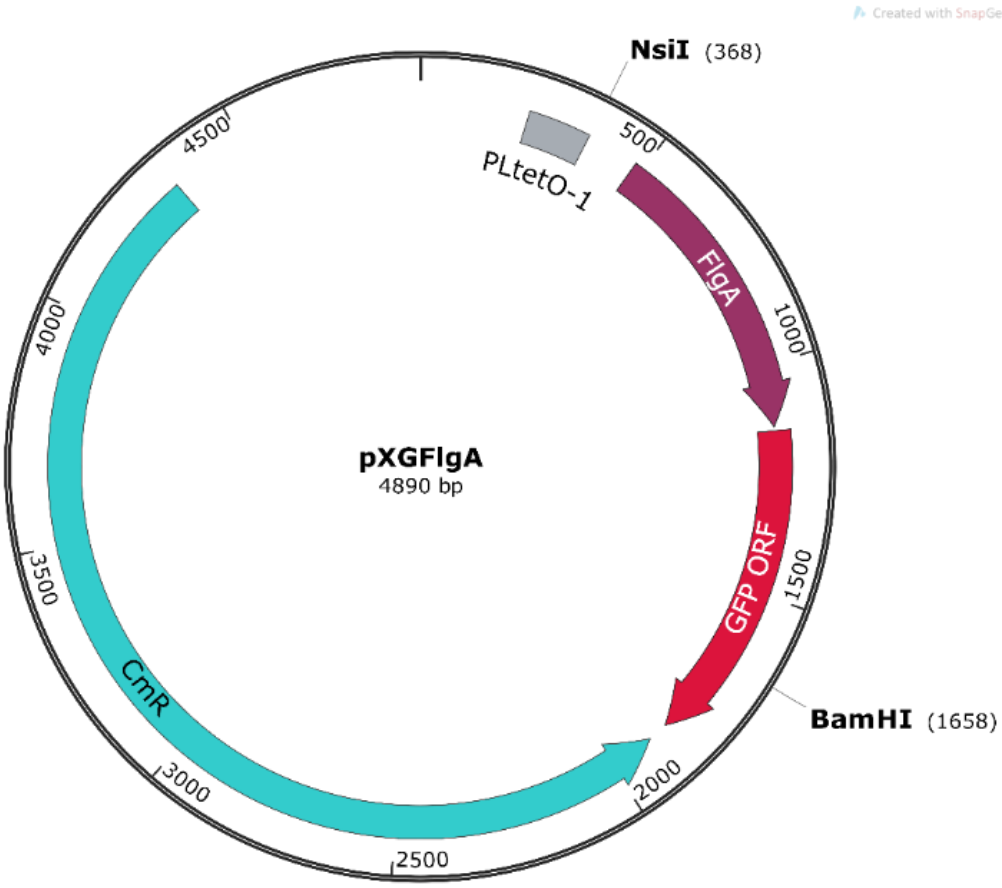


Figure 15. pXGFlgA map

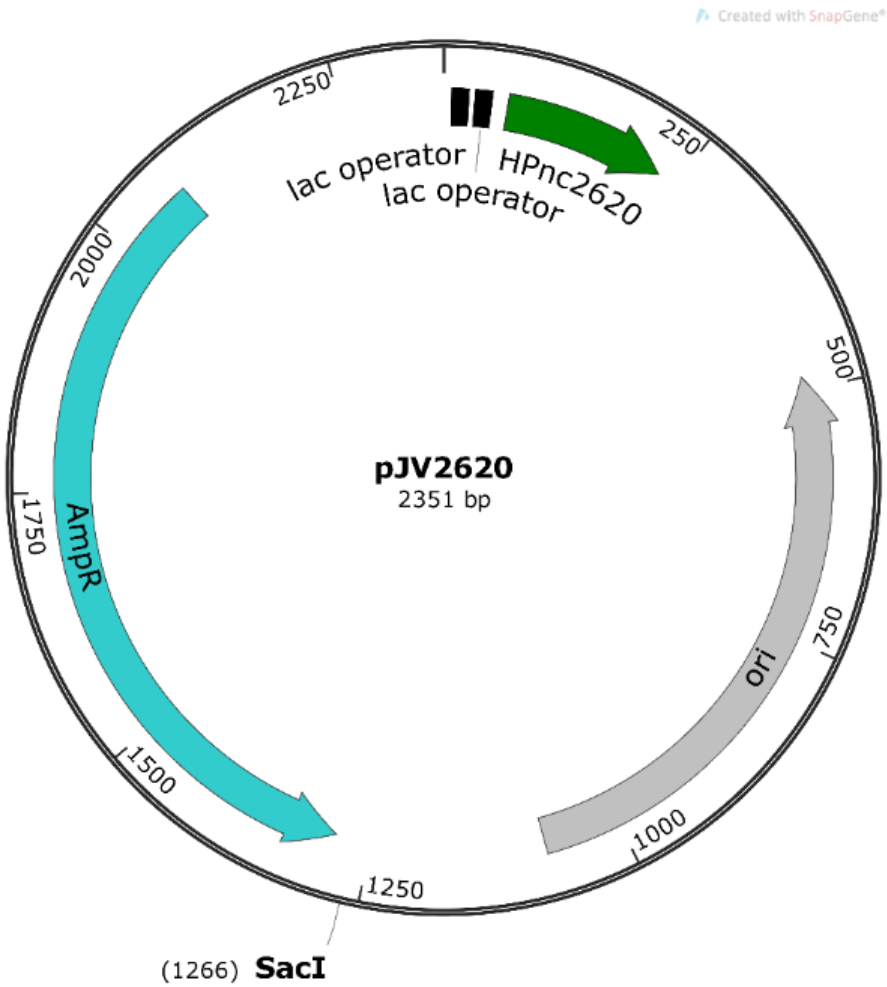


Figure 16. pJV2620 map

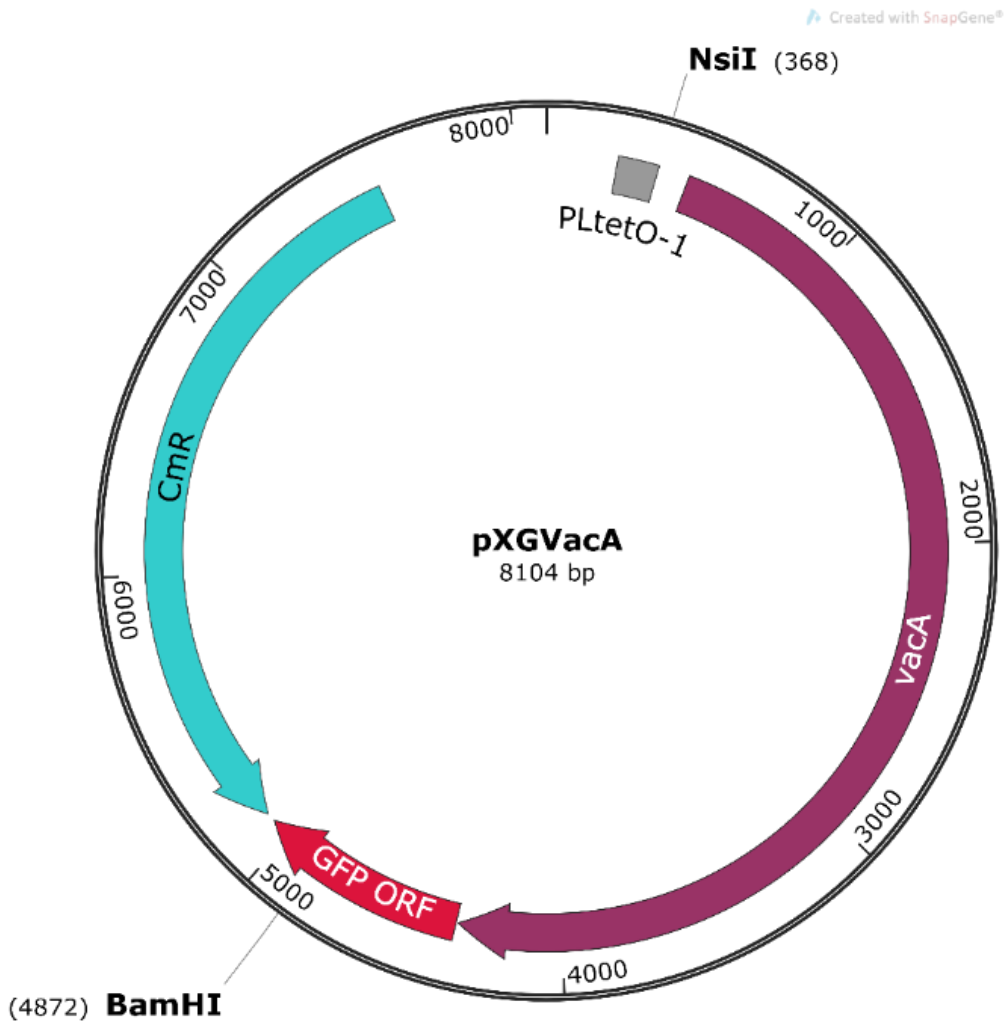


Figure 17. pXGVacA map

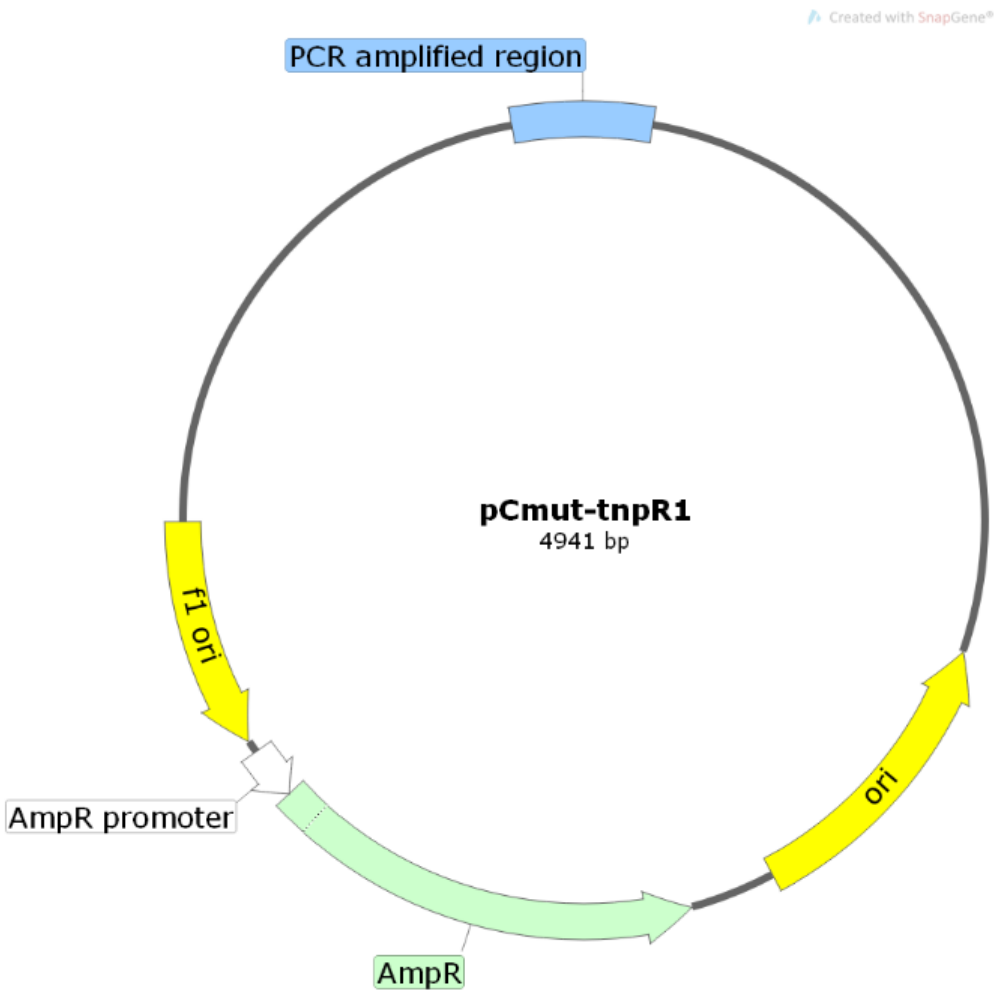


Figure 18. pcmut-tnpR1 map

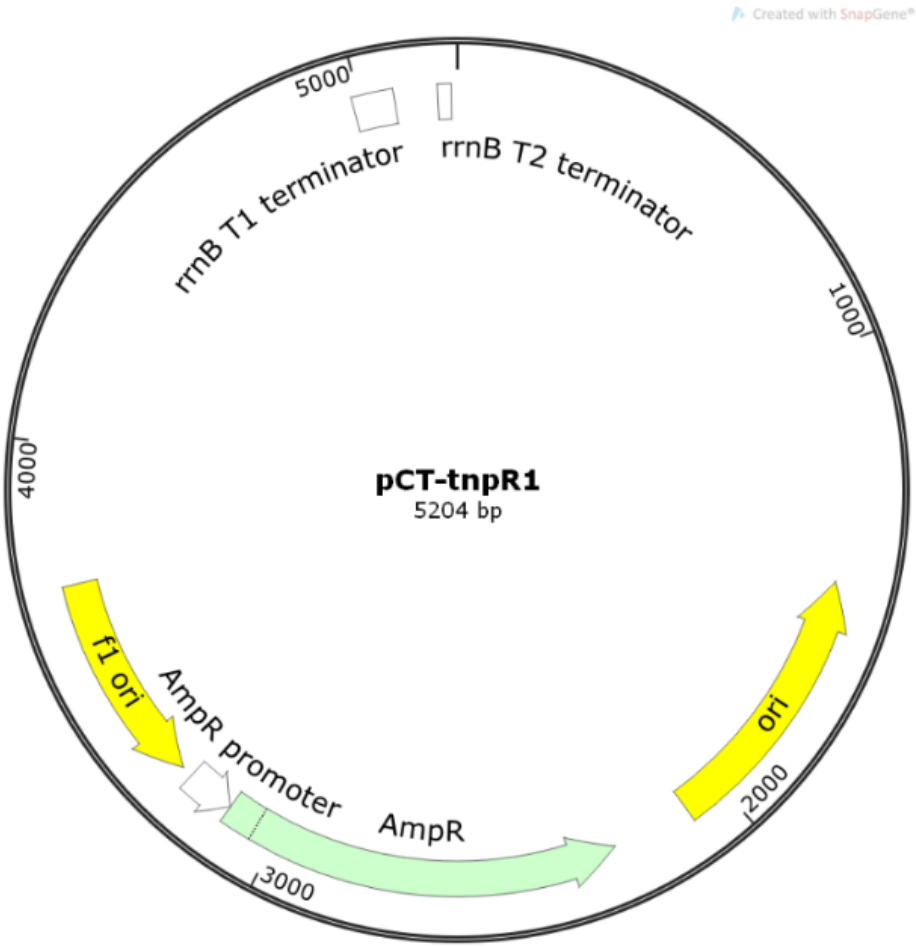


Figure 19. pCT-tnpR1 map

III. Digestion protocols

i. Set up reaction as follows:

Table 12. Reaction for a single digestion.

Restriction enzyme 1	1 µl
DNA	1 µg
Buffer	2 µl
Deionized water	to 20 µl
Total	20 µl

Table 13. Reaction for a double digestion.

Restriction enzyme 1	1 µl
Restriction enzyme 2	1 µl
DNA	1 µg

Buffer	2 µl
Deionized water	to 20 µl
Total	20 µl

ii. Incubate at 37°C for 30 minutes

b. Check by agarose gel electrophoresis (Appendix XII)

IV. Digestion cleanup

a. Gel extraction and cleanup performed using Nucleospin Gel and PCR Clean-up Kit

b. Using the agarose gel electrophoresis used to check the digestion went to completion:

i. Take a clean scalpel to excise the DNA fragment from an agarose gel. Remove all excess agarose. Determine the weight of the gel slice and transfer it to a clean tube. For each 100 mg of agarose, double the volume of Buffer NTI. Incubate sample for 5–10 min at 50 °C. Vortex the sample briefly every 2–3 min until the gel slice is completely dissolved!

ii. Place a NucleoSpin® Gel and PCR Clean-up Column into a Collection Tube (2 mL) and load up to 700 µL sample. Centrifuge for 30 s at 11,000 x g. Discard flow-through and place the column back into the collection tube. Load remaining sample if necessary and repeat the centrifugation step.

iii. Add 700 µL Buffer NT3 to the NucleoSpin® Gel and PCR Clean-up Column. Centrifuge for 30 s at 11,000 x g. Discard flow-through and place the column back into the collection tube.

iv. Recommended: Repeat previous washing step to minimize chaotropic salt carry-over

v. Centrifuge for 1 min at 11,000 x g to remove Buffer NT3 completely. Make sure the spin column does not touch the flow-through while removing it from the centrifuge and the collection tube.

vi. Place the NucleoSpin® Gel and PCR Clean-up Column into a new 1.5 mL microcentrifuge tube (not provided). Add 15–30 µL Buffer NE and incubate at room temperature (18–25 °C) for 1 min. Centrifuge for 1 min at 11,000 x g.

V. Polymerase chain reaction protocols

a. Polymerase chain reaction using One Taq

i. Set up the reactions on ice.

Table 14. PCR reaction mix for One Taq.

10 µM Forward oligonucleotide	1 µl
10 µM Revers oligonucleotide	1 µl
One *Taq* 2X Master Mix with Standard Buffer	10 µl
DNA	1 µg
Deionized water	to 20 µl
Total	20 µl

Table 15. Thermocycling conditions for One Taq

Initial Denaturation	94°C	2 min
30 cycles	94°C 45-68°C 68°C	15-30 s 15-60 s 1 minute/kb
Final Extension	68°C	5 minutes
Hold	4-10°C	

ii. Check products by agarose gel electrophoresis

b. Polymerase chain reaction using Phusion

i. Set up the reactions on ice.

Table 16. PCR reaction mix for Phusion.

10 µM Forward oligonucleotide	2 µl
10 µM Revers oligonucleotide	2 µl
Phusion 5X buffer	4 µl
Phusion high-fidelity DNA polymerase	0.2 µl
2.5 µM dNTPs	2 µl
DNA	1 µg
Deionized water	to 20 µl
Total	20 µl

Table 17. Thermocycling conditions for Phusion.

Initial Denaturation	94°C	
30 cycles	94°C 45-68°C 68°C	15-30 s 15-60 s 1 minute/kb
Final Extension	68°C	5 minutes
Hold	4-10°C	

c. Check products by agarose gel electrophoresis

VI. PCR cleanup

a. Nucleospin Gel and PCR Clean-up Kit was used as follows:

i. Mix 1 volume of sample with 2 volumes of NTI buffer.

ii. Place a Nucleospin Gel and PCR Clean-up Column into a 2 ml collection tube and load up to 700 µl sample

iii. Centrifuge for 30 s at 11000 X g. Discard the flow-through and place column back into collection tube.

iv. Load remaining sample if needed.

v. Add 700 µl of NT3 buffer to the column. Centrifuge for 30 s at 11000 X g. Discard the flow-through and place column in the collection tube.

vi. Recommended: repeat step v to minimize chaotropic salt carry-over and improve A260/A280 values.

vii. Centrifuge for 1 min at 11000 X g to remove NT3 buffer completely. Spin column should not touch the flow-through.

viii. Place the spin column into a new 1.5 ml microcentrifuge tube. Add 15-30 µl NE buffer and incubate at room temperature for 1 min. centrifuge for 1 min at 11000 X g.

ix. Check concentration either by Nanodrop or gel electrophoresis.

VII. Ligation

a. DNA assembly using NEBuilder HiFi Assembly Master Mix

i. Set up the reactions on ice as follows:

Table 18. Ligation reaction mix

	Fragment assembly	Positive control	Negative control (minus insert)
Recommended DNA molar ratio	Vector:insert= 1:2		
Total amount of fragments	0.03-0.2pmol	0.03-0.2pmol	0.03-0.2pmol
NEBuilder HiFi DNA assembly master mix	5 µl	5 µl	5 µl
Deionized water	To 10 µl	To 10 µl	To 10 µl
Total	10	10	10

ii. Incubate samples in a thermocycler at 50°C for 15 minutes.

iii. Store samples on ice or at -20°C

iv. Transform into NEB 5-alpha *E. coli* provided by the kit. See appendix D2 for chemical transformation protocol.

VIII. Plasmid extraction (Qiagen miniprep)

a. Resuspend pelleted bacterial cells in 250 µl Buffer P1 and transfer to a microcentrifuge tube. Ensure that RNase A has been added to Buffer P1. No cell clumps should be visible after resuspension of the pellet. If LyseBlue reagent has been added to Buffer P1, vigorously shake the buffer bottle to ensure LyseBlue particles are completely dissolved. The bacteria should be resuspended completely by vortexing or pipetting up and down until no cell clumps remain.

b. Add 250 µl Buffer P2 and mix thoroughly by inverting the tube 4–6 times. Mix gently by inverting the tube. Do not vortex, because this will result in

shearing of genomic DNA. Do not allow the lysis reaction to proceed for more than 5 min. If LyseBlue has been added to Buffer P1, the cell suspension will turn blue after addition of Buffer P2.

c. Add 350 µl Buffer N3. Mix immediately and thoroughly by inverting the tube 4–6 times. The solution should become cloudy. If LyseBlue reagent has been used, the suspension should be mixed until all trace of blue has gone and the suspension is colorless.

d. Centrifuge for 10 min at 13,000 rpm (~17,900 x g) in a table-top microcentrifuge. A compact white pellet will form.

e. Apply 800 µl of the supernatant from step 4 to the QIAprep 2.0 spin column by pipetting.

f. Centrifuge for 30–60 s. Discard the flow-through.

g. Recommended: Wash the QIAprep 2.0 spin column by adding 0.5 ml Buffer PB and centrifuging for 30–60 s. Discard the flow-through.

h. Wash QIAprep 2.0 spin column by adding 0.75 ml Buffer PE and centrifuging for 30–60 s. 9. Discard the flow-through, and centrifuge at full speed for an additional 1 min to remove residual wash buffer.

i. Place the QIAprep 2.0 column in a clean 1.5 ml microcentrifuge tube. To elute DNA, add 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) or water to the center of each QIAprep 2.0 spin column, let stand for 1 minute and centrifuge for 1 minute.

IX. Promoter assay

a. *Helicobacter pylori* strain mG27 was transformed by natural transformation with pCT-PHPnc2620-tnpR1 or pCT-PHPnc2665-tnpR1 and selected for on cm (13 μg/ml) imbued Colombia blood agar.
b. Plates were incubated for 48 hrs at 37°C in microaerophilic conditions.
c. Total number of transformants were counted and recorded.
d. Cm resistant Cm transformants were plated on Kan (15 μg/ml) + CBA and incubated for 48 hrs at standard conditions.
e. The total number of Kan transformants were counted and recorded.
f. The promoter efficiency was determined using the following equation: (total Cm transformants/total Kan transformants) * 100

X. Terminator assay

a. *Helicobacter pylori* strain mG27 was transformed by natural transformation with pCmut-THPnc2620-tnpR1 or pCmut-THPnc2665-tnpR1 and selected for on cm (13 μg/ml) imbued Colombia blood agar.
b. Plates were incubated for 48 hrs at 37°C in microaerophilic conditions.
c. Total number of transformants were counted and recorded.
d. Cm resistant Cm transformants were plated on Kan (15 μg/ml) + CBA and incubated for 48 hrs at standard conditions.
e. The total number of Kan transformants were counted and recorded.
f. The promoter efficiency was determined using the following equation: (total Kan transformants/total Cm transformants) * 100

XI. GFP plasmid based expression protocol

a. Transform *E. coli* DH5α with pJV2620 or pJV300 as follows:

i. Electroporation

1. Took electrocompetent *E. coli* cells from the -80°C freezer to ice and let it thaw on ice.
2. In a properly labeled microcentrifuge tube, 50µl of cells were mixed with the plasmid of interest.
3. The cell mixture was then transferred to a chilled (on ice) electroporation cuvette with a 2mm gap. And placed back on ice.
4. Check the liquid in the cuvette to ensure it is at the bottom and wipe off any liquid from the sides with a KimWipe.
5. Using a Bio-Rad pulser set to Ec2, place the cuvette into the machine and pulse.
6. As quickly as possible, add 950 µl of SOB media to the cuvette, pipette up and down to mix.
7. Transfer the cell culture to a recovery tube and incubate for 30 min at 37°C.
8. Plate the cells on LB+Amp plates.
9. Grow overnight at 37°C.

b. Make the bacteria from (a) electrocompetent to transform them again with the target plasmids:

i. Use the bacteria from (a) to inoculate 2 ml of LB+Amp broth and allow to grow overnight

ii. Use 1 mL overnight culture of *E. coli* to inoculate 100mL of media (100 mL media should be in 500 mL flask)

iii. Set large centrifuge temperature to 4°C; it will cool down while your culture grows.

iv. Incubate culture for 2 hours in a 37°C shaking incubator until the Optical Density (OD) 600 is between 0.4 and 0.7

1. check the OD600 at 2 hours using a Bio-Rad spectrophotometer
2. Remove 0.5 mL culture using sterile technique and place in cuvette
3. Make a blank also, 0.5mL LB in cuvette
4. Carefully wipe down the outside of the cuvette with a KimWipe to remove debris that could impact the optical density
5. Blank the machine using the blank made in step iii.
6. Measure the OD600, if it is between 0.4 and 0.7 proceed to step 5. If the OD600 is below 0.4 incubate the culture for another 30 minutes and repeat steps i-vi.
7. If the culture has an OD above 0.7, dilute it down with sterile LB.

v. When between 0.4-0.7, pour approximately 45mL of culture into each of 2, sterile 50mL falcon tubes. Make sure they are balanced. STORE TUBES ON ICE. CELLS MUST always be kept on ice now.

vi. Pellet the cells by centrifugation for 5 minutes at 6,000 RPM. Remove promptly when done spinning and pour off supernatant

vii. Add 45 mL ice cold sterile water to each pellet and vortex vigorously to resuspend.

viii. Pellet the cells by centrifugation for 5 minutes at 6,000 RPM. Remove promptly when done spinning and pour off supernatant.

ix. Add 45 mL ice cold sterile water to each pellet and vortex vigorously to resuspend.

x. Pellet the cells by centrifugation for 5 minutes at 6,000 RPM. Remove promptly when done spinning and pour off supernatant.

xi. Add 25 mL ice cold sterile 10% glycerol to each pellet and vortex vigorously to resuspend.

xii. Pellet the cells by centrifugation for 5 minutes at 6,000 RPM. Remove promptly when done spinning and GENTLY pour off supernatant.

xiii. Resuspend pellet in 1mL of ice cold sterile 10% glycerol. Transfer to a microfuge tube—this will make it easier to aliquot. Keep this tube on ice!

xiv. Aliquot 100ul of cells to microcentrifuge tubes.

xv. Note—you can check competency right away or future.

    1. If you check the competency later, put your cells in the -80^{o}C freezer.

2. To check the competency of cells, do a transformation

c. Transform the *E. coli* above with the target plasmids (pXGFlgA, pXGVacA, or pXG0) using the protocol listed for (a)

d. Bacterial strains should be:

Table 19. Bacterial strains for GFP plasmid based expression system.

Strains	Plasmids present	Function
E. coli 300_ureB	(pJV300)(pXGUreB)	control, fluorescence of UreB:GFP
E. coli 2620_0	(pjv2620)(pXG0)	Control, autofluorescence change by pJV2620
E. coli 2665_0	(pJV2665)(pXG0)	Control, autofluorescence change by pJV2665
E. coli 300_0	(pJV300)(pXG0)	Control, Autofluorescence
E. coli 300_flgA	(pJV300)(pXGFlgA)	Control, fluorescence of FlgA:GFP
E. coli 300_tlpB	(pJV300)(pXGTlpB)	control, fluorescence of TlpB:GFP
E.coli 300_vacA	(pJV300)(pXGVacA)	Control, Fluorescence of VacA:GFP
E. coli 2620_flgA	(pJV2620)(pXGFlgA)	HPnc2620 regulation on flgA
E. coli 2620_vacA	(pJV2620)(pXGVacA)	HPnc2620 regulation on vacA
E. coli 2665_tlpB	(pJV2665)(pXGTlpB)	HPnc2665 regulation on tlpB
E. coli 2665_ureB	(pJV2665)(pXGUreB)	HPnc2665 regulation on ureB

e. All bacterial strains are grown in 2 ml LB + Amp + Chl broth overnight at 37°C in a shaking incubator.

f. The next morning inoculate 20 ml LB + Amp +Chl broth with 0.5 ml culture that grew overnight.

g. Incubate for 2 hours at 37°C in a shaking incubator.

h. Measure OD600 as mentioned above. Cultures should be between 0.5 and 0.6.

i. Using a Corning general assay microplate lid with raised lips over the wells (Figure X) pipet 10 µl of culture into “wells”. Do this in triplicate for all bacterial strains.

Figure 20. microplate lid

j. Promptly image fluorescence using an imager such as ChemStudio Touch from Analytik Jena

k. Save the image and analyze fluorescence using imageJ.

XII. Agarose gel electrophoresis

a. Measure out 50 ml of 1X Tris-acetate-EDTA.

b. Weigh out agarose. 0.5g for 1% gel, 1g for 2% gel, 0.4g for 0.8% gel.

c. Combine in a flask and microwave until the agarose is completely dissolved.

d. Pour into the gel box with comb and allow to harden (5-10 minutes)

e. Turn gel and cover with 1X Tris-acetate-EDTA

f. Add loading dye to samples and mix thoroughly.

g. Add 5 µl of appropriate ladder to a well, to other wells add samples

h. Apply electric current. 70 volts for 2.5 hours, 80 volts for 1.5-2 hours, 90 volts for 1-1.5 hours.

i. Turn off electric current and move the gel from the gel box to a box for post staining.

j. Add enough water to the box with the gel for the gel to be about half covered.

k. Add 5 µl ethidium bromide.

l. Stain for 15 minutes on a tilting table then image gel.

Vitae

Veronica Janette Albrecht
Department of Biology
Eastern Washington University

Education

2018-2020 **Master of Science in Biology (MS)**
Eastern Washington University, Cheney, Washington, USA
Area of study: Microbiology

2014-2018 **Bachelor of Science in Biology (BS)**
Eastern Washington University, Cheney, Washington, USA
Area of study: Biology

2010-2014 **Undergraduate Coursework**
Spokane Falls Community College, Spokane, Washington, USA
Area of Study: General Coursework

Research Experience

2018-2020 **Master's thesis project**
Department of Biology, Eastern Washington University
Advisor: Dr. A. Castillo
Helicobacter pylori gene regulation by virulence region located sRNAs
Techniques used: RT-PCR, gel electrophoresis, Rapid amplification of cDNA ends (RACE), cloning, bioinformatics, and R analysis.

Summer 2017 **Biological SIGNALS summer internship**
Department of Microbiology, University of Wisconsin Madison
Advisors: Dr. J.M. Ane and M. Pearson-Keller
Tested an *in vitro* system for analyzing if arbuscular mycorrhizal fungi (AMF) provide water to their plant host.
Techniques used: Grow and maintain plant roots and fungi in culture, qRT-PCR, root and fungal staining and dissection, compound and confocal microscopy, ELISA, and measuring root hair length and branching.

2016-2017 **Ronald E. McNair research project**
Department of Biology, Eastern Washington University
Advisor: Dr. A. Castillo

Analyzed if fecal coliforms increased following rain events in a local pond that receives storm water runoff filtered through a biofiltration cell.
Techniques used: Water sampling, membrane filtration, culturing and identifying fecal coliforms, bacterial enumeration techniques, and utilizing selective and differential media.

2016-2017 **Student research assistant for master's thesis project**
Department of Biology, Eastern Washington University
Advisors: Dr. J. Joyner-Matos and C. Higbee
Tested whether amphipods could be used as a sentinel species of mining pollution.
Techniques used: Titration, identifying macroinvertebrates to the family, proper water and catch per unit effort collection, and observing the fitness of amphipods with weight, size and swimming capacity in 10 minutes.

2015-2016 **Student research assistant for master's thesis project**
Department of Biology, Eastern Washington University
Advisors: Dr. J. Joyner-Matos and J. Yri
Analyzed if the biodiversity of the wetlands at Turnbull National Wildlife Refuge is affected by an invasive fish, brook stickleback.
Techniques used: identification of macroinvertebrates and aquatic plants down to family, and collection of water and catch per unit effort.

Teaching Experience

2018-current Graduate teaching assistant, Eastern Washington University
Biol 438 Molecular Biology (active learning classroom), two quarters
Biol 421 Medical Bacteriology, two quarters
Biol 353 Microbial Physiology, two quarters
Biol 301 Microbiology Lab, three quarters
Biol 235 Elementary Medical Microbiology Lab, two quarters
Biol 171 Biology I Lab, one quarter

2015 Undergraduate teaching assistant, Eastern Washington University
Biol 235 Elementary Medical Microbiology Lab, one quarter

Funding and Scholarships

2019-2020 Graduate Service Appointment, Eastern Washington University **($18,624)**

2018-2019 Graduate Service Appointment, Eastern Washington University **($18,624)**

2017 Daniel and Margaret Carper Foundation Specified Academic Interest (Social/Hard Sciences) Scholarship. **($10,000)**

2016 Ronald E. McNair Post-Baccalaureate Achievement Program **($4,185)**

Publications

Albrecht, V. 2017. Fecal coliforms increased in a storm drain fed pond following rain events.
NCUR Proceedings.

Presentations

Albrecht, V. and Castillo, A. 2020. *Helicobacter pylori* gene regulation by virulence region located sRNAs. Eastern Washington University Student Research and Creative Works Symposium 2020, Cheney WA. **Speaker**

Albrecht, V., Flatgard, B., and Castillo A. 2019. Characterizing a novel small RNA found within the cytotoxin associated gene pathogenicity island in *Helicobacter pylori*. American Society of Microbiology Northwest branch meeting, Seattle WA. **Poster**

Albrecht, V., Flatgard, B., and Castillo, A. 2019. Characterizing the role of sRNA cagPAI-I in *Helicobacter pylori* gene regulation. Eastern Washington University Student Research and Creative Works Symposium, Cheney WA. **Poster**

Flatgard, B. **Albrecht, V.**, Baber, C., and Castillo, A. 2019. Analysis of *Helicobacter pylori* sRNA cagII and sRNA-cagIII and the identification of genes they regulate. Eastern Washington University Student Research and Creative Works Symposium 2019, Cheney WA. **Poster**

Flatgard, B., **Albrecht, V**., and Castillo, A. 2019. Gene expression regulation in clinically relevant *Helicobacter pylori*. Spokane Area Microbiology Meeting, Spokane WA. **Speaker**

Albrecht, V. and Castillo, A. 2018. Fecal Coliforms Increase in a Storm Drain Fed Pond After Rain Events. Regional American Society of Microbiology Meeting, Pullman, Wa. **Poster**

Albrecht, V. and Castillo A. 2017. Fecal Coliforms Increase in a Storm Drain Fed Pond After Rain Events. Eastern Washington University Research and Creative Works Symposium, Cheney WA. **Speaker**

Albrecht, V. and Castillo A. 2017. Fecal Coliforms Increase in a Storm Drain Fed Pond After Rain Events. 42nd Annual West Coast Biological Sciences Undergraduate Research Conference (WCBSURC), Silicon Valley, CA. **Poster**

Albrecht, V. and Castillo A. 2017. Fecal Coliforms Increase in a Storm Drain Fed Pond After Rain Events. National Conference on Undergraduate Research, Memphis, TN. **Poster**

Higbee, C., **Albrecht, V**., Clinkenbeard, J., Davies, G., Davies, C., Johnston, L., Kenney, J., Shultz, A., Wolkenhauer, B., McNeely, C., Nezat, C., Joyner-Matos, J. 2016. Elevated Zn and Pb levels in the chain lakes of the Coeur d'Alene River, ID may contribute to the low abundance of an amphipod (Hyalella azteca). 7th SETAC World Congress/North America 37th Annual Meeting, Orlando, FL. **Coauthor**

Outreach

2019-2020	Biology Graduate Student Organization Officer. **Spokane, WA**
2019	Introducing Bacteria Event at YMCA EWU Childcare Center. **Spokane, WA**
2019	Fifth Grade Career Fair. **Spokane, WA**
2019	Spooky Science Event by Eastern Washington University. **Spokane, WA**
2019	Book It, Wiggly Science Community Event. **Spokane, WA**
2019	Eastern Washington University microbiology table at **March for Science event Spokane, WA**

Relevant Work Experience

2016-2018	Group Tutor for Program Leading to University Success (PLUS) Eastern Washington University, Cheney, WA Courses tutored: Genetics, Microbiology, and Elementary Medical Microbiology I tutored 15 people per week on average.
2016-2018	One-on-one Tutor for Program Leading to University Success Eastern Washington University, Cheney, WA Courses tutored: Genetics and Microbiology I tutored 2 people weekly.
2012-2013	Greenhouse Assistant Spokane Falls Community College, Spokane, WA My duties included cleaning pots and floors, transplanting plants, pruning, checking for pests or disease, and other general maintenance.