Characterization of Type IV Pilus System Genes and Their Regulation in *Clostridium perfringens*

Samantha R. Murray

Thesis submitted to the faculty of Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Biological Sciences

Stephen B. Melville, Chair
Ann M. Stevens
Clayton C. Caswell

April 24, 2017

Blacksburg, VA

Keywords: *Clostridium perfringens*, Type IV Pili, gene regulation, RNA sequencing

Characterization of Type IV Pilus System Genes and Their Regulation in *Clostridium perfringens*

Samantha Murray, M.S.

Virginia Polytechnic Institute and State University, 2017

ABSTRACT

Clostridium perfringens is a Gram-positive (Gr+) anaerobic pathogen that was found to contain Type IV pilus (T4P) system genes within the genomes of all its sequenced strains. T4P are widely used in Gram-negative organisms for aggregation, biofilm formation, adherence, and DNA uptake. Because few examples of T4P-utilizing Gram-positive bacteria are studied to date, we wanted to characterize the T4P system in this Gr+ bacterium.

To understand the regulation of T4P genes and therefore better understand their expression, we employed the highly powerful next-generation sequencing tool RNA-seq in a variety of conditions. RNA-seq uncovered previously unknown regulatory mechanisms surrounding T4P genes as well as provided transcriptional information for most of the genes in the *C. perfringens* strain 13 genome. We also utilized reporter gene assays to look at post-transcriptional regulation of T4P promoters.

The wealth of RNA-seq data acted as a jumping-off point for many smaller projects involving transcriptional regulators that may influence T4P expression. We investigated a novel small RNA in close proximity to the major T4P operon, as well as two little-characterized transcriptional regulators that function in the same conditions as T4P genes. RNA-seq also provided data to develop a method for protein purification from *C. perfringens* without induction.

Characterization of Type IV Pilus System Genes and Their Regulation in *Clostridium perfringens*

Samantha Murray, M.S.

Virginia Polytechnic Institute and State University, 2017

GENERAL AUDIENCE ABSTRACT

Clostridium perfringens is a ubiquitous bacterium that causes many diseases that negatively impact the public, including gas gangrene and food poisoning. This bacterium is able to infect through its ability to adhere to muscle or intestinal cells, and its infection results in breakdown of muscle tissue or severe diarrhea. In order to investigate how this bacterium senses its environment and consequently infects human beings, we looked at which genes the bacteria used in different environments, particularly on solid surfaces and in liquids. We also looked at a profile of different nutrients in order to determine which conditions cause the bacterium to use genes that start the infection process. This study impacts the literature on *Clostridium perfringens* by highlighting what physical cues signal this bacterium to start infecting, in hopes of disrupting this process and provide relief from *C. perfringens* infections in the medical community in the future.

DEDICATIONS

To *Ryan*, my future husband,

For bettering me in every way & for being at my side through the toughest challenge of my life.

To my loving and supporting parents, *Zachery* and *Carmela*,

For believing in me when I couldn't believe in myself & for teaching me to persevere and be me.

To my sister, *Danielle*,

For being my best friend through miles of distance, you were always here for me in spirit.

ACKNOWLEDGEMENTS

My biggest thanks goes to my mentor and advisor, Dr. Stephen Melville, for always giving me a direction to go in. I am a better scientist for your advice and guidance.

My deepest thanks is also extended to Dr. Ann Stevens and Dr. Clay Caswell, my graduate advisory committee. They both played a pivotal role in the success of this project and thesis. On the same note, Dr. Rick Jensen helped me understand transcriptional regulation and analysis in a way I never knew was possible. Many of these analyses are thanks to his insight and teachings.

To members of the lab, past and present, I am beyond thankful to you. Gary Camper and Angie Saadat, you are a large part of the reason I got up to go to work every day. The relationship and dynamic of our lab made the graduate experience more fun and rewarding, and I know that you will both remain great scientists after our roads part.

To the Life Sciences I collective, you have always been welcoming, passionate, and helpful in every way. I feel that you have all attributed to my success in one way or another, by offering support, advice, criticism, laughs, and an attentive ear.

TABLE OF CONTENTS

CHAPTER I

Introduction & Literature Review

1.1 Clostridium perfringens

Clostridium perfringens exists as a Gram-positive, spore-forming anaerobic bacillus of low G+C content (1). It is ubiquitous in the environment, in part due to its aerotolerance. It is found in the intestinal tracts of mammals; spores of this organism persist in the soil, sediments, and areas with exposure to human and animal excreta (2). This ancient organism is one of the most widely occurring pathogens on Earth with a rapid doubling time of nine minutes in optimal conditions (3, 4). *C. perfringens* may infect endogenously, as an opportunistic pathogen from the normal flora of the gut, or exogenously via oral infection by contaminated foods or cutaneous infection by wounds infected from non-sterile objects or soil. Diseases caused by this bacterium include clostridial myonecrosis, anaerobic cellulitis, endometritis, necrotizing enterocolitis, enteritis necroticans, and food poisoning (5, 6). This multiplicity of infectious modes may stem from the sixteen different virulence factors employed by *C. perfringens,* including fifteen toxins, the highest number produced by a single known organism (1). Types and toxotypes of *C. perfringens* are listed in Tables 1 and 2.

As an anaerobic chemoheterotrophic fermenter, it lacks genes encoding components of the tricarboxylic acid cycle and electron transport chain. Instead, it utilizes carbohydrates by glycolysis before converting pyruvate to lactate or acetyl-CoA, producing large amounts of hydrogen gas and carbon dioxide (7). It is naturally auxotrophic for eleven amino acids (arginine, glutamate, histidine, isoleucine, leucine, methionine, phenylalanine, threonine, tryptophan, tyrosine, valine) and two vitamins (pantothenate and pyridoxamine) and requires either glycine

or serine supplementation to grow (8). In nature, this saprophytic bacterium gains these nutrients by breaking down the tissues of both living and dead organisms.

Table 1 Diseases caused by *Clostridium perfringens*. Adapted from (6)

C. perfringens type	Disease produced
A	Gas gangrene (clostridial myonecrosis), food poisoning, necrotic enteritis of infants, necrotic enteritis of poultry
B	Lamb dysentery, enterotoxemia of sheep, foals, and goats
C	Enterotoxemia of sheep (struck), necrotic enteritis of animals, human enteritis necroticans (Darmbrand in Germany/pigbel in Guinea)
D	Enterotoxemia of sheep (pulpy kidney disease)
E	Enteritis of rabbits

Table 2 Toxins produced by *Clostridium perfringens*. Adapted from (9)

C. perfringens type	Toxins produced												
	α	**β**	**ε**	**ι**	**δ**	**θ**	**κ**	**λ**	**μ**	**ν**	**CPE**	**Nm$^+$**	**En***
A	+++	-	-	-	-	+	+	-	+	+	+	+	+
B	+	++	+	-	+	+	+	+	+	+	-	+	+
C	+	++	-	-	+	+	+	-	+	+	+	+	+
D	+	-	++	-	-	+	+	+	+	+	+	+	+
E	+	-	-	+	-	+	+	+	-	+	+	+	+

$^+$Nm, neuraminidase or sialidase

*En, enterotoxin

1.2 MAJOR DISEASES OF *Clostridium perfringens*

Most of the toxins produced by *C. perfringens* act on healthy cells by cell membrane disruption and pore formation. All of the toxins and hydrolytic enzymes created by this organism leave the cell through various secretion methods, excluding CPE enterotoxin, which is released upon bacterial cell lysis (10). The extent of the commonality between these toxins ends there, however, as is described in further detail through the diseases they cause.

1.2.1 Clostridial myonecrosis

More commonly known as gas gangrene, clostridial myonecrosis is a highly lethal soft tissue infection of skeletal muscle by *C. perfringens* type A. Generally, gas gangrene manifests 10 – 48 hours from exposure and presents as severe pain and pressure at the site of the wound; the disease later progresses rapidly with edema, tenderness, pallor, and hemorrhagic bullae at the site of infection (5). This fast-acting process can spread at a rate of 2 cm/h and results in systemic toxicity, overwhelming shock, and renal failure that can be fatal in as little as twelve hours (11). In the United States, approximately 1000 cases of clostridial gas gangrene are reported each year, with far more unreported cases existing in developing countries with a lack of access to healthcare. If properly treated, 70-80% of cases are resolved; unfortunately, this disease is 100% fatal if left untreated (11).

The high fatality of this disease is primarily due to the production of enormous amounts of α-toxin, a phospholipase C sphingomyelinase that hydrolyzes phospholipids and causes cell membrane disorganization; it is the major pathogenic factor causing gas gangrene (9). Alpha-toxin induces shock by affecting myocardial function, blood vessel contraction, vascular permeability, and platelet aggregation, causing hypotension and bradycardia that is often fatal in gangrenous cases (3, 9). Taken together, these actions serve the bacteria by creating an anaerobic environment and providing nutrients through cell destruction, thereby allowing the bacteria to continue living on even after the patient's death.

1.2.2 Clostridial food poisoning

The Centers for Disease Control and Prevention (CDC) states that *C. perfringens* is one of the most common causes of food poisoning, accounting for 1 million cases each year in the United States alone. It is the third most common pathogen associated with foodborne illness, contributing 10% of cases reported in 2011 (12). Meat-containing dishes like stews and gravies have the potential risk of spreading this bacterium if not properly heated to 100°C or are improperly stored without refrigeration (13). These strict heating requirements are necessary to destroy *C. perfringens* spores, which surpass the acidic stomach and become vegetative cells that produce enterotoxin in the gut (5). Affected individuals notice nausea, abdominal pain, and diarrhea typically 8 – 24 hours post-ingestion of infected foods, and due to the self-limiting nature of this infection, healthy individuals resolve poisoning within 24 hours (13). Only elderly and immunocompromised individuals suffer more severe cases; it is rarely fatal.

Symptoms from *C. perfringens* food poisoning are linked to the action of enterotoxin in the colon. Enterotoxin binding in cell membranes creates beta-barrel pores, resulting in calcium influx into the infected cells and subsequent cell death (14). Major cell death results in intestinal damage that causes fluid and electrolyte loss into the lumen. These steps in the infection process directly cause the diarrheal symptom of this disease. Interestingly, only some *C. perfringens* cells of types A, C, D, and E carry the *cpe* gene encoding enterotoxin, comprising only 5% of all *C. perfringens* isolates (14). However, it still causes widespread disease in the world today.

1.2.3 Enteritis necroticans

While the production of toxins in *C. perfringens* is of utmost importance in causing disease, the environment in which the bacterium and its toxins exist also play a major role. This is demonstrated in the uncommon cases of enteritis necroticans, also called pigbel in Papua New Guinea and Darmbrand in Germany (meaning "fire bowels") (13). A rare form of enteritis, incidence of this disease is dramatically higher in underdeveloped countries where protein deprivation, poor food hygiene, irregular meat eating, and staple diets containing trypsin inhibitors exist (15). Afflicted individuals experience vomiting, bloody diarrhea, and severe abdominal pain upon mass proliferation of *C. perfringens* type C in the gut. This disease was epidemic in Germany post-World War II and is endemic in the children of New Guinea, many times proving fatal (16).

C. perfringens type C utilizes the extracellular pore-forming β-toxin during enteritis infections. While little is known about the exact mechanism of action, it is characterized as a lethal necrotizing toxin that causes hemorrhagic necrosis and destruction of the intestinal villi

(16). Typically, β-toxin is inactivated in the gastrointestinal tract by trypsin, a digestive enzyme that catalyzes the hydrolysis of peptide bonds to amino acids to be absorbed into the bloodstream (13). However, a low protein diet reduces trypsin levels and therefore β-toxin inactivation. Trypsin levels can also decline from ingestion of trypsin inhibitors, found in high quantities in the sweet potatoes that are traditionally eaten with pork during feasts in New Guinea (17). Taken together, the extremely reduced levels of trypsin in the systems of the New Guinea population leave them highly susceptible to β-toxin's effects. Fortunately, an immunization program in Papua New Guinea to β-toxin has reduced instances of hospitalization to 20% of levels reported in the 1980s before immunization began (18).

1.3 TYPE IV PILI

Type IV pili (T4P) are long, flexible surface-exposed filaments that are homopolymers of a pilin subunit. The polymerization of these monomers by an assembly ATPase extends the pilus further through the membranes and cell wall of a Gram-negative organism (or through the cell membrane and thicker cell wall in Gram-positive organisms). This provides the cell with many advantages, including increased adherence to surfaces and DNA uptake, as well as twitching motility and biofilm formation (19-21). This system is well studied in Gram-negative organisms, which until a decade ago, were widely believed to be the only prokaryotes with T4P. However, the increasing abundance of genome sequencing has facilitated the identification of T4P gene homologs in all sequenced clostridial genomes due to their sequence similarity to the Gram-negative T4P components (22). This has led to more extensive research on Gram-positive

organisms to identify the existence of T4P and their uses, extending this motility and adhesive mechanism to many Gram-positive organisms as well. Figure 1 provides a look at how some Gram-positive bacteria utilize T4P.

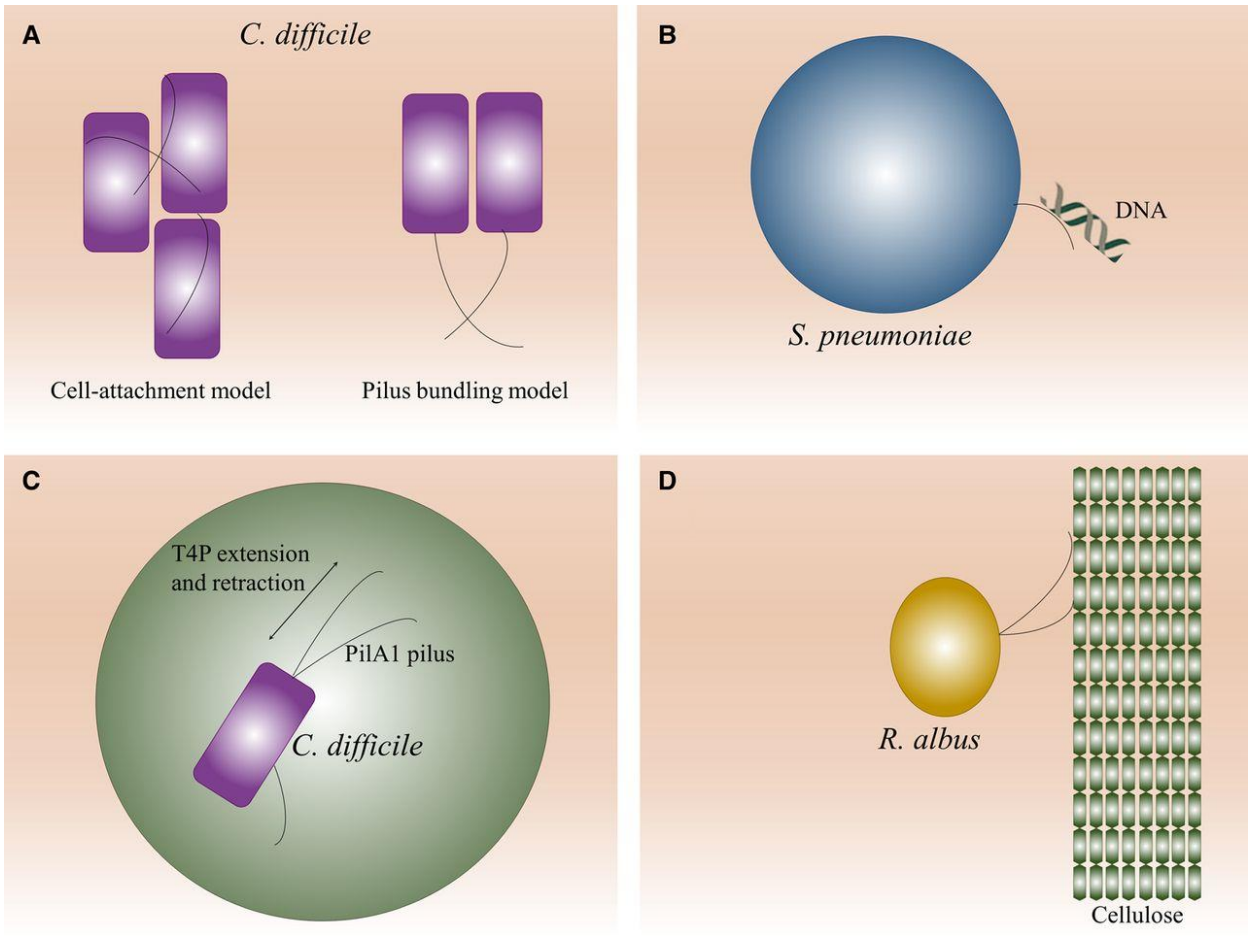


FIGURE 1. Utilization of Type IV pili for adherence and motility in Gram-positive organisms. (23) (A) *Clostridium difficile* utilizes T4P for cell-cell adherence in the formation of biofilms, either by attachment of cells in the cell-attachment model or by attachment of pili by the pilus-bundling model. (B) *Streptococcus pneumoniae* adheres to DNA during horizontal gene transfer. (C) *Clostridium difficile* also uses pili for tradition twitching motility, by extending the pilus, attaching the end to a surface, and retracting the pili to move towards the point of adherence. (D) *Ruminococcus albus* can adhere to crystalline cellulose along the entire length of the pilus.

As part of the clostridial lineage, the genome of *C. perfringens* includes genes homologous to those found in Gram-negative organisms with T4P. The possession of T4P may provide this bacterium with the capabilities necessary to infect its tropic tissues, as it is a noted virulence factor in many other pathogenic species. For example, this microbe's T4P have a specific adherence profile to myoblasts and fibroblasts (24), both of which are found in muscle tissue that *C. perfringens* necrotizes during infection. The adherence gained from T4P expression may also allow the bacteria to remain adhered during muscle contraction and relaxation (24). Three strains of this organism whose genomes had been completely sequenced were found to be capable of motility, while lacking chemotaxis and flagellar genes. These findings were the first evidence of a Gram-positive organism utilizing T4P for motility and provided new avenues for potential research on T4P (22).

Figure 2 illustrates the genomic configuration of the genes believed to be involved in T4P formation in *C. perfringens* (22). The functional designations of each gene signify the proposed function as based on genetic homology to Gram-negative organisms following the Pil nomenclature. As originally proposed, five key operons in three chromosomal loci in the genome contain groupings of genes functioning in this system. The encoded gene products are illustrated in a schematic detailing the predicted T4P system in Gram-positive organisms, as shown in Figure 3 (25).

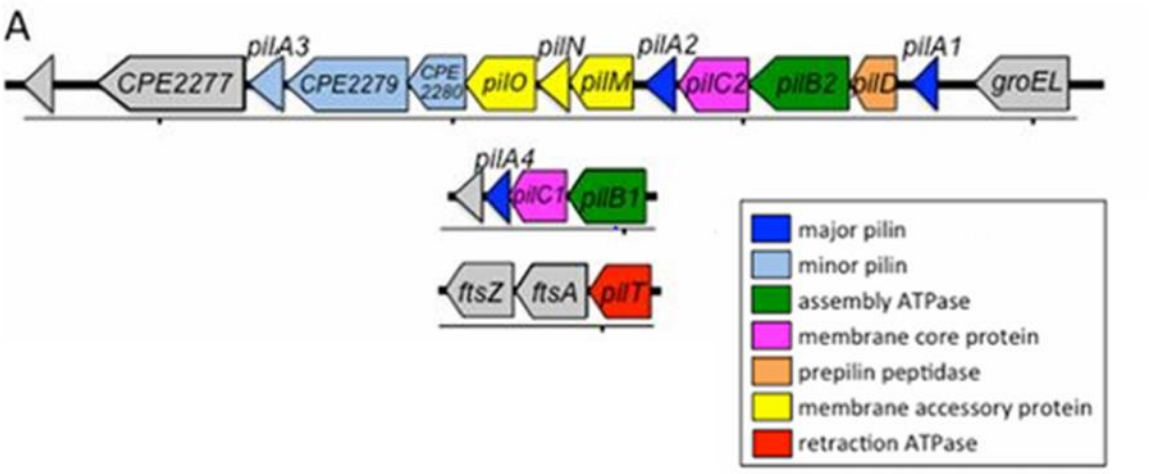


FIGURE 1. T4P gene clusters in *C. perfringens* (25). The primary T4P operon is shown as the top operon, and the secondary gene clusters are shown underneath. Protein annotations are based on sequence homology with proteins of known function, gene organization, and functional analysis. The legend identifies the supposed main function of each gene product.

The proposed Gram-positive T4P system is amenable to study due to its simplicity. While few of these functions have been experimentally proven, this model embodies the current understanding of the Gram-positive T4P system based off of Gram-negative examples. In *Pseudomonas aeruginosa*, the pre-pilin peptidase, PilD, cleaves the N-terminal sequence from the precursor pilin proteins to process them for polymerization (26). The major pilin subunit is believed to be PilA2, due to homology of the PilA2 protein structure to major pilins found in *Neisseria gonorrhoeae* and *Pseudomonas aeruginosa* (22); three other minor pilin subunits (PilA1, PilA3, and PilA4) exist and have unknown functions in this system. The assembly ATPase (PilB) hydrolyzes ATP in order to assemble the pilins into a pilus, in concert with the integral membrane core protein PilC (27). As the pilus is polymerized, it is pushed out of the cell where the terminal end can bind surfaces or cells for adherence. In the case of motility, the retraction ATPase PilT retracts the pilus while the tip is still bound, allowing for the cell to be

pulled towards the point of attachment and create forward movement. Other membrane proteins provide support for the apparatus (PilM, PilN, and PilO) or act as the secretin in the peptidoglycan that allows for the pilus to extend through it (possibly CPE2277).

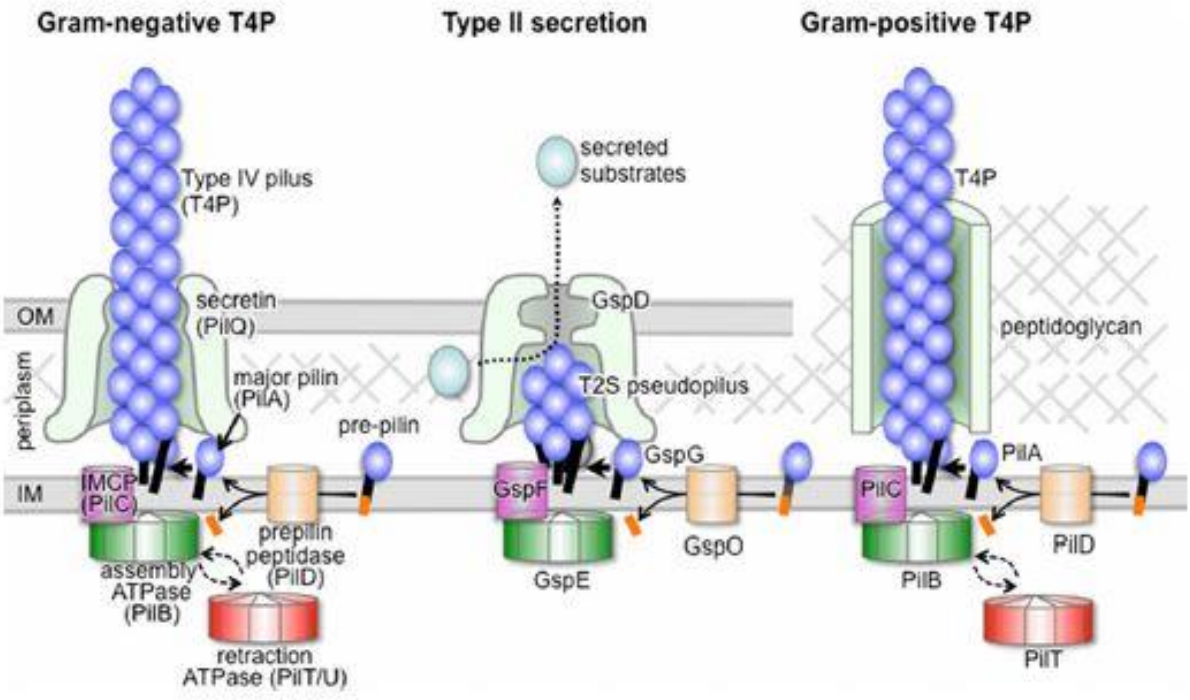


FIGURE 2. T4PS and T2SS schematic (25). Illustration of key components in the Gram-negative T4P and T2S systems and the Gram-positive T4P system and their localization in the bacterial envelope. The descriptive name for each component is shown for the Gram-negative T4P system, and common names for each component are listed for all three systems.

1.4 GENE REGULATION

Regulation of gene expression is of paramount importance to all bacteria. The ability to finely tune gene expression allows for adaptability, efficiency, and survival during environmental changes, pathogenesis, and stress conditions. For these reasons, transcriptional regulation in bacteria acts on promoters for transcribed regions that perform cooperating functions, usually in close proximity to each other, and transcribed together on a single polycistronic mRNA. These regions, termed operons, can be under the control of a single or multiple promoters which dictate the start site of transcription to the RNA polymerase holoenzyme. Operons can be finely tuned by activating and repressing elements, and termination sites exist at the ends of operons to halt transcription of downstream genes or transcribed regions (28). All of these transcriptional regulatory mechanisms must work in concert in order to generate the appropriate amount of mRNA transcript that the bacteria requires at any one time.

1.4.1 Operons

Bacterial operons are generally composed of genes whose products are linked by a common purpose, such as multiple genes in an enzymatic pathway or components needed to form a single complex. Furthermore, genes tend to be arranged in the order of their biochemical reactions. Scientists have speculated many models in the last half century as to why and how genes cluster themselves into patterns of usefulness; however, the fitness conferred by gene

clustering is demonstrated most logically by the selfish operon model proposed by Lawrence and Roth (29). In this model, genes are separated into two groups, critical and useful. Merely useful genes that only provide sporadic benefits to the cell have weak evolutionary selection to remain in the genome and are more easily lost by mutation and genetic drift. Once several accessory genes are abandoned in this way, the only reconstitution of function can occur in the event of horizontal transfer of multiple genes, and the probability of multiple gene transfer increases when genes are physically proximate. If this proximity confers increased fitness to the progeny cells, the gene cluster is more likely to spread continually between bacterial genomes and remain clustered. Thus, genes selfishly organize themselves into operons to avoid extinction.

Genes need not be clustered into operons to be transcribed; in fact, many genes exist independently in the bacterial genome with an individual promoter sequence. However, operon organization confers some advantages over co-regulated gene transcription. Upstream regions of multi-gene promoters have 10% more regulatory sequences than single gene promoters, allowing heightened control over transcription (30). Sharing regulatory sequences saves space on the chromosome and could be energetically favorable for chromosomal replication and cell division (31). This may also be the case for co-translation of proteins, due to immediate translation of polycistronic mRNAs after transcription by ribosomes. Furthermore, the stoichiometry of proteins in a complex is favored by polycistronic transcription when proteins interact on a 1:1 basis. The most conserved bacterial operons code for proteins that interact physically (32), but generally genes are needed in different quantities compared to their neighbors, which is where individual promoters and increased regulation come into play.

1.4.2 Promoters

The placement and composition of the promoter sequence upstream of a gene or operon gives the cell transcriptional power and control over gene expression. The core enzyme of the RNA polymerase (RNAP), composed of $\beta\beta\alpha_2\omega$ subunits, is capable of RNA synthesis but unable to locate the starting point of transcription without the σ subunit. Together, the core RNAP enzyme and the σ factor form the holoenzyme, which initiates transcription at the promoter and transcribes DNA to the termination site where it falls off of the DNA (33).

A promoter exists upstream of the first gene of the operon at a specific distance. Promoters in enteric bacteria have four major components: the UP sequence, the -35, the extended -10, and the -10 sequences (33). Each family and subfamily of σ factors has specific promoter sequences that it recognizes; the specificity of this recognition, and competition between σ factors from non-specific DNA-binding, determines the level of transcription by mediating how often the promoter is bound and if the RNAP can progress downstream to transcribe mRNA (34). For example, σ^{70} is a well characterized housekeeping σ factor that directs the RNAP to a multitude of promoters, due to its universal binding to many promoter regions on the genome.

1.4.3 Extracytoplasmic function (ECF) sigma factors

The σ^{70} family of σ factors comprises four groups of proteins, all of which have differing domains that correspond to areas of the promoters they recognize. Group 4, called σ^{ECF} factors, are responsible for transcribing genes in response to extracellular stimuli, such as changes in osmolality, barometric pressure, nutrient limitation, oxidative stress, and toxin molecules (35). These σ factors have a partnering protein called an anti-σ, usually transcribed and translated with the σ factor, which antagonizes its function by holding the σ at the cell membrane until an external stimulus is sensed. Once freed from the membrane with the help of a protease (36), the σ^{ECF} can interact with DNA and the RNAP to promote transcription of regions whose products are pertinent to adjusting to the stimulus.

1.5 TRANSCRIPTOMIC ANALYSES

Understanding transcription of ORFs and how they are regulated can provide essential information about infrequently transcribed regions of the bacteria chromosome, i.e. genes that are not constitutively transcribed or genes that are transcribed at low levels. Various methods currently exist for quantifying gene expression and regulation, with applications for measuring relative levels of transcripts between conditions or absolute numbers of RNA molecules in a cell at a given time. Whichever method is primarily used, it is most important that a secondary method is used for validation and that the data agree between methodologies.

1.5.1 RNA-sequencing

In the last decade, the expansion of next-generation sequencing has allowed numerous researchers to examine the transcriptome of both eukaryotes and prokaryotes using RNA sequencing (RNA-seq). The transcriptome is simply the complete set of transcripts and their quantities in a cell or population. RNA-seq is a multipurpose tool that allows for quantification of transcripts as well as determination of transcriptional structure in operons, presence of small regulatory RNAs, and differences in transcription across multiple conditions. The RNA-seq method is novel in its deep-sequencing ability but actually builds upon pre-existing technologies for quantifying transcripts. To start, a population of RNA transcripts are converted to more stable complementary DNA (cDNA) fragments with adaptors attached to the ends. High-

throughput sequencing is then performed on one end (single-end) or both ends (paired-end) of the short, double-stranded cDNA fragments. The resulting DNA fragments can be aligned to the reference genome of the organism in which the RNA was extracted, and multiple processing tools can compute the number and integrity of the transcripts aligning to particular regions (37).

RNA sequencing technology is superior to most other current technologies for RNA quantification for a few key reasons. RNA-seq is not confined to identifying transcripts that correspond to an existing genome, and it is frequently used to perform *de novo* assemblies in which transcripts from multiple organisms can be pieced together separately within a single sample, even when a reference genome is not known. RNA-seq can determine transcriptional boundaries at the resolution of a single base pair. It has little background due to the nature of its cDNA library creation and has no upper limit in the number of transcripts it can detect (37). This is ideal for comparing samples in varying conditions, as it allows for quantification of large copies of transcripts and high fold-changes between samples. However, this is not to mean that RNA-seq is without flaws; long RNA fragmentation and resulting library biases can occur, and PCR artifacts can resemble high transcript numbers at certain locations (38). Shorter transcripts can also be lost if they fall below the chosen read length, which has continually gotten longer in recent years due to dropping sequencing costs and the notion that longer reads are more informative (39). These reasons strengthen the need for validation from other techniques, such as quantitative reverse transcription polymerase chain reaction (qRT-PCR).

1.5.2 Quantitative reverse transcription polymerase chain reaction

Once accepted as the "gold standard" for the analysis of gene expression, qRT-PCR allows for amplification and quantification of DNA targets as a polymerase chain reaction progresses (40). Quantification of an initial DNA template is done by comparing the number of PCR amplification cycles it takes for a target molecule to reach the threshold of detection. Targets with higher starting concentrations reach the threshold sooner, thus allowing relative quantification to occur. qRT-PCR quantifies DNA products by monitoring the intensity of fluorescent probes or dyes within the reaction (41). In many research labs, RNA-seq has replaced the use of qRT-PCR as the first means of identifying differential expression of genes. qRT-PCR is a much cheaper alternative that provides results that have been consistently shown to align well with RNA-seq data and acts to validate many RNA-seq experiments (10, 42, 43). The key to validation is careful selection of experimental parameters, particularly reference genes (44). Relative quantification is based on the relative expression of a reference gene, typically a chosen gene that maintains consistently moderate to high expression across samples, allowing for simple normalization and comparison to genes of interest which may have varying expression in different conditions. It is important to note that sample preparation is a critical element in the validation process. Validation using the same samples assayed during an RNA-seq experiment will only validate the technology itself. A new set of samples identical to those used for RNA-seq will validate the biological conclusions drawn from the data and provide a stronger argument for differential expression between samples (45).

1.5.3 Reporter gene assays

A faster, simpler technique for evaluating transcriptional and sometimes translational activity of certain promoters is the use of reporter gene assays. Typically, a promoter region or transcribed region of interest is fused via PCR to the full sequence of a reporter gene, a gene encoding an enzyme whose activity can be easily quantified via assays (46). Transcription through the promoter can then be calculated using the activity of the enzyme as related to the optical density of the sample. This method can also be extended to give translational information as well if the native ribosomal binding site is left intact during construction of the promoter fusion. There are multiple genes that are regularly used as reporter genes for measuring transcriptional activity, such as β-galactosidase, luciferase, and β-glucuronidase, but no enzyme is the best choice for all applications. The best reporter enzyme for a particular use is dependent on the basal level of the enzyme's activity in the organism being studied, which ideally is lowly expressed normally, not expressed in the organism, or has only one copy of the gene that can be mutated before assaying. While limitations to this technique include varying sensitivities of enzymes and assays as well as endogenous activity, the versatility and simplicity of these assays are most attractive (47). Once the initial fusion strain is constructed, assays can be performed on cells grown in a number of conditions and screened for transcription through the promoter in question. Moreover, these assays can provide translational information as well if the native ribosomal binding site is maintained upstream of the promoter fusion.

1.6 INTRODUCTION TO THESIS

This project grew from an interest to quantify Type IV pili on the surface of pilin gene deletion mutants in order to assess which genes were necessary for pilus formation and which genes encoded accessory proteins that were non-essential. Initial experiments were unable to quantify pili at high enough levels in wild type HN13 cells to see significant differences when particular genes were deleted, especially *pilA2*.

In order to rectify this problem, the course of the project changed to investigate the basis of Type IV pilus formation, the transcription of T4P themselves. Chapter 2 introduces our solution to this problem, a carefully designed RNA-sequencing experiment that was aimed at distinguishing the difference in T4P expression in cells grown in liquid culture and on agar surfaces. To extend our understanding of T4P expression beyond the transcript level, we also characterize translational activity through T4P promoters by way of transcriptional and translational reporter gene fusions. The end of Chapter 2 elucidates the beginning of our hunt for regulators that control when T4P are expressed, such as a highly regulated small RNA.

Chapter 3 goes further into depth in our discovery of T4P regulators. We utilized the wealth of RNA-seq data at our disposal to find possible regulators of T4P expression, as well as pursuing another well-known regulator of T4P. Chapter 3 outlines some smaller projects performed to characterize T4P proteins better, through purification of the T4P accessory protein PilM, creation of strains for crystallization of the minor pilins, and investigation of the floating phenotype that was abolished in the *pilA3* deletion mutant. Chapter 4 summarizes the results of these efforts and provides insight into our findings.

CHAPTER II

Differential gene expression analyses of Type IV Pilus System genes in *Clostridium perfringens*

INTRODUCTION

Bacteria have evolved numerous ways to be motile, and types of bacterial motility differ based on the utilized appendage performing the movement and the physical condition in which it occurs. For example, flagellar motility can simply describe any net movement from the use of a rotating flagella, but swimming and swarming motility are quite different (48). Swimming implies the bacteria are immersed in a liquid environment, while swarming motility is only observed when bacteria are cultured on varying concentrations of solid agar (49). Other types of surface motility include gliding motility, the active surface movement of cells along the long axis without any appendage, and twitching motility, a surface-dependent movement powered by the extension and retraction of Type IV pili (T4P) (25, 50). Because many bacterial species have multiple types of motility they utilize for specific environments, an underlying mechanism must cause the switch between types of motility when the external environment changes.

In order to switch between planktonic and surface-associated lifestyles such as biofilms, bacteria have evolved the ability to detect the presence of a surface. Bacteria that possess flagella can sometimes sense surfaces by inhibition of their flagellar rotation (51). One such case is with the Gram-negative bacterium *Pseudomonas aeruginosa*, which couples its viscosity-dependent flagellar inhibition with increased synthesis of cyclic-di-GMP via the SadC enzyme (52). This activates processes involved in biofilm formation by initiating reversible cell attachment and inhibiting flagellar biosynthesis.

A Gram-positive example of surface sensing occurs in *Bacillus subtilis*, in which the two-component regulatory system DegS and DegU senses increased motor torque in the basal body of the flagella in response to increased viscosity from a surface. The autophosphorylation of DegS and subsequent transfer of the phosphate group to DegU results in transcriptional repression of flagellar and chemotaxis genes and activation of biofilm matrix biosynthesis (53).

Non-flagellar surface sensing mechanisms have also been investigated and characterized. *Proteus mirabilis*, an enteric bacterium, uses the Rcs phosphorelay to link contact of the O-antigen to surfaces as a signal to switch to a swarming phenotype, resulting in elongated cells with hundreds of flagella (54). In *Pseudomonas aeruginosa*, the Wsp system responds to an unknown surface signal resulting in a methylation event of membrane-bound methyl-accepting chemotaxis protein WspA (55). Phosphorylation of the response regulator, WspR, leads to increased cyclic-di-GMP levels and further irreversible microcolony formation, the beginning of mature biofilm formation in this organism (55, 56).

The Gram-positive anaerobic pathogen *Clostridium perfringens* presents a unique opportunity to study alternative methods of surface sensing, as it lacks both flagellar and chemotaxis genes (22). On plates, *C. perfringens* cells exhibit gliding motility by forming long, thin filaments of cells linked end-to-end, a structure that resists bending and maximizes movement away from the larger colony. These gliding cells are highly elongated, flexible, and express T4P. Liquid-grown cells have the ability to float, not swim, in liquid media by a currently unknown mechanism (see Chapter 4.4); they remain suspended in the fluid column as individual cells and exist as shorter cells than *C. perfringens* grown on agar plates (3-4 μm versus 5-6 μm, respectively). *C. perfringens* forms biofilms in static liquid cultures, a mechanism dependent on T4P and carbon catabolite protein CcpA (57). However, the

mechanism that induces formation of a biofilm in liquid is not currently known, much like the mystery of this bacterium's surface sensing mechanism and how T4P play a role in the transition between floating cells in liquid and gliding or twitching cells on plates.

In order to investigate the surface sensing mechanism of *C. perfringens* and how T4P are regulated, we designed an RNA-sequencing experiment using three media with differing metabolic profiles and grew wild type cells on agar plates and liquid media of all three. In search of transcriptional regulators that flip the switch between planktonic and surface-associated phenotypes, we compared the transcription levels of nearly all of the genes, including T4P genes, and found commonalities between transcriptomes of plate-grown cells and differences from liquid-grown cells. We also conducted translational promoter fusions to known and newly discovered T4P promoters to assess the rate of synthesis of T4P in different media, in plates versus liquid, and at differing temperatures.

MATERIALS & METHODS

Bacterial strains and culture conditions. Bacterial strains, plasmids, and primers used in this study are listed in Tables 3 and 4. *Escherichia coli* strain DH10B was grown in Luria Bertani broth at 37°C for all transformations. When necessary, kanamycin and chloramphenicol were added to the media at a concentration of 100 μg/ml and 20 μg/ml, respectively. *Clostridium perfringens* strain HN13, a Δ*galKT* derivative of strain 13 (58), was used as the wild type strain in this study. *C. perfringens* strains were grown anaerobically in PGY (30 g proteose peptone #3, 20 g glucose, 10 g yeast extract, 1 g sodium thioglycolate per liter), BHI (brain-heart infusion, Thermo Fisher), or FABG (LAB M fastidious anaerobe broth + 2% glucose) in an anaerobic chamber (Coy Laboratory Products, Inc.). Strains derived from strain AH2 (59) which also contained the β-glucuronidase reporter gene vector pJV50, required 30 μg/ml erythromycin and 20 μg/ml chloramphenicol to maintain chromosomal insertion constructs.

Promoter-*gusA* fusions in the chromosome. Introduction of chromosomal promoter fusions to the reporter gene *uidA* has been described elsewhere (60). Briefly, the promoter regions of T4P operons and the reporter gene *uidA* from *E. coli* were amplified and joined together using overlapping PCR. The resulting PCR product and suicide vector pJV50 were digested with SalI-HF and PstI-HF, and the vector and PCR product were ligated using T4 DNA ligase (Promega). Transformants were screened for correct constructs using agarose gel electrophoresis, and a high concentration of plasmid DNA was extracted from resulting strains using a ZymoPure Midi Prep

Kit (Zymo Research) following the manufacturer's protocol. Forty micrograms of suicide plasmid were electroporated into *C. perfringens* strain AH2 using the method described previously (59), and the chromosomal DNA from resulting strains was checked for homologous recombination via PCR using primer pairs OSRM24+39, OSRM50+39, OSRM53+39, OSRM27+39, OSRM30+39, and OSRM36+39 for the respective strains (see Table 4).

β-glucuronidase assays. Reporter gene promoter fusion strains were utilized to assess promoter activity of T4P operons at their chromosomal loci. The β-glucuronidase assay was performed as previously described (60). Briefly, cells were harvested from either the edges of colonies on plates or pelleted from a liquid culture in mid-log phase and resuspended in 1 ml DPBS (Dulbecco's phosphate-buffered saline). The OD_{600} was obtained, and cells were microcentrifuged for 5 minutes to pellet cells before resuspending them in 0.8 ml assay buffer (50 mM $NaHPO_4$ [pH 7.0], 1 mM EDTA, 5 mM dithiothreitol). Cells were vortexed for 1 minute with 8 µl toluene and put on ice for 10 minutes. Samples then sat in a 37°C water bath with caps open for 30 minutes. The assay was initiated by addition of 160 µl 6mM 4-nitrophenyl D-β-glucuronide (Sigma Chemical Co.), and after further incubation, the reaction was halted by addition of 400 µl 1M Na_2CO_3. Cellular debris was pelleted for 10 minutes, and A_{405} was measured in a Genesys 10S UV-VIS spectrophotometer (Thermo Scientific). The specific activity of β-glucuronidase enzyme in each sample was calculated using the following equation: specific activity (in Miller units) = (A_{405} x 1000)/(OD_{600} x time [minutes] x culture volume [milliliters]).

Isolation of total RNA. *C. perfringens* HN13 cells were grown in duplicate in six conditions: liquid culture to mid-log phase and 1% agar plates of BHI, PGY, and FABG media for 24 hours. Liquid cultures were standardized to OD_{600} = 0.5 and cells were scraped from the outer edges of plate grown colonies and resuspended in DPBS to OD600 = 0.5. Bacteria were pelleted and frozen in liquid nitrogen prior to RNA purification. Cells were lysed in 500 µl Tri Reagent (Zymo Research) using high-impact zirconium beads in a Mini-Beadbeater (Biospec) for one minute and kept on ice. RNA was then purified from cell lysate using a Direct-Zol RNA Mini-Prep Plus Kit according to the manufacturer's protocol (Zymo Research), including an on-column DNA digestion. RNA integrity was measured using an Agilent BioAnalyzer 2100 (Biocomplexity Institute of Virginia Tech), and samples with a RIN of 8.3 or greater were used for RNA sequencing.

RNA-sequencing and analysis. Library construction was performed for Illumina sequencing by the Biocomplexity Institute of Virginia Tech. All samples were processed using HiSeq Illumina sequencing creating 100-bp paired-end reads, and the resulting data was received in fastq format and aligned to *C. perfringens* strain 13 reference chromosome (NCBI, accession number NC_003366) using the bioinformatics mapper Geneious version 9 with the Geneious mapping algorithm on low sensitivity. Transcripts were aligned to all annotated coding regions. Relative expression levels of annotated coding regions were calculated in Geneious with units of transcripts per million (TPM). TPM, as proposed by Wagner (61), is calculated as: TPM = (coding sequence read count x mean read length x 10^6)/(coding sequence length x total transcript count). Differentially expressed genes were identified by a stringent cutoff of fold-change > 4 when media conditions were compared against one another for each gene.

qRT-PCR. A new set of RNA was extracted from wild type samples using the same growth and extraction conditions described above for RNA-sequencing. RNA was quantified using a NanoPhotometer (Implen) and checked for quality using an Agilent BioAnalyzer 2100 (VBI). All RIN values were above 8.2. The extracted RNA was converted to cDNA using a High Capacity cDNA Reverse Transcription kit (Life Technologies) per the manufacturer's instructions. The cDNA was quantified using a NanoPhotometer (Implen), tested for purity by measuring absorbance ratios at 260/280 nm and 260/230 nm, and used as the template in a 7300 Real-Time PCR System (Applied Biosystems/Life Technologies). Primer pairs (Table 4) for five pilin genes of interest plus the control gene *lon* were designed using the software Primer Express, version 3 (Life Technologies) and optimized to 100% ± 10% efficiency using cloned coding regions of each gene as the template (Table 3). Parameters for qRT-PCR primer design were as follows: 20-30 bp in length, 80-120 bp amplicon, 35% GC content or higher, and Tm=58°C or higher (pairs not deviating by more than 1°C). Template DNA (either plasmid or cDNA) was used at concentrations of 0.001 ng to 20 ng per 20 µl reaction containing 0.8 µl 10 µM specific forward and reverse primer, 10 µl 2x qPCRBIO SyGreen Mix Hi-ROX (PCRBiosystems), and 7.4 µl dH2O. Reactions were performed in MicroAmp Optical 96-well reaction plates in triplicate for each cDNA sample. Thermal cycler settings were programmed as follows: 95°C for 2 min, 40 cycles at 95°C for 5 seconds and 60°C for 30 seconds, followed by a dissociation stage during primer optimization to confirm specific product amplification. Data was collected during stage 2 and analyzed through 7300 System SDS RQ software, version 1.4 (Life Technologies), using an automated cycle threshold, and relative expression level ranges were calculated using the ΔΔCt method as described in the Applied Biosystems user bulletin "Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR".

Northern blotting. RNA was isolated from plate grown cells and liquid cultures in a manner identical to that used for the RNA-seq experiments (above). Ten micrograms of RNA was loaded onto acrylamide gels and processed for Northern blots as previously described (62). Northern blotting primer used for *sr79* was OSRM97.

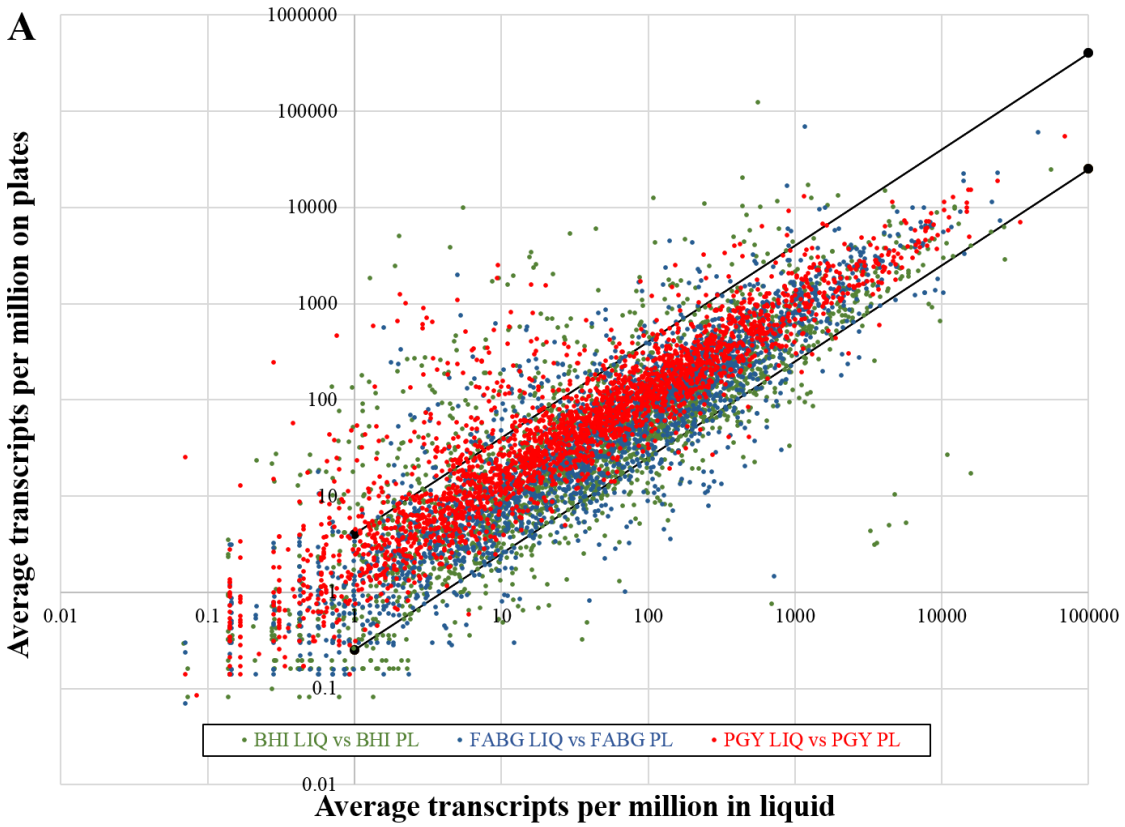
Cell length measurements. For plate-grown cells, bacteria were grown on BHI plates with 20 μg/ml chloramphenicol and 1 mM lactose for 18 h and cells at the edge of the colonies were scraped off and suspended in DPBS. For liquid growth, overnight cultures grown in BHI with 20 μg/ml chloramphenicol were diluted 1:50 in fresh BHI with 20 μg/ml chloramphenicol and 1 mM lactose and grown to an OD600 of ~0.5 (mid-log phase of growth). Cell length measurements were obtained using phase contrast images obtained from an Olympus IX81 upright microscope linked to DeltaVision imaging software package (Applied Precision). Images were analyzed with image analysis program ImageJ (63).

RESULTS

Hundreds of transcript levels significantly increase due to plated versus liquid media. From previous adherence assays, PilA2 protein levels were known to be greater in plate-grown *C. perfringens* cells than in cells harvested from liquid culture (data not shown). We wanted to investigate if this change in protein levels was due to changes in transcript levels, so we used RNA-sequencing to identify the transcripts of all genes in the *C. perfringens* HN13 genome.

The experimental design was as follows: duplicate samples of cells were grown in liquid and on 1% agar plates of three media, BHI, PGY, and FABG. RNA was extracted from all twelve samples and sequenced. Figure 4A shows the four-fold cutoff lines for average normalized transcripts; every point represents a single gene, and genes outside of the four-fold lines indicate upregulation in that condition. To identify specific genes that were upregulated on agar plates when compared to liquid media, we took the ratio of normalized transcript levels on plates over liquid media for each gene and used a four-fold difference as the cut-off for significance. This method indicated that hundreds of genes were upregulated in response to the 1% agar of the plate for each media (Fig. 4B). Comparisons of transcripts in BHI plate-grown bacteria over liquid-grown showed that 431 genes were upregulated, 549 in PGY, and 286 in FABG. We also found commonalities between pairs of media and between all agar media; for example, 49 genes were always highly upregulated on plates. There were far fewer genes upregulated in liquid than in their plate counterparts, with only four genes always having higher amounts of transcripts in each liquid media (Fig. 4C). Upregulated genes on plates and in liquid

(and the gene annotation as reported by MicrobesOnline.org) are listed in Table 5. Tables for two-way comparisons can also be found following this chapter (Tables 6, 7, and 8). Most genes upregulated on plates in all three media are metabolic in nature and aid in carbon utilization. None of the genes upregulated in either of the three-way comparisons stood out as transcriptional regulators, however. While repressed genes were considered as possible regulators, the small number of genes that were upregulated in liquid cultures and the unknown nature of their gene products led to a narrowing of this search to purely plate-upregulated genes.



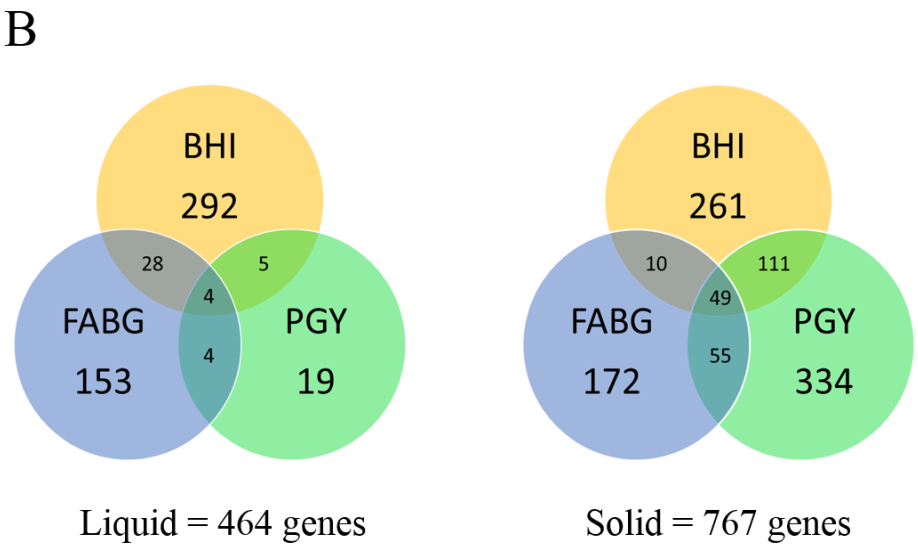


FIGURE 3. Comparisons of RNA-seq datasets. (A) Duplicate samples were averaged and normalized transcript levels in liquid media (x-axes) and plate media (y-axes) were plotted against each other on a logarithmic graph. Black lines indicate four-fold boundaries; each point is representative of a single gene. Points lying far outside of the four-fold lines indicate upregulated genes favoring a particular condition. (B) Number of genes upregulated four fold or greater in each plate condition. (C) Number of genes upregulated four fold or greater in each liquid condition.

Multiple T4P genes are upregulated on plated media containing glucose. We compared the normalized transcript levels (TPM) of genes in the major pilin operon, running from *cpe2277* to *cpe2288* (*pilA1*). There was no significant difference in transcription of any of the genes between the three liquid conditions. (Figure 5A). Overall, transcript levels on plates of these genes was about twice as high as the liquid counterparts (compare Fig. 5A and 5B).

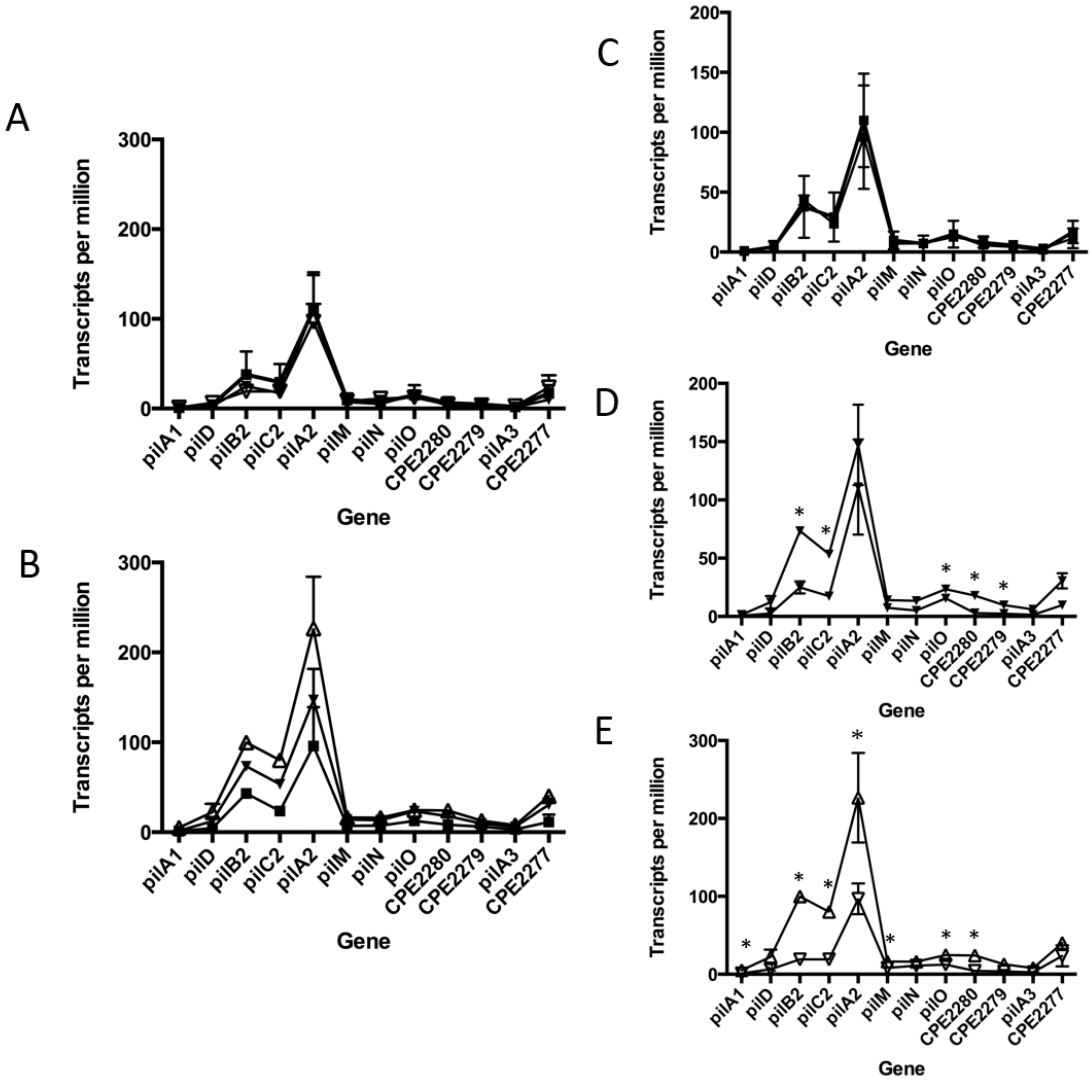


FIGURE 5. Normalized transcript levels of each gene in the major pilin operon, running from *cpe2277* to *pilA1* (*cpe2288*). (A) Transcripts per million (TPM) of each gene in three liquid conditions. (B) TPM of each gene in three plate conditions. (C) BHI plate and liquid (D) PGY plate and liquid (E) FABG plate and liquid. Asterisks (*) represent significantly different levels between samples using Student's t-test ($p<0.05$). In CDE, the upper line is the plate condition.

A comparison of each media in the solid and liquid state revealed that increased transcription of the pilin genes is likely glucose-dependent. The BHI condition showed no differences between samples (Fig. 5C). In contrast, PGY and FABG showed multiple pilin genes that had higher transcripts per million in plate-grown cells; particularly, *pilB2*, *pilC2*, and *pilA2* follow this trend (Fig. 5D and 5E). *pilA2* was only significantly higher on FABG plates compared to the other five conditions, and it had the highest transcription levels of all the genes in the operon, followed by *pilB2* and *pilC2* (Fig. 5E).

Operons containing T4P genes have complex regulatory patterns that correspond to transcript levels in RNA-seq. Previous studies from our lab predicted operon structures for the seventeen T4P-associated genes found in the *C. perfringens* strain 13 genome (22). RNA-sequencing has shown a much more complex regulatory environment surrounding these genes than previously predicted. Using transcriptional boundaries and coverage depth within genes, we were able to construct a new operon map for the T4P genes (Figure 6). Adjustments to our previous operon map are discussed in upcoming figures.

pilD, encoding for a pre-pilin peptidase, is the first gene in a long operon of T4P genes and is controlled by an upstream promoter. Figure 7 shows that the next gene in the operon, *pilB2*, also has a promoter directly upstream of its TTG start site. Coverage graphs (Figure 7A) show a sharp increase in transcripts aligning to the strain 13 chromosome directly after the sequencing of the predicted promoter region (bp 2631846-2631878). In one condition (BHIPL2), average transcripts per base were 33 times higher after the promoter sequence than before it (Figure 7B). Coverage between the genes falls to zero in four of twelve samples examined (Fig. 7A).

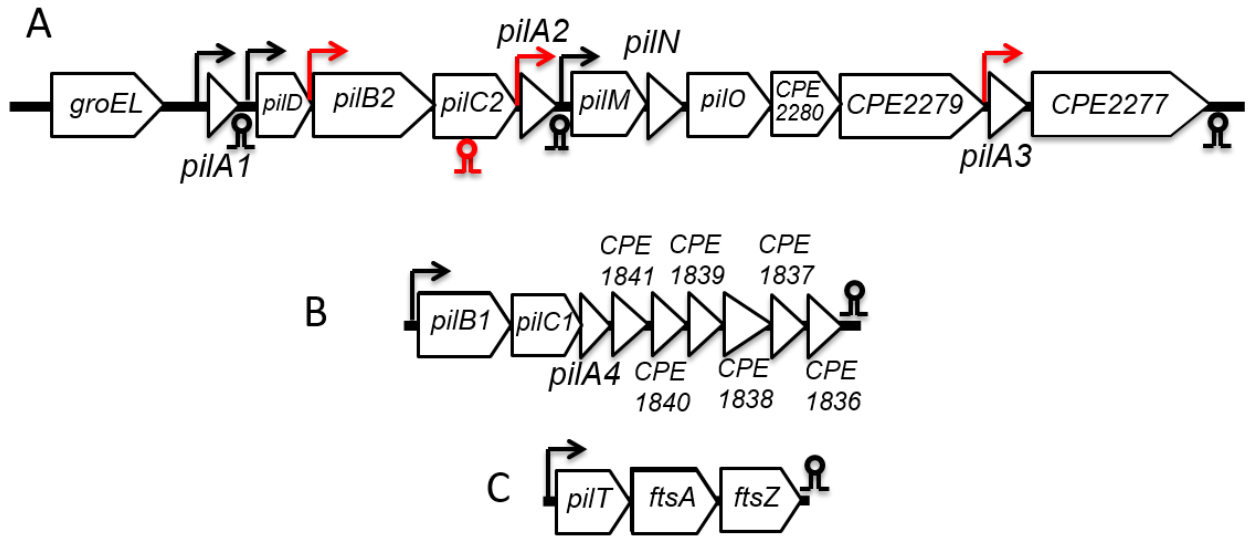


FIGURE 6. Revised Type IV pilus operon maps. (A) Operon map of major T4P operon in strain 13. (B) Operon map of minor pilin operon in strain 13. (C) *pilT*, the retraction ATPase necessary for T4P-mediate motility, lies in an operon with known cell-division genes *ftsA* and *ftsZ*. Terminators are predicted rho-independent terminators. Red color indicates new regulatory features found by RNA-seq.

The *pilC2* gene is also regulated in a complex manner due to the presence of a terminator in the middle of the coding region of the gene, corresponding to a sharp decrease in transcripts aligning within the gene (Figure 8A). The drop in transcripts corresponds directly to the end of the terminator sequence, at bp 2629539 (Fig. 8C). Transcript levels fell to 15% of pre-termination levels in some samples (Fig. 8B). This mid-gene termination pattern was present in all samples, regardless of plate or liquid condition. Interestingly, the terminated *pilC2* sequence encodes for a truncated PilC2 protein that maintains a full cytoplasmic domain as well as one full transmembrane domain (Fig. 8D). Previous purification methods of PilC2 in our lab uncovered a truncated version of PilC2 using Western blotting from protein preparations out of *Pseudomonas aeruginosa* (data not shown), but mid-gene termination was not considered as a source of this shortened protein until this RNA-sequencing data analysis.

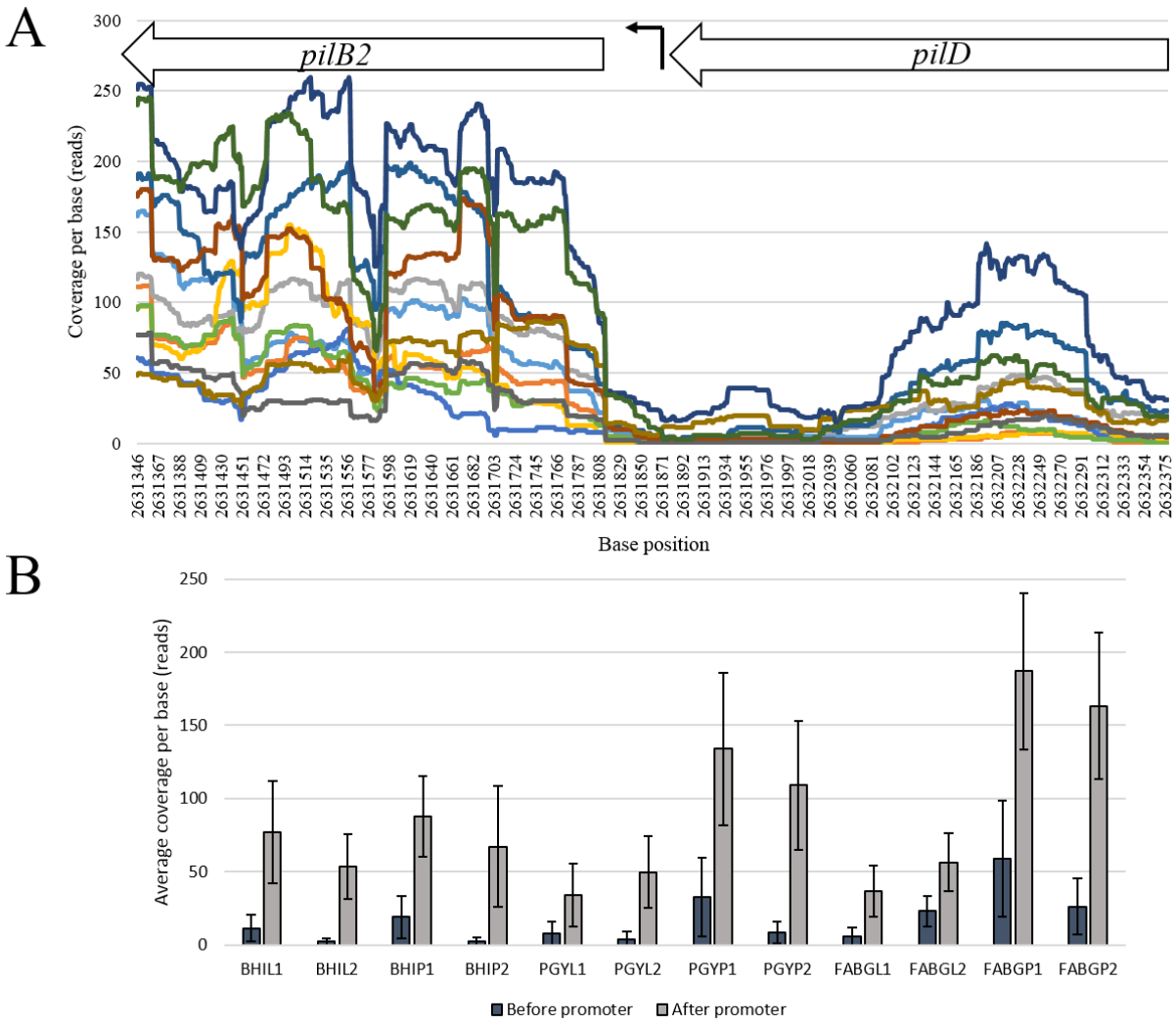


FIGURE 7. Promoter region located upstream of *pilB2*. (A) Coverage map of transcripts aligning 500 bp upstream and downstream of the suspected promoter between *pilD* and *pilB2* for all twelve samples. (B) Average number of transcripts aligning per base 500 bp upstream of *pilB2* promoter (before, blue) and 500 bp downstream of promoter (after, grey) for twelve samples. Error shown is standard deviation. All difference are significantly different using Student's t-test ($p<0.001$).

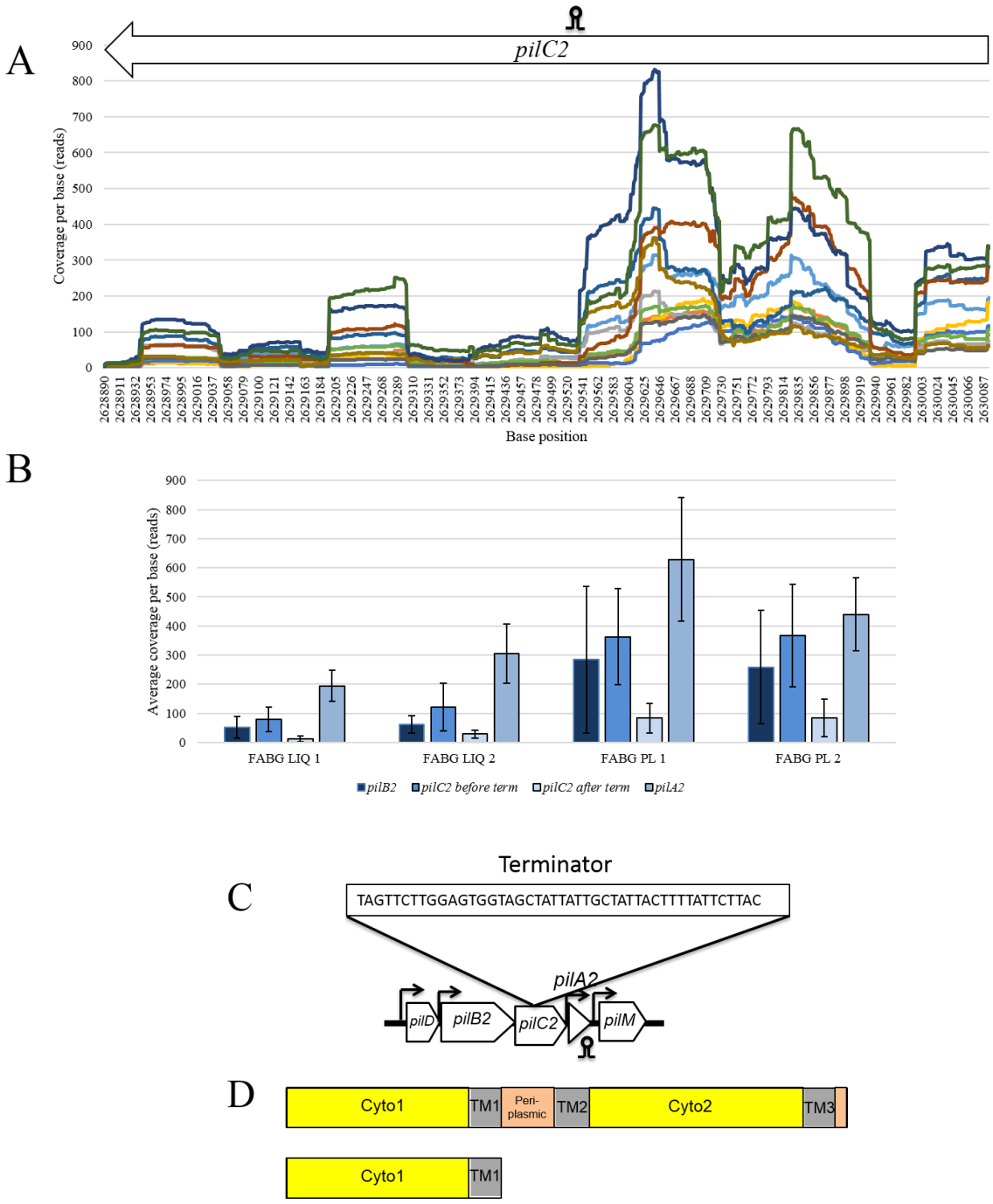


FIGURE 8. Termination during coding sequencing of *pilC2*. (A) Coverage map of transcripts aligning upstream and downstream of the suspected terminator interrupting *pilC2* coding region for all twelve samples. (B) Average number of transcripts aligning per base in *pilB2*, *pilC2* coding region before terminator, *pilC2* coding region after terminator, and in *pilA2*. Only FABG samples are shown, although all samples show this trend (see A). Error shown is standard deviation. Difference between *pilC2* regions is significantly different in all samples using Student's t-test ($p<0.001$). (C) Operon map and terminator sequencing inside *pilC2*. (D) Top: protein domains of full length PilC2, bottom: protein product of terminated *pilC2* gene.

Pilin genes *pilA2* and *pilA3* have transcription initiation sites. The alignment software Geneious utilized in the processing of the RNA-sequencing experiments creates volcano plots of coverage per base. Figure 9 illustrates the log-scaled volcano plots, or coverage maps, of the DNA surrounding *pilA2* (Fig. 9A) and *pilA3* (Fig. 9B). Both of these genes encode for pilins that are polymerized into T4P. The breaks in transcription, shown by gaps in the coverage maps upstream of these pilin genes, could be transcriptional termination and restarts or RNA processing sites. Bacterial promoter prediction software BPROM (64) also predicts promoters in these gaps of transcription. Interestingly, evidence of transcripts begin 40 – 100 bases into *pilA3* in all conditions, even with the promoter outside of the coding region itself.

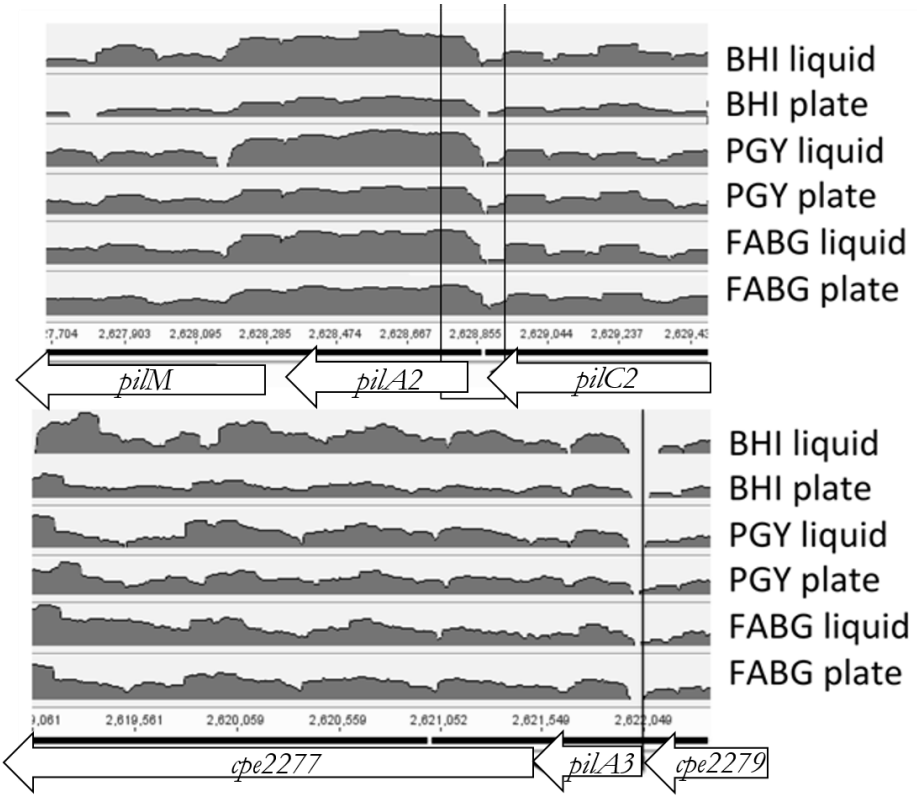


FIGURE 9. Coverage maps of *pilA2* and *pilA3* promoters. (A) Upregulation of transcription at predicted promoter region upstream of major pilin gene *pilA2*. (B) Restart of transcription after coverage gap upstream of *pilA3*. Sold line indicates start site of *pilA3*. One representative sample from each condition is shown. Volcano plots represent the number of transcripts aligning per base on a log scale.

The *pilB1* operon is transcribed at low levels and may contain nine genes. The original annotation suggested that *pilB1* was the first gene in a four gene operon with *pilC1*, *pilA4*, and *cpe1841* (22). RNA-sequencing data show few to no reads aligning in these genes in either liquid (Figure 10A) or plate-grown cells (Fig. 10B). The neighboring five genes downstream of this operon are ORFs on the same strand of DNA and are also transcribed at this very low level. (Fig 10A and B). The operon prediction program DOOR predicts in strain 13 that these genes exist as two separate operons, but in *C. perfringens* strains SM101 and ATCC 13124, DOOR predicts that these nine genes comprise a single operon (65). From DOOR prediction, strain 13 is the only strain of *C. perfringens* sequenced that these genes could be two operons. Due to the lack of an identifiable terminator after *cpe1841*, and the low levels of transcripts aligning across the region, we hypothesize that all nine genes from *pilB1 to cpe1836* are part of the same operon.

***pilT-ftsA-ftsZ* comprises one regulated operon.** Previous qRT-PCR experiments from our lab suggested that *pilT* was co-transcribed with the *ftsA* gene from *C. perfringens* strain SM101 (66). The RNA-sequencing data collected for strain 13 indicate similar results using coverage maps. In all samples, these three genes are highly transcribed and show no termination or gaps between the pilin gene *pilT* and the two cell division genes *ftsA* and *ftsZ* (Figure 10C). These data support the previous findings that *pilT* is co-transcribed with *ftsA* and *ftsZ*, a gene synteny that is conserved in all of the *C. perfringens* strains that have been sequenced.

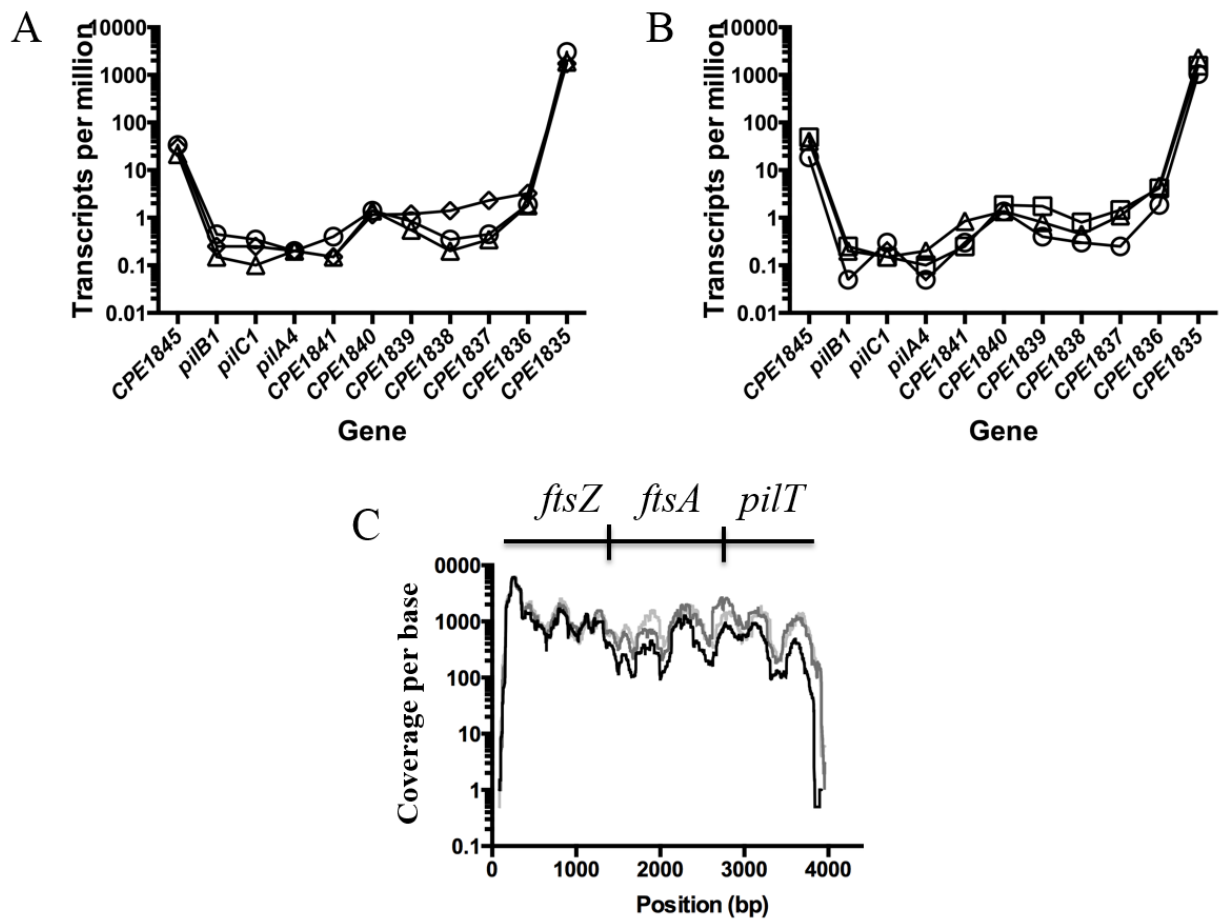


FIGURE 10. Transcription of minor T4P operons. (A) Normalized transcript levels of *pilB1-pilC1-pilA4-cpe1841* operon and surrounding genes, including predicted operon *cpe1840 – cpe1836*. Values shown are averaged for each liquid condition. (B) Normalized transcript levels of *pilB1-pilC1-pilA4-cpe1841* operon and surrounding genes, including predicted operon *cpe1840 – cpe1836.* Values shown are averaged for each plate condition. (C) Coverage map of *pilT-ftsA-ftsZ* operon in three representative plate conditions.

Transcriptional and translational reporter gene fusions indicate increased regulation at the translational level. In order to determine if translation of T4P proteins was comparable to the levels of transcripts indicated by RNA-seq, we designed constructs for each of the seven putative promoters to report translational activity on the chromosome. Each fusion was designed to allow for retention of the native ribosomal binding site of the first gene downstream of each promoter, and the reporter enzyme gene *uidA* from *E. coli* was placed downstream to measure transcriptional and translational activity of the first gene (Figure 11A). Mutant strains and the wild type, an empty vector strain, were grown in identical conditions as those used for RNA-sequencing: BHI, PGY, and FABG plates and liquid. Plate-grown cells exhibited a much higher degree of variation in expression than those grown in liquid (Figure 11B & C). The *pilB1* and *pilT* promoters showed the highest variation, with 20 and 30-fold lower levels of β-glucuronidase activity in BHI versus FABG plated cells (Fig. 11C). Additionally, the *pilA1* promoter showed significantly higher activity on PGY plates than the other two media (Fig 11C).

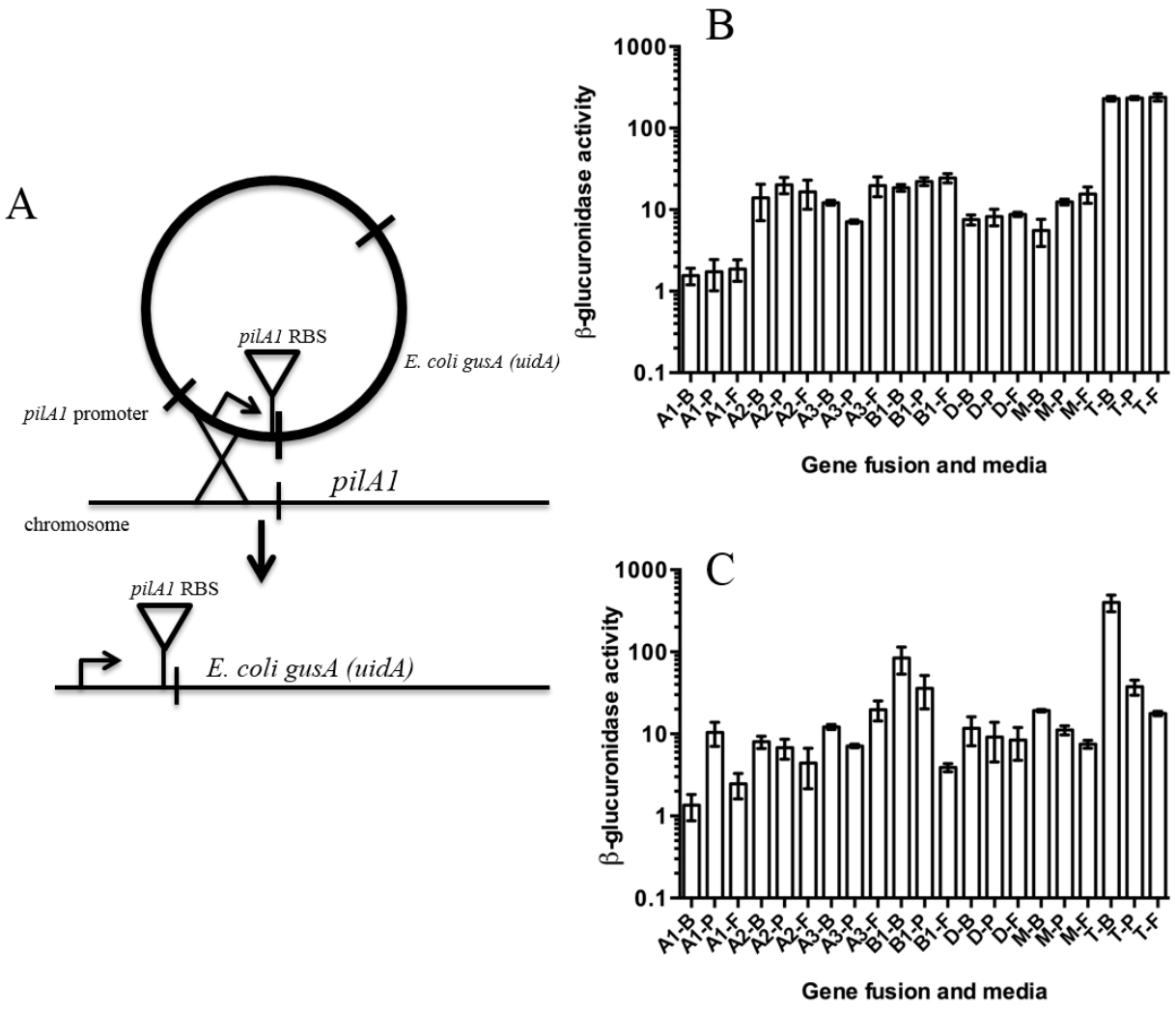


FIGURE 11. Translational β-glucuronidase fusions. (A) Schematic of method used for constructing promoter and ribosomal binding site (RBS) fusions to *E. coli* reporter gene *uidA* on the chromosome of *gusA*- strain of *C. perfringens*. β-glucuronidase activity of seven T4P promoters discovered by RNA-seq experiment from liquid-grown cells (B) and plate-grown cells (C). Samples are named for the *pil* gene located downstream of the promoter, B for BHI, P for PGY, F for FABG. (A1-B is the *pilA1* promoter fusion strain grown in BHI). Activity reported in Miller units.

Plate-grown bacteria show a significant increase in post-transcriptional regulation than liquid-grown cells. Due to the nature of the reporter gene constructs we created, there should be a linear relationship between transcript levels and translation in the absence of post-transcriptional effects. To determine this, normalized transcript levels (TPM) were plotted against the β-glucuronidase activity of each of the seven promoters that were examined (Figure 12). Independent plots were constructed for liquid culture cells and agar plate cells. For the liquid grown cells, there is a linear relationship between the amount of transcripts from each gene and the reporter gene activity for the corresponding medium. The R^2 value of the line, the closeness of the fit of the data to the linear regression line, was 0.8966 (Figure 12A). In contrast, plate-grown bacteria did not have a linear relationship between their transcript levels and reporter gene fusions, as evident by the R^2 value of 0.0533 and a y-intercept of 22.79 (Figure 12B). The large y-intercept indicates that significant reporter gene activity was observed in cells that had low levels of transcripts for at least some of the promoters. This discrepancy is most dramatic in the case of the *pilB1* promoter: normalized transcripts were under 0.1 TPM while the β-glucuronidase activity was near 100 units. With these findings, we conclude that post-transcriptional activity plays a role in T4P gene regulation in plate-grown cells much more than liquid cultured bacteria.

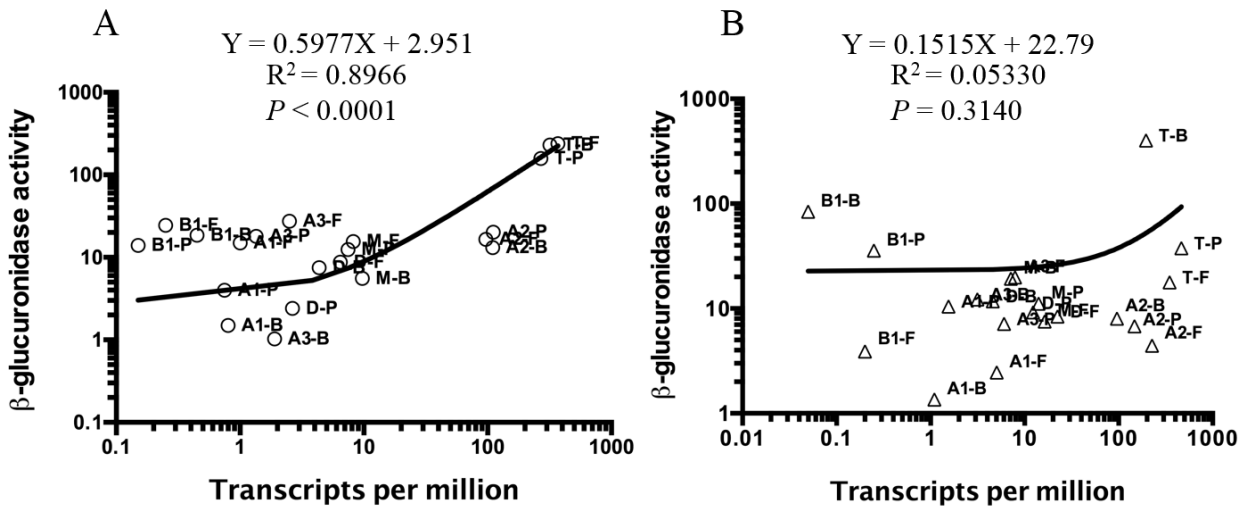


FIGURE 12. Linear regression of reporter assay data. Normalized transcript levels (TPM) were plotted against reporter gene activity (in Miller units) of the seven promoter fusion strains in *C. perfringens* cells grown in liquid (A) and on plates (B) of BHI, PGY, and FABG. A linear regression was found for each scatterplot, and closeness of fit was calculated. P-values were obtained using ANOVA.

The *pilT-ftsA-ftsZ* operon is controlled by temperature. *C. perfringens* cells can be found in the intestines of mammals, in decaying organisms at higher temperatures, and in the soil which can fluctuate in temperature by the hour (1). Because of the ubiquity and diversity of the environments *C. perfringens* can be found in, we subjected the reporter gene fusion strains to growth in three different temperature conditions (25°C, 37°C, 45°C) to detect any effect on pilin gene expression due to temperature. A large decrease in reporter gene activity from the *pilT* promoter (302, 27, and 4 units) was noted when the cells were grown on plates and the temperature was increased from 25 to 37 to 45°C (Figure 13B). The *pilT* promoter also showed a similar but less dramatic (380, 239, 193 units) decrease in activity as the temperature increased when the cells were grown in liquid at 25°C to 37°C to 45°C (Fig. 13A). There was some temperature dependent variation seen with the other tested promoters, but none of them held the

specific temperature-controlled pattern that the *pilT* promoter exhibited. Because *ftsA* and *ftsZ* are in an operon with *pilT*, and because these cell division genes are known to be regulated by temperature in other organisms (*fts* stands for filamentous temperature sensitive) (67), it should be noted that the effects we observed could be due to changes in growth rate at these temperatures.

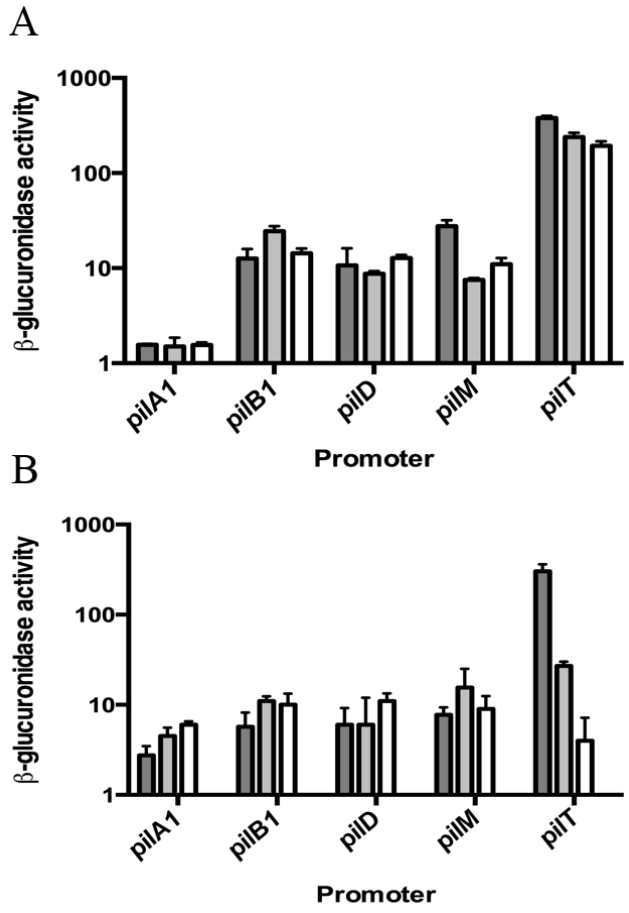


FIGURE 13. Temperature regulation of *pilT* promoter. Five T4P promoters were tested for regulated expression due to changes in temperatures. Bacteria assessed from liquid cultures (A) and agar plates (B) were grown at 25°C (dark grey), 37°C (light grey), and 45°C (white). Error shown is standard error of the mean.

A novel small RNA adjacent to the major T4P operon is highly transcribed on plated media. Because there were no identifiable transcriptional regulators present in the gene products table of upregulated genes in all plate-grown cells, we examined coverage levels of small regulatory RNAs that had been previously computationally predicted in strain 13 (68, 69) (see Chapter 4.5). Only three of the verified small RNAs were upregulated in all three plate conditions, one of which is a small RNA called *sr79* (69) (previously called sCPE2634). *sr79* lies immediately upstream of the major T4P pilin operon, making it a candidate for further study as a potential regulator of T4P genes (Figure 14A). Terminators were identified upstream and downstream of *sr79*, suggesting it is independently transcribed in the region (Fig. 14A). It was upregulated 71-, 41-, and 10-fold higher in BHI, PGY, and FABG plate-grown cells, respectively (Fig. 14B). Northern blotting confirmed that *sr79* is indeed transcribed and upregulated in plate-grown cells over liquid equivalents (Fig. 14C).

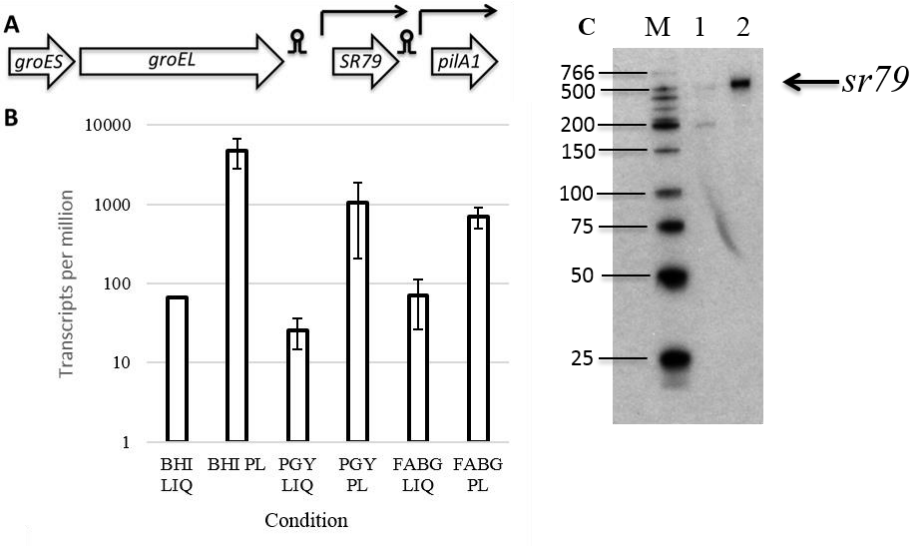


FIGURE 14. Identification of a plate-regulated small RNA. (A) Map of *sr79* location and predicted regulatory mechanisms on the chromosome of *C. perfringens* strain 13. The *sr79* coding region is flanked on either side by terminators. (B) Normalized average transcript levels of *sr79* in all six RNA-seq conditions. Error bars represent standard deviation. (C) Northern blot confirming the presence of RNA transcript of *sr79*. Lane M: molecular size marker; 1: RNA prepped from BHI liquid culture cells; 2: RNA prepped from BHI agar plate cells.

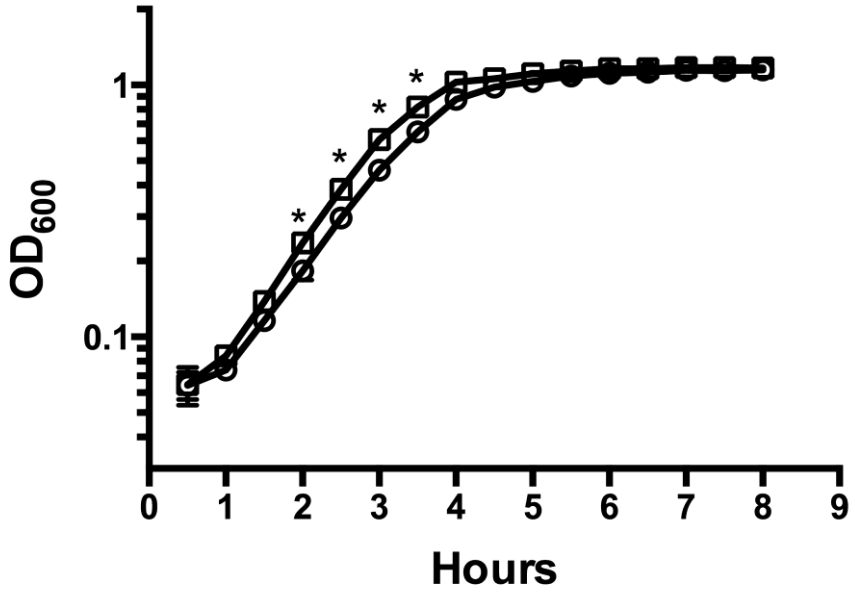


FIGURE 15. Growth curves of overexpressed small RNA *sr79*. HN13 cells with the empty pKRAH1 vector (circle markers) and *sr79* coding region overexpressed in pKRAH1 (square markers) were grown in 100-ml BHI with lactose induction every hour after the two hour time point. Bacterial density was recorded every half hour using a spectrophotometer measure A_{600}. Asterisks indicate a significant difference in log phase cells with a p-value <0.05 using the Student's t-test.

Overexpression of *sr79* increased growth rate in log phase cells grown in liquid media. To determine if the proximity of *sr79* suggests it has a regulatory effect on T4P genes, we attempted to delete the chromosomal region containing *sr79*. We were unable to establish a full deletion of this region using our standard in-frame deletion methods, instead obtaining heterozygotes in which both wild type and mutant versions of the region could be obtained without any antibiotic selection for the mutant copy (data not shown). As an alternative method, we chose to overexpress the *sr79* coding region lacking the promoter region upstream. We placed the *sr79* coding region in a lactose-inducible expression vector pKRAH1 to assess for any phenotypic effects from overexpression in liquid media and on plates. In liquid-grown cells with *sr79* overexpressed by lactose induction, in which the small RNA is not as actively transcribed from RNA-seq data, bacterial density increased significantly faster during log phase growth in

comparison to an empty vector wild type strain (Figure 15). We found no significant difference in the amount of PilA2 detectable on the surface of cells with *sr79* overexpressed in plate-grown bacteria using a whole-cell pilin ELISA (data not shown) (see Chapter 3.1 for method).

Overexpression of *sr79* results in longer plate-grown cells. The same *sr79* overexpression strain, containing plasmid pSRM35, was grown on BHI plates containing lactose as an inducing agents. Cells were scraped from the edges of colonies and viewed using light microscopy to detect any changes in cell morphology as a result of small RNA overexpression. Using the image analysis tool ImageJ, we found that cells overexpressing *sr79* were 19% longer than wild type cells (Figure 16). In liquid, no difference was found between the empty vector wild type strain and the *sr79* overexpressing strain (Fig. 16).

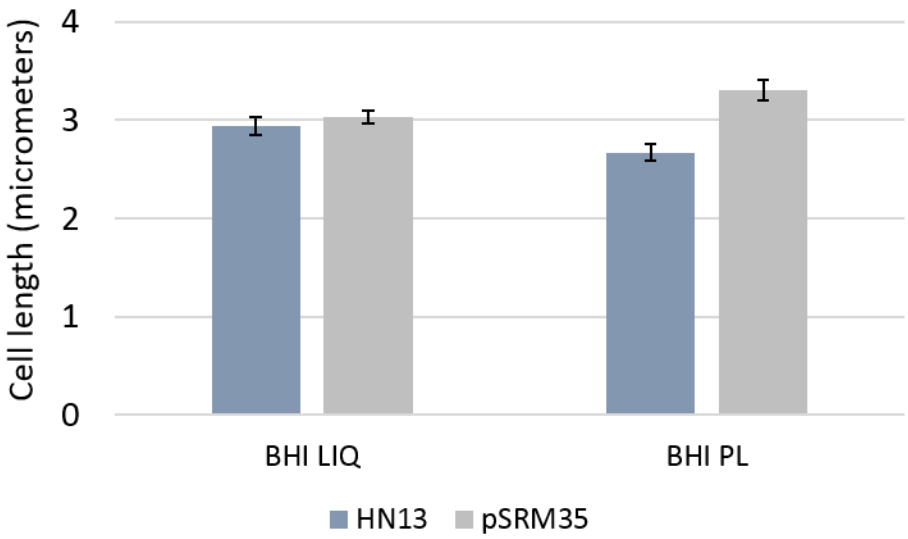


FIGURE 16. Cell length response of overexpressed *sr79*. Cell length changes in response to *sr79* overexpression. pSRM35 vector contains *sr79* behind a lactose-inducible promoter in pKRAH1. HN13 wild type strain contains empty pKRAH1 vector. Cells were induced with 1 mM lactose until log phase for BHI liquid cultures, or for 24 hours on BHI plates. Error bars represent standard error of the mean. Asterisks indicate a p-value < 0.01 using Student's t-test.

DISCUSSION

Together, these experiments provide a much clearer picture of T4P gene regulation than was previously known in *C. perfringens*. Necessary T4P genes, such as *pilA2*, *pilB2*, and *pilC2* were highly expressed in all conditions tested by RNA-seq, yet seem to be upregulated in response to agar plates or glucose, or both. Other less characterized T4P genes in the same operon, *cpe2279* and *cpe2280*, share this upregulation trend. qRT-PCR was used to confirm the RNA-sequencing data set (data not shown), which can be mined for numerous future experiments and uses.

A complex operon structure was also found utilizing RNA-seq and reporter gene data. Up to three promoters were found in a single operon, from *pilD* to *pilA2*, each seeming to correspond in strength to the amount of transcript required for the single downstream gene, after which other regulatory mechanisms take effect and change transcription levels further (such as the case of the *pilB2* promoter and subsequent *pilC2* mid-gene termination). *pilA2*, the major pilin and stoichiometrically the most abundant protein needed in formation of T4P, was the highest transcribed gene; plate-grown bacteria, in which *pilA2* transcripts were the highest, also had the highest overall amount of transcripts for the T4P genes than liquid grown cells.

The case of *pilC2* is highly perplexing yet fascinating. There is little literature about mid-gene termination occurring in other organisms; however, *C. perfringens* strain 13 has two genes that are known to terminate during the coding region. The genes *cpe0515* and *cpe0516* are annotated as one continuous gene in other strains of *C. perfringens*, yet strain 13 shows termination mid-gene creating two separate coding sequences. *pilC2* was found to have a rho-independent terminator mid-sequence, which created a termination product that has a fully intact transmembrane domain. In some samples, there were six times more transcripts before the

terminator sequence than after, meaning there could be six times more truncated product than functional PilC2. The truncated protein product could hypothetically still be inserted into the membrane and may act to obstruct the role of functional PilC2 by binding the Type IV pilus machinery but lacking the necessary domains to perform its usual function. From ELISA assays measuring polymerized PilA2 on the surface of cells (see Chapter 3.1), we have evidence that some functional PilC2 is made. T4P detected on the surface were significantly less in a *pilC2* deletion mutant, proving that functional PilC2 is necessary for T4P polymerization.

Interestingly, we were able to confirm the operon structure of *pilT-ftsA-ftsZ*. These three genes were highly transcribed in all conditions, which is necessary for cell division to occur normally with high levels of FtsA and FtsZ. PilT, which acts as the retraction ATPase for the Type IV pilus system, may also play an important role in cell division to warrant being transcribed at commensurate levels with *ftsA* and *ftsZ*. From previous transposon mutagenesis experiments, *pilT* insertion mutants were shown to lack the gliding motility phenotype. It was also observed by a previous student that PilT localized to the center of the cell away from Type IV pilus protein PilB2 at the poles during cell division. These two experiments along with our RNA-seq findings suggest that motility and cell division may be linked in some way.

Reporter gene assays uncovered unique transcriptional patterns at the post-translational level. In plate-grown cells, β-glucuronidase activity held less of a linear relationship to transcript levels from RNA-seq than liquid-grown cells did, pointing to some modification or degradation of mRNA between transcription and translation. We found that the *pilT* operon transcription and translation are inversely proportional to the temperature at which bacterial growth occurred. The high activity of the *pilB1-uidA* fusion operon with low transcript levels of *pilB1* is still not well understood.

We found hundreds of genes that are upregulated in each condition that we tested, noting that the highest amount of regulation always occurred in bacteria grown on plates. Many metabolic genes aiding in carbon starvation and nutrient acquisition were upregulated in all of the plate conditions. Also upregulated very strongly in all conditions was a novel small RNA found upstream of the major pilin operon, next to *pilA1*, that was unable to be deleted from the chromosome and therefore may be necessary for survival. In overexpression studies, we found that increasing transcription of the sRNA, *sr79*, caused cells to divide significantly faster during log phase growth in liquid and become longer on plated media than wild type cells.

Taken together, the results summarized here will be consequential in future studies on T4P in *C. perfringens*, due to an increased understanding of their regulation, their expression, and their gene organization. Future students can use these data to investigate the regulation of the *pilB1* operon and its high translational activity with a lack of transcripts, the truncated transcriptional pattern of the *pilA3* gene after the start site, or the curious termination mid-gene of *pilC2*.

Table 3 Strains and plasmids used in RNA-seq study.

Strain or plasmid (parent)	Relevant genotype	Selection/ Resistance	Source
Escherichia coli			
DH10B	Electrocompetent strain	None	Gibco/BRL
Clostridium perfringens			
HN13	Wild type; *galKT*-	None	(58)
AH2	Mutant in *bglR*; *uidA* fusion parent strain	Erm	(60)
SRM4	*pilA1* operon promoter fused to *uidA*	Erm Cm	This study
SRM9	*pilA2* operon promoter fused to *uidA*	Erm Cm	This study
SRM10	*pilA3* operon promoter fused to *uidA*	Erm Cm	This study
SRM5	*pilB1* operon promoter fused to *uidA*	Erm Cm	This study
SRM6	*pilD* operon promoter fused to *uidA*	Erm Cm	This study
SRM7	*pilM* operon promoter fused to *uidA*	Erm Cm	This study
SRM8	*pilT* operon promoter fused to *uidA*	Erm Cm	This study
Plasmids			
pSRM18	*cpe1788* in pKRAH1	Cm	This study
pSRM8	*pilA1::uidA* in pJV50	Cm	This study
pSRM22	*pilA2::uidA* in pJV50	Cm	This study
pSRM23	*pilA3::uidA* in pJV50	Cm	This study
pSRM10	*pilB1::uidA* in pJV50	Cm	This study
pSRM12	*pilD::uidA in* pJV50	Cm	This study
pSRM14	*pilM::uidA in* pJV50	Cm	This study
pSRM16	*pilT*::*uidA in* pJV50	Cm	This study
pJV50	Suicide vector used for *uidA* fusion strains	Cm	(60)
pKRAH1	Lactose inducible expression vector	Cm	(59)
pKAH1-*pilA2*	*pilA2* in pKRAH1	Cm	Andrea Hartman
pKRAH1-*pilB2*	*pilB2* in pKRAH1	Cm	Andrea Hartman
pKRAH1-*pilC2*	*pilC2* in pKRAH1	Cm	Andrea Hartman
pKRAH1-*pilT*	*pilT* in pKRAH1	Cm	Andrea Hartman
pKRAH1-*cpe2277*	*cpe2277* in pKRAH1	Cm	Andrea Hartman
pSRM24	*lon* in pKRAH1	Cm	This study
pSRM35	Promoterless *sr79* in pKRAH1	Cm	This study

Table 4 Primers used in RNA-seq study.

Primer	Sequence (5' – 3')
pilA1::*uidA*	
OSRM24	5' gtaGTCGACttattttgtagacaactgggaaaaactgtg 3' (SalI site)
OSRM25	5' cacgggttggggtttctacaggacgtaacatcagtaataacatttatggtcctcc 3'
OSRM26	5' ggaggaccataaatgttattactgatgttacgtcctgtagaaaccccaacccgtg 3'
pilA2::*uidA*	
OSRM50	5' cactagGTCGACtattagataaaacagcagatttttatgatgatgaac 3' (SalI site)
OSRM51	5' ggaaaaccaatgaatacaatgttacgtcctgtag 3'
OSRM52	5' ctacaggacgtaacattgtattcattggttttcc 3'
pilA3::*uidA*	
OSRM53	5' ggaGTCGACactggaaatccagattctacttacttattagtg 3' (SalI site)
OSRM54	5' ctacaggacgtaacataaattttcatttaatcacc 3'
OSRM55	5' ggtgattaaatgaaaaatttatgttacgtcctgtag 3'
pilB1::*uidA*	
OSRM27	5' gatGTCGACggaagctatactttagaaaagcaaatag 3' (SalI site)
OSRM28	5' cacgggttggggtttctacaggacgtaacatatatttcatatatcaacctcctc 3'
OSRM29	5' gaggaggttgatatatgaaatatatgttacgtcctgtacaaaccccaacccgtg 3'
pilD::*uidA*	
OSRM30	5' ggatagGTCGACcgctaaggcaaaagtagataacaaaattg 3' (SalI site)
OSRM31	5' cacgggttggggtttctacaggacgtaacattatagacacattgcctcccaaatcatcc 3'
OSRM32	5' ggatgatttgggaggcaatgtgtctataatgttacgtcctgtagaaaccccaacccgtg 3'
pilM::*uidA*	
OSRM33	5' catcaGTCGACgctttggttgctgaaggtaagataataggagg 3' (SalI site)
OSRM34	5' cacgggttggggtttctacaggacgtaacattattgccaaaactttccctcctttacc 3'
OSRM35	5' ggtaaaggagggaaagttttggcaataatgttacgtcctgtagaaaccccaacccgtg 3'
pilT::*uidA*	
OSRM36	5' caaGTCGACgggaatacgacttgcaacatatgcagc 3' (SalI site)
OSRM37	5' cacgggttggggtttctacaggacgtaacataxtttgcataaatctcctcctaaac 3'
OSRM38	5' gtttaggaggagatttatgcaaagtatgttacgtcctgtagaaaccccaacccgtg 3'
All *uidA* fusion C-termini	
OSRM39	5' gttCTGCAGgattcattgtttgcctccc 3' (PstI site)
pilA2 qRT-PCR	
OSRM62	5' tcaaatgctaaagacgacgttacag 3'
OSRM63	5' tctactgcaaatacaccaccatcaa 3'
pilB2 qRT-PCR	
OSRM78	5' tgtgcgtccgtatatagctgt 3'
OSRM79	5' gcttttgttaattcttccgctgc 3'
pilC2 qRT-PCR	
OSRM76	5' tgaaaaagagaxaagataagaggaaaga 3'
OSRM77	5' gcaataatagctaccactccaagaact 3'
pilT qRT-PCR	
oRM64	5' gaatgagaacgatggacatggaa 3'
OSRM65	5' gcacaattcattgccgtttctt 3'
cpe2277 qRT-PCR	
OSRM68	5' tctgtaggaagttggaagtggagtt 3'
OSRM69	5' tcttacattgggtaaatctccaacac 3'
lon qRT-PCR	
OSRM70	5' ggtagttggtcaagagagagctgtta 3'
OSRM71	5' caacccctggaggtccataa 3'
pSRM35	
OSRM108	5' GTCGACgaaaagtttatgaaaacaggttatttagaaattaattg 3' (SalI site)
OSRM95	5' gctaaaacaacagttaattctattaatgtGGATCCccc 3' (BamHI site)
OSRM97 (Northern blot probe)	5' CTCATATGTAACTTTAATATCTTTCGTTCTATGAATTC 3'

Table 5 Upregulated genes from three-way RNA-seq comparisons.

Genes upregulated in all plate-grown bacteria		
cpe0113		hypothetical protein, alone in operon
cpe0149		probable oxidoreductase
cpe0150		KDPG/KHG aldolase
cpe0151	*uxuA*	D-mannonate dehydrolase
cpe0152	*uxaC*	glucuronate isomerase
cpe0153	*uidB*	glucuronide permease
cpe0154		probable beta-hexosamidase A
cpe0166		membrane-spanning transporter protein
cpe0167	*pbg*	beta-galactosidase
cpe0168	*arcA*	arginine deiminase
cpe0169	*arcB*	ornithine carbamoyl transferase
cpe0170	*arcD*	arginine ornithine antiporter
cpe0171	*arcC*	carbamate kinase
cpe0172	*argR*	arginine repressor
cpe0310		probable lactate permease
cpe0311	*etfB*	electron transfer flavoprotein beta-subunit
cpe0312	*etfA*	electron transfer flavoprotein alpha-subunit
cpe0313		probable glycolate oxidase subunit
cpe0317	*fucK*	rhamnulokinase
cpe0318	*fucI*	L-fucose isomerase
cpe0320		phosphotransferase system, mannose/fructose-specific component IIA
cpe0371		conserved hypothetical protein, ABC-type sugar transport system
cpe0389	*arcD*	probable amino acid permease
cpe0390	*dchS*	histidine decarboxylase
cpe0426		conserved hypothetical protein, six-hairpin glycosidase
cpe0455		alkaline phosphatase-like protein, SNARE associated Golgi protein
cpe0743	*cstA*	probable carbon starvation protein
cpe0806		conserved hypothetical protein, serine dehydratase-like, alpha subunit
cpe0808		hypothetical protein, alone in operon
cpe0856		alpha-mannosidase
cpe0876		hypothetical protein, acyl-CoA N-acyltransferase
cpe0929		coenzyme B12-dependent glycerol dehydrogenase large subunit
cpe0930		coenzyme B12-dependent glycerol dehydrogenase medium subunit
cpe0931		coenzyme B12-dependent glycerol dehydrogenase small subunit
cpe0932		probable glycerol dehydratase large subunit
cpe0935		uncharacterized protein, glycolate and propanediol utilization
cpe0982		hypothetical protein, likely alone in operon
cpe1344		aldose 1-epimerase
cpe1428	*clpB*	ATPase with chaperonin activity
cpe1806	*dacF*	serine-type D-Ala-D-Ala carboxypeptidase
cpe1876		conserved hypothetical protein, alpha-L-fucosidase
cpe2078		conserved hypothetical protein, ABC-type sugar transport system
cpe2100		hypothetical protein, in operon with *cpe2101* and *cpe2102*
cpe2107		hypothetical protein, in operon with *cpe2108*
cpe2108		hypothetical protein in operon with *cpe2107*

cpe2549		conserved hypothetical protein, in operon with *cpe2550* and *cpe2551*
cpe2550		probable oxidoreductase
cpe2551	*glpA*	probable glycerol-3-phosphate dehydrogenase
cpe2630	*manZ*	probable mannose permease IIm
Genes upregulated in all liquid-grown bacteria		
cpe0449	*adh*	alcohol dehydrogenase
cpe1284		pyrimidine nucleoside transporter
cpe1539		hypothetical protein, alone in operon
cpe2585		conserved hypothetical protein, alone in operon

Table 6 Upregulated genes in BHI & PGY two-way comparisons.

Genes Upregulated >4x on BHI and PGY (but not FABG) Plates		
cpe0085	*mdh*	probable alcohol dehydrogenase
cpe0086	*alf1*	fructose-1,6-bisphosphate aldolase
cpe0087	*iolC*	myo-inositol catabolism protein
cpe0088	*iolB*	myo-inositol catabolism protein
cpe0089	*iolD*	myo-inositol catabolism protein
cpe0090	*mviM*	probable dehydrogenase
cpe0091	*iolE*	myo-inositol catabolism protein
cpe0092		Na+/myo-inositol cotransporter
cpe0093	*mviM*	probable dehydrogenase
cpe0094	*focA*	formate/nitrite family transporter
cpe0095	*caiD*	crotonase
cpe0096		proprionate CoA-transferase
cpe0097	*acdS*	acyl-CoA dehydrogenase
cpe0099	*gldA*	glycerol dehydrogenase
cpe0105		hypothetical protein
cpe0114		hypothetical protein
cpe0217	*ackA*	acetate kinase
cpe0319	*fucA*	L-fuculose-phosphate aldolase
cpe0321		probable PTS system
cpe0322	*manY*	probable PTS system
cpe0323	*manZ*	probable PTS system
cpe0324		probable glycosyl hydrolase
cpe0325		hypothetical protein
cpe0326	*lacA*	galactose-6-phosphate isomerase
cpe0327	*lacB*	galactose-6-phosphate isomerase
cpe0328	*lacC*	tagatose-6-phosphate kinase
cpe0329		hypothetical protein
cpe0372	*ugpA*	probable ABC transporter
cpe0373	*ugpE*	probable sugar ABC transporter
cpe0374	*aga*	alpha-galactosidase
cpe0375	*skn1*	endo-beta-galactosidase
cpe0553	*nanJ*	exo-alpha-sialidase
cpe0554		hypothetical protein
cpe0566	*ribB*	riboflavin biosynthesis protein
cpe0578	*ugpE*	probable sugar ABC transporter
cpe0643		hypothetical protein
cpe0660	*pflA*	pyruvate-formate lyase-activating enzyme
cpe0669		hypothetical protein

cpe0725	*nanI*	exo-alpha-sialidase
cpe0726		uncharacterized protein conserved in bacteria
cpe0742	*lipB*	probable lipase
cpe0775		regulator of cell morphogenesis and NO signaling
cpe0778	*hmp*	probable flavohemoprotein
cpe0782	*ahpC*	alkyl hydrogen peroxide reductase
cpe0783	*trxB*	probable thioredoxin reductase
cpe0792	*fepB*	probable iron(III) dicitrate ABC transporter
cpe0793	*fepD*	probable iron(III) dicitrate ABC transporter
cpe0794	*fepG*	probable iron(III) dicitrate ABC transporter
cpe0818		endo-beta-N-acetylglucosaminidase
cpe0845		hypothetical protein
cpe0863		hypothetical protein
cpe0881	*nagI*	hyaluronidase
cpe0892	*eutG*	NADPH-dependent butanol dehydrogenase
cpe0897	*eutA*	ethanolamine utilization protein
cpe0898	*eutB*	ethanolamine ammonia lyase heavy chain
cpe0899	*eutC*	ethanolamine ammonia lyase light chain
cpe0900	*eutL*	ethanolamine utilization protein
cpe0901	*pduJ*	propanediol utilization protein
cpe0902	*adhE*	probable alcohol dehydrogenase
cpe0903	*pduJ*	propanediol utilization protein
cpe0904	*eutT*	ethanolamine utilization cobalamin adenosyltransferase
cpe0905	*pduL*	probable propanediol utilization protein
cpe0906		hypothetical protein
cpe0907	*eutN*	ethanolamine utilization protein
cpe0908	*pduT*	propanediol utilization protein
cpe0909	*eutH*	ethanolamine utilization protein
cpe0910	*eutQ*	ethanolamine utilization protein
cpe0919		hypothetical protein
cpe0920		hypothetical protein
cpe0981		probable beta-N-acetylhexosaminidase
cpe0987	*pfoS*	probable regulatory protein
cpe0989	*sdhA*	L-serine dehydratase alpha subunit
cpe1024		hypothetical protein
cpe1030	*ugpB*	ABC-type sugar transport system, periplasmic component
cpe1151	*sfcA*	probable malate oxidoreductase
cpe1152	*citN*	probable citrate/sodium symporter
cpe1165	*thd*	threonine dehydratase
cpe1167	*gltP*	probable glutamate/aspartate transporter

cpe1185	*pfk*	6-phosphofructokinase
cpe1250		hypothetical protein
cpe1257	*ugpB*	ABC-type sugar transport system, periplasmic component
cpe1293		conserved hypothetical protein
cpe1294		hypothetical protein
cpe1368		conserved hypothetical protein
cpe1414	*pheB*	chorismate mutase
cpe1415	*ams1*	alpha-mannosidase
cpe1513		hypothetical protein
cpe1523	*nagL*	hyaluronidase
cpe1626	*purR*	probable transcriptional regulator
cpe1627	*rbsE*	probable ribose ABC transporter
cpe1629		probable ribose ABC transporter
cpe1630	*rbsC*	probable ribose ABC transporter
cpe1631	*rbsA*	probable ribose ABC transporter
cpe1875		conserved hypothetical protein
cpe1920		hypothetical protein
cpe1951	*cpdA*	2',3'-cyclic nucleotide 2'-phosphodiesterase
cpe2011	*podK*	pyruvate phosphate dikinase
cpe2077	*nagC*	probable transcriptional regulator
cpe2081	*ugpE*	probable ABC transporter
cpe2082	*lplB*	probable ABC transporter
cpe2089		hypothetical protein
cpe2104		predicted membrane protein
cpe2105		uncharacterized protein conserved in bacteria
cpe2106	*apbE*	probable thiamine biosynthesis lipoprotein
cpe2262		hypothetical protein
cpe2263		hypothetical protein
cpe2295	*lepW*	signal peptidase type I
cpe2344		conserved hypothetical protein
cpe2553	*glpP*	probable glycerol uptake operon antiterminator
cpe2625	*agaS*	probable tagatose-6-phosphate aldose/ketose isomerase
cpe2626	*gatY*	probable tagatose-bisphosphate aldolase
Genes Upregulated >4x in BHI and PGY (but not FABG) Liquid		
cpe0768		hypothetical protein
cpe1050	*pfs*	5'-methylthioadenosine/S-adenosylhomocysteine nuclosidase
cpe1529		hypothetical protein
cpe2531	*adhE*	aldehyde-alcohol dehydrogenase E

Table 7 Upregulated genes in PGY & FABG two-way comparisons.

Genes Upregulated 4x or Greater on PGY and FABG Plates (but not BHI)		
cpe0072	*abgT*	putative p-aminobenzoyl-glutamate transporter
cpe0103	*ldh*	L-lactate dehydrogenase
cpe0129		predicted membrane protein
cpe0145		hypothetical protein, *bglR* operon
cpe0146		2-keto-3-deoxygluconate kinase
cpe0147	*bglR*	beta-glucuronidase
cpe0155		conserved hypothetical surface protein
cpe0175	*ycgJ*	methylase involved in ubiquinone/menaquinone biosynthesis
cpe0183		probable NAD(P)H-flavin oxidoreductase
cpe0211		conserved hypothetical protein, alone in operon
cpe0262		hypothetical protein, alone in operon
cpe0286	*galE*	UDP-glucose 4-epimerase
cpe0336	*rhaT*	Drug/metabolite permease protein
cpe0411		probable monooxygenase
cpe0530	*ldhD*	D-lactate dehydrogenase
cpe0538		hypothetical protein, in operon with *cpe0537* (transposase)
cpe0550	*speE*	spermidine synthase
cpe0551	*speB*	probable agmatinase
cpe0559	*sigV*	RNA polymerase ECF-type sigma factor
cpe0560		hypothetical protein, in operon with *sigV*
cpe0632	*brnQ*	sodium-coupled branched-chain amino acid carrier protein
cpe0771	*bgaL*	beta-galactosidase
cpe0819	*ddlB*	D-alanine-D-alanine ligase
cpe0860		probable gluconate permease
cpe0917	*nrdD*	probable anaerobic ribonucleotide reductase
cpe0918		hypothetical protein, in operon with *nrdD*
cpe0933		conserved hypothetical protein, glycerol dehydrogenase operon
cpe0934		conserved hypothetical protein, glycerol dehydrogenase operon
cpe1012		probable translation initiation inhibitor
cpe1013		hypothetical protein, in operon with *cpe1012*
cpe1078		hypothetical protein
cpe1079	*nfo*	endonuclease IV
cpe1118		probable phage-related terminase
cpe1322	*cysK*	o-acetylserine sulfhydrylase
cpe1472		hypothetical protein, in operon with *cpe1473*
cpe1536	*asrC*	anaerobic sulfite reductase subunit C
cpe1537	*asrB*	anaerobic sulfite reductase subunit B
cpe1538	*asrA*	anaerobic sulfite reductase subunit A

cpe1573		probable glutamate-cysteine ligase
cpe1782		conserved bacterial protein, in operon with *yhhT* (permease)
cpe1783	*trmU*	probable tRNA methyltransferase
cpe1784	*nifU*	probable nitrogen fixation protein
cpe1785	*nifS*	iron-sulfur cofactor synthesis protein
cpe1786		predicted transcriptional regulator
cpe1970	*potC*	probable spermidine/putrescine ABC transporter
cpe1980		putative effector of murein hydrolase LrgA
cpe1981	*lrgB*	putative effector of murein hydrolase
cpe1989	*pldB*	lysophospholipase
cpe2092	*glnQ*	probable amino acid ABC transporter
cpe2093	*hisJ*	probable amino acid ABC transporter
cpe2151		probable mercuric ion-binding protein
cpe2158		probable adhesin
cpe2280		hypothetical protein, in *pilA2* operon
cpe2319		hypothetical protein, alone in operon
cpe2506	*add*	probable adenosine deaminase
Genes Upregulated 4x or Greater in PGY and FABG Liquid (but not BHI)		
cpe0556		NAD(P)H dehydrogenase
cpe0561		phosphotransferase system
cpe1051	*cfa*	probable cyclopropane-fatty-acylphospholipid synthase
cpe1751		conserved hypothetical permease

Table 8 Upregulated genes from BHI & FABG two-way comparisons.

Genes Upregulated >4x on BHI and FABG (but not PGY) Plates		
cpe0098		conserved hypothetical protein
cpe0238		hypothetical protein
cpe0333	*mgtA*	probable cation-transporting ATPase
cpe0769	*gutA*	probable sugar transport protein
cpe1092	*pac*	probable penicillin acylase
cpe1535		hypothetical protein
cpe1592		hypothetical protein
cpe2058	*gadB*	glutamate decarboxylase
cpe2059		hypothetical protein
cpe2631	*manY*	probable PTS system
Genes Upregulated >4x in BHI and FABG (but not PGY) Liquid		
cpe0117		conserved hypothetical protein
cpe0174	*mscL*	large-conductive mechanosensitive channel homologue
cpe0244		hypothetical protein
cpe0246		conserved hypothetical protein
cpe0511		hypothetical protein
cpe0557	*opuBA*	probable glycine betaine/carnitine/choline ABC transporter
cpe0558	*opuBC*	probable glycine betaine/carnitine/choline ABC transporter
cpe0600	*hisJ*	probable amino acid ABC transporter
cpe0601	*hisM*	probable amino acid ABC transporter
cpe0602	*glnQ*	probable amino acid ABC transporter
cpe0616	*rfbA*	glucose-1-phosphate thymidylyltransferase
cpe0630		hypothetical protein
cpe0690	*argH*	argininosuccinate lyase
cpe0691	*argG*	argininosuccinate synthase
cpe0710	*lrp*	probable transcriptional regulator
cpe0711	*aspC*	aspartate aminotransferase
cpe0760	*def*	polypeptide deformylase
cpe0945		hypothetical protein
cpe1053		hypothetical protein
cpe1176		conserved hypothetical protein
cpe1506	*uup*	probable ABC transporter
cpe1655		conserved hypothetical protein
cpe2037	*rplY*	ribosomal protein L25, general stress protein
cpe2191	*atpF*	ATP synthase B chain
cpe2311		conserved hypothetical protein
cpe2569	*glnA*	probable glutamine synthetase
cpe2597		hypothetical protein
cpe2601		hypothetical protein

CHAPTER III

Additional Projects & Collaborations

3.1 CONTROL OF TYPE IV PILIN POLYMERIZATION BY C-DI-GMP IN *Clostridium perfringens*

Text adapted from

Hendrick, William A., Mona W. Orr, Samantha R. Murray, Vincent T. Lee, and Stephen B. Melville. "Cyclic-di-GMP binding by an assembly ATPase (PilB2) and control of Type IV pilin polymerization in the Gram-positive pathogen *Clostridium perfringens*." *Journal of Bacteriology* (2017): JB-00034.

CO-AUTHOR CONTRIBUTIONS TO THIS SECTION

William Hendrick created the *pilC2* in-frame deletion strain. Andrea Hartman created the *pilB2* deletion strain as well as complementation plasmids pAH10 and pAH11. Stephen Melville created the figures shown in this section and co-wrote the manuscript of the publication with William Hendrick.

Clostridium perfringens has four identifiable pilin proteins, named PilA1-PilA4 and two pairs of PilB/PilC proteins, PilB1/PilC1 and PilB2/PilC2 (22, 25). We consider PilA2 to be the major pilin involved in adherence to host cells, since PilA2 is on the surface of *C. perfringens* (22) and a strain of *Neisseria gonorrhoeae* expressing the *pilA2* gene but lacking its endogenous pilin, PilE, changed its adherence profile from human genital-urinary tract epithelial cells to mouse myoblasts and fibroblasts (24). Given its location on the chromosome, immediately downstream of the *pilB2* and *pilC2* genes, it seems likely that these are the core assembly proteins responsible for polymerizing PilA2 into a pilus.

Recently, a direct connection to c-di-GMP dependent regulation was reported in which a family of PilB assembly ATPases proteins bound c-di-GMP directly (70). c-di-GMP is synthesized from 2 GTP molecules by diguanylate cyclase (DGC) enzymes (71). In the case of *Vibrio cholerae*, this leads to pilus biogenesis and changes in surface-based motility (72). c-di-GMP is believed to regulate, amongst other functions, the switch from flagella-mediated swimming motility to surface colonization and biofilm formation (71), so having a direct effect on PilB activity fits within the confines of that model. In the Gram-positive bacterium *Clostridioides* (*Peptoclostridium*) *difficile,* there is a c-di-GMP type II riboswitch directly upstream of the main TFP-encoding locus (73, 74). Expression of a DGC in *C. difficile* led to increased c-di-GMP and pilus biogenesis (75, 76) but this riboswitch is absent in *C. perfringens*. We demonstrate that PilB2 binds both ATP and the second messenger molecule, c-di-GMP. PilB2 is stimulated to polymerize PilA2 pilin by expression of a native DGC, thereby extending the role of c-di-GMP in regulating PilB activity to Gram-positive bacteria.

MATERIALS & METHODS

Bacterial strains and culture conditions. All bacterial strains used in this study are listed in Table 9. *E. coli* was grown in Luria-Bertani (LB) medium with antibiotics added as necessary (100 μg/ml kanamycin or 20 μg/ml chloramphenicol) and 1% agar for plates. *C. perfringens* was grown in a Coy anaerobic chamber in brain heart infusion (BHI) medium (20 μg/ml chloramphenicol) and/or 1% agar when necessary, or in fastidious anaerobe broth with 2% glucose added (20 μg/ml chloramphenicol) and/or 1% agar.

Construction of *pilC1* in-frame deletion mutant. A mutant strain of *C. perfringens* with an in-frame deletion in *pilC1* was constructed according to the method of Nariya et al (58). To create plasmid pSRM2, the region upstream of the *pilC1* gene, including the first three codons, was amplified from *C. perfringens* strain 13 chromosomal DNA with primers OSRM21 and OSRM2. The region downstream of the *pilC1* gene, including the codons coding for the last two amino acids and a stop codon, was amplified from *C. perfringens* strain 13 chromosomal DNA with primers OSRM3 and OSRM23. Primers were created to exclude the *cpe1841* coding region. These two partially overlapping products were then used as the template, allowing for amplification of the *pilC1* coding region with primers OSRM2 and OSRM3. Phusion polymerase was used to give this product blunt ends to allow for direct ligation into pCR-Blunt-TOPO. The insert was then digested out of this plasmid with the enzymes BamHI-HF and PstI-HF, and ligated into pCM-GalK digested with the same enzymes to form pSRM2. This plasmid was introduced into strain HN13 by electroporation. The first crossover event was selected for by the addition of chloramphenicol to the growth medium. An overnight culture of HN13(pSRM2) was

grown at 37 °C in BHI with 20 μg/ml chloramphenicol and subcultured into TY + 3% galactose, and grown overnight at room temperature. This culture was then serially diluted and plated onto TY + 3% galactose plates. Resultant colonies were patched on to TY + 3% galactose and BHI Cm^{20}. Clones that underwent a second recombination event, and thus were sensitive to chloramphenicol, were screened by PCR with primers OSRM21 and OSRM22 to confirm deletion of *pilC1*. The mutant retained, in-frame, the first three and final three codons of the *pilC1* gene.

Expression of a diguanylate cyclase enzyme in *C. perfringens*. The gene encoding the DGC CPE1788 was amplified using PCR from *C. perfringens* chromosomal DNA using primers OSRM48 and OSRM49. The PCR product, which contained the entire coding region and the putative ribosomal binding site for the gene, was ligated to the PCR cloning vector pCR-Blunt-TOPO to create plasmid pSRM16, which was digested with the restriction enzymes PstI and BamHI ligated to PstI-BamHI digested plasmid pKRAH1 to produce plasmid pSRM18.

Measurements of surface-exposed PilA2. HN13 and T4P gene in-frame deletion mutants Δ*pilA2*, Δ*pilB2*, and Δ*pilC2* were transformed with pSRM18 using standard electroporation techniques described previously. Cell suspensions of WT and each mutant were spotted onto FABG plates containing 20 μg/ml chloramphenicol and 5 mM lactose and grown anaerobically for 24 hours at 37°C. Cells were scraped from the edges of colonies and fixed anaerobically in 2% glutaraldehyde in DPBS for 30 minutes. Cells were pelleted by centrifugation and washed once in DPBS before normalizing OD_{600} to 0.5. Wells of a 96-well untreated opaque white microplate (Nalge Nunc Inc.) were coated in triplicate with 100 μl cell suspension before

incubating for 40 minutes at 37°C. Wells were washed four times with DPBS before being blocked with 200 µl blocking solution containing 4% milk and 2% bovine serum albumin in DPBS. Blocking occurred at 37°C for 90 minutes. Wells were washed three times with DPBS, and 100 µl affinity-purified rabbit antibody against PilA2 diluted to 0.5 µg/ml was incubated in wells at 37°C for 30 minutes. Wells were washed again and 100 µl peroxidase conjugated goat anti-rabbit secondary antibody (Thermo Scientific) was applied to wells at 0.1 µg/ml at 37°C for 30 minutes. Detection was performed using Quanta Red HRP Substrate (Thermo Scientific) as per the manufacturer's instructions. Peroxidase activity was stopped after 5 minutes incubation at room temperature, and colorimetric values of each well were read in a clear 96 well plate with a SpectraMax M5 plate reader at 576 nm.

RESULTS

Expression of a DGC leads to increased polymerization of PilA2 in a PilB2- and PilC2-dependent manner. Since we had determined that PilB2 can bind c-di-GMP but did not observe an increase in ATPase activity in vitro with c-di-GMP present, we increased the c-di-GMP levels in vivo by expressing the gene encoding CPE1788, an endogenous protein with a GGDEF domain. As a measurement of PilB2 activity, we used an ELISA-based assay to quantify the amount of PilA2 on the surface of the bacteria. Expression of CPE1788 led to a significant increase in surface exposed PilA2 (Fig. 17). When the CPE1788 expression vector was placed in a Δ*pilB2* or Δ*pilC2* strain, the increase was lost. Complementation of the Δ*pilB2* with the YFP-PilB2 fusion protein restored the phenotype, suggesting the YFP-PilB2 fusion was

fully functional for polymerization of PilA2. Complementation of the Δ*pilC2* with the PilC2-6His protein also fully restored the PilA2 phenotype, indicating the His-tag did not interfere with PilC2 functions as they relate to PilA2 polymerization.

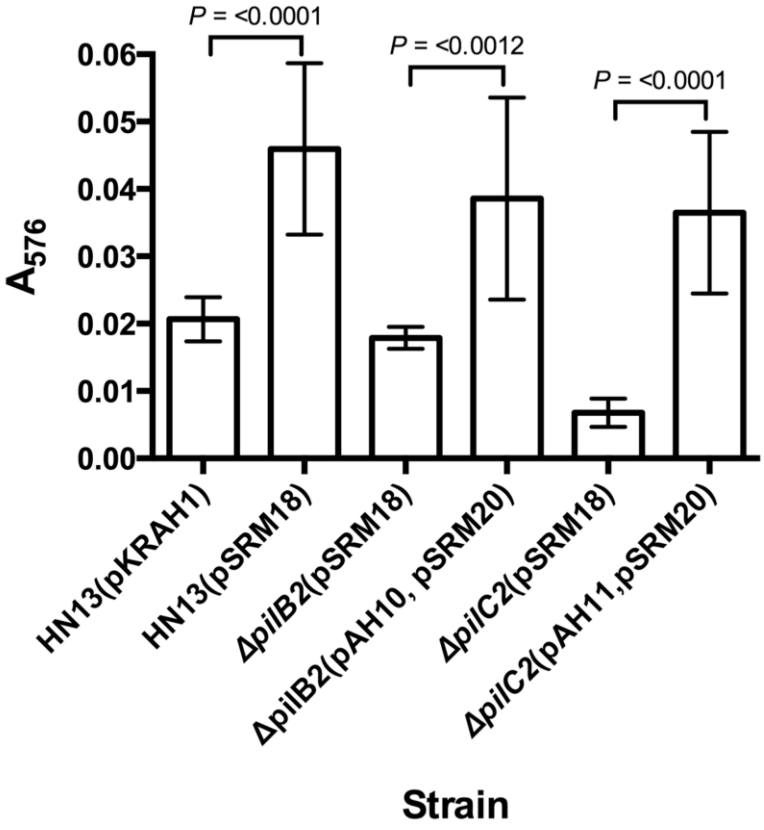


Figure 17: ELISA assay using anti-PilA2 antibodies to detect surface-exposed PilA2. pKRAH1, expression vector; pSRM18, pKRAH1 with *cpe1788* (a DGC); pAH10, pKRAH1 with a YFP-PilB2 fusion; pAH11, pKRAH1 with a YFP-PilB1 fusion; pSRM20, *cpe1788* in pXEH, a xylose-inducible promoter containing vector (77).

DISCUSSION

We found evidence that PilB2 responds to increased c-di-GMP in vivo to polymerize PilA2, as seen with a whole cell ELISA assay using anti-PilA2 antibodies (Fig. 17). *C. perfringens* has six proteins that contain a GGDEF domain, which have the potential to function as DGCs, along with three EAL domain- and one HD-GYP domain-containing protein which can function as c-di-GMP phosphodiesterases (PDEs) (Fig. 18). CPE0245 has both a GGDEF and EAL domain while CPE1560 has both a GGDEF and HD-GYP domain. We chose to use CPE1788 to increase the intracellular levels of c-di-GMP and stimulate PilB2 activity because it is small, lacks other recognizable domains, and is not membrane bound (Fig. 18). Expression of CPE1788 did lead to a significant increase in surface-localized PilA2 (Fig. 17), but this does not prove that CPE1788 actually controls PilB2 activity under normal physiological conditions. Deletions of the *pilB2* and *pilC2* genes led to a complete loss of an increase in PilA2 polymerization (Fig. 17) in the presence of CPE1788 and complementation with genes encoding a YFP-PilB2 and His-tagged PilC2 protein restored activity. We deliberately used the YFP-PilB2 fusion to test whether it could function as a tracer of PilB2 localization and still complement a *pilB2* mutation. The PilC2-6His fusion that we used for complementation, although not identical to the purified HisFlgPilC2, was used for similar reasons. Successful complementation with each protein does support our hypothesis that PilB2 and PilC2 interact, albeit indirectly, perhaps through another protein known to be located at the base of Type IV pilus, such as PilM.

If c-di-GMP binding by PilB2 does not increase the ATPase activity but does stimulate PilA2 polymerization, the question arises as to what mechanism underlies these findings. It is possible that, in vivo, c-di-GMP binding does lead to increased ATP hydrolysis although this is difficult to measure directly. Since the c-di-GMP binding and ATP binding sites are in different parts of the protein, another option is that c-di-GMP binding changes the conformation of the N-terminal domain such that it becomes more effective at polymerizing PilA2, even if the rate of ATP hydrolysis remains unchanged.

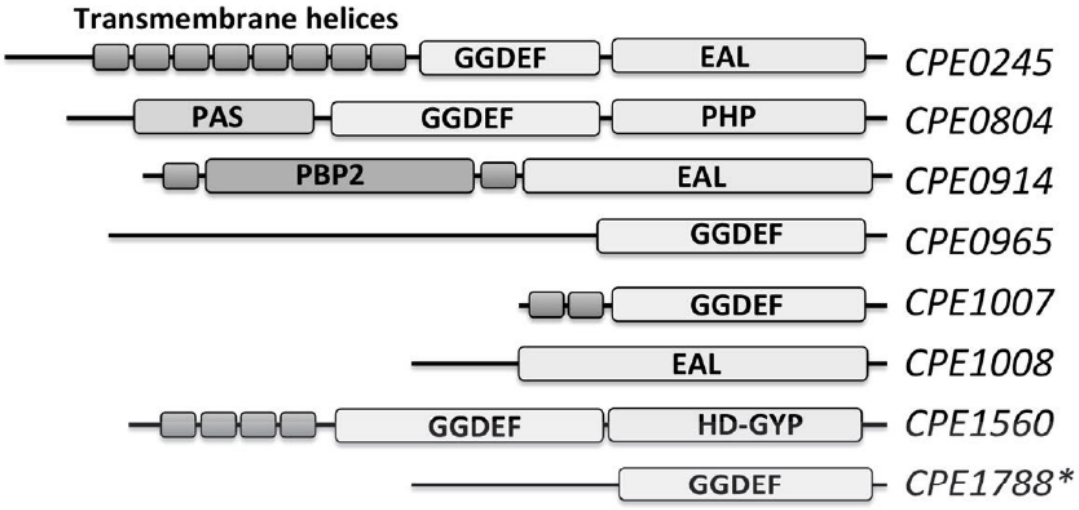


Figure 18: The functional domains of each protein in *C. perfringens* strain 13 that possesses a putative DGC (GGDEF) or PDE (EAL, HD-GYP) activity. PAS, signaling domain; PBP2, periplasmic binding protein type 2. CPE1788 (asterisk) was used to stimulate PilB2 activity in vivo.

Table 9 Strains, plasmids, and primers used in cyclic-di-GMP study.

Strains		
Name	**Relevant characteristics**	**Source**
E. coli		
DH10B	F$^-$ *mcrA*Δ *mrr-hsdRMS mcrBC* ϕ80d*lacZ*ΔM15 *lacX74 deoR recA1 araD139*Δ *ara, leu7697 galU galK*Δ*rpsL endA1 nupG*	Gibco/BRL
C. perfringens		
HN13	*galKT* deletion strain of parent, strain 13, used for constructing in-frame deletions	(58)
SRM1	In-frame deletion of the *pilC1* gene in strain HN13	This study
Plasmids		
pCR-BluntII-TOPO	Blunt PCR cloning vector	Invitrogen
pKRAH1	Lactose-inducible expression vector	(59)
pCM-GALK	Contains a *C. beijerinckii galK* gene under control of a ferredoxin promoter from *C. perfringens*	(58)
pAH10	*yfp-pilB2* in pKRAH1	This study
pAH11	*pilC2-His* in pKRAH1	This study
pSRM2	Vector, based on pCM-GALK, for making a *pilC1* in-frame deletion	This study
pSRM16	pCR-Blunt II-TOPO with *cpe1788*	This study
pSRM18	*cpe1788* in pKRAH1	This study
pXEH	Vector with a xylose inducible promoter for expression in *C. perfringens*	(77)
pSRM20	CPE1788 in pXEH	This study
Primers		
Name	**Sequence (5' to 3')**	
OSRM21	GGGCATAAAGTATATTCAACAATACACGCTAACAGTGG	
OSRM2	GGGTAAATCCTTTTCTTTTACTATACATATTTCCCAAACATCTAACCCCTCACATTC	
OSRM3	GAATGTGAGGGGTTAGATGTTTGGGAAATATGTATAGTAAAAGAAAAGGATTTACCC	
OSRM23	CCATATCTATATTTCCTCATAAATCACTTAACATTAATTTTCCATATGCC	
OSRM48	GTTCTAGGATCCGGAGGAAAGAATATTGGAAGTTTTACTAG	
OSRM49	CTAAAACTGCAGATAATACTAAAATACAGTTTATTTCTTCC	

3.2 CRYSTALLIZATION OF MINOR TYPE IV PILINS FROM *Clostridium perfringens*

Clostridium perfringens may utilize Type IV pili for a multitude of functions, such as motility, cellular aggregation, biofilm formation, and DNA uptake (25). In this organism, the major pilin protein involved in the formation of a Type IV pilus is believed to be PilA2, due to its ability to be detected on the surface of cells using immunofluorescence microscopy and ELISA methods and also its position in an operon with other major Type IV pilus proteins, such as PilD, PilC2, and PilB2. However, the genome of *C. perfringens* also contains multiple minor pilin genes of currently unknown function, such as *pilA1*, *pilA3* and *pilA4*. Two other genes, *cpe2279* and *cpe2280*, encode for pilin-like hypothetical proteins and are proximal to other T4P genes. In order to characterize the minor pilins of the organism and elucidate their function, this project is a collaboration with Dr. Paula Salgado of Newcastle University to overexpress truncated pilin proteins for crystallization and analysis by X-ray diffraction.

MATERIALS & METHODS

Bacterial strains and culture conditions. All bacterial strains and vectors used in this study are listed in Table 10. *Escherichia coli* was grown in Luria-Bertani (LB) medium with 100 μg/ml kanamycin and 1% agar for plates.

Creation of pilin protein overexpression vectors. Primers for PCR amplification of N- and C-terminal domains of five minor pilin genes (*pilA1*, *pilA3*, *pilA4*, *cpe2279*, and *cpe2280*) were provided by Dr. Stephen Melville and listed in Table 10. Phusion High-Fidelity DNA polymerase (New England Biolabs) was used to create PCR products with blunt ends to allow for direct ligation into pCR-Blunt-TOPO. Ligations were electroporated into an electrocompetent *E. coli* strain DH10B using standard electroporation methods, and transformants were selected by growth on LB kanamycin agar plates. Plasmids were extracted and digested out of pTOPO with the enzymes NcoI-HF and XhoI and checked by Sanger sequencing using the same primers as were used for the initial PCR. Correct gene fragments were then ligated into either pET-28a or pET-M11 digested with the same enzymes using T4 DNA ligase (Promega) to create pSRM24 through pSRM33. These plasmids were introduced into *E. coli* via electroporation, and transformants were selected by growth on LB kanamycin plates. Plasmids were extracted using a Mini Prep kit (Qiagen), and the constructs were confirmed via restriction digest.

RESULTS

Currently, Dr. Salgado has only partially crystallized CPE2280 with resolution to 2 angstroms, and efforts are concentrated on optimizing expression conditions for future crystallization of the other four minor pilin proteins.

Table 10 Strains, plasmids, and primers used in pilin crystallization study.

Strains		
Name	**Relevant characteristics or sequence**	**Source**
E. coli		
DH10B	F- *mcrA*Δ *mrr-hsdRMS mcrBC* ϕ80d*lacZ*ΔM15 *lacX74 deoR recA1 araD139*Δ *ara, leu7697 galU galK*ΔrpsL *endA1 nupG*	Gibco/BRL
Plasmids		
pCR-Blunt II-TOPO	Blunt PCR cloning vector	Invitrogen
pET-28A	Protein expression vector for pilin crystallization, kan^R	EMD
pET-M11	Protein expression vector for pilin crystallization, kan^R	EMD
pSRM24	*pilA1* in pET28A, kan^R	This study
pSRM25	*pilA1* in pETM11, kan^R	This study
pSRM26	*pilA3* in pET28A, kan^R	This study
pSRM27	*pilA3* in pETM11, kan^R	This study
pSRM28	*pilA4* in pET28A, kan^R	This study
pSRM29	*pilA4* in pETM11, kan^R	This study
pSRM30	*cpe2279* in pET28A, kan^R	This study
pSRM31	*cpe2279* in pETM11, kan^R	This study
pSRM32	*cpe2280* in pET28A, kan^R	This study
pSRM33	*cpe2280* in pETM11, kan^R	This study
Primers		
*pilA1::*pET28a		
oSM273	5’ ctccaaCCATGGcatatgtaaggatagcgc 3’ (NcoI site)	
oSM274	5’ ctttaattaaagtatatCTCGAGtatagacaccg 3’ (XhoI site)	
pilA1::pETM11		
oSM273	5’ ctccaaCCATGGcatatgtaaggatagcgc 3’ (NcoI site)	
oSM275	5’ ctttaattaaaCTCGAGttatactatagacaccg 3’ (XhoI site)	
*pilA3::*pET28a		
oSM276	5’ ctttaCCATGGattatgtaactttagataagacaagc 3’ (NcoI site)	
oSM277	5’ cacctccaacccatcttCTCGAGtttatttctaaaag 3’ (XhoI site)	
pilA3::pETM11		
oSM276	5’ ctttaCCATGGattatgtaactttagataagacaagc 3’ (NcoI site)	
oSM278	5’ cacctccaaccCTCGAGttatttttttatttctaaaag 3’ (XhoI site)	
*pilA4::*pET28a		
oSM279	5’ caacatCCATGGtgattataaaagatagggggg 3’ (NcoI site)	
oSM280	5’ ctatatttcctcataaaCTCGAGaacattaattttcc 3’ (XhoI site)	
pilA4::pETM11		
oSM279	5’ caacatCCATGGtgattataaaagatagggggg 3’ (NcoI site)	
oSM281	5’ ctatatttcctCTCGAGtcacttaacattaattttcc 3’ (XhoI site)	
*cpe2279::*pET28a		
oSM282	5’ cagctaCCATGGcactttctttagcaaattataaagc 3’ (NcoI site)	
oSM283	5’ ccatctttctcttaatacCTCGAGtacttccc 3’ (XhoI site)	
cpe2279::pETM11		
oSM282	5’ cagctaCCATGGcactttctttagcaaattataaagc 3’ (NcoI site)	
oSM284	5’ ccatctttctctCTCGAGttagtttacttccc 3’ (XhoI site)	
*cpe2280::*pET28a		
oSM285	5’ cccctaggtaCCATGGctttagcagctgttaaaac 3’ (NcoI site)	
oSM286	5’ cttcatgtataacatattaCTCGAGtcctataaattc 3’ (XhoI site)	
cpe2280::pETM11		
oSM285	5’ cccctaggtaCCATGGctttagcagctgttaaaac 3’ (NcoI site)	
oSM287	5’ cttcatgtataacCTCGAGtcactctcctataaattc 3’ (XhoI site)	

3.3 INVESTIGATION OF POSSIBLE TYPE IV PILUS GENE REGULATORS FROM *Clostridium perfringens*

Clostridium perfringens utilizes Type IV pili for adherence to muscle cells, as well as for twitching and gliding motility (24). From previous RNA-sequencing data analyses (see Chapter 2), Type IV pilus system genes were found to be transcriptionally upregulated in certain growth conditions. *pilA2*, the major pilin gene in an operon with other necessary T4P proteins, had significantly more transcription in plate-grown cells than in their liquid counterparts. Transcriptional and translational assays with β-glucuronidase as the reporter enzyme support the differential activity of the *pilA2* promoter from plate-grown cells and cells grown in nutrient-rich media. *C. perfringens* cells were also shown to adhere more readily to muscle cells when taken from plates than when harvested from liquid. To better understand the regulatory mechanisms that cause transcriptional upregulation of T4P genes in glucose-containing plate conditions, we utilized RNA-sequencing data to identify transcriptional regulators that were highly expressed in the same environment (see Table 7). *sigV* and *anti-sigV*, an extra-cytoplasmic sigma factor and its antagonist partner, were highly regulated in plate-grown bacteria and were not differentially expressed in liquid media. Because this regulation pattern is also seen in *pilA2*, we chose to delete *cpe0559* (*sigV*) and *cpe0560* (*anti-sigV*) to observe the change in cell morphology and surface-associated PilA2 in hopes of uncovering which regulators affect T4P production.

MATERIALS & METHODS

Bacterial strains and culture conditions. All bacterial strains used in this study are listed in Table 11. *E. coli* was grown in Luria-Bertani (LB) medium with antibiotics added as necessary (100 μg/ml kanamycin or 20 μg/ml chloramphenicol) and 1% agar for plates. *C. perfringens* was grown in a Coy anaerobic chamber in brain heart infusion (BHI) medium (20 μg/ml chloramphenicol) and/or 1% agar when necessary, or in fastidious anaerobe broth with 2% glucose added (20 μg/ml chloramphenicol) and/or 1% agar.

Construction of *cpe0559* in-frame deletion mutant. The protocol from Chapter 3.1 was followed with some minor changes. To create plasmid pSRM51, the regions surrounding the *cpe0559* gene, including the first three codons and the stop codon, were amplified from *C. perfringens* strain 13 chromosomal DNA with primers OSRM84 through OSRM86. The same method was repeated for deletion of *cpe0560*, with primer pairs OSRM88 and OSRM89, OSRM90 and OSRM91, and OSRM89 and OSRM90. Vector pSMART-HC Kan was used in lieu of pTOPO for *cpe0560* cloning in *E. coli*.

Cell-length imaging and measurements. *C. perfringens* HN13, Δσ^V and Δ*anti*-σ^V were grown in BHI, PGY, and FABG liquid media at 37°C to log phase or on plates at 37°C overnight. Cells were resuspended in PBS and spotted onto polylysine slides sealed with a coverslip. The cells were then imaged on an Olympus IX71 using the DeltaVision SoftWorx program (Applied

Precision) with an Olympus 60x or 100x UPLS Apo phase objective. Lengths of cells were obtained using the ROI manager function of ImageJ (63). The number of cells examined for each sample ranged from 100 to 335.

RNA-sequencing analyses. Method for RNA preparation and analysis can be found in Chapter 2. Differentially expressed genes were found by taking the ratio of average TPM of each gene in the plated condition over the liquid condition, and the ratios for BHI, PGY, and FABG media were compared. Genes with a fold change of 4 times or greater are listed in Chapter 2.

RESULTS

***σ^V* and *anti-σ^V* are highly upregulated in PGY and FABG plate-grown cells.** Figure 19 illustrates the average transcripts per million for three T4P genes and two transcription factor genes, σ^V and *anti-σ^V*. *pilC2* and *pilB2* were transcribed significantly more in plate-grown cells over broth-grown cells in two of the three media conditions tested, PGY and FABG. Both of these media contain glucose while BHI has no added carbohydrate. The major pilin gene, *pilA2*, also has more transcripts on plates of FABG. Another gene in the T4P operon, coding for a pilin-like protein CPE2280, has six times more reads on PGY and FABG plates than in the liquid cultures (Figure 5). Similarly, σ^V and *anti-σ^V* transcription was upregulated 40-fold on PGY plates and 14-fold on FABG plates.

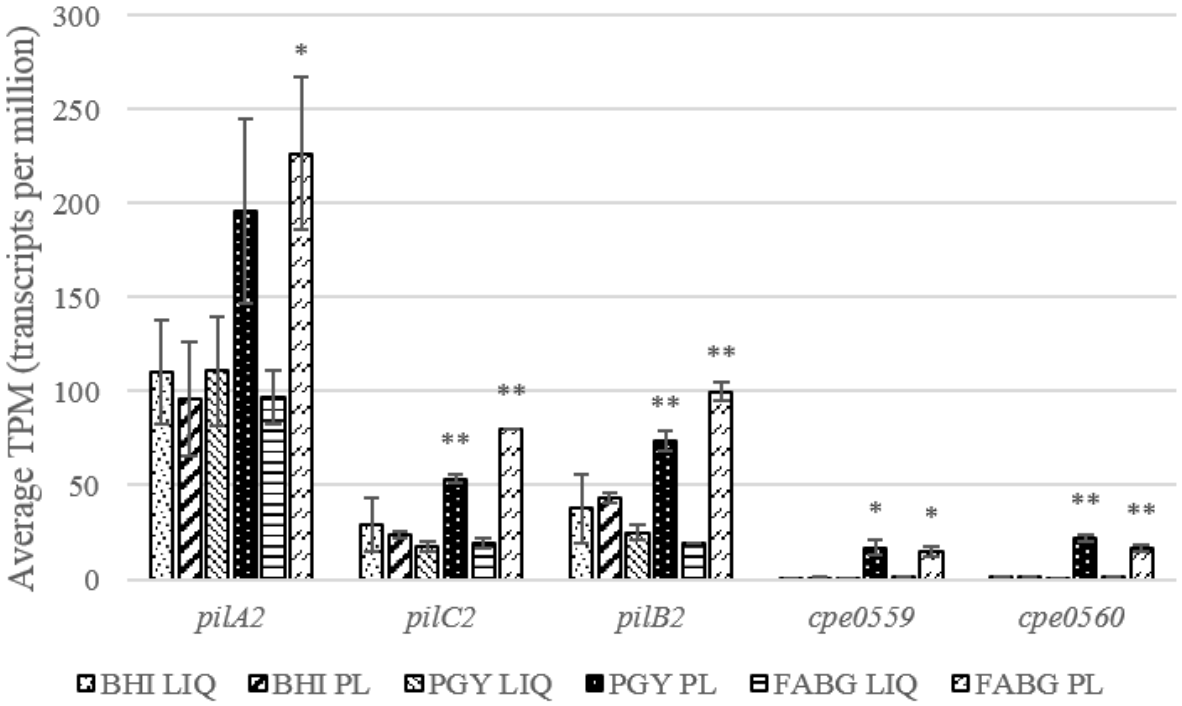


Figure 19: Normalized transcripts of Type IV pilus genes and transcription factors σ^V and *anti-*σ^V. *pilC2*: integral membrane core protein, *cpe2285*. *pilB2*: assembly ATPase, *cpe2286*. σ^V: *cpe0559*. *anti-*σ^V: *cpe0560*. P-values: <0.5 (*), <0.01 (**). P-values obtained using Student's t-test Error shown is standard error of the mean.

Δσ^V and Δ*anti*-σ^V have reduced growth rate in liquid culture. In PGY liquid media, Δσ^V and Δ*anti*-σ^V were assessed for growth deficiency by recording density of cells in absorbance over time. Both mutant strains have decreased growth rate in log phase compared to wild type cells, but total yield in stationary phase was not affected by these deletions (Figure 20).

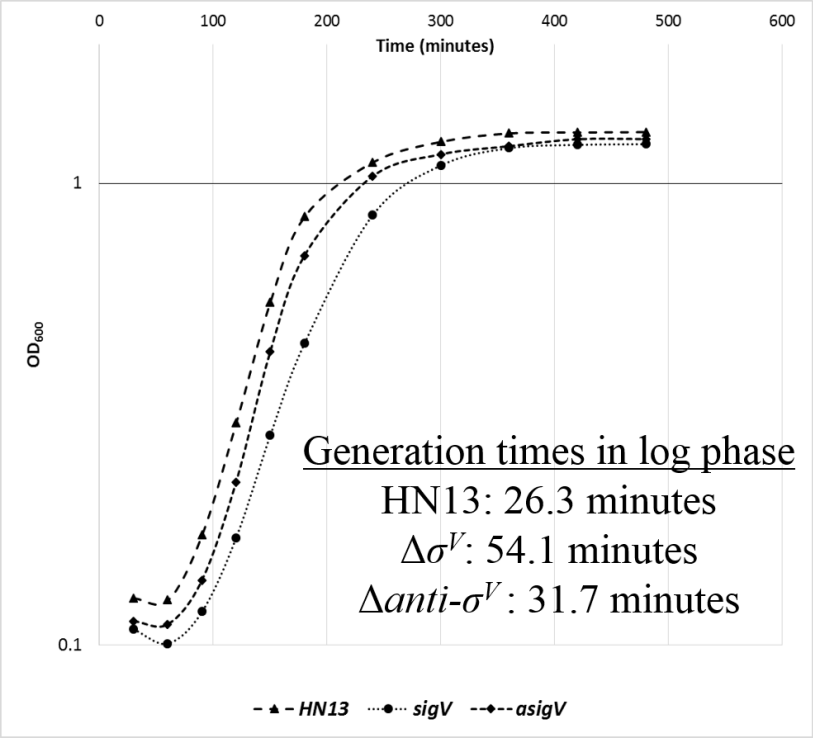


FIGURE 20. Growth curve and growth rate comparison of wild type, Δσ^V and Δ*anti*-σ^V. Cells were grown in 100-ml serum bottles for eight hours, with OD_{600} (y-axes) recorded every thirty minutes (x-axes) for the first two hours post-inoculation, then every hour for eight total hours of growth (*left*). Three points from log phase were used to calculate growth rate equation (*right*). Growth curves are representative of duplicate trials.

Δσ^V has irregular cell division in PGY and FABG plate-grown cells. Because increased transcription is observed in plate conditions, the largest effects from deletion of these genes was expected to be in plate-grown cells. Wild type (HN13) cells and transcription factor deletions were grown on plates and imaged with phase microscopy to determine irregular cell shape or size. Δσ^V cells were significantly longer than wild type cells when grown in all plated conditions tested, up to 18% longer. The largest differences in cell length were observed on PGY and FABG plates (Figure 21).

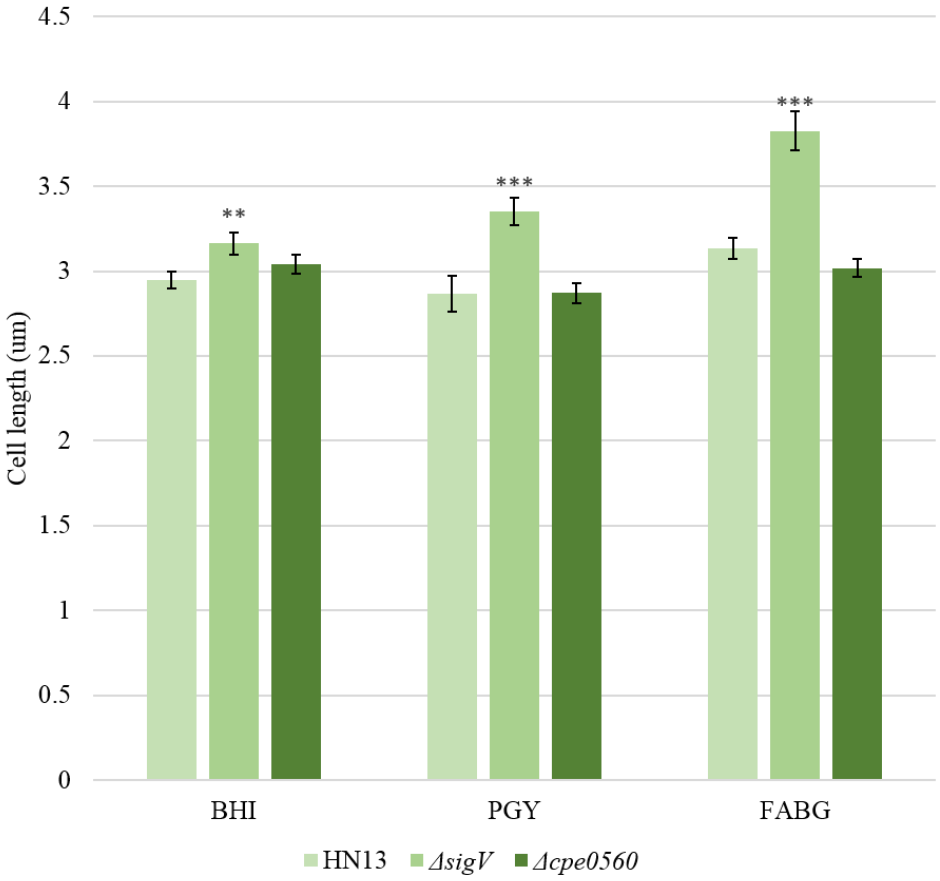


FIGURE 21. Cell lengths of HN13, Δ*sigV*, and Δ*cpe0560* cells. Bacteria were grown on plates of three media. Lengths were obtain by measuring distance between poles in ImageJ.

$\Delta\sigma^V$ cells were observed to have irregular cell division depending on the media of growth. On PGY plates, cells lacking σ^V had micro-cells attached to their terminal ends (Figure 22B). In contrast, some $\Delta\sigma^V$ cells were highly elongated when grown on FABG plates (Fig. 22D). These morphologies were each unique to a single condition.

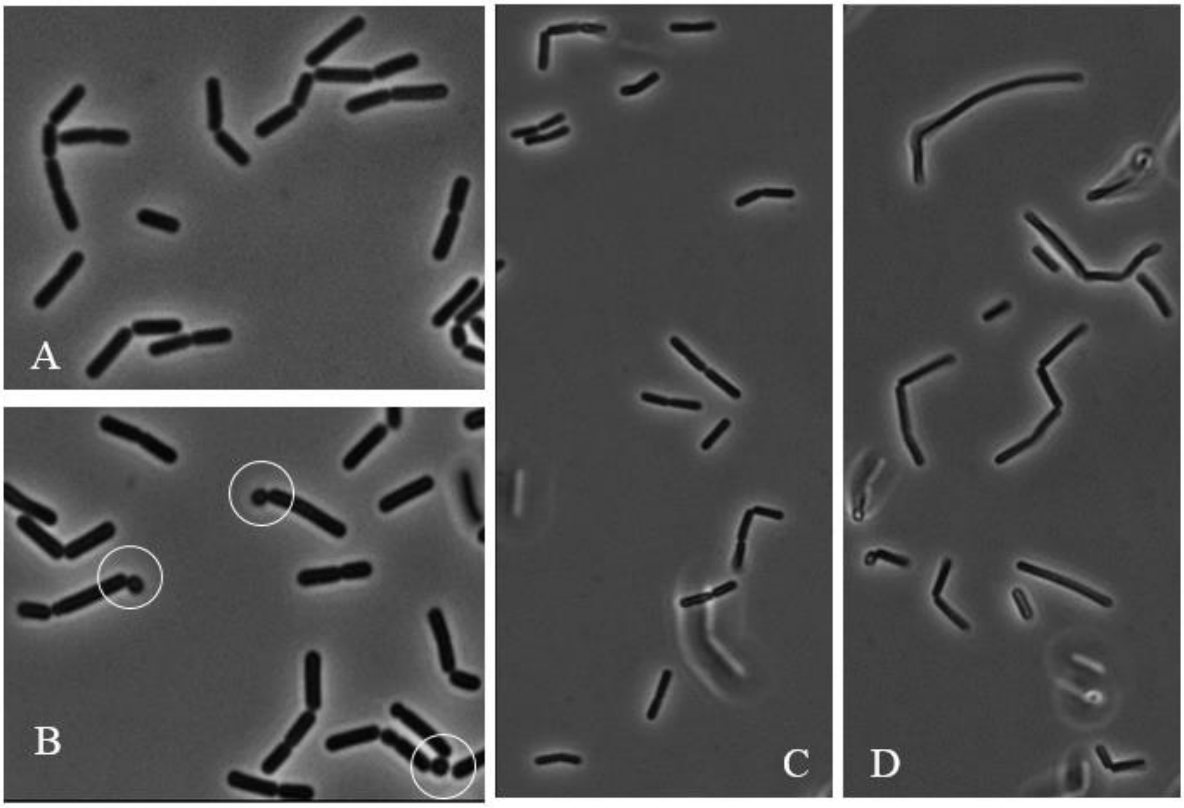


FIGURE 22. Irregular cell morphologies of $\Delta\sigma^V$ plate-grown cells. Cells were grown at 37°C overnight, scraped from the edges of colonies, and visualized by phase microscopy. (A) HN13, PGY plate, 100x. (B) $\Delta\sigma^V$, PGY plate, 100x. Circles indicate microcells. (C) HN13, FABG plate, 60x. (D) $\Delta\sigma^V$, FABG plate, 60x.

DISCUSSION

ECF-sigma factors regulate gene networks in response to an external stimulus, with the help of an antagonist protein, the anti-sigma factor. Once an external stimulus is detected, the ECF-sigma factor is released from the antagonist by a protease and binds DNA to direct the RNA polymerase to transcribe genes in response (35). RNA-sequencing analyses and adherence assays reveal that T4P are differentially transcribed in *C. perfringens* cells grown on solid media. This preference for solid versus liquid in the external environment indicates that the regulatory mechanism occurs in two steps: the external environment is being sensed and a message is being relayed to the internal environment of the cell to change gene expression. By comparing solid versus liquid state transcription levels of all genes on the chromosome, it was possible to narrow down a long list of regulators capable of this function. Δσ^V and Δ*anti-*σ^V were very highly upregulated in conditions that also showed increased transcription of T4P genes and, due to the mechanism of ECF-sigma factor regulation, seemed to be possible candidates for transcriptional regulators of T4P genes. While it is very possible that the regulator of T4P genes may not show any differential expression between conditions and therefore unable to be identified in this manner, investigation of differentially transcribed regulators acted as the best first means of narrowing down a large pool of possible regulatory mechanisms that could act on T4P.

Currently literature on σ^V indicates no relationship to T4P or their regulation. In *Bacillus subtilis*, the anti-sigma factor RsiV directly binds lysozyme and initiates proteolytic degradation by a signal peptidase, degrading RsiV and releasing σ^V to aid in transcription of genes for

lysozyme resistance (36, 78). Similar lysozyme-sensitive phenotypes have also been documented in $\Delta\sigma^V$ mutants in *Enterococcus faecalis* and *Peptoclostridium difficile* (79, 80). Utilizing similar methods to experiments performed on these organisms, we were unable to find any relationship to lysozyme resistance in $\Delta\sigma^V$ in *C. perfringens* (data not shown). However, $\Delta\sigma V$ did have significantly altered cell morphologies that could be effects of disordered cell wall degradation, such as the increased cell length and the presence of microcells at the poles of some cells. To date, no evidence has conclusively shown that σ^V directly controls T4P gene transcription, and it is possible that this regulator acts in a cascade that can go on further to effect T4P production. Future experiments could elucidate this relationship by quantification of PilA2 on the surface of $\Delta\sigma^V$ cells.

Table 11 Strains, plasmids, and primers used in ECF sigma factor study.

Strains		
Name	**Relevant characteristics or sequence**	**Source**
E. coli		
DH10B	F$^-$ *mcrA*Δ *mrr-hsdRMS mcrBC* ϕ80d*lacZ*ΔM15 *lacX74 deoR recA1 araD139*Δ *ara, leu7697 galU galK*Δ*rpsL endA1 nupG*	Gibco/BRL
C. perfringens		
HN13	*galKT* deletion strain of strain 13	(58)
SRM11	Δ*sigV* (*cpe0559*)	This study
SRM12	Δ*αsigV* (*cpe0560*)	This study
Plasmids		
pCR-Blunt II-TOPO	Blunt PCR cloning vector	Invitrogen
pCM-GALK	Contains a *C. beijerinckii galK* gene under control of a ferredoxin promoter from *C. perfringens*	(58)
pSMART-HC Kan	High copy PCR blunt cloning vector with terminators flanking MCS	Lucigen
pSRM50	Δ*sigV* in pTOPO	This study
pSRM51	Δ*sigV* in pCM-GALK	This study
pSRM52	Δ*αsigV* in pSMART	This study
pSRM55	Δ*αsigV* in pCM-GALK	This study
Primers		
Name	**Sequence**	
OSRM84	5’ ggtGTCGACccagaagaggttgctaaagc 3’ (SalI site)	
OSRM85	5’ gctaggaggggaaagatggaaaaaactatgggctatttaaatgaataaaaaagacttag 3’	
OSRM86	5’ ctaagtctttttattcatttaaatagcccatagtttttccatctttcccctcctagc 3’	
OSRM87	5’ ccgtaaaacttactttagcttctgcagcatagGGATCCtgc 3’ (BamHI site)	
OSRM88	5’ ggcGTCGACatatgttaaagatgagcaggatgctcttgaag 3’ (SalI site)	
OSRM89	5’ ggagggctatttaaatgaataaaaaagacggaaataaatttagaataaacaaggtttat 3’	
OSRM90	5’ ataaaccttgtttattctaaatttatttccgtctttttattcatttaaatagccctcc 3’	
OSRM91	5’ ctcGGATCCgtagcacagtgagtaactgcagctacg 3’ (BamHI site)	

3.4 IDENTIFYING THE FLOATING MECHANISM OF *Clostridium perfringens*

Clostridium perfringens is a Gram-positive pathogen known for being the causative agent of gas gangrene, a highly invasive infection causing muscle breakdown and large amounts of CO_2 and H_2 gas production under the skin (9). It is currently unknown what role gas production plays in the rapid spread of infection in this disease. One possibility is that gas production allows for neutral buoyancy in liquid media. An examination of the *C. perfringens* strain 13 genome indicated that this organism lacks known chemotaxis protein-encoding genes as well as flagella-encoding genes (22). In previous experiments, *C. perfringens* has been observed maintaining a neutral buoyancy the full length of a liquid column, although they have no obvious proteins enabling them to swim in liquid media.

The ability to maintain buoyancy in liquid without the energy expenditure would be highly advantageous to *C. perfringens*. Ecologically, this bacterium is ubiquitous in the environment, including anaerobic freshwater sediments where remaining suspended may make it more competitive (81). Suspended bacteria have greater access to increased levels of nutrients due to diffusion, with some bacteria even abandoning protective biofilms to become planktonic in starvation situations (82). In order to achieve this, the bacterium may use one of two mechanisms: trapping gas in vacuoles to maintain buoyancy, or holding gas bubbles on the surface of the cell to maintain buoyancy. Gas vacuoles are usually readily visible using light or electron microscopy, and no such structures have been observed in *C. perfringens* cells. Previously, a T4P minor-pilin deletion mutant Δ*pilA3* was shown to lack the ability to remain

suspended in liquid when subjected to slow centrifugation (data not shown). With this evidence and the lack of gas vacuoles present, we hypothesize that buoyancy in *C. perfringens* cells is due to a surface factor that is either PilA3 itself or an undiscovered surface protein. We used a sinking phenotype enrichment on a mutant library constructed by random insertions of a mariner-transposon. Mutants that lack the ability to float sink to the bottom of long glass tubes filled with media, and multiple enrichments will select for mutants with the largest defect in floating. Genes are then easily identified by sequencing around the point of the transposon insertion, and "floating genes" can be further studied for their physiological contribution to the floating phenotype. With this method, we are able to investigate this bacterium's unique ability to float and the effect of this ability on gangrenous disease progression.

MATERIALS & METHODS

Bacterial strains and culture conditions. All bacterial strains and vectors used in this study are listed in Table 12. *C. perfringens* mutants were grown in brain heart infusion medium (BHI) with 30 μg/ml erythromycin and 1% agar for plates. Incubations were performed at 37°C in a Coy anaerobic chamber for anaerobic growth. *Escherichia coli* strains were grown in Luria-Bertani broth (LB) with 100 μg/ml erythromycin and 100 μg/ml ampicillin and 1% agar for plates.

Enrichment for mariner-transposon mutants with sinking phenotype. Fifteen Pyrex glass tubes (22 cm long) were stopped with a rubber stopper at the bottom and autoclaved, then filled with 20 ml sterile BHI plus 30 μg/ml erythromycin in a Coy anaerobic chamber and allowed to warm to 37°C. The first tube was inoculated with 500 μl of an overnight culture of a *C. perfringens* strain HN13 mariner-transposon library constructed in previous experiments (83). This tube was allowed to incubate for twelve hours at 37°C, after which 0.5 ml culture was collected by syringe through the rubber stopper and transferred to the top of the second tube. The second tube was incubated for twelve hours, and the process was repeated until all fifteen tubes had been incubated for twelve hours anaerobically. From the fifteenth tube, cells were collected from the bottom and diluted to a final 10^{-4} dilution. Ten microliters of the last dilution was placed on two BHI erythromycin plates and incubated overnight. Twenty-five colonies were patched to BHI erythromycin from each plate and grown overnight, totaling 50 colonies. Each

of the colonies was individually inoculated into Pyrex tubes and incubated, and five tubes showing the strongest sinking phenotype were chosen for sequencing.

Sequencing analysis of sinker mutants. The protocol for identifying the transposon insertion points has been previously published (83). The Quick Extract chromosomal DNA extraction kit (Epicentre) was used to extract chromosomal DNA from the five sinker mutants. Chromosomal DNA and vector pBSII were incubated overnight at 16°C with HindIII-HF. Cut chromosomal DNA was ligated into cut pBS11 using T4 DNA ligase per manufacturer's protocol (Promega), and ligations were electroporated into *E. coli* electrocompetent strain DH10B. Transformants with vectors containing the transposon in pBS11 were selected for by growth on LB plates containing erythromycin and ampicillin.

RESULTS & DISCUSSION

Five mutants were selected for greatest sinking phenotype. Of the fifty patched colonies from the sinking enrichment process, five mutants were chosen and given number designations: 9, 10, 17, 8A, and 16A. Mutants will be sequenced at the site of the transposon insertion to identify genes that directly affect the ability of *C. perfringens* to float in liquid media.

Once the transposon-inserted genes that affect floating are identified, in-frame deletions will be constructed to observe the full effect of the sinking phenotype for each gene. The genes will also be complemented on a lactose-inducible vector pKRAH1 (59), and the floating phenotype should be restored. "Floating genes" will then be further investigated to uncover the mechanism of *C. perfringens* floating in media without flagella.

Table 12 Strains, plasmids, and primers used in floating phenotype study.

Strains		
Name	**Relevant characteristics or sequence**	**Source**
E. coli		
DH10B	F$^-$ *mcrA*Δ *mrr-hsdRMS mcrBC* ϕ80d*lacZ*ΔM15 *lacX74 deoR recA1 araD139*Δ *ara, leu7697 galU galK*Δ*rpsL endA1 nupG*	Gibco/BRL
C. perfringens		
HN13	*galKT* deletion strain of strain 13	(58)
Plasmids		
pHLL24	Mariner transposon mutant library, in HN13	(83)
pBluescriptII KS+	Ampicillin resistance	Stratagene
Primers		
Name	**Sequence**	
OHLL21	5' GCAATGAAACACGCCAAAGTAAACAATTTAAGTACCG 3'	(83)
OHLL22	5'GTTTTATTATTTGGTTGAGTACTTTTTCACTCG 3'	(83)

3.5 COMPUTATIONAL VERIFICATION OF PREDICTED SMALL RNAS IN *Clostridium perfringens* STRAIN 13

The recent development of next-generation sequencing has brought small RNA research to the forefront of biological inquiry today, not only proving vital in understanding eukaryotic cells but prokaryotic pathogens as well. Until this technology expanded, small RNAs (sRNAs) were often overlooked in genetic analyses, due to their small coding size and intergenic or antisense location to other important genes, like elements seen in phages and plasmids (84, 85). Small RNAs are described as RNA transcripts ranging from 50 – 500 base pairs that do not encode for proteins and have target sequences that are usually located in 5′ untranslated regions (UTRs) as short as 6–7 nucleotides (86). These RNAs act as regulatory elements of translation initiation or messenger RNA stability, either by binding an RNA target and marking it for degradation, by binding with an RNA-binding protein, or by inhibiting translation by the ribosome at the RBS or within the protein-coding region. Small RNA regulators can even be multifunctional and provide answers to a number of previously mysterious regulatory effects seen before deep sequencing existed (87).

Small regulatory RNAs are often synthesized in the cell as a stress response, such as during glucose starvation or iron starvation, oxidative stress, temperature shock, and radiation (88). The complementary nature of small RNAs to their mRNA counterparts makes them perfect for specific, targeted regulation before translation of proteins occurs. Bound mRNAs are either destroyed due to their double-stranded nature (typical of phages), or mRNAs can reach their half-life before the ribosome is allowed to bind and translate proteins. Thus, small RNAs have transcriptional and translational downstream effects. Decay of sRNA-bound mRNA is overseen

by RNase III, which cleaves mRNA-sRNA duplexes, or by RNase E, which cleaves single-stranded RNA not protected by the massive ribosome during translation (88). These finely tuned regulatory mechanisms are internally monitored by the cell due to their destructive and powerful nature and their ability to greatly alter cellular physiological responses.

In 2011, Chen et al computationally predicted the small RNAs in all clostridial species and strains sequenced up to 2009 (68). The group's approach integrated genetic features of sequence conservation between species, and predictions of transcriptional terminators and clostridial promoters, to predict 193 small RNAs in strain 13 of *C. perfringens*. A random set of predicted sRNAs were validated in *C. acetobutylicum* and three strains of C. *botulinum* using quantitative real-time PCR and Northern blotting analyses. Thirty of these small RNAs were also observed in RNA-seq datasets by Low et al. Of the 93 novel small RNAs this group discovered in *virR* and *revR* regulator mutants, four were validated by qRT-PCR and one by Northern blotting (69). This project further expands upon these validations by providing evidence of expression of predicted small RNAs in another strain, *C. perfringens* strain 13.

First, a lower cut-off of transcription will be established, due to the small coding size of small RNAs. Evaluating this cut-off will correct for false positives due to any random sequence alignment in intergenic regions on the chromosome. Single-end high-throughput Illumina sequences of cDNA converted from RNA of twelve samples from six biological conditions will be assessed for transcription at the predicted small RNA sites in the strain 13 genome using the alignment program Geneious. If the small RNA has transcripts aligning at the computationally predicted site in the correct orientation and above the random noise cut-off, it will be called verified. Due to the diversity of the conditions selected, small RNAs rejected by this data may be transcribed in very select conditions or not at all in this organism. This method can use

bioinformatics to confirm the existence of many small RNAs for future work without the need for specific sequencing experiments to be performed prior to Northern blotting, and it can be a valuable tool for future regulatory studies in *C. perfringens*.

MATERIALS & METHODS

Data retrieval. RNA-sequencing data from experiments performed in Chapter 2 were utilized in this project; methods, bacterial strains, growth conditions, and RNA extraction can be found there. Computationally predicted small RNAs were annotated into the genome of strain 13 in Geneious as detailed in Table S1 from Chen et al (68). Single-end reads were mapped to the modified annotated genome using the low sensitivity Geneious mapper to distinguish transcript orientation.

Intergenic background noise. Because RNAs are transcribed at low levels, the background noise of the technology from sequencing errors was calculated. To do this, ten intergenic regions of 100 base pairs were selected across the genome of strain 13, and mean coverage per base was extracted from Geneious for two different samples, BHI LI 1 and FABG PL 1. The average was calculated for each sample, totaling 1.5 transcripts per base (Table 13). This average was then corrected by calculated the expected error using the Poisson distribution for random sampling. This corrected the background noise with a higher limit of 2.75 transcripts, which was used as the conservative cut off for confirmation of sRNA expression (Table 14).

Bioinformatic confirmation of predicted small RNAs. Raw read counts of all predicted small RNAs were extracted from Geneious for each of the twelve samples, and biological replicates were averaged. Small RNAs were considered confirmed if they had an average of 2.75 transcripts aligned to them in at least two of six conditions tested.

Differential expression of small RNAs in plate-grown bacteria versus liquid cultures. Confirmed small RNAs were then assessed for differential expression in different media conditions. For each type of media used (BHI, PGY, and FABG), the transcripts per million (TPM) of each small RNA in plates was divided by the TPM for the liquid culture. Small RNAs were considered upregulated if the ratio of TPM on plates was 2 or greater. This ratio was recalculated to determine upregulation in liquid cultures by dividing mean liquid transcripts by mean plate transcripts. Small RNAs were considered upregulated on agar plates over liquid media if at least two media showed two-fold regulation, and the same was regarded for liquid cultures.

Differential expression of small RNAs in media of differing nutritional content. Two-fold upregulation was screened for each confirmed small RNA by calculating ratios of TPM between BHI, PGY, and FABG averages in both plate and liquid grown cultures. For example, upregulation of FABG was considered for plate grown bacteria by dividing the TPM of each small RNA in FABG versus BHI and FABG versus PGY. If there was two-fold regulation of the small RNA in a certain media versus both of the other two, that small RNA was deemed upregulated in that media in a particular condition (plate or liquid).

RESULTS

With a background noise level of 2.75 transcripts calculated from the intergenic regions of strain 13 (Table 13), 134 small RNAs were confirmed out of 193 that were computationally predicted, a 69% acceptance rate. Eighty-five small RNAs were expressed in all conditions that were tested and may be constitutively expressed. It is important to note that predicted small RNAs not confirmed in this study may still encode for expressed small RNAs that may be transcribed in other conditions not tested by this study.

Analyses for differential expression on plates revealed that 40 confirmed small RNAs are upregulated in at least two types of media when normalized plate culture transcripts were compared to their liquid counterparts (Table 14). In contrast, only 6 small RNAs were upregulated in at least two liquid media over their respective plate grown equal. This may mean that more regulation is required for processes in plate-grown bacteria than in liquid, which may include motility, biofilm formation, and carbon starvation mechanisms. A list of the identified upregulated small RNAs are listed in Table 15.

Table 13 Calculation of intergenic noise for small RNA confirmation cutoff.

Intergenic Regions	Start bp	End bp	Mean Coverage Per Base (BHI LI 1)	Mean Coverage Per Base (FABG PL 2)
IR1	93497	93596	3.0	0.0
IR2	396470	396569	1.7	1.0
IR3	660803	660902	0.0	0.0
IR4	976367	976466	0.3	0.0
IR5	1219580	1219679	5.9	8.5
IR6	1581009	1581108	0.0	1.8
IR7	1808426	1808525	0.7	0.1
IR8	2127264	2127363	2.8	1.8
IR9	2425461	2425560	0.0	0.2
IR10	2805494	2805593	0.8	1.6
		Average (n)	1.5	1.5
	Range of error (+/-√n)		Lower limit 0.287117199	Upper limit 2.752882801
			sRNA Cutoff	2.75

The mean coverage per base of ten intergenic regions of 100 base pairs in length were averaged for two samples. Both samples had the same average coverage. A range of error was calculated, and the upper error limit of basal expression that could exist was taken as the cutoff for confirming sRNA expression as more than background noise.

Table 14 Comparison of upregulated small RNAs across samples.

	Liquid culture	1% agar culture	Both liquid & plate cultures
FABG	52	23	9
PGY	4	8	0
BHI	5	0	0
Total upregulated sRNAs	61	31	9

Ratios of transcripts per million (TPM) were calculated between plate and liquid grown cultures to determine upregulation. A cutoff of 2-fold regulation was considered a significant increase in transcripts. The total number of small RNAs being transcribed was highest in liquid cultures, and the media with the highest nutritional content had the most expressed sRNAs between the media.

Table 15 Upregulated verified small RNAs in *C. perfringens* strain 13.

FABG PL	FABG LIQ	FABG LIQ (cont.)	PGY PL	PGY LIQ	BHI LIQ	FABG PL & LIQ
sCPE1040	*sCPE102*	*sCPE2516*	*sCPE150*	*sCPE1536*	*sCPE1165*	*sCPE1202*
sCPE1202	*sCPE103*	*sCPE2570*	*sCPE1505*	*sCPE249*	*sCPE1550*	*sCPE1868*
sCPE135	*sCPE1039*	*sCPE2601.1*	*sCPE2516*	*sCPE2812*	*sCPE1662*	*sCPE2036*
sCPE1494	*sCPE1080*	*sCPE2615.1*	*sCPE2642*	*sCPE665.1*	*sCPE2169*	*sCPE2615.1*
sCPE1538	*sCPE1118*	*sCPE2642*	*sCPE2784*		*sCPE966*	*sCPE2653*
sCPE168	*sCPE1128*	*sCPE2653*	*sCPE539.1*			*sCPE2754*
sCPE1799	*sCPE1136*	*sCPE2754*	*sCPE555*			*sCPE2821*
sCPE1868	*sCPE116*	*sCPE2784*	*sCPE631*			*sCPE2872*
sCPE2036	*sCPE1200*	*sCPE281*				*sCPE313*
sCPE2228	*sCPE1202*	*sCPE281.1*				
sCPE252	*sCPE1257*	*sCPE2821*				
sCPE2615.1	*sCPE1380*	*sCPE2872*				
sCPE2653	*sCPE1438*	*sCPE289*				
sCPE2754	*sCPE150*	*sCPE313*				
sCPE2805	*sCPE1505*	*sCPE345.1*				
sCPE2821	*sCPE1522*	*sCPE361*				
sCPE2872	*sCPE1540*	*sCPE369*				
sCPE313	*sCPE1598*	*sCPE465*				
sCPE434	*sCPE1705*	*sCPE539*				
sCPE665.1	*sCPE1751*	*sCPE555*				
sCPE701	*sCPE1789*	*sCPE614*				
sCPE844	*sCPE1868*	*sCPE657*				
sCPE869	*sCPE1967*	*sCPE665*				
	sCPE2036	*sCPE705*				
	sCPE2460	*sCPE732*				
	sCPE2481	*sCPE980*				

*Names pulled from Chen et al (68). Small RNAs were renamed by Low et al (69).

3.6 APPLIED TRANSCRIPTOMICS: NATIVE PROTEIN PURIFICATION FROM *Clostridium perfringens* USING RNA-SEQ

Previous efforts in our lab to purify T4P proteins from *C. perfringens* for downstream binding experiments and antibody production have been fruitless. T4P proteins lack high enough endogenous expression for purification directly from this organism, and overexpression in the lactose-inducible expression vector pKRAH1 has not yielded enough protein for downstream applications. Without a better alternative to purify proteins from *C. perfringens,* previous students have utilized *E. coli* and *P. aeruginosa* for overexpression and purification of T4P proteins (89). While this method does yield usable amounts of protein, the exact characteristics of the native protein may not be preserved in other organisms. Possible protein adulterations from other organisms could include protein misfolding, protein insolubility, and lack of typical post-translational modifications. This is especially true in *C. perfringens,* which has been found to modify prototoxin intracellularly before secretion (90).

In order to bypass the use of non-native organisms for *C. perfringens* protein purification, we developed a method to overexpress native proteins using highly active native promoters. We inserted histidine-tagged proteins into the cloning site of newly developed condition-specific overexpression vectors pARCA and pPYRB. These vectors, which utilize promoters that are upregulated in specific conditions as determined by RNA-sequencing data, allow for natural induction by normal bacterial growth instead of inducing agents or metabolites. pARCA can be used for plate-specific proteins while pPYRB is ideal for proteins secreted in liquid media.

MATERIALS & METHODS

Bacterial strains and culture conditions. Bacterial strains, plasmids, and primers used in this study are listed in Table 16. *Escherichia coli* strain DH10B was grown in Luria Bertani broth at 37°C for all transformations. When necessary, chloramphenicol was added to the media at a concentration of 20 μg/ml, respectively. *Clostridium perfringens* strain HN13, a Δ*galKT* derivative of strain 13 (58), was used as the wild type strain in this study. *C. perfringens* strains were grown anaerobically in BHI (brain-heart infusion, Thermo Fisher) in an anaerobic chamber (Coy Laboratory Products, Inc.), supplemented with chloramphenicol to maintain plasmids. Strains derived from strain AH2 (59) which also contained the β-glucuronidase reporter gene vector pSM240, required 30 μg/ml erythromycin and 20 μg/ml chloramphenicol to maintain chromosomal insertion constructs.

Assessing promoter strength of ParcA and PpyrB. RNA-sequencing data from Chapter 2 was used for comparative analysis. Upregulated genes were chosen as possible candidates for protein overexpression in *C. perfringens* due to high transcript levels in particular types of media while little to no expression was shown in the other conditions. β-glucuronidase activity of the promoter upstream of selected genes was assessed by creation of promoter fusion vectors pSRM49 and pSRM47 (see Table 16 for strains and primers) using the protocol from Chapter 2 with minor modifications. Instead of suicide vector pJV50, pSM240 was used.

Creation of pARCA and pPYRB vectors. An existing vector, pKRAH1, was used as the backbone for the pARCA and pPYRB protein overexpression vectors. pKRAH1 was digested with restriction enzymes KpnI and PstI to remove to previous promoter sequences, and P*arcA* or P*pyrB* was inserted into the vector and verified by PCR and Sanger sequencing using primer pairs OSRM80 and OSRM81 for pARCA and OSRM82 and OSRM83 for pPYRB.

Creation of overexpression strains for PilM-6H and ColA-6H.

Upstream and downstream primers were used to amplify the full sequence of each gene without its native stop site, with six AT-rich histidine residues and two stop sites attached at the C-terminal end (OSRM104 and OSRM105 for PilM-6H, OSRM106 and OSR107 for ColA-6H). Primers are listed in Table 16. Blunt-end PCR products were ligated into blunt-end shuttle vector pTOPO and electroporated into *Escherichia coli* strain DH10B using standard methods described previously. Transformants were screened via restriction digest with enzymes SacII and SacI. His-tagged genes were ligated into pARCA and pPYRB respectively using T4 DNA ligase (Promega) and transformed into *E. coli*. Correctly ligated vectors were subsequently transformed into wild type *Clostridium perfringens* strain HN13 using previously described methods. Transformants were verified by PCR and Sanger sequencing with original primers listed.

Purification of PilM-6H. Protein overexpression was induced through P*arcA* by growth on BHI plates with 1% agar. An overnight culture of PilM-6H in pARCA was grown anaerobically in a Coy anaerobic chamber at 37°C in BHI liquid supplemented with 20 μg/ml chloramphenicol. Using sterile cotton swabs, lawns were streaked onto ten BHI Cm plates and allowed to grow for 24 hours. Bacteria were harvested from the plates with PBS and centrifuged at 3500 x g for 15 minutes. The cells were washed once in 5 ml PBS and finally resuspended up to 5 ml with PBS. Peptidoglycan was removed by lysis with 10 mg lysozyme at 37°C for 1.5 hours. Spheroplasts were collected by centrifugation at 15,000 x g for 15 minutes at 4°C, then washed once in PBS to remove residual lysozyme, and resuspended in 5 ml PBS after repeated centrifugation. To complete lysis, spheroplasts were sonicated in a Model 500 Sonic Dismembrator (Fisher) at 20% amplitude for 20 minutes in 30 second bursts, followed by 30% amplitude for 20 minutes in 30 second bursts. The lysate was cleared by centrifugation at 25,000 x g for 30 minutes at 4°C. The clarified lysate was applied to a 5 ml HisTrap FF column (GE Healthcare Life Sciences) in PBS with 20 mM imidazole, and eluted with a linear gradient of 0.5 x PBS with 1 M imidazole. Fractions containing PilM-6H (as evaluated by SDS-PAGE) were pooled and dialyzed overnight in cellulose dialysis tubing suspended in PBS at 12°C. The dialyzed sample was concentrated and applied to a Superose 6 366 10/300 gel filtration column (GE Healthcare Life Sciences) equilibrated with 1x PBS. All chromatography procedures were performed using an ÄKTApurifier 10 FPLC system (GE Healthcare Life Sciences) at 4°C. Fractions were evaluated for purity by SDS-PAGE, and those that contained >90% PilM-6H were stored at 4 °C in 1x PBS + 10% glycerol. Protein concentration was determined with the Micro BCA Protein Assay kit (Thermo).

RESULTS

P*arcA* and P*pyrB* are highly active promoters in certain conditions. Transcript levels from RNA-sequencing were used to identify highly expressed genes that were only differentially expressed in few conditions, thus giving us the ability to selectively "turn on" protein expression by growth in a single condition while still using other conditions for culturing and growth without induction (Figure 23). In order to assess the strength of the upstream promoters of the genes selected, β-glucuronidase assays were performed with the non-native ribosomal binding site from the *cpe* gene. ParcA and PpyrB showed high levels of transcription through the promoter in some conditions corresponding to the RNA-sequencing (Figure 24). In cases where β-glucuronidase activity differs from RNA-sequencing transcript levels, we conclude the effect may be caused by saturation of regulatory elements that control gene expression in these conditions.

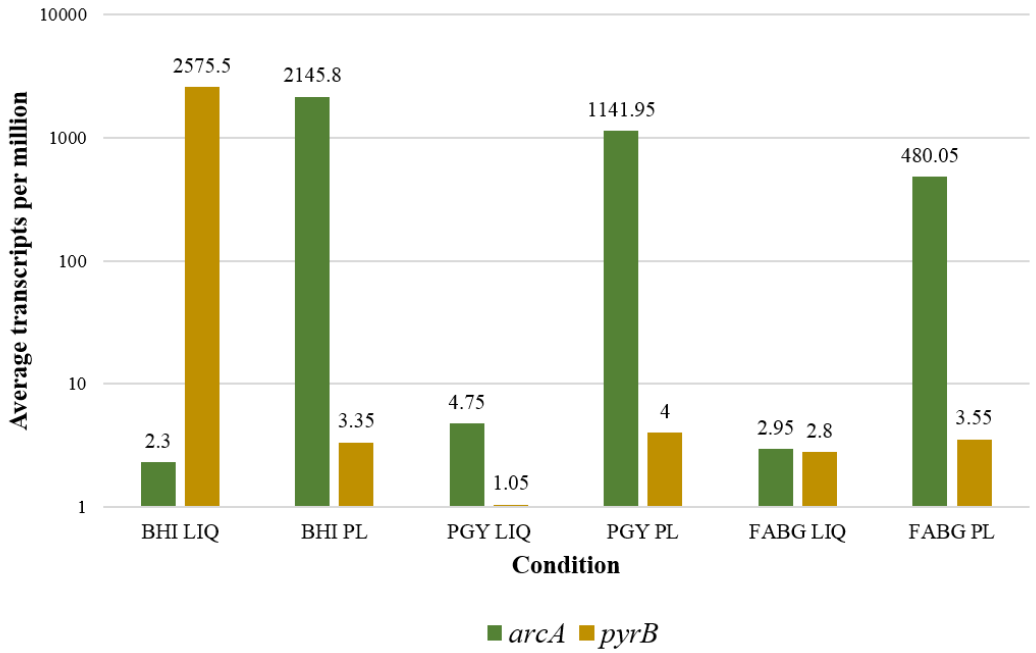


FIGURE 23. Normalized transcript levels of *arcA* and *pyrB* genes. The *arcA* gene, encoding for an arginine deiminase, is upregulated 640 times, 285 times, and 135 times on BHI, PGY, and FABG plates, respectively. The *pyrB* gene, encoding for an aspartate carbamoyltransferase, is only upregulated in BHI liquid culture, with a 1,119 fold increase.

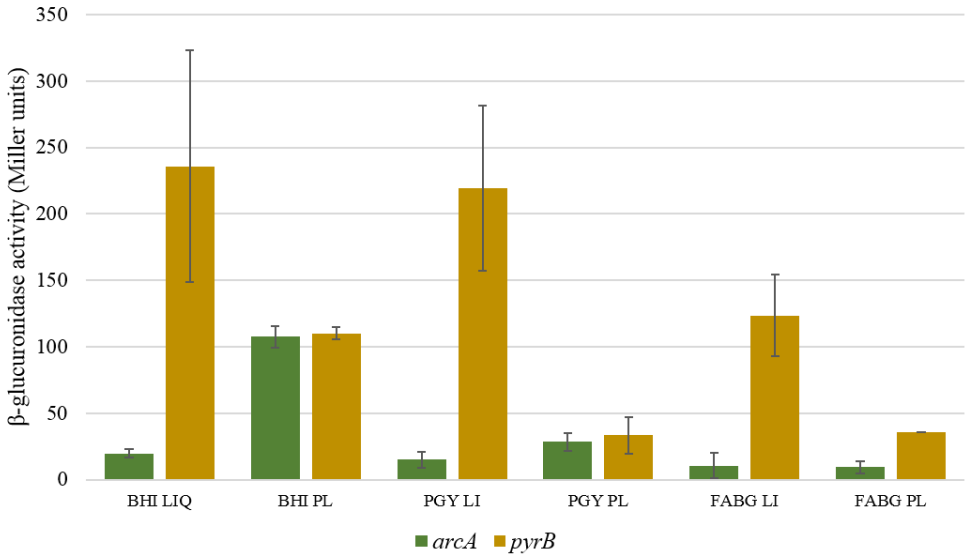


FIGURE 24. β-glucuronidase activity of highly transcribed genes in *C. perfringens*. Upstream promoters of *arcA* and *pyrB* were assessed for their transcriptional activity using reporter gene assays. Error bars report standard error.

PilM-6H can be purified easily from *C. perfringens*. Growth on ten BHI plates allowed us to purify usable amounts of PilM-6H for future experiments. Purification by nickel-affinity chromatography yielded fractions containing suspected PilM-6H with only a few other contaminating proteins (Figure 25, lanes 4 and 5). Further purification using size-exclusion chromatography yielded three fractions with PilM-6H being the dominant protein present, about 90% pure. Because we currently do not have a PilM-specific antibody for detection, we speculate our protein is PilM-6H due to visibility with Invision His-stained SDS-PAGE gels and correct molecular weight including the histidine tag. Micro BSA assay revealed that the most pure fraction of PilM-6H contained 3.8 mg of protein.

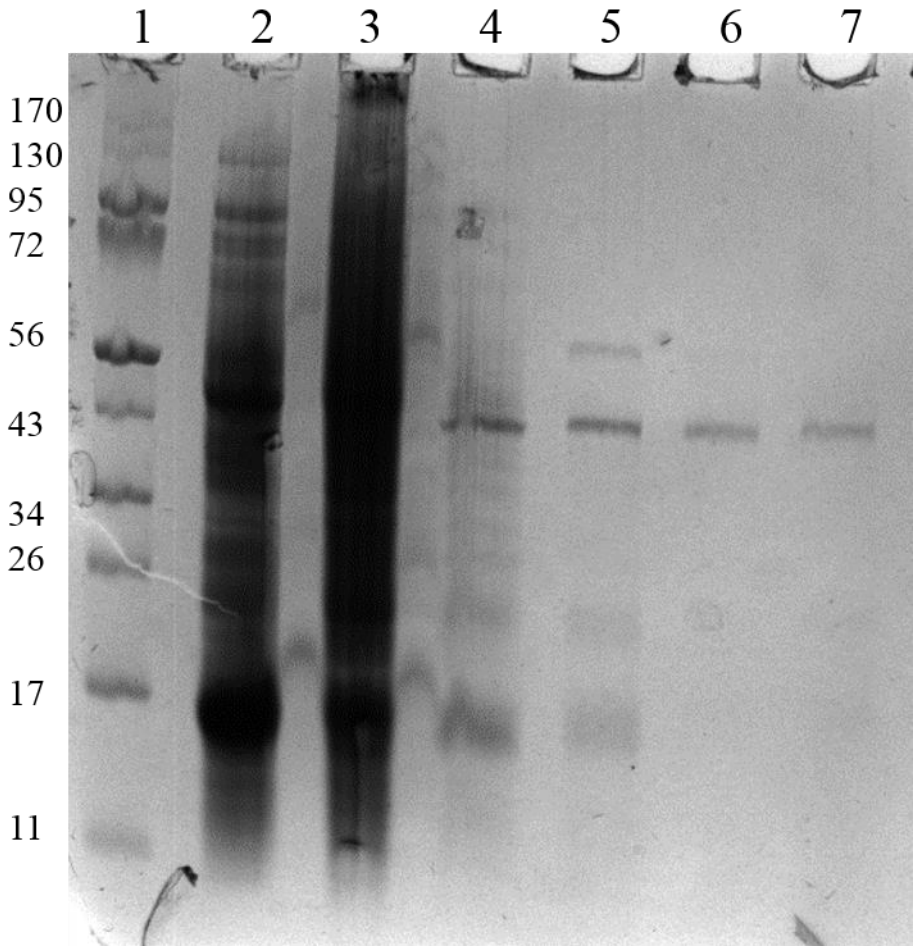


Figure 25: SDS-PAGE of PilM-6H protein purification.

Lane 1: molecular weight protein standards.

Lane 2: P*arcA*:PilM-6H uninduced (BHI LIQ)

Lane 3: P*arcA*:PilM-6His induced, soluble fraction (BHI PL)

Lanes 4 & 5: Nickel affinity chromatography fractions eluted with imidazole gradient

Lanes 6&7: Gel filtration chromatography fractions (purified PilM-6His)

PilM-6H size: 44.5 kDa

DISCUSSION

We have developed a method that can be easily applied to use with purifying other proteins from *C. perfringens*. From these experiments, we now know that *C. perfringens* cells grown in BHI plate and liquid have few proteins that bind to a nickel column in the conditions used, making our purification method highly efficient. This method allows for overexpression to be induced without the addition of metabolites or inducers that could affect cell growth or overall yield of cells or protein. The creation of vectors pARCA and pPYRB allow for easy insertion of other histidine-tagged protein constructs and future purification from the native organism, *C. perfringens*, which has been difficult in previous studies up to this point.

Future work will characterize PilM complex formation, as well as binding assays with purified PilB2. Purification of the native collagenase from *C. perfringens* will also be performed to assess the ability of our overexpression system to be utilized in liquid culture using the pPYRB vector. Collagenase assays will also be performed to study the effect of the histidine tag on enzyme function.

This method could be indispensable in the industry setting for purifying large toxins from *C. perfringens* for medical research without the addition of inducing agents. It will be an integral part of our lab's future work, as it makes protein purification from *C. perfringens* much simpler.

Table 16 Strains, plasmids, and primers used in protein purification study.

Strain or plasmid (parent)	**Relevant genotype**	**Selection/ Resistance**	**Source**
Escherichia coli			
DH10B	Electrocompetent strain	None	Gibco/BRL
Clostridium perfringens			
HN13	Wild type; *galKT*-	None	(58)
AH2	Mutant in *bglR*; parent for β-glucuronidase assays	Erm	(60)
Plasmids			
pTOPO	Suicide vector used for *uidA* fusion strains	Kan	Invitrogen
pKRAH1	Lactose inducible expression vector	Cm	(59)
pSM240	Contains *cpe-gusA*	Cm	(66)
pSRM49	ParcA in pSM240	Cm Erm	This study
pSRM47	PpyrB in pSM240	Cm Erm	This study
pARCA	Protein overexpression vector induced on BHI plates	Cm	This study
pPYRB	Protein overexpression vector induced in BHI liquid	Cm	This study
pSRM58	PilM-6H in pARCA	Cm	This study
pSRM68	ColA-6H in pPYRB	Cm	This study
Primers			
OSRM80	5’ gaaGTCGACatatgagacagttaagtgaaaacttacctgtc 3’		
OSRM81	5’ GTTCTGCAGCTTCCTATTTCAGAAGTAACATTTAATGCTCTGTC 3’		
OSRM82	5’ gtgGTCGACctaactcagtctacgggactg 3’		
OSRM82	5’ CCTCTGCAGTGAAATCTGATGGATCTAATAAGTGTTTG 3’		
OSRM104	5’ggacctacaaagttataaaagttaaaaggaggaagggtatattTTGGCAATAGAAAATAAAG AGAAGAAGAGC 3’		
OSRM105	5’GTCGACgaaaaagttaattattaatgatgatgatgatgAAGTCTTATTAAGGCTCCAA TAGCG 3’		
OSRM106	5’ CCGCGGGATAGAGGAGGCATTTAAACATGAAGAAAAACTTAAAAA GGG 3’		
OSRM107	5’ GAGCTCgatacagcctctttattaatgatgatgatgatgTTTGTTTTGTAAATTCAC 3’		

CHAPTER IV

Final Discussion

What once started as a research project to quantify Type IV pili (T4P) has blossomed into an in-depth analysis of the regulation surrounding these genes in *Clostridium perfringens*. T4P are an important part of pathogenesis in many bacteria, and *C. perfringens* is a ubiquitous Gram-positive pathogen with the genes to utilize T4P. Detailed in Chapter I are a few of the diseases that *C. perfringens* causes and the possession of T4P may play an integral role in the progression of these diseases. For this reason, they are of great interest to study as they help clarify the picture of *C. perfringens* pathogenesis.

Because little is known about the T4P system in Gram-positive organisms, it was difficult to quantify a surface-associated appendage that had went undiscovered in Gr+ bacteria until about a decade ago. We had little knowledge about what conditions induced T4P expression in the lab setting, and this proved to be a great obstacle in this research project. To rectify this, we first narrowed our investigation of T4P to the transcriptional level. We utilized RNA-sequencing in multiple conditions to observe transcript levels of T4P genes (Chapter 2). Our RNA-seq experiment gave us a plethora of data explaining the transcriptional regulation of T4P genes, including unknown sites of promoter sequences upstream of *pilA2*, *pilA3* and *pilB2*, a termination site in the coding region of *pilC2,* operon structures, and possible regulators, such as an uncharacterized small RNA *sr79* found in close proximity to the major T4P operon. We found that T4P are induced when *C. perfringens* are grown on plates, in media that contained glucose. This information is paramount to the success of future experiments that rely on quantification of T4P, as we now have a better understanding of when and how T4P genes are transcribed and expressed.

We also utilized reporter gene assays to expand our knowledge of T4P expression by creating translational fusions of T4P promoters. By creating translational fusions, we gathered a

more detailed picture of the regulation of T4P genes beyond the transcript level. Reporter gene assays also allowed us to observe activity of T4P promoters in a more diverse set of conditions than allowed by the expensive RNA-seq experiment. For example, the *pilT* promoter shows decreased activity in response to rising temperatures, and *pilA2* promoter activity indicates that *pilA2* regulation occurs post-translationally, as disproportionate amounts of enzyme activity were observed from conditions with large amounts of transcripts.

Using a whole-cell pilin ELISA method developed by our lab, we were able to observe the effect that diguanylate cyclase overexpression has on PilA2 polymerization on the surface of *C. perfringens* wild type cells and T4P gene in-frame mutants (Chapter 3.1). This method was unable to find a link between *sr79* overexpression and PilA2 polymerization, however.

The vast amount of RNA-seq data we obtained created multiple new avenues for research in *C. perfringens*. The observation of highly expressed genes in select conditions allowed us to create a new protein overexpression method in *C. perfringens* that has ultimately led to purification of the pilin protein PilM from the native organism without the addition of inducing agents. Our new method is important for future protein studies in our lab due to the usable amount of protein obtained from purification and its simplicity and efficiency.

Our RNA-seq dataset also led us to a little-characterized ECF sigma factor σ^V and its anti-sigma factor, which were both highly upregulated in conditions where T4P genes are also upregulated. While no direct link between T4P and σ^V exists at this time, creating in-frame deletions of *sigV* and *cpe0560* allowed us to further characterize a largely unstudied ECF sigma factor.

This thesis is broad in the sense that it covers a variety of topics and experimental procedures. However, the main goal throughout this work has always been characterization of

the Type IV pilus system. The work detailed herein lays the foundation for future experiments in T4P quantification, T4P protein purification from *C. perfringens*, and further analysis of T4P regulation. The ultimate goal remains to elucidate the link between T4P and *C. perfringens* pathogenesis, in hopes of inhibiting adherence of these cells during infection and mitigating the effects of *C. perfringens* diseases.

BIBLIOGRAPHY

1. **Dürre P.** 2005. Handbook on Clostridia. Chapter 1, pg 5. CRC Press.
2. **Center for Food Safety and Applied Nutrition (CFSAN) of the Food and Drug Administration (FDA) USDoHaHS.** 2012. Bad Bug Book: Foodborne Pathogenic Microorganisms and Natural Toxins Handbook.
3. **Hatheway CL.** 1990. Toxigenic clostridia. Clin Microbiol Rev **3:**66-98.
4. **Labbe RGH, T. H.** 1995. Generation times and modeling of enterotoxin-positive and enterotoxin-negative strains of *Clostridium perfringens* in laboratory media and ground beef. Journal of Food Protection **12:**1303-1306.
5. **Ryan KJR, C. George.** 2004. Sherris Medical Microbiology : an Introduction to Infectious Diseases, 4th ed. McGraw-Hill, New York.
6. **Rood JI, Cole ST.** 1991. Molecular genetics and pathogenesis of *Clostridium perfringens*. Microbiol Rev **55:**621-648.
7. **Shimizu T, Ohtani K, Hirakawa H, Ohshima K, Yamashita A, Shiba T, Ogasawara N, Hattori M, Kuhara S, Hayashi H.** 2002. Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. Proc Natl Acad Sci U S A **99:**996-1001.
8. **Sebald M, Costilow RN.** 1975. Minimal growth requirements for *Clostridium perfringens* and isolation of auxotrophic mutants. Appl Microbiol **29:**1-6.
9. **Petit L, Gibert M, Popoff MR.** 1999. *Clostridium perfringens*: toxinotype and genotype. Trends Microbiol **7:**104-110.
10. **Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M.** 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science **320:**1344-1349.
11. **Don R Revis J, MD; .** July 16 2014. Clostridial Gas Gangrene. http://emedicine.medscape.com/article/214992-overview#a6. Accessed October 10.
12. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis **17:**7-15.
13. **Rood JI.** 1998. Virulence genes of *Clostridium perfringens*. Annu Rev Microbiol **52:**333-360.
14. **Freedman JC, Shrestha A, McClane BA.** 2016. *Clostridium perfringens* enterotoxin: action, genetics, and translational applications. Toxins (Basel) **8**.
15. **Murrell TG, Walker PD.** 1991. The pigbel story of Papua New Guinea. Trans R Soc Trop Med Hyg **85:**119-122.

16. **Nagahama M, Ochi S, Oda M, Miyamoto K, Takehara M, Kobayashi K.** 2015. Recent insights into *Clostridium perfringens* beta-toxin. Toxins (Basel) **7:**396-406.
17. **Lawrence G, Walker P.** 1976. Pathogenesis of enteritis necroticans in Papua New Guinea. The Lancet **307:**125-126.
18. **Lawrence GW, Lehmann D, Anian G, Coakley CA, Saleu G, Barker MJ, Davis MW.** 1990. Impact of active immunisation against enteritis necroticans in Papua New Guinea. Lancet **336:**1165-1167.
19. **Aas FE, Wolfgang M, Frye S, Dunham S, Lovold C, Koomey M.** 2002. Competence for natural transformation in *Neisseria gonorrhoeae*: components of DNA binding and uptake linked to type IV pilus expression. Mol Microbiol **46:**749-760.
20. **Mattick JS.** 2002. Type IV pili and twitching motility. Annu Rev Microbiol **56:**289-314.
21. **O'Toole GA, Kolter R.** 1998. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. Mol Microbiol **30:**295-304.
22. **Varga JJ, Nguyen V, O'Brien DK, Rodgers K, Walker RA, Melville SB.** 2006. Type IV pili-dependent gliding motility in the Gram-positive pathogen *Clostridium perfringens* and other Clostridia. Mol Microbiol **62:**680-694.
23. **Piepenbrink KH, Sundberg EJ.** 2016. Motility and adhesion through type IV pili in Gram-positive bacteria. Biochem Soc Trans **44:**1659-1666.
24. **Rodgers K, Arvidson CG, Melville S.** 2011. Expression of a *Clostridium perfringens* type IV pilin by *Neisseria gonorrhoeae* mediates adherence to muscle cells. Infect Immun **79:**3096-3105.
25. **Melville S, Craig L.** 2013. Type IV pili in Gram-positive bacteria. Microbiol Mol Biol Rev **77:**323-341.
26. **Strom MS, Lory S.** 1992. Kinetics and sequence specificity of processing of prepilin by PilD, the type IV leader peptidase of *Pseudomonas aeruginosa*. J Bacteriol **174:**7345-7351.
27. **Bischof LF, Friedrich C, Harms A, Sogaard-Andersen L, van der Does C.** 2016. The Type IV pilus assembly ATPase PilB of *Myxococcus xanthus* interacts with the inner membrane platform protein PilC and the nucleotide-binding protein PilM. J Biol Chem **291:**6946-6957.
28. **Balleza E, Lopez-Bojorquez LN, Martinez-Antonio A, Resendis-Antonio O, Lozada-Chavez I, Balderas-Martinez YI, Encarnacion S, Collado-Vides J.** 2009. Regulation by transcription factors in bacteria: beyond description. FEMS Microbiol Rev **33:**133-151.
29. **Lawrence JG, Roth JR.** 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics **143:**1843-1860.
30. **Price MN, Huang KH, Arkin AP, Alm EJ.** 2005. Operon formation is driven by co-regulation and not by horizontal gene transfer. Genome Res **15:**809-819.
31. **Rocha EP.** 2008. The organization of the bacterial genome. Annu Rev Genet **42:**211-233.
32. **Dandekar T, Snel B, Huynen M, Bork P.** 1998. Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci **23:**324-328.
33. **Lee DJ, Minchin SD, Busby SJ.** 2012. Activating transcription in bacteria. Annu Rev Microbiol **66:**125-152.

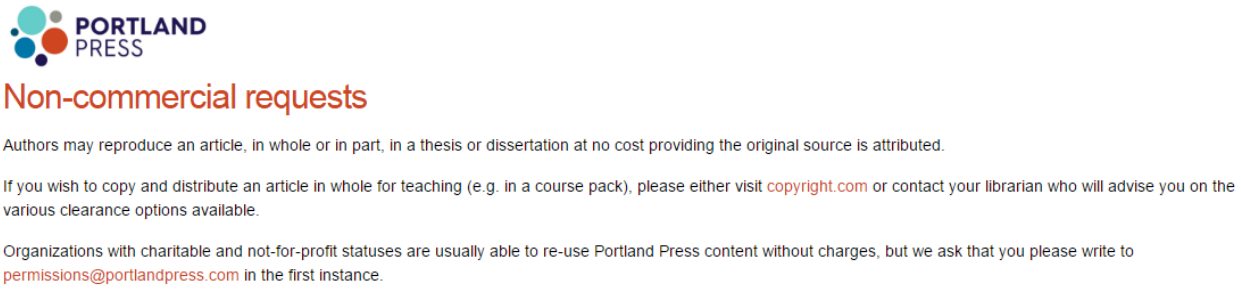
34. **Grigorova IL, Phleger NJ, Mutalik VK, Gross CA.** 2006. Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. Proc Natl Acad Sci U S A **103:**5332-5337.
35. **Souza BM, Castro TL, Carvalho RD, Seyffert N, Silva A, Miyoshi A, Azevedo V.** 2014. sigma(ECF) factors of gram-positive bacteria: a focus on *Bacillus subtilis* and the CMNR group. Virulence **5:**587-600.
36. **Hastie JL, Williams KB, Ellermeier CD.** 2013. The activity of sigmaV, an extracytoplasmic function sigma factor of *Bacillus subtilis*, is controlled by regulated proteolysis of the anti-sigma factor RsiV. J Bacteriol **195:**3135-3144.
37. **Wang Z, Gerstein M, Snyder M.** 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet **10:**57-63.
38. **Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B.** 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods **5:**621-628.
39. **Chhangawala S, Rudy G, Mason CE, Rosenfeld JA.** 2015. The impact of read length on quantification of differentially expressed genes and splice junction detection. Genome Biol **16:**131.
40. **Higuchi R, Dollinger G, Walsh PS, Griffith R.** 1992. Simultaneous amplification and detection of specific DNA sequences. Biotechnology (N Y) **10:**413-417.
41. **Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonak J, Lind K, Sindelka R, Sjoback R, Sjogreen B, Strombom L, Stahlberg A, Zoric N.** 2006. The real-time polymerase chain reaction. Mol Aspects Med **27:**95-125.
42. **Core LJ, Waterfall JJ, Lis JT.** 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science **322:**1845-1848.
43. **Feng L, Liu H, Liu Y, Lu Z, Guo G, Guo S, Zheng H, Gao Y, Cheng S, Wang J, Zhang K, Zhang Y.** 2010. Power of deep sequencing and agilent microarray for gene expression profiling study. Mol Biotechnol **45:**101-110.
44. **Pabinger S, Dander A, Fischer M, Snajder R, Sperk M, Efremova M, Krabichler B, Speicher MR, Zschocke J, Trajanoski Z.** 2014. A survey of tools for variant analysis of next-generation genome sequencing data. Brief Bioinform **15:**256-278.
45. **Fang Z, Cui X.** 2011. Design and validation issues in RNA-seq experiments. Brief Bioinform **12:**280-287.
46. **Rosenthal N.** 1987. Identification of regulatory elements of cloned genes with functional assays. Methods Enzymol **152:**704-720.
47. **Schenborn E, Groskreutz D.** 1999. Reporter gene vectors and assays. Mol Biotechnol **13:**29-44.
48. **Henrichsen J.** 1972. Bacterial surface translocation: a survey and a classification. Bacteriol Rev **36:**478-503.
49. **Verstraeten N, Braeken K, Debkumari B, Fauvart M, Fransaer J, Vermant J, Michiels J.** 2008. Living on a surface: swarming and biofilm formation. Trends Microbiol **16:**496-506.
50. **Mignot T.** 2007. The elusive engine in *Myxococcus xanthus* gliding motility. Cell Mol Life Sci **64:**2733-2745.
51. **Lee YY, Patellis J, Belas R.** 2013. Activity of *Proteus mirabilis* FliL is viscosity dependent and requires extragenic DNA. J Bacteriol **195:**823-832.

52. **Merritt JH, Brothers KM, Kuchma SL, O'Toole GA.** 2007. SadC reciprocally influences biofilm formation and swarming motility via modulation of exopolysaccharide production and flagellar function. J Bacteriol **189:**8154-8164.
53. **Cairns LS, Marlow VL, Bissett E, Ostrowski A, Stanley-Wall NR.** 2013. A mechanical signal transmitted by the flagellum controls signalling in *Bacillus subtilis*. Mol Microbiol **90:**6-21.
54. **Morgenstein RM, Clemmer KM, Rather PN.** 2010. Loss of the *waaL* O-antigen ligase prevents surface activation of the flagellar gene cascade in *Proteus mirabilis*. J Bacteriol **192:**3213-3221.
55. **Hickman JW, Tifrea DF, Harwood CS.** 2005. A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. Proc Natl Acad Sci U S A **102:**14422-14427.
56. **Valentini M, Filloux A.** 2016. Biofilms and cyclic di-GMP (c-di-GMP) signaling: lessons from *Pseudomonas aeruginosa* and other bacteria. J Biol Chem **291:**12547-12555.
57. **Varga JJ, Therit B, Melville SB.** 2008. Type IV pili and the CcpA protein are needed for maximal biofilm formation by the gram-positive anaerobic pathogen *Clostridium perfringens*. Infect Immun **76:**4944-4951.
58. **Nariya H, Miyata S, Suzuki M, Tamai E, Okabe A.** 2011. Development and application of a method for counterselectable in-frame deletion in *Clostridium perfringens*. Appl Environ Microbiol **77:**1375-1382.
59. **Hartman AH, Liu H, Melville SB.** 2011. Construction and characterization of a lactose-inducible promoter system for controlled gene expression in *Clostridium perfringens*. Appl Environ Microbiol **77:**471-478.
60. **Melville SB, Labbe R, Sonenshein AL.** 1994. Expression from the *Clostridium perfringens* cpe promoter in *C. perfringens* and *Bacillus subtilis*. Infect Immun **62:**5550-5558.
61. **Wagner GP, Kin K, Lynch VJ.** 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci **131:**281-285.
62. **Sheehan LM, Budnick JA, Blanchard C, Dunman PM, Caswell CC.** 2015. A LysR-family transcriptional regulator required for virulence in *Brucella abortus* is highly conserved among the alpha-proteobacteria. Mol Microbiol **98:**318-328.
63. **Schneider CA, Rasband WS, Eliceiri KW.** 2012. NIH Image to ImageJ: 25 years of image analysis. Nature methods **9:**671.
64. **Solovyev V, Salamov A.** 2011. Automatic annotation of microbial genomes and metagenomic sequences. Metagenomics and its applications in agriculture, biomedicine and environmental studies**:**61-78.
65. **Mao F, Dam P, Chou J, Olman V, Xu Y.** 2009. DOOR: a database for prokaryotic operons. Nucleic acids research **37:**D459-D463.
66. **Harry KH, Zhou R, Kroos L, Melville SB.** 2009. Sporulation and enterotoxin (CPE) synthesis are controlled by the sporulation-specific sigma factors SigE and SigK in *Clostridium perfringens*. J Bacteriol **191:**2728-2742.
67. **Taschner PE, Huls PG, Pas E, Woldringh CL.** 1988. Division behavior and shape changes in isogenic *ftsZ, ftsQ, ftsA, pbpB,* and *ftsE* cell division mutants of *Escherichia coli* during temperature shift experiments. J Bacteriol **170:**1533-1540.

68. **Chen Y, Indurthi DC, Jones SW, Papoutsakis ET.** 2011. Small RNAs in the genus Clostridium. MBio **2:**e00340-00310.
69. **Low LY, Harrison PF, Lin YH, Boyce JD, Rood JI, Cheung JK.** 2016. RNA-seq analysis of *virR* and *revR* mutants of *Clostridium perfringens*. BMC Genomics **17:**391.
70. **Roelofs KG, Jones CJ, Helman SR, Shang X, Orr MW, Goodson JR, Galperin MY, Yildiz FH, Lee VT.** 2015. Systematic identification of cyclic-di-GMP binding proteins in *Vibrio cholerae* reveals a novel class of cyclic-di-GMP-binding ATPases associated with Type II secretion systems. PLoS Pathog **11:**e1005232.
71. **Hengge R.** 2009. Principles of c-di-GMP signalling in bacteria. Nat Rev Microbiol **7:**263-273.
72. **Jones CJ, Utada A, Davis KR, Thongsomboon W, Zamorano Sanchez D, Banakar V, Cegelski L, Wong GC, Yildiz FH.** 2015. C-di-GMP regulates motile to sessile transition by modulating MshA pili biogenesis and near-surface motility behavior in *Vibrio cholerae*. PLoS Pathog **11:**e1005068.
73. **Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR.** 2010. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. Science **329:**845-848.
74. **Soutourina OA, Monot M, Boudry P, Saujet L, Pichon C, Sismeiro O, Semenova E, Severinov K, Le Bouguenec C, Coppee JY, Dupuy B, Martin-Verstraete I.** 2013. Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. PLoS Genet **9:**e1003493.
75. **Bordeleau E, Purcell EB, Lafontaine DA, Fortier LC, Tamayo R, Burrus V.** 2015. Cyclic di-GMP riboswitch-regulated type IV pili contribute to aggregation of *Clostridium difficile*. J Bacteriol **197:**819-832.
76. **Purcell EB, McKee RW, Bordeleau E, Burrus V, Tamayo R.** 2015. Regulation of Type IV pili contributes to surface behaviors of historical and epidemic strains of *Clostridium difficile*. J Bacteriol **198:**565-577.
77. **Nariya H, Miyata S, Kuwahara T, Okabe A.** 2011. Development and characterization of a xylose-inducible gene expression system for *Clostridium perfringens*. Appl Environ Microbiol **77:**8439-8441.
78. **Ho TD, Hastie JL, Intile PJ, Ellermeier CD.** 2011. The *Bacillus subtilis* extracytoplasmic function sigma factor sigma(V) is induced by lysozyme and provides resistance to lysozyme. J Bacteriol **193:**6215-6222.
79. **Varahan S, Iyer VS, Moore WT, Hancock LE.** 2013. Eep confers lysozyme resistance to *Enterococcus faecalis* via the activation of the extracytoplasmic function sigma factor SigV. J Bacteriol **195:**3125-3134.
80. **Woods EC, Nawrocki KL, Suarez JM, McBride SM.** 2016. The *Clostridium difficile* Dlt pathway Is controlled by the extracytoplasmic function sigma factor SigmaV in response to lysozyme. Infect Immun **84:**1902-1916.
81. **Lisle JT, Smith JJ, Edwards DD, McFeters GA.** 2004. Occurrence of microbial indicators and *Clostridium perfringens* in wastewater, water column samples, sediments, drinking water, and Weddell seal feces collected at McMurdo Station, Antarctica. Appl Environ Microbiol **70:**7269-7276.
82. **Schleheck D, Barraud N, Klebensberger J, Webb JS, McDougald D, Rice SA, Kjelleberg S.** 2009. *Pseudomonas aeruginosa* PAO1 preferentially grows as aggregates in liquid batch cultures and disperses upon starvation. PLoS One **4:**e5513.

83. **Liu H, Bouillaut L, Sonenshein AL, Melville SB.** 2013. Use of a mariner-based transposon mutagenesis system to isolate *Clostridium perfringens* mutants deficient in gliding motility. J Bacteriol **195:**629-636.
84. **Wagner EG, Simons RW.** 1994. Antisense RNA control in bacteria, phages, and plasmids. Annu Rev Microbiol **48:**713-742.
85. **Wagner EG, Brantl S.** 1998. Kissing and RNA stability in antisense control of plasmid replication. Trends Biochem Sci **23:**451-454.
86. **Bossi L, Figueroa-Bossi N.** 2016. Competing endogenous RNAs: a target-centric view of small RNA regulation in bacteria. Nat Rev Microbiol **14:**775-784.
87. **Gottesman S.** 2004. The small RNA regulators of *Escherichia coli*: roles and mechanisms. Annu Rev Microbiol **58:**303-328.
88. **Murina VN, Nikulin AD.** 2015. Bacterial small regulatory RNAs and Hfq protein. Biochemistry (Mosc) **80:**1647-1654.
89. **Hendrick WA, Orr MW, Murray SR, Lee VT, Melville SB.** 2017. Cyclic-di-GMP binding by an assembly ATPase (PilB2) and control of Type IV pilin polymerization in the Gram-positive pathogen *Clostridium perfringens*. J Bacteriol doi:10.1128/jb.00034-17.
90. **Harkness JM, Li J, McClane BA.** 2012. Identification of a lambda toxin-negative *Clostridium perfringens* strain that processes and activates epsilon prototoxin intracellularly. Anaerobe **18:**546-552.

COPYRIGHT PERMISSIONS



PORTLAND PRESS

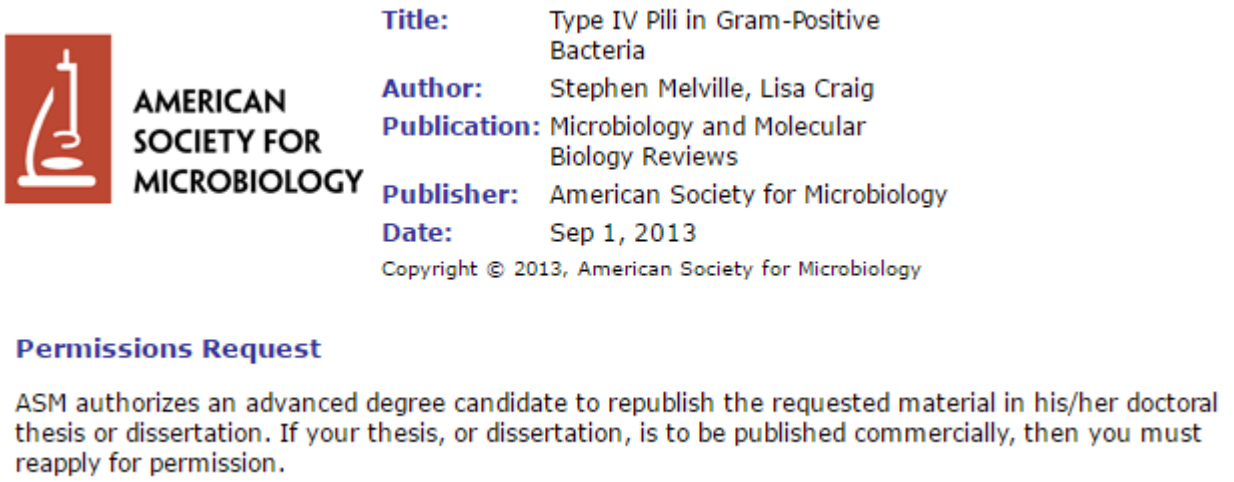
Non-commercial requests

Authors may reproduce an article, in whole or in part, in a thesis or dissertation at no cost providing the original source is attributed.

If you wish to copy and distribute an article in whole for teaching (e.g. in a course pack), please either visit copyright.com or contact your librarian who will advise you on the various clearance options available.

Organizations with charitable and not-for-profit statuses are usually able to re-use Portland Press content without charges, but we ask that you please write to permissions@portlandpress.com in the first instance.

Figure A1. Reproducibility permission for Figure 1.



AMERICAN SOCIETY FOR MICROBIOLOGY

Title:	Type IV Pili in Gram-Positive Bacteria
Author:	Stephen Melville, Lisa Craig
Publication:	Microbiology and Molecular Biology Reviews
Publisher:	American Society for Microbiology
Date:	Sep 1, 2013

Copyright © 2013, American Society for Microbiology

Permissions Request

ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation. If your thesis, or dissertation, is to be published commercially, then you must reapply for permission.

Figure A2. Reproducibility permission for Figures 2 and 3.

LAST UPDATED: January 2017

ASM Journals Statement of Authors' Rights

Authors may post their articles to their institutional repositories

ASM grants authors the right to post their accepted manuscripts in publicly accessible electronic repositories maintained by funding agencies, as well as appropriate institutional or subject-based open repositories established by a government or non-commercial entity. Since ASM makes the final, typeset articles from its primary-research journals available free of charge on the ASM Journals and PMC websites 6 months after final publication, ASM recommends that when submitting the accepted manuscript to PMC or institutional repositories, the author specify that the posting release date for the manuscript be no earlier than 6 months after the final publication of the typeset article by ASM. Please note that the information outlined above, in addition to the **new author fee structure**, allows authors to comply with the **Policy for open access in the post-2014 Research Excellence Framework**.

Authors may post their articles in full on personal or employer websites

ASM grants the author the right to post his/her article (after publication by ASM) on the author's personal or university-hosted website, but not on any corporate, government, or similar website, without ASM's prior permission, provided that proper credit is given to the original ASM publication.

Authors may make copies of their articles in full

Corresponding authors are entitled to 10 free downloads of their papers. Additionally, all authors may make up to 99 copies of his/her own work for personal or professional use (including teaching packs that are distributed free of charge within your own institution). For orders of 100 or more copies, you should seek ASM's permission or purchase access through Highwire's Pay-Per-View option, available on the ASM online journal sites.

Authors may republish/adapt portions of their articles

ASM also grants the authors the right to republish discrete portions of his/her article in any other publication (including print, CD-ROM, and other electronic formats) of which he or she is author or editor, provided that proper credit is given to the original ASM publication. ASM authors also retain the right to reuse the full article in his/her dissertation or thesis. "Proper credit" means either the copyright lines shown on the top of the first page of the PDF version, or "Copyright © American Society for Microbiology, [insert journal name, volume number, year, page numbers and DOI]" of the HTML version. For technical questions about using Rightslink, please contact Customer Support via phone at (877) 622-5543 (toll free) or (978) 777-9929, or e-mail Rightslink customer care at **customercare@copyright.com**.

Please note that the ASM is in full **compliance with NIH Policy**.

Figure A3. Reproducibility permission for content in Chapter 3.1, including Figures 17 and 18.