CHARACTERIZATION OF BROADLY CONSERVED AVCID TOXIN-ANTITOXIN SYSTEM AND ITS MECHANISM TO INHIBIT PHAGE BY DISRUPTING NUCLEOTIDE METABOLISM

By

Brian Yifei Hsueh

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Microbiology and Molecular Genetics – Doctor of Philosophy

2022

ABSTRACT

CHARACTERIZATION OF BROADLY CONSERVED AVCID TOXIN-ANTITOXIN SYSTEM AND ITS MECHANISM TO INHIBIT PHAGE BY DISRUPTING NUCLEOTIDE METABOLISM

By

Brian Yifei Hsueh

The prevalence of antiphage defense systems, which have recently been shown to be located on mobile genetic elements in bacteria, have sparked interest to understand the coevolutionary arms race of bacteria and bacteriophage (phage). Bacteria and phages have coexisted for billions of years, and phages are widely distributed in different environmental niches populated by their bacterial hosts, including the human intestine and marine environment. The evolutionary pressure imposed by phages have led bacteria to evolve diverse strategic systems to protect themselves from phage predation, including CRISPR-Cas, restriction-modification, and abortive infection. Recent studies have begun to reveal that toxin-antitoxin (TA) system are associated with antiphage defense systems. *Vibrio cholerae* El Tor, the causative agent of current cholera pandemics, has acquired two unique genomic islands of unknown origins, known as Vibrio Seventh Pandemic Islands 1 & 2 (VSP-1 & 2). It is hypothesized that the acquisition of VSP islands increase environmental fitness of El Tor. While both islands encode approximately 36 open reading frames, yet many remain largely uncharacterized.

In this work, I characterize a novel TA antiphage system encoded on VSP-1 of *V. cholerae*, here named AvcID. Chapter 2 describes the biological function of AvcD toxin by which it possesses deoxycytidylate deaminase (DCD) activity and produces dUMP

as the final product. Further experiments identify the AvcI antitoxin as a small RNA and determine that it post-translationally inhibits the activity of AvcD. Moreover, AvcD consists of two domains—a N-terminal P-loop NTPase and a C-terminal DCD—and mutations in conserved features of each domain abrogate its activity. AvcD is widely conserved across kingdoms, and virtually all bacteria that encode AvcD also have AvcI homologs. Notably, chromosomal AvcID can solely be activated by transcriptional shutoff in *V. cholerae*, demonstrating that AvcID is a type III TA system. Unlike canonical type III TA systems, in which the toxin is an endoribonuclease, the AvcD toxin is a deaminase. Importantly, the AvcID system provides antiphage defense in *Escherichia coli* that lacks this system by corrupting nucleotides for phages to utilize to reduce coliphage replication efficiency.

In Chapter 3, I explore the activation mechanism of the AvcID system as well as the consequences to phages after encountering AvcID. During infection, virtually all lytic phages induce transcription shutoff of the host by hijacking host transcription machinery to make virion progeny. I uncover that phage-induced transcriptional shutoff leads to turnover of labile AvcI antitoxin and concomitantly activates the deaminase activity of AvcD, leading to a disruption of nucleotide levels. This disruption of nucleotide levels is shown in both susceptible phages (*ex.* T5) and resistant phages (*ex.* T7). Through an unknown mechanism, AvcID also increases the abundance of defective phages that are susceptible to AvcID. In summary, this work has made contributions in the field of TA systems and its association with the antiphage defense paradigm by uncovering the biological function and mechanism in response to phage infection.

This dissertation is dedicated to my parents, my siblings, my grandpa, and my girlfriend.
Thank you for all your love and support. I love you all.

ACKNOWLEDGEMENTS

Completing this dissertation has been by far one the most challenging and exciting journeys I have embarked on in my life, and I am utterly grateful for the ones who has helped me on this journey. First, I want to thank my scientific mentor, Christopher Waters Ph.D., for your continuous support, guidance, and encouragement over the past five years. I sometimes use the metaphor that graduate programs like MMG are like a forest: They are made up of a million different plants, shrubs, and trees. We, the students, are these plants, graduating and moving on to hopefully greater things in life. But Chris is like an *evergreen*, rain or shine that keeps growing because it has that kind of beauty and purposes that will never wither and wilt. And that is what I have always seen in Chris. As an image, you are indelible as a mentor.

I would also like to thank my guidance committee, Victor DiRita Ph.D., Lee Kroos Ph.D., Beronda Montgomery Ph.D., and George Sundin Ph.D. I want to thank you all for your insightful ideas and critiques that allow me to be part of something important in science and as well as helping me to improve as an individual.

I also want to thank Roseann Bills and the people working behind the scenes in the MMG department for all the questions I had during my graduate career. You all are the unsung heroes to make this department one of the best.

I have been fortunate to work with several fantastic collaborators on my Ph.D. journey. I owe a huge thank you to Roshni Kharadi Ph.D. from the Department of Plant, Soil, and Microbial Sciences, MSU, Evan Waldron M.D., Ph.D., Abhiruchi Kant Ph.D., and Mathew Neiditch Ph.D. from Rutgers University, Clinton Elg and Eva Top Ph.D. from the University of Idaho, Janani Ravi Ph.D. here at MSU, Sundharraman

Subramanian Ph.D., John Dover Ph.D., and Kristin Parent Ph.D. from the Department of Biochemistry. I also like to thank Lijun Chen Ph.D. and Tony Schilmiller Ph.D. from the Mass Spectrometry Core at MSU for taking their valuable time and energy to answer my trivial questions and troubleshoot numerous technical troubles I encountered that you would probably never get back. Thank you for your perseverance.

To all the Waters lab members, past and present. Thank you, Geoff Severin Ph.D., and Nico Fernandez Ph.D., for everything you have done to train me to become a better scientist. You two are one of the most brilliant people I know. I may not mount to your levels of scientific achievement, but I appreciate the great time we spent at the lab, conferences, and social events. Thank you Soo Hun Yoon for the wonderful time we have at the lab and all the puppy dates. It has a been a pleasure working alongside you. Thank you Jacob Hieber for allowing me to be your mentor. You were a joy to mentor, and I admire your perseverance and radiant attitude even when some of the experiments failed. I wish you the best. Thank you Amber Bedore M.S., for being the best lab manager and always having an open ear and positive outlook on life as well as helping me with experiments in Chapter 3. I also very much enjoyed our discussions on daily events, which could be traced back to me. Thank you Kaylee Wilburn M.S., Alex Wessel, and Jasper Gomez for being great lab mates, and I am sure you will achieve great things in your graduate career and in the future. Thank you Micah Ferrell Ph.D. and Aathmaja Rangarajan Ph.D. for joining the Waters lab to expand our research scope and train us young scientists.

I would like to thank my parents Yenyi Hsueh DDS and Shu Fang Chuang from the bottom of my hearts for their sacrifices by immigrating to the United States and

working hard to provide our family opportunities to succeed. My siblings, Peter Hsueh Ph.D., Evan Hsueh M.S., and May Hsueh for always being there for me. My grandpa, Mouchi Hsueh, for supporting and believing me that I will do remarkable things in life since I was a kid. Lastly, I would like to thank my girlfriend, Gillian Gahn MSW, for supporting me throughout my graduate career. My life would be lost without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
APOBEC	Apolipoprotein B mRNA editing catalytic polypeptide like
BLAST	Basic local alignment search tool
BP	Base pair
CBASS	Cyclic oligonucleotide-based antiphage signaling systems
CDA	Cytosine deaminase
c-di-GMP	Cyclic dimeric guanosine monophosphate
cDNA	Complementary DNA
CFU	Colony forming units
cGAMP	Cyclic dimeric guanosine-adenosine monophosphate
dATP	Deoxyadenosine triphosphate
DCD	Deoxycytidylate deaminase
dCMP	Deoxycytidine monophosphate
dCTP	Deoxycytidine triphosphate
DGC	Diguanylate cyclase
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DRaCALA	Differential radial capillary action of ligand assay
dTMP	Deoxythymidine monophosphate
dTTP	Deoxythymidine triphosphate

dUMP	Deoxyuridine monophosphate
dUTP	Deoxyuridine triphosphate
EBP	Enhancer binding protein
ECF	Extracytoplasmic function
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
EOP	Efficiency of plaquing
EPS	Exopolymeric substances
ESI	Electrospray ionization
GDP	Guanosine diphosphate
GMP	Guanosine monophosphate
GTP	Guanosine triphosphate
HA	Hemagglutinin tag
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGT	Horizontal gene transfer
HIS	Six histidine tag
HRP	Horseradish peroxide
IPTG	Isopropyl-D-1-thiogalactopyranoside
LB	Luria-Bertani
MGE	Mobile genetic element
MOI	Multiplicity of infection
MRS	Maximum related subnetworks
NT	Nucleotide

NTP	Nucleotide triphosphate
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFU	Plaque forming units
PDE	Phosphodiesterase
pGpG	5'-phosphoguanylyl-(3',5')-guanosine
PLN	P-loop NTPase
qRT-PCR	Quantitative real-time polymerase chain reaction
QS	Quorum sensing
RBS	Ribosomal binding site
RC	Reverse complement
REC	Receiver domain
RFP	Red fluorescent protein
RM	Restriction modification
RNA	Ribonucleic acid
RPM	Revolution per minute
RT	Room temperature/Reverse transcription
SDS	Sodium dodecyl sulfate
SEM	Standard error of mean
sRNA	Small ribonucleic acid

SSC	Saline-sodium citrate
T3SS	Type 3 secretion system
TA	Toxin-antitoxin
TBE	Tris borate EDTA
TCEP	Tris (2-carboxyethyl) phosphine
TCP	Toxin-coregulated pili
TLC	Thin layer chromatography
TLD	Thymine-less death
tRNA	Transfer RNA
TS	Thymidylate synthase
UMP	Uridine monophosphate
UPLC	Ultra performance liquid chromatography
VSP Island	Vibrio Seventh Pandemic Island
WT	Wild type

CHAPTER 1 – Introduction of Vibrio Seventh Pandemic Islands in *Vibrio cholerae* El Tor

Horizontal gene transfer (HGT) is frequently mediated by mobile genetic elements (MGEs), which may result in the acquisition of genes that contribute to bacterial fitness, diversity, and evolution [1–3]. The genes encoded on MGEs not only encode virulence factors and antimicrobial resistance but also antiphage defense [3–6]. Some refer to these bacterial antiphage defense systems as the prokaryotic "immune system." Different MGEs from various bacteria encode distinct sets of antiphage defense systems. These defense systems frequently cluster in the vicinity of one another and are commonly known as "defense islands" [7–9]. *Vibrio cholerae*, the etiological agent of cholera, is a Gram-negative, rod-shaped pathogen that is enriched with prophages and MGEs on its chromosomes [10]. Out of more than 200 serogroups of *V. cholerae*, only the serogroups O1 and O139 have been the causative agents of current cholera epidemics, and specifically serogroup O1 *V. cholerae* is the major infectious agent [11]. Serogroups are further subclassified into two major biotypes: classical and El Tor. The first six pandemics of *V. cholerae* from the years 1817 to 1921 were most likely caused by the classical biotype, whereas the seventh and current pandemic that began in 1961 was caused by the El Tor biotype [12]. Virtually all modern-day cholera infections are caused by El Tor, and environmental sampling identifies only El Tor, suggesting that the classical biotypes are no longer prevalent [13]. One of the major genetic differences between the classical and El Tor biotypes is El Tor has acquired two unique genomic islands of unknown origins, named the Vibrio Seventh Pandemic Islands 1 and 2 (VSP-1 and 2) [14, 15]. This led to the hypothesis that the two VSP islands played a crucial role in El Tor's evolution to pandemicity and the displacement of the classical biotype in modern cholera infections [16].

VSP-1 and -2 are encoded on chromosome I of *V. cholerae*. VSP-1 is integrated into the site between *vc0174* and *vc0186* and VSP-2 is integrated into the site between *vc0489* and *vc0517* [14, 17]. Unlike VSP-1, VSP-2 is integrated at a tRNA-methionine locus and is flanked by direct repeats. However, both VSP islands encode either an integrase or transposase. The association with the tRNA gene, direct repeats and the presence of integrase and transposase are all features of MGEs [18–20]. Moreover, it was demonstrated that both VSP islands are capable of excising themselves, dependent on their cognate integrase, from the chromosome and forming circular intermediates, which is likely a first step in their horizontal transfer, whether by transformation, conjugation or transduction [4, 17, 21, 22].

VSP-1 and VSP-2 account for approximately 39 kilobases (kb) of DNA and encode nearly thirty-six putative open reading frames (ORFs), many of which remain to be characterized. It is hypothesized that the biological functions they encode may contribute to environmental persistence and/or the pathogenicity of El Tor infections [23, 24]. In 2012, the Mekalanos group was the first to link the connection between toxin-coregulated pilus (TCP) pathogenicity island and VSP-1 and later discovered that VSP-1 encodes a dinucleotide synthase named DncV that synthesizes the novel second messenger cyclic GMP-AMP (cGAMP) from ATP and GTP [24]. The expression of *dncV* is negatively regulated by a transcriptional regulator encoded upstream of *dncV* termed VspR [24]. VspR is not only a negative regulator of *dncV* but also the genes upstream and downstream of *dncV*, *vc0178* and *vc0180,* respectively [24]. Additionally, DncV contributes to colonization in an infant mouse model through an unknown mechanism [24]. In 2018, the Ng and Waters groups demonstrated that cGAMP can allosterically

activate a phospholipase encoded by the gene VC0178, now named CapV, to degrade the host membrane [25]. Like DncV, *capV* expression is also modulated by repressor VspR [24]. Another *V. cholerae* virulence regulator, ToxR, was also shown to directly negatively regulate *capV* expression [26]. In addition to transcriptional regulation, DncV has been co-crystalized with folate-like molecules that serve as inhibitors by modifying its active site conformation and preventing it from synthesizing cGAMP [27].

Though the VSP-2 island is much larger and encodes more ORFs than VSP-1, its link to host fitness is still poorly understood. Up until 2020, the function of only two ORFs on VSP-2 had been validated; *vc0516,* an integrase, and *vc0503,* a peptidoglycan endopeptidase [17, 28]. In 2020, the Dörr group uncovered a transcriptional network encoded by the *vc0513-vc0515* operon that is induced under zinc-deprivation [29]. The operon encodes a transcriptional activator, VerA, that increases expression of VSP-2 chemotaxis and motility-related genes, including the AerB chemotaxis receptor, which mediates oxygen-dependent congregation and energy taxis, under zinc-deprived condition, implicating the roles of VSP-2 in response to the environment. While both islands are unique to *V. cholerae* El Tor, VSP-2, however, has several variations in other El Tor strains. For instance, C6706, a Peruvian strain, lacks *vc0511-vc515,* which encodes the Zur (zinc uptake regulator)-regulated chemotaxis pathway, while a strain isolated from the recent Haitian outbreak lacks *vc0495-vc0512*, which encodes part of the chemotaxis pathway and several putative proteins, including a transcriptional regulator (*vc0497*), a ribonuclease H (*vc0498*), a type IV pilin *(vc0502*), and a DNA repair protein *(vc0510*) [29]. The variations in the VSP-2 island among the El

Tor strains may be a beneficial adaptation in the ever-changing environments where these strains reside.

One challenge that bacteria encounter under both host and environmental conditions is bacteriophage infection. The relationship between bacteriophages (henceforth phages) and bacteria has been studied extensively for decades mainly because bacteria constantly face challenges from phages, as they are outnumbered approximately 10:1 [30]. As the Red Queen Hypothesis posits, this has led to a coevolutionary arms race that resulted in tremendous diversity in both bacterial and phage defensive and offensive strategies [6, 8, 31, 32]. *V. cholerae* is endlessly being challenged by phages in the environment, leading to hypotheses stating that the displacement of the classical biotype and the emergence of El Tor biotype may have been driven by phages by selecting the phage-resistant strains [33, 34]. In 2019, the Sorek group discovered that the operon encompassing the genes *capV, dncV*, *vc0180* (*E1-E2*) and *vc0181* (*Jab*) is an antiphage defense system, now recognized as the cyclic-oligonucleotide-based antiphage signaling system (CBASS) [7, 35]. Upon phage infection, CBASS is activated and limits phage invasion of bacterial populations via abortive replication [7, 24, 25]. Though the activation mechanism of CBASS is still under investigation, CapV and DncV were shown to be indispensable for this system to provide phage defense. Despite their putative accessory role to CapV/DncV signaling, the contributions of VC0180 and VC0181 to CBASS function remains mysterious. Interestingly, when testing the *V. cholerae* CBASS system for phage infection in a naïve laboratory strain of *Escherichia coli* MG1655, it solely conferred phage resistance against two out of ten *E. coli* phages while the CBASS derived from *E. coli* TW1168

provided resistance to six [7]. This discrepancy is partly due to the lack of the regulatory factors from *V. cholerae* needed to properly function in *E. coli*. Similarly, the Kranzusch group found that DncV-like enzymes belong to a large family of cyclic-nucleotide/oligonucleotide cyclases and are capable of synthesizing diverse sets of cyclic-oligonucleotide molecules, including aforementioned cGAMP, as well as cyclic UMP-AMP, cyclic-di-UMP, and even cyclic trinucleotide AMP-AMP-GMP [35]. More importantly, the CBASS operons were also found to be frequently clustered in the vicinity of known antiphage defense systems, such as CRISPR-Cas or Restriction Modifications (RMs) [7].

To expand the current understanding of VSP islands, we collaborated with Eva Top's group at the University of Idaho where they developed a bioinformatics analysis pipeline that determined the co-occurrence of gene networks within the VSP islands across sequenced bacterial genomes. The results indicate that along with the previously described CBASS system, *dncV* not only co-occurs in bacterial genomes with *capV, vc0180,* and *vc0181,* but also with *vc0175,* renamed herein as **a**nti**v**iral-**c**ytidine **D**eaminase (*avcD*), suggesting a shared biological purpose. Enzymes of the deaminase superfamily are widely conserved in all kingdoms in diverse biological contexts. We demonstrated AvcD exhibits deoxycytidylate deaminase (DCD) activity, catalyzing the deamination of cytidine-containing bases to uridine-containing bases. DCD is part of the broader Zn-dependent cytosine deaminase (CDA) family of enzymes, which are primarily involved in the pyrimidine and purines salvage metabolism pathways [36–38]. The activity of DCD enzymes play an important role in the de novo synthesis of deoxythymidine triphosphate (dTTP) by supplying the dUMP required by thymidylate

synthase (TS) to yield deoxythymidine monophosphate (dTMP), the building block for dTTP synthesis [36]. Besides their role in homeostatic regulation of dNTP pools, CDA enzymes are implicated in other biological functions. For instance, unlike AvcD, the APOBEC (Apolipoprotein B mRNA editing enzyme catalytic polypeptide-like) family plays a vital role in viral immunity in eukaryotes where their catalytic activity is deaminating minus-strand DNA of retroviruses, leading to viral genome instability [39–42]. While the regulation of the APOBEC family is poorly understood, AvcD activity is post-translationally inhibited by a unique sRNA named AvcI (**Avc**D **I**nhibitor), which is encoded immediately upstream of the *avcD* locus in a manner that is similar to the Type III TA systems.

The central aim of my thesis is to understand the function of AvcID and the activation of AvcID in response to phage infection. In Chapter 2, I found that AvcD consists of two domains—a P-loop NTPase and a DCD domain— both of which are required for the deamination of dCTP and dCMP. Overproduction of AvcD results in a cell filamentation phenotype and requires conserved features of both domains. AvcD-induced filamentation can also be abolished in the presence of the sRNA AvcI in a post-translational manner. Furthermore, AvcID forms a novel class of Type III TA system, unlike most proteinaceous toxins in Type III TA systems which are endonucleases, and I demonstrated that the AvcID system confers phage protection. In Chapter 3, I explore the mechanisms by which AvcID is activated upon phage infection and find that AvcD is released from AvcI inhibition and thereby depletes dC pools and increases the dUMP level. The exhaustion of nucleotides by AvcD demonstrates this biological utility as a combat strategy against phage infection.

Chapter 2 – A Broadly Conserved Deoxycytidine Deaminase Protects Bacteria from Phage Infection

2.1: PREFACE

The content of this chapter is in preprint at *biorxiv* (2021.03.31.437871) and the manuscript is accepted at Nature Microbiology as of 04/07/2022. Dr. Severin and I contributed equally to this work. Mr. Clint Elg from Dr. Eva Top group and Dr. Benjamin Ridhenhour developed and performed the correology bioinformatic pipeline. Dr. Janani Ravi performed the phylogenetic tree and domain analysis of AvcD and homologs. Dr. Evan Waldron and Dr. Abhiruchi Kant from Dr. Matthew Neiditch's group purified AvcD and performed EMSAs. Dr. John Dover and Dr. Kristin Parent provided insights for phage experiments. Alex Wessel made plasmids for Figure 2.5 and Christopher Rhoades made VSP island knockout strains in *V. cholerae*.

2.2: ABSTRACT

The El Tor biotype of *Vibrio cholerae* is responsible for perpetuating the longest cholera pandemic in recorded history. The genomic islands VSP-1 and -2 are understudied genetic features that distinguish El Tor from previous pandemic *V. cholerae*. To understand the role of VSP genes, we calculated the co-occurrence of VSP genes across bacterial genomes. This analysis predicted that the previously uncharacterized gene *vc0175*, herein renamed **a**nti-**v**iral **c**ytidine **d**eaminase (*avcD*), is in a gene network with *dncV*, a cyclic GMP-AMP synthase involved in phage defense. AvcD consists of two domains; a P-loop NTPase and a deoxycytidylate deaminase, both of which are required for the deamination of dCTP and dCMP. We found that homologs of *avcD* are broadly conserved across the three domains of life. Additionally, AvcD activity is post-translationally inhibited by a unique noncoding RNA named AvcI located

immediately upstream of the *avcD* locus, in a manner analogous to Type III toxin-antitoxin systems, and we demonstrate that AvcID protects bacteria from phage infection. Activation of AvcD upon inhibition of transcription or phage infection significantly alters cellular nucleotides by depleting dC substrates and increasing dUMP. Our results show that AvcID protects against bacteriophage infection by combining aspects of two eukaryotic anti-viral strategies; cytosine deamination (*e.g.*, APOBEC) and the depletion of cellular deoxynucleotides (*e.g.*, SAMHD1).

2.3: INTRODUCTION

Vibrio cholerae, the etiological agent responsible for the diarrheal disease cholera, is a monotrichous, Gram-negative bacterium found ubiquitously in marine environments [43]. There have been seven recorded pandemics of cholera, beginning in 1817, and the fifth and sixth pandemics were caused by strains of the classical biotype. The seventh pandemic, which began in 1961 and continues today, was initiated and perpetuated by circulating strains of the El Tor biotype. Numerous phenotypic and genetic characteristics are used to distinguish the classical and El Tor biotypes [44]. It is hypothesized that El Tor's acquisition of two unique genomic islands of unknown origins, named the Vibrio Seventh Pandemic Islands 1 and 2 (VSP-1 and 2) [14], played a pivotal role in El Tor's evolution to pandemicity and the displacement of the classic biotype in modern cholera disease [16].

Combined, VSP-1 and VSP-2 encode ~36 putative open reading frames (ORFs) within ~39 kb (Figs. 1A-B) [14, 17, 45, 46]. While the majority of the genes in these two islands remain to be studied, it is hypothesized that the biological functions they encode

may contribute to environmental persistence [23, 28] and/or the pathogenicity [24] of the El Tor biotype. In support of this idea, VSP-1 encodes a phage defense system encompassing the genes *dncV, capV*, *vc0180* and *vc0181* called the cyclic-oligonucleotide-based antiphage signaling system (CBASS) [7] (Fig. 1A). CBASS limits phage invasion of bacterial populations via a process termed abortive replication whereby upon phage infection cyclic GMP-AMP (cGAMP) synthesis by DncV activates cell lysis by stimulating the phospholipase activity of CapV [7, 25]. During our search for VSP-1 and 2 gene networks, we determined that the gene *vc0175,* renamed herein as **a**nti**v**iral **c**ytidine **d**eaminase (*avcD*), co-occurs in bacterial genomes with *dncV*, suggesting a common function.

We show that AvcD exhibits deoxycytidylate deaminase (DCD) activity and is part of the broader zinc-dependent cytosine deaminase (CDA) family of enzymes [36–38]. The activity of DCD enzymes play a vital role in the de novo synthesis of deoxythymidine triphosphate (dTTP) by supplying the dUMP required by thymidylate synthase (TS) to form deoxythymidine monophosphate (dTMP) [36]. CDA enzymes belonging to the APOBEC (Apolipoprotein B mRNA editing enzyme catalytic polypeptide-like) family also play an important role in viral immunity in higher organisms where their catalytic activity is utilized for the deamination of nucleic acids rather than free nucleotide substrates to restrict several types of viruses, such as retroviruses, and retroelements [39–42, 47].

A primary challenge faced by lytic phage is to rapidly replicate many copies of its genome, which requires sufficient nucleotide substrates [48]. During DNA phage infection, total DNA within a bacterium can increase 5-10 fold, illustrating the vast

amount of DNA replication that must occur in a short window of time [49, 50]. To accomplish this feat, invading DNA phage often corrupt the delicate balance of enzymatic activity across a host's deoxynucleotide biosynthetic pathways by deploying their own DCD, dUTPase, TS, and ribonucleotide reductase to ensure the appropriate ratio and abundance of deoxyribonucleotides [51–55].

In this chapter, we show that AvcD is a dual domain protein consisting of a putative N-terminal P-loop NTPase (PLN) and C-terminal DCD domain, and this novel domain architecture is present across the tree of life. Overexpression of AvcD promotes cell filamentation, which has hallmarks of nucleotide starvation resembling thymine-less death (TLD) toxicity [56–59]. Our results demonstrate that ectopic expression of AvcD indeed corrupts the intracellular concentrations of deoxynucleotides by depleting dCTP and dCMP and this activity protects bacteria from phage infection. Moreover, we demonstrate that AvcD activity is negatively regulated by a non-coding RNA located 5' of the *avcD* locus [renamed herein as **AvcD I**nhibitor (AvcI)] which resembles a toxin-antitoxin (TA) system. Furthermore, *avcID* systems are widely encoded in bacteria and we show that a subset of them function similarly, establishing cytidine deaminase enzymes as antiphage defense systems in bacteria.

2.4: MATERIALS AND METHODS

2.4.1: Bacterial Strains, Plasmids, and Growth Conditions

The strains, plasmids, and primers used in this study are listed in Appendix 2 (Tables 1-3). Unless otherwise stated, cultures were grown in Luria-Bertani (LB) at 35°C and supplemented with the following as needed: ampicillin (100 µg/mL), kanamycin

(100 µg/mL), and isopropyl-β-D-thiogalactoside (IPTG) (100 µg/mL). *E. coli* BW29427, a diaminopimelic acid (DAP) auxotroph, was additionally supplemented with 300 µg/mL DAP. The *V. cholerae* El Tor biotype strain C6706str2 [60] was utilized in this study and mutant strains were generated using the pKAS32 suicide vector [61] using three fragments: 500 bp of sequence upstream of the gene of interest, 500 bp of sequence downstream of the gene of interest and cloned into the KpnI and SacI restriction sites of pKAS32 using by Gibson Assembly (NEB). P_{tac} inducible expression vectors were constructed by Gibson Assembly with inserts amplified by PCR and pEVS143 [62] or pMMB67EH [63] each linearized by EcoRI and BamHI, as well as pET28b digested with NcoI and XhoI for the C-terminal His tags. To generate the N-terminal His tag AvcD, pAvcD (4-532), *avcD* (corresponding to residues 4-532) was PCR amplified from $pAvcD^{6xHis}$ using Phusion High-Fidelity DNA Polymerase (NEB) with EWAvcDFwd and EWAvcDRev primers. Finally, In-Fusion® Snap Assembly (Takara Bio US Inc.) was used to integrate the purified insert into pET28b that had been linearized using the restriction enzymes NdeI and XhoI. pEVS141 [64] is used as an empty vector control for experiments using pEVS143 derived constructs. Site-directed mutagenesis was performed using the SPRINP method [65]. Plasmids were introduced into *V. cholerae* through biparental conjugation using an *E. coli* BW29427 donor. Transformation of *E. coli* for ectopic expression experiments was performed using electroporation with DH10b for expression of pEVS143 and pMMB67EH derived plasmids. Transformation of *E. coli* for protein production experiments was performed using either electroporation or heat shock at with BL21(DE3) for expression with pET28b based constructs.

2.4.2: GeneCoOccurrence Bioinformatics Analysis

Our GeneCoOccurrence software package is built on Kim and Price's approach [66] to of creating maximum related subnetworks (MRS) using the co-occurrence of genes (in this case, the genes within genomic islands VSP-1 and 2) to calculate genetic co-occurrence. The source code, documentation, and a Docker container for this Python3 package are available at https://github.com/clinte14/GeneCoOccurrence. While VSP-1 is used to simplify the description of the method detailed below, both VSP-1 and 2 were independently analyzed in the same fashion. First, BLASTP was used to find homologs for each VSP-1 gene against the NCBI non-redundant protein database with an E-value cutoff of 10^{-4}. The BLAST results were limited to bacterial genomes, and all taxa belonging to the genus *Vibrio* were removed to avoid bias from closely related vertical inheritance. The BLAST results were used to generate a presence or absence matrix of VSP-1 homologues with all species along one axis and VSP-1 genes along the other axis. Next, a pairwise Pearson correlation value was calculated between all VSP-1 genes *i* and *j* using binary data from the above-mentioned presence/absence matrix*:*

$$r_{ij} = \frac{C_{ij}N - E_i E_j}{\sqrt{E_i E_j (N - E_i)(N - E_j)}},$$

where N is the total number of unique species returned from the BLAST search and C_{ij} the number of species with co-occurrence of genes *i and j.* While a Pearson correlation is warranted for a normally distributed binary data set, it does not account for indirect correlation. For example, if genes *i* and *j* individually associate with a third gene, a Pearson correlation will incorrectly calculate a correlation between *i* and *j.* To help correct for indirect correlation we calculate a partial correlation w_{ij} from the Pearson r_{ij}:

$$w_{ij} = \frac{P_{ij}}{\sqrt{P_{ii}P_{ij}}},$$

where the (*i, j*) element of the inverse matrix of Pearson r_{ij} is P_{ij} [66].

The partial correlation correction w_{ij} has the advantage of generating a normalized output that ranges between -1 to 1. For example, a w_{ij} of -1 reveals genes *i* and *j* never co-occur in the same species, while a value of 1 demonstrates genes *i* and *j* always co-occur in the same species. A w_{ij} of 0 is the amount of co-occurrence expected between unrelated genes *i* and *j* drawn from a normal distribution. Using the above-mentioned approach, a partial correlation value w_{ij} was calculated for all genes *i* to *j* in VSP-1 and VSP-2 (Supplemental Files 1 and 2). The single highest w_{ij} value for each VSP-1 gene was represented as an edge (i.e., line) in our visualization (Figs. 1A-B). Any set of genes that contains no further edges were assigned to a unique MRS that suggests functional association of the gene products within a unique gene network.

2.4.3: Genomic Identification, Structural, and Sequence Analyses of AvcD & AvcI Homologs

AvcD from *V. cholerae* El Tor N16961 (AAF93351.1) was identified as locus tag *vc0175*. AvcD and homologs profiles are performed using translated BLAST tblastn and run against the nucleotide collection (nr/nt) in the National Center for Biotechnology Information (NCBI) database, using >40% similarities cutoff. For previously annotated domains, the Pfam feature in KEGG [67, 68] was utilized to determine AvcD homologs. Out of all the AvcD homologs, AvcD homologs from *Vibrio parahaemolyticus* O1: Kuk str. FDA_R31 [69] (WP_020839904.1), *Proteus mirabilis* AR_379 [70]

(WP_108717204.1), and *E. coli* O78:H11 H10407 [71] (ETEC) (WP_096882215.1) were analyzed in this study. Genomic contextual information from prokaryotic gene neighborhoods was retrieved from NCBI genome graphics feature to uncover *avcI*-like genes located as a hypothetical ORF 5' of the *avcD* locus. If unannotated, the ORFinder feature from NCBI was used to determine the location and size of the putative *avcI* locus. To predict the structure of AvcD from *V. cholerae,* the amino acid sequence was submitted to Phyre2 [72] and structural visualization was performed using PyMol (https://pymol.org). The amino acid and nucleotide alignments were analyzed using ClustalW Omega from EMBL-EBI web services [73] and LocARNA [74], respectively.

2.4.4: Identification and Characterization of Protein Homologs

Homology searches: To ensure the identification of a comprehensive set of homologs (close and remote), we started with six representative proteobacterial AvcD proteins from *V. cholerae*, *V. parahaemolyticus*, *P. mirabilis*, and *E. coli* ETEC described above along with *E. cloacae* (WP_129996984.1), and *A. veronii* (WP_043825948.1) and performed homolog searches using DELTABLAST [75] against all sequenced genomes across the tree of life in the NCBI RefSeq database [76–78]. Homology searches were conducted for each protein and the search results were aggregated; the numbers of homologs per species and of genomes carrying each of the query proteins were recorded. These proteins were clustered into orthologous families using the similarity-based clustering program BLASTCLUST [75].

Characterizing homologous proteins: Phyre2, InterProScan, HHPred, SignalP, TMHMM, Phobius, Pfam, and custom profile databases [72, 79–86] were used to

identify signal peptides, transmembrane (TM) regions, known domains, and secondary structures of proteins in every genome. Custom scripts were written to consolidate the results [87–92], and the domain architectures and protein function predictions were visualized using the MolEvolvR web application [93] (https://jravilab.github.io/phage_defense_avcd/).

Phylogenetic analysis (MSA and Tree): Thousands of homologs from all six starting points for AvcD proteins were consolidated and representatives were chosen from distinct Lineages and Genera, containing both the N- and C-terminal AvcD domains (PLN and DCD domains). Multiple sequence alignment (MSA) of the identified homologs was performed using Kalign [90] and MUSCLE [94, 95] (msa R package [96]). The phylogenetic trees were constructed using FastTree [97] FigTree [98] and the R package, ape [99].

All our molecular evolution and phylogenetic analyses for protein characterization were done using the MolEvolvR webapp: http://jravilab.org/molevolvr.

2.4.5: Growth Curve Assays

Overnight cultures were diluted 1:1000 into LB supplemented with antibiotics and IPTG in a 96-well microplate (Costar®). Growth was monitored by measuring OD_{600} every 15 min for 15 hours (h) using a BioTek plate reader with continuous, linear shaking.

2.4.6: Fluorescence Microscopy and Analysis

Cells were imaged as previously described [100]. Briefly, overnight cultures were diluted 1:1000 into LB supplemented with antibiotics and IPTG. Cultures were grown and induced for 7-8 h, at which point cells were diluted to an OD_{600} of 0.5 in 1X PBS, then membrane stain FM4-64 dye (ThermoFisher Scientific) was added to a final concentration of 20 µg/mL. One percent agarose pads in deionized water were cut into squares of approximately 20 x 20 mm and placed on microscope slides. 2 µl of diluted cultures were spotted onto a glass coverslip and then gently placed onto the agarose pad. FM4-64 signal was visualized using a Leica DM5000b epifluorescence microscope with a 100X-brightfield objective under RFP fluorescence channel. Images were captured using a Spot Pursuit CCD camera and an X-cite 120 Illumination system. Each slide was imaged with at least 20 fields of view for each biological replicate. Cell lengths were processed using the Fiji plugin MicrobeJ [101], and data were visualized and analyzed using R [91] by quantifying the length of the curvilinear (medial) axis of detected cells.

2.4.7: Construction and Screening of Mutant Gene Libraries

AvcI-insensitive AvcD constructs were generated by error-prone PCR (epPCR) using pAvcD (pCMW204) as the template. Three different concentrations of $MnCl_2$ (12.5 mM, 1.25 mM, and 125 µM) were used in triplicate using Taq polymerase (Invitrogen) and reactions containing the same $MnCl_2$ concentration were pooled. The PCR products were purified, using The Wizard® SV Gel and PCR Clean-Up Kit (Promega), and ligated to pEVS143 via Gibson Assembly. The assembled reactions were

electroporated into *E. coli* DH10b and plasmid libraries were collected from ~ 30,000 representative colonies for each MnCl2 concentration. Plasmid libraries were harvested using the Wizard® Plus SV Minipreps DNA purification Kit (Promega). Plasmid libraries were subsequently electroporated into *E. coli* BW29427 which were again plated and pooled to contain ~ 30,000 representative colonies. The *E. coli* BW29427 random mutant pAvcD libraries were conjugated with Δ*avcD V. cholerae* on LB agar plates for 8 h, harvested, diluted, and spread on LB agar plates containing 1 mM IPTG and antibiotics, and grown overnight. ~ 5,000 colonies were screened in each library and all colonies exhibiting a wrinkled and small colony morphology, indicative of cell filamentation, were isolated and filamentation was confirmed by fluorescence microscopy. Mutant pAvcD plasmids recovered from cells exhibiting cell filamentation were sequenced by Sanger sequencing. Mutations were reintroduced individually into the WT pAvcD construct using SPRINP mutagenesis [65] and reevaluated using fluorescence microscopy to confirm the AvcD variant's ability to remain constitutively active in *ΔavcD V. cholerae*.

2.4.8: RNA Isolation, qRT-PCR, and Co-transcription Analysis

RNA isolation and qRT-PCR analysis were carried out as previously described [102]. Briefly, triplicate overnight cultures were subcultured 1:1000 in 10 mL LB and grown to three different OD_{600}: 0.2 (Early Exponential), 1.0 (Late Exponential), and 2.5 (Stationary). 1 mL of each replicate was pelleted, and RNA was extracted using TRIzol® reagent following the manufacturer's directions (Thermo Fischer Scientific). RNA quality and quantity were determined using a NanoDrop spectrophotometer (Thermo Fischer

Scientific). 5 µg of purified RNA was treated with DNase (Turbo™ DNase, Thermo Fischer Scientific). cDNA synthesis was performed using SuperScript™ III Reverse Transcriptase (Thermo Fischer Scientific). cDNA was diluted 1:64 into molecular biology grade water and amplification was quantified using 2x SYBR Green (Applied Biosystems™). For measuring gene expressions or determining *ori*/*ter* ratios, 25 µL reactions consisted of 5 µL each of 0.625 µM primers 1 and 2, 12.5 µL of 2X SYBR master mix, and 2.5 µL of template (0.78 ng/µL cDNA for gene expression and 0.25 ng/µL *V. cholerae* genomic DNA for *ori*/*ter* [103]). qRT-PCR reactions were performed in technical duplicates for biological triplicate samples and included no reverse transcriptase reaction controls ("no RT") to monitor for contaminating genomic DNA in purified RNA samples. qRT-PCR reaction thermo profile was 95°C for 20 seconds (s) then 40 cycles of 95°C for 2 s and 60°C for 30 s in the QuantStudio 3 Real-Time PCR system (Applied Biosystems™). The *gyrA* gene was used as an endogenous control to calculate relative quantification (ΔC_t).

To determine the co-transcription of *avcI* and *avcD*, PCR amplification was performed in 25 µL volumes using Q5 polymerase (NEB), 0.5 µM each of the forward and reverse primers as indicated, 0.2 mM dNTPs, and 3.5 µL of cDNA or no RT control templates (0.78 ng/µL) from RNA purified from WT and Δ*ig*^{222} *V. cholerae* grown to late exponential-phase in biological triplicate. The thermal profile was 98°C for 30 s, 30 cycles of 98°C for 10 s, 55°C for 30 s, 72°C for 10 sec and one cycle of 72°C for 2 min. PCR products were loaded on a 1% agarose gel and stained with EZ-Vision® (VWR). Images were taken using the GelDoc system (Bio-Rad).

2.4.9: Protein Purification

pAvcD(4-532) was transformed via heat shock at 42°C into *E. coli* strain BL21(DE3) and grown at 37°C and 200 RPM to OD_{600} = 0.8-0.9 in LB medium containing 30 μM kanamycin. The medium was then supplemented with 100 μM $ZnCl_2$, and AvcD expression was induced with 500 μM IPTG. Following induction, the cells were grown overnight at 18°C and 200 RPM. The cells were then pelleted at 7,000 x g for 15 min. Cell pellets were resuspended in Buffer A (500 mM NaCl, 50 mM HEPES pH 7.5, 40 mM imidazole pH 7.5, 1 μg/μL DNase, and 1 mM phenylmethanesulfonyl fluoride (PMSF)) and lysed by two passages through a French press at approximately 25,000 PSI. Lysate was clarified at 35,000 x g at 4°C for 45 min. The clarified lysate was passed over His60 superflow Ni resin (Takara Bio US Inc.), and the protein-bound resin was washed with Buffer B (500 mM NaCl, 50 mM HEPES pH 7.5, 70 mM imidazole pH 7.5). The resin was then resuspended in Buffer C (500 mM NaCl, 50 mM HEPES pH 7.5, 40 mM imidazole pH 7.5) and the slurry was nutated overnight at 4°C in the presence of 3.2 μg/mL thrombin (BioPharm Laboratories, LLC.). Following overnight incubation, the resin was repacked into a column and eluted by gravity. SDS-PAGE of the eluate revealed that overnight digestion at 4°C resulted in complete cleavage of AvcD from the His6 affinity tag. Following thrombin digestion, AvcD (residues 4-532) contained two heterologous N-terminal residues (Gly-Ser) derived from the thrombin cleavage sequence. To remove the majority of thrombin, the digested AvcD fraction was combined with p-aminobenzamidine-agarose (Millipore-Sigma) and nutated for 30 min at 4°C. The protein-resin mixture was then repacked into a column and AvcD was eluted by gravity. To inactivate residual thrombin, benzamidine was then added to the

eluted AvcD to a final concentration of 50 mM. AvcD was then concentrated using a 10 kDa cut-off Vivaspin concentrator (Sartorius). The concentrated AvcD was loaded onto a Superdex 200 16/70 column (GE Healthcare) equilibrated in Buffer D (100 mM NaCl, 20 mM HEPES pH 7.5, and 1 mM tris (2-carboxyethyl) phosphine (TCEP)). S200 peak fractions were analyzed by SDS-PAGE to assess purity. The purest AvcD-containing fractions were combined and concentrated using a 10 kDa cut-off Vivaspin concentrator to approximately 4.38 mg/mL (as evaluated using the Bradford method).

2.4.10: RNA Synthesis and Purification

The method for RNA production was modified from previously described [104–107]. The AvcI DNA template for *in vitro* transcription was PCR amplified from pAvcI using Q5 High-Fidelity DNA Polymerase (NEB) and the oligonucleotide pair EJW002 and EJW003. To incorporate the T7 promoter into the final AvcI DNA template, forward primer, EJW002, included the T7 promoter sequence prior to the homologous sequence for AvcI. Additionally, the first two residues of the reverse primer, EJW003, were 2′-OMe modified to reduce 3′-end heterogeneity of the transcript [108]. The PCR reaction was analyzed using a 1% agarose gel, and the band corresponding to the AvcI DNA template was excised and gel purified. AvcI RNA was synthesized by *in vitro* transcription using the T7-AvcI DNA template and the HiScribe™ T7 High Yield RNA Synthesis Kit (NEB). The transcription reactions were incubated at 37°C for 4 h. Following transcription, DNase I (NEB) was added to a final concentration of 1X per reaction and incubated at 37°C for an additional 15 min. AvcI was then purified using a guanidinium thiocyanate-phenol-chloroform extraction with Trizol® reagent (Thermo

Fisher Scientific). Extracted RNA was subsequently precipitated with isopropanol. Precipitated RNA was pelleted via centrifugation and subsequently re-solubilized in RNA storage buffer (300 mM NaCl, 10 mM Tris pH 8.0, 2 mM EDTA) for 4 h at 65°C. Purity of product was evaluated using a denaturing 7 M urea PAGE. Individual aliquots of AvcI were flash frozen using liquid nitrogen and stored long-term at -80°C. Reverse complement AvcI was generated as described above using the oligonucleotide pair EJW016 and EJW017.

2.4.11: Electrophoretic Mobility Shift Assay (EMSA)

10 µL binding reaction mixtures containing 60 µM, 30 µM, 15 µM, 7.5 µM or 3.75 µM AvcD or 357.6 µM RpfR with 0.5 µM AvcI or AvcI-RC RNA were incubated at 30°C for 30 min in binding buffer (20 mM HEPES pH 7.5, 100 mM NaCl, and 1 mM TCEP). The controls contained either 60 µM AvcD, 0.5 µM AvcI, or 0.5 µM AvcI-RC in Buffer D. EMSA loading buffer (40% sucrose, 10 mM Tris pH 8.0, 60 mM EDTA and 0.03% bromophenol blue) was added to each sample in a 1:1 ratio prior to loading 10 µL of sample onto the native 1X TBE 6% polyacrylamide gel. Native PAGE was carried out at 300 V for 40 min at 4°C. The gel was subsequently stained with SYBR Gold (ThermoFisher Scientific) and visualized using UV trans illumination and a 602/50 emission filter on a ChemiDoc MP Imaging System (Bio-Rad Laboratories, Inc.).

2.4.12: Denaturing Urea PAGE

Denaturing urea PAGE of AvcI and AvcI-RC was performed using 1X TBE 8% polyacrylamide 7 M urea denaturing gels. 10 µL AvcI or AvcI-RC at 0.5 µM in RNA

storage buffer was mixed 1:1 with 2X RNA loading dye (NEB). Samples were heated for 5 min at 95°C and loaded onto the gel. Denaturing PAGE was carried out at 300 V until the dye front had traveled 3/4th the length of the gel. The gel was then stained with ethidium bromide, and the RNA products were visualized using UV trans illumination and a 602/50 emission filter on a ChemiDoc MP Imaging System.

2.4.13: In vitro Nucleic Acid Deamination Assay

Cell Lysate Preparation: Overnight cultures were subcultured 1:333 and grown to an OD_{600} of ~0.5 - 1.0. Cultures were induced with 1 mM IPTG, supplemented with 100 µM $ZnSO_4$, and grown for an additional 3 h. Cell pellets from 100 mL of induced cultures were harvested in two successive 15 min centrifugation steps at 4,000 x g and 4°C. Supernatants were decanted and pellets were flash frozen in an ethanol and dry ice bath and stored at -80° C. Pellets were thawed on ice and suspended in 2 mL of lysis buffer E (50 mM $NaPO_4$, pH 7.3, 300 mM NaCl, 2 mM β-mercaptoethanol, 20% glycerol and Roche cOmplete™ protease inhibitor (1 tablet per 10 mL)). 1 mL of cell suspension was transferred to a microcentrifuge tube and sonicated on ice using a Branson 450 Digital Sonifier (20% amplitude, 20 sec total, 2.5 sec on, 2.5 sec off). Crude lysates were centrifuged at 15,000 x g for 10 min at 4°C and clarified lysates were transferred to fresh microcentrifuge tubes on ice. Clarified lysates were normalized for total protein to a concentration of 1.9 mg/mL as measured by Bradford reagent and a BSA standard. 26.5 µL reactions composed of lysis buffer E, nucleic acid substrates, and 3.5 µL of normalized clarified lysates were assembled in PCR strip tubes, mixed by gentle pipetting, and incubated at room temperature (~23°C) for 1 h. NH_4Cl solutions at

the indicated concentration were dissolved in lysis buffer E and substituted for nucleic acid substrates as positive controls.

Ammonia Detection: The evolution of ammonia from the deamination of the nucleic acid substrates was observed using a phenol-hypochlorite reaction to produce indophenol in a clear 96-well microtiter plate and modified from Dong et al. [109]. The work of Ngo et al. [110] was considered when designing the lysis buffer so as not to interfere with the phenol-hypochlorite reaction. 50 µL of Reagent A (composition below) was added to each well followed by 20 µL of the completed in vitro deamination reaction described above. The phenol-hypochlorite reaction was initiated by the addition and gentle mixing of 50 µL Reagent B (composition below) to the wells. The reaction was incubated at 35°C for 30 min and the ABS_{630} was measured using a plate reader.

Reagent A = 1:1 (v/v), 6% (w/v) sodium hydroxide (Sigma) in water: 1.5% (v/v) sodium hypochlorite solution (Sigma, reagent grade) in water.

Reagent B = 1:1:0.04 (v/v/v), water: 0.5% (w/v) sodium nitroprusside (Sigma) in water: phenol solution (Sigma, P4557)

2.4.14: Western Blot

Strains containing AvcD- and variant- C-terminal 6x-histidine fusions were grown, induced, and harvested as described previously above (See In vitro Nucleic Acid Deamination Assay: Cell Lysate Prep), except for the His-tag fusion (pGBS98) which are induced for only 2 h with 100 µM IPTG and not subjected to sonication. The cell pellets were resuspended in 2 mL of chilled 1X PBS and subsequently normalized to OD of 1.0. 1 mL aliquots were collected by centrifugation at 15k x g for 1 min. Cell

pellets were subsequently resuspended in 90 µL of lysis buffer A and 30 µL of 4x Laemmli buffer, denatured for 10 min at 65°C, and centrifuged at 15k x g for 10 min. 5 µL of samples were loaded into a precast 4-20% SDS-PAGE gels (Mini-PROTEAN TGX Precast Protein Gels, Bio-Rad) alongside size standards (Precision Protein Plus, Bio-Rad). Gels were run at room temperature for 90 min at 100 V in 1x Tris/glycine/SDS running buffer. Proteins were transferred to nitrocellulose membranes (Optitran). The membranes were blocked using 5% skim milk and incubated with 1:5000 THE™ His Tag Antibody, mAb, Mouse (GenScript) followed by 1:4000 Goat Anti-Mouse IgG Antibody (H&L) [HRP], pAb (GenScript), treated with Pierce™ ECL Western Blotting Substrate, and imaged using an Amersham™ Imager 600.

2.4.15: UPLC-MS/MS Quantification of Deoxynucleotides

Deoxynucleotide concentrations were determined as previously described [111] with minor modifications. For measuring in vivo intracellular deoxynucleotide concentrations, overnight cultures were subcultured 1:1000 and grown to OD_{600} of ~1.0. Plasmid expression was induced by the addition of 1 mM IPTG for 1 h, and 1 mL of cultures were collected by centrifugation at 15,000 x g for 1 min. Pellets were resuspended in 200 µL of chilled extraction buffer [acetonitrile, methanol, ultra-pure water, formic acid (2:2:1:0.02, v/v/v/v)]. To normalize in vivo nucleotide samples, an additional cell pellet was collected from 1 mL of culture by centrifugation at 15,000 x g for 1 min, resuspended in 200 µL lysis buffer F (20 mM Tris·HCl, 1% SDS, pH 6.8), and denatured for 10 min at 60°C. Denatured lysates were centrifuged at 15,000 x g for 1 min to pellet cellular debris, and the supernatant was used to quantify the total protein

concentration in the sample using the DC protein assay (Bio-Rad) and a BSA standard curve [100]. The concentrations of deoxynucleotides detected by UPLC-MS/MS were then normalized to total protein in each sample.

For the quantification of deoxynucleotides in vitro *E. coli* BL21(DE3) clarified lysates were prepared as described for the deamination experiment above and normalized to 20 mg/mL of total protein and 200 µL of normalized clarified lysates were assembled in PCR strip tubes. To measure abundance of dUMP and dUTP prior to the addition of 1 µM dCTP, 20 µL of normalized clarified lysates were added to 200 µL of chilled extraction buffer. 20 µL of 10 µM dCTP was then added to the remaining clarified lysate and 20 µL lysates aliquots were removed 1, 5, 10, and 30 min after the addition of dCTP and mixed in 200 µL chilled extraction buffer.

All samples resuspended in extraction buffer, in vivo and in vitro, were immediately incubated at -20°C for 30 min after collection and centrifuged at 15,000 x g for 1 min. The supernatant was transferred to a new tube, dried overnight in a speed vacuum, and finally resuspended in 100 µL ultra-pure water. Experimental samples and deoxynucleotides standards [1.9, 3.9, 7.8, 15.6, 31.3, 62.5, and 125 nM of dATP (Invitrogen), dGTP (Invitrogen), dTTP, (Invitrogen), dCTP (Invitrogen), dCMP (Sigma), dUTP (Sigma), and dUMP (Sigma)] were analyzed by UPLC-MS/MS using an Acquity Ultra Performance LC system (Waters) coupled with a Xevo TQ-S mass spectrometer (Waters) with an ESI source in negative ion mode. The MS parameters were as follows: capillary voltage, 1.0 kV; source temperature, 150°C; desolvation temperature, 400°C; cone gas, 120 L/h. Five microliter of each sample was separated in reverse phase using Acquity UPLC Premier BEH C18, 2.1 x 100 mm, 1.7 µm particle size, VanGuard FIT at

a flow rate of 0.3 mL/min with the following gradient of solvent A (8mM DMHA (N,N-dimethylhexylamine) + 2.8 mM acetic acid in water, pH~9) to solvent B (methanol): *t* = 0 min; A-100%:B-0%, *t* = 10 min; A-60%:B-40%, *t* =10.5; A-100%:B-0%, *t* = 15 min; A-100%:B-0% (end of gradient). The conditions of the MRM transitions were as follows [cone voltage (V), collision energy (eV)]: dATP, 490 > 159 (34, 34); dCTP, 466 > 159 (34, 34); dGTP, 506 > 159 (15, 46); dTTP, 481 > 159 (25, 34); dUTP, 467 > 159 (25, 34); dCMP, 306 > 97 (43, 22); dUMP, 306 > 111 (22, 22).

2.4.16: In vivo dNTP Quantification Following Termination of Transcription and Translation

Overnight cultures were subcultured 1:1000 into a pair of matched sister cultures and grown at 35° C with aeration. 2 mL of each sister culture was collected at time zero (OD600 ~1.0) with 1 mL for dNTPs quantification and 1 mL for total protein quantification (See 'UPLC-MS/MS Quantification of Deoxynucleotides'). Following the initial culture sampling, each sister culture was treated with either rifampicin (250 µg/mL) to stop transcription or spectinomycin (200 µg/mL) to stop translation. Following the initial antibiotic treatments, cultures were sampled for both dNTP and total protein quantifications for the duration of the experiment at indicated timepoints.

2.4.17: Phage Infection and Plaque Assays

Coliphages were propagated on *E. coli* MG1655 in LB, and their titer was determined using the small drop plaque assay method, as previously described [7]. Briefly, 1 mL of overnight cultures were mixed with 50 mL of MMB agar (LB + 0.1 mM

MnCl2 + 5 mM MgCl2 + 5 mM CaCl2 + 0.5% agar), tenfold serial dilutions of phages in MMB were dropped on top of them and incubated overnight at room temperature. The viral titer is expressed as plaque forming units per mL (pfu/mL).

E. coli MG1655 cells transformed with empty vector (pBRP15) and each associated pAvcI-AvcD plasmids were grown in LB overnight at 37°C. Overnight cultures are subcultured 1:1000 in melted MMB agar and let to solidify at room temperature. Tenfold serial dilutions of coliphages in MMB medium were dropped on top of them and incubated overnight at room temperature. Efficiency of plaquing (EOP) was determined for each coliphage by comparing the plaque forming units (PFUs) on each of the AvcI-AvcD containing strains to the control pBRP15 containing strain.

To measure the nucleotides after phage infection, cells were grown in LB overnight at 37°C. Overnight cultures are subcultured 1:1000 in LB and grown to OD_{600} of ~0.3. 3 mL of culture was collected for a time zero reading: 1.5 mL for dNTPs quantification and the 1.5 mL for total protein quantification (See 'UPLC-MS/MS Quantification of Deoxynucleotides'). The cultures are then infected with phage (T3, MOI of 5; SECΦ18, MOI of 10), and additional 3 mL was removed at each indicated subsequent time point.

2.4.18: DNA Replication Assay

Overnight cultures were subcultured 1:100 in LB at 35° C and grown to OD_{600} of ~0.3. The cultures are then infected with T5 phage at the final MOI of 1. 1.5 mL of culture was collected at 10-, 20-, 30-, and 40-min post infection. Culture aliquots were centrifuged at 15k x g for 1 min and the pellets were flash frozen in a dry ice-ethanol

slurry. DNA was extracted using Wizard® Genomic DNA Purification Kit (Promega), using the Gram-negative bacteria protocol and purified DNA from each sample was uniformly resuspended in 50 µL of DNA dehydration solution. DNA quality and quantity were determined using a NanoDrop spectrophotometer (Thermo Fischer Scientific). Using primers were targeting the T5 phage-tail fiber (ORF124) the abundance of T5 phage genome in each sample was quantified using qPCR as described above (See 'RNA Isolation, qRT-PCR, and Co-transcription Analysis**).** The relative abundance of T5 genome was calculated using the difference of C_t between 10 min and each subsequent timepoints [2^-(ΔC_t)] for each strain.

2.4.19: Statistical Analysis

As specified in the figure legends, all of the statistical analyses for the violin plots were performed with R statistical computing software [91], while other data were analyzed in GraphPad Prism Software. Statistically significances denote as the following: a single asterisk (*) indicates $p < 0.05$; double asterisks (**) indicate $p < 0.01$; triple asterisks (***) indicate $p < 0.001$; and quadruple asterisks (****) indicate $p < 0.0001$. Means ± SEM and specific n values are reported in each figure legend.

2.5: RESULTS

2.5.1: *dncV* and *avcD* cooccur in bacterial genomes

To help identify functional interactions within the largely unclassified VSP-1 & 2 genes, VSP island genes were classified into putative "gene networks" or sets of genes that form a functional pathway to accomplish a biological task. Since gene networks

often share deep evolutionary history among diverse taxa, we hypothesized that the set of genes in a gene network would co-occur together in the genomes of diverse taxa at a higher frequency than chance alone would predict. Our software package was named 'GeneCoOccurrence inspired by [66] and is described in detail in the materials and methods.

We calculated a Pearson correlation followed by a partial correlation correction between each of the VSP island genes from the same island across the sequenced bacterial domain. This resulting partial correlation correction "w_{ij}" has an output normalized to a range of -1 to 1, with a w_{ij} of -1 revealing homologs of genes *i* and *j* never occur in the same species as opposed to a value of 1 in which homologues of genes *i* and *j* always co-occur in the same species. Previous research using well-classified *Escherichia coli* gene networks showed that partial correlation values w_{ij} > 0.045 were highly correlated with shared biological functions [66]. Using the above-mentioned approach, we calculated a partial correlation value w_{ij} for all genes *i* to *j* in VSP-1 and VSP-2. From there, we generated a visualization of the Maximum Relatedness Subnetworks (MRS) showing the single highest w_{ij} value for each VSP gene (Figs. 2.1A-B).

One of our VSP-1 gene networks centered on *dncV* and identifies the experimentally validated CBASS anti-phage system (Fig. 2.1A) [7]. Curiously, the putative deoxycytidylate deaminase encoded by *vc0175,* which we renamed *avcD*, was also found to co-occur with *dncV* (w_{ij} = 0.147) but not with any of the other CBASS members (w_{ij} < 0.045) (Fig. 2.1A). Recognizing that co-occurrence of *dncV* with *avcD*

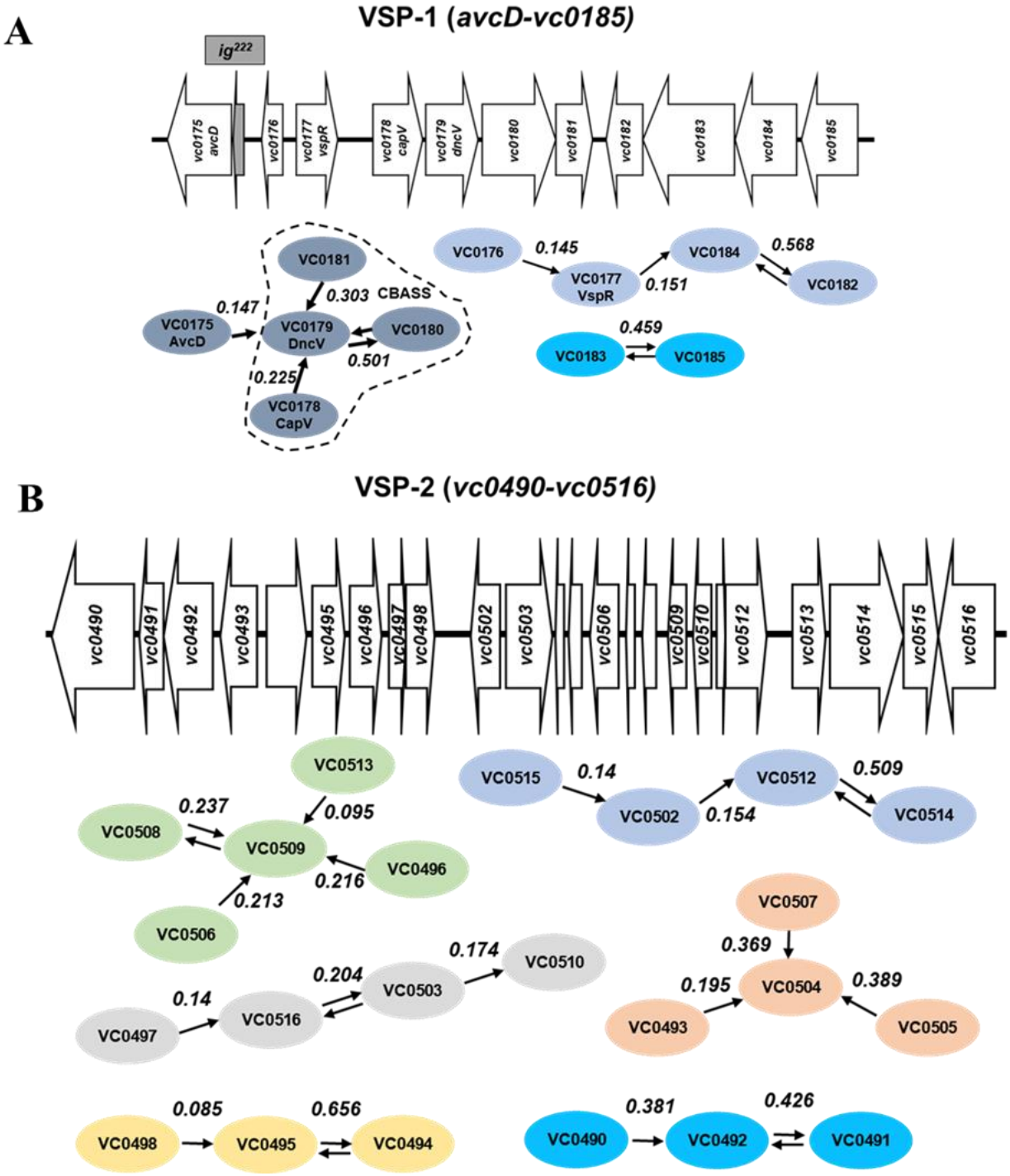


Figure 2.1. VSP-1 and VSP-2 schematic and predicted gene networks (MRS). Cartoon of VSP-1 (**A**) and VSP-2 (**B**) from El Tor *V. cholerae* N16961 and gene network predictions from GeneCoOccurrence. Arrows indicate the highest partial correlation $\boldsymbol{W_{ij}}$ each gene has to another (ovals). Two arrows are presented pointing in opposing directions where the highest correlation $\boldsymbol{W_{ij}}$ is reciprocal between two genes. MRS = maximum relatedness subnetwork

may indicate a shared or common biological function, we sought to understand the biological activity of *avcD*.

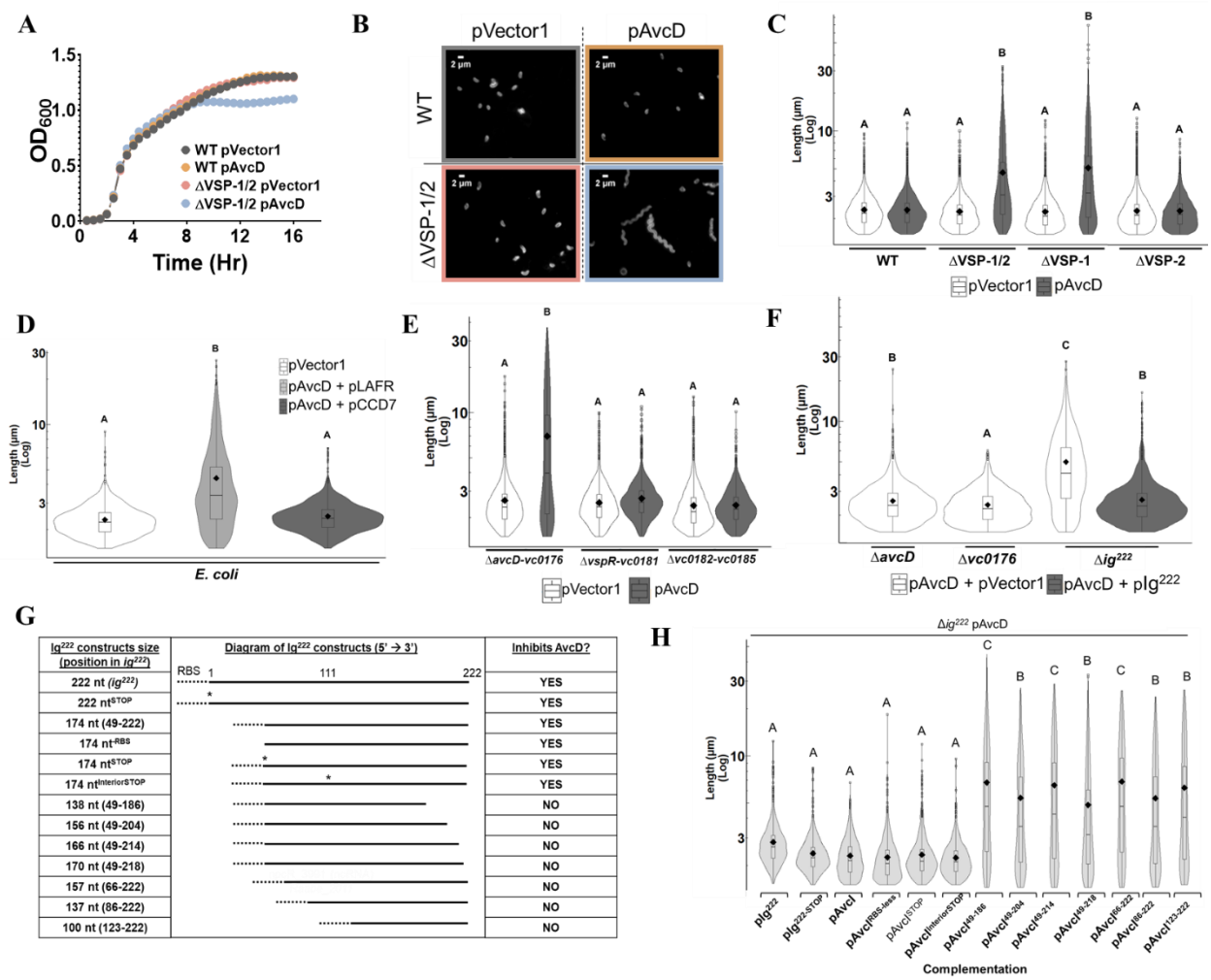


Figure 2.2. Expression of AvcD induces cell filamentation and is inhibited by sRNA AvcI.
(**A**) Growth curves and (**B**) representative images of WT El Tor *V. cholerae* and ΔVSP-1/2 strains expressing *avcD* from a P_{tac}-inducible plasmid (pAvcD) or an empty vector control (pVector1). Cells were stained with FM4-64 prior to imaging. Error bars represent standard error of the mean from three biological replicates. Scale represents 2 µm. (**C**) Cell length distributions of WT *V. cholerae* and VSP island mutants expressing pAvcD or pVector1. (**D**) Distribution of cell lengths replicates of *E. coli* cultures carrying an pVector1 or pAvcD in addition to either an empty vector single copy cosmid (pLAFR) or pLAFR containing VSP-1 (pCCD7). (**E**) Cell length distributions of gene deletions within VSP-1 (Δ*avcD-vc0176*, Δ*vspR-vc0181*, and Δ*vc0182-vc0185*) maintaining either pVector1 or pAvcD. (**F**) Cell length distributions of VSP-1 gene locus mutants expressing pAvcD in combination with either pIg^{222} or a vector control

Figure 2.2 (cont'd) (pVector2). (**G**) Table reporting the capacity of various pIg^{222} constructs to prevent AvcD induced cell filamentation when expressed in combination with pAvcD in Δ*ig*^{222} *V. cholerae*. Dotted line denotes a non-native ribosomal binding site (RBS), "*****" indicates a putative start codon mutated to a stop. (**H**) Cell length distributions of Δ*ig*^{222} *V. cholerae* expressing pAvcD. All cell length distributions represent ~750-3000 cells measured per strain with summary statistics: mean (diamonds), median (horizontal black line), interquartile range (box), and data below and above the interquartile range (vertical lines). Different letters indicate significant differences at $p < 0.05$, according to Tukey's post-hoc test.

2.5.2: Expression of AvcD induces cell filamentation and is abrogated in the presence of sRNA AvcI

We found that over-expressing *avcD* (pAvcD) in wild type (WT) *V. cholerae* did not impact growth, but growth was impaired in a double VSP island deletion mutant (ΔVSP-1/2) (Fig. 2.2A). Expression of pAvcD in the ΔVSP-1/2 background yielded filamentous cell morphologies, suggesting these cells have a defect in cell division that manifests in a reduced growth yield (Fig. 2.2B). We performed the same image analysis in single island mutants and found that cell lengths only increased in strains lacking VSP-1 (Fig. 2.2C). Likewise, overexpression of pAvcD in a laboratory strain of *E. coli* DH10b also induced cell filamentation that was inhibited by provision of a single copy cosmid containing VSP-1 (pCCD7) but not the vector cosmid control (pLAFR) (Fig. 2.2D). To identify the negative regulator of AvcD activity encoded in VSP-1, we generated partial VSP-1 island deletions and quantified cell filamentation following AvcD expression. Of the three partial VSP-1 deletion strains, expression of pAvcD only induced filamentation in the Δ*avcD-vc0176* mutant (Fig. 2.2E). Individual gene deletion mutants of *avcD* and *vc0176* maintained WT cell morphology following expression of AvcD (Fig. 2.2F), suggesting the 504 nt intergenic region between *avcD* and *vc0176* is the source of AvcD inhibition. We identified a 222 nucleotide (nt) open reading frame we

named *ig*^{222} encoded in the same orientation immediately 5' of *avcD* as a possible candidate for the AvcD regulation (Fig. 2.1A). Expression of pAvcD in Δ*ig*^{222} strain produced a filamentous cell morphology that was abrogated when *ig*^{222} was provided in trans (Fig. 2.2F).

To determine whether the AvcD inhibiting component encoded in *ig*^{222} was a small peptide or an untranslated small regulatory RNA (sRNA), we deconstructed the *ig*^{222} locus in an inducible plasmid by truncating the locus, removing the plasmid-encoded ribosome binding site (RBS), and mutating potential start codons ("*"). We then explored whether these constructs retained the capacity to inhibit AvcD induced filamentation in Δ*ig*^{222} when co-expressed with pAvcD. Mutation of the *ig*^{222} rare CTG start codon to a TAG stop codon (222 nt^{STOP}) did not abrogate the ability of this construct to inhibit AvcD activity in trans when co-expressed in the Δ*ig*^{222} strain (Fig. 2.2G). We then examined a 174 nt ORF completely encoded within *ig*^{222} (174 nt) and found it was also sufficient to prevent AvcD induced filamentation (Fig. 2.2G). Additionally, expression of this 174 nt ORF from constructs either lacking a ribosome binding site (174 nt^{-RBS}) or where the native ATG start codon was mutated to a TAA stop codon (174 nt^{STOP}) each retained the ability to inhibit AvcD activity (Fig. 2.2G). We also identified an ATG start codon on the interior of the 174 nt ORF corresponding to an alternative reading frame and mutation of this interior start codon to a TAA stop codon (174 nt^{InteriorSTOP}) also failed to abrogate AvcD activity. To identify the minimum functional size of inhibitor, we further truncated this 174 nt segment from both the 5' and 3' ends and found that removal of either 18 bp from the 5' end or 4 bp from the 3' end was sufficient to abolish inhibitor's activity (Fig. 2.2G). Collectively, these results

suggest that the AvcD inhibitory component of *ig*^{222} is a novel sRNA, and we refer to the 174 nt locus within *ig*^{222} as AvcI, for **Avc**D **I**nhibitor.

2.5.3: AvcI post-translationally regulates the activity of AvcD and interacts with AvcD in vitro

To determine whether AvcI regulates AvcD at the level of pre- or post-translation, we expressed an *avcD* C-terminally tagged 6x histidine construct (AvcD^{6xHIS}) in WT and Δ*ig*^{222} *V. cholerae* and measured the cellular abundance of AvcD^{6xHIS} using Western blotting. Expression of AvcD^{6xHIS} manifested in a filamentous cell phenotype in Δ*ig*^{222} that was not observed in the WT strain, indicating the tagged protein maintained activity and was negatively regulated by AvcI (Fig. 2.3A). However, the cellular abundance of AvcD^{6xHIS} was not significantly different between these two strains (Figs. 2.3B-C). Additionally, incubation of purified AvcD with in vitro transcribed AvcI RNA revealed robust AvcD-AvcI complex formation (Fig. 2.3D). While AvcD also interacted with the AvcI reverse complement sequence (AvcI-RC), this interaction was not as strong as AvcI (Fig. 2.3E). A negative control protein, RpfR-FI [112] (Fig. 2.3D) did not interact with AvcI. Native gels were used for the binding reactions, which results in different migration between AvcI and AvcI-RC; however, the sizes of AvcI and AvcI-RC were identical as determined using denaturing polyacrylamide gel electrophoresis (PAGE) (Fig. 2.3F). This result is consistent with the formation of a distinct 3-dimensional RNA structure by AvcI but not AvcI-RC (Figs. 2.3D-E). Together, we conclude that AvcI RNA interacts with AvcD to suppress AvcD-dependent cell filamentation although the RNA binding specificity parameters of AvcD require further analysis.

2.5.4: Conservation and evolution of AvcI and AvcD

To identify if AvcD activity and its regulation by AvcI are conserved, we evaluated the activity of three Proteobacterial AvcD homologs from *Vibrio parahaemolyticus, Proteus mirabilis,* and *E. coli* ETEC (Fig. 2.4A). Like *V. cholerae avcD*, all three *avcD* homologs induced filamentation in *E. coli* (Fig. 2.5A). While there was no strong nucleotide sequence similarity for the sequence encoded 5' of each *avcD* homolog (Fig. 2.4B), we hypothesized these sequences also encoded *avcI* homologs. Indeed, co-expression of each AvcI homolog with its corresponding AvcD suppressed cell filamentation (Fig. 2.5A). We then challenged each *avcD* with each *avcI* and found cross-species inhibition only between *V. cholerae* and *V. parahaemolyticus*, while AvcI from *P. mirabilis* and ETEC only inhibited the activity of their native AvcD partner (Figs. 2.5B-C). These data demonstrate that while AvcI inhibition of AvcD activity is conserved across species the specific molecular interactions that mediate this process are not.

Using a selection of proteobacteria AvcD homologs as starting points, we performed a homology search across the tree of life (Methods) to determine the breadth of AvcD's phylogenetic distribution. This analysis revealed a conserved two-domain AvcD core architecture consisting of an N-terminal P-loop NTPase (PLN) [113] and a C-terminal deoxycytidylate deaminase (DCD) with homologs found in all three domains of life (Figs. 2.6A-B, Table 4). In support of functional homology among distant *avcD* homologs, expression of *avcD* from the eukaryotic *Saccharomyces cerevisiae* [114] *dcd1* (~30% similarity) also led to cell filamentation in *E. coli* (Figs. 2.5D, 2.6A).

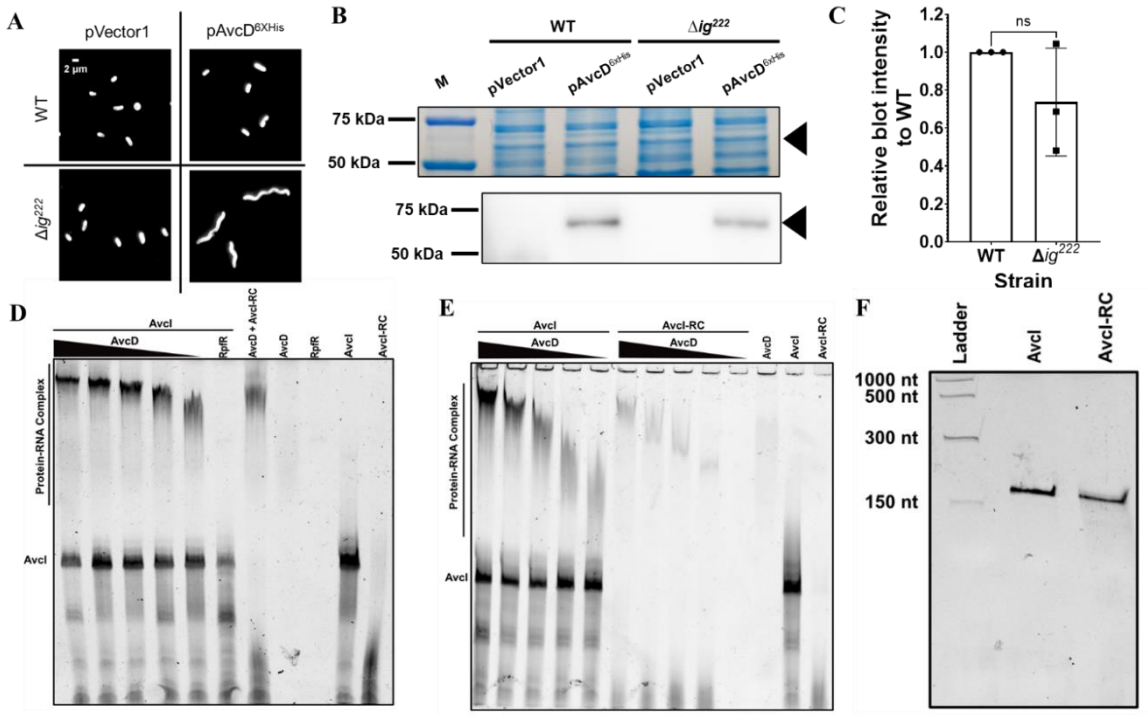


Figure 2.3. The presence of *avcI* does not reduce the abundance of AvcD and AvcI and AvcD forms a complex formation in solution.
(**A**) Representative images of WT *V. cholerae* and Δ*ig*^{222} cultures maintaining an empty vector plasmid (pVector1) or P_{tac}-inducible *avcD-6xHIS* plasmid (pAvcD^{6xHis}) grown in the presence of 100 µM IPTG for 2 h. Cells were stained with FM4-64 prior to imaging and performed in biological triplicate. (**B)** Representative Coomassie stained PAGE gel (top) and matched anti-6x His antibody Western blot (bottom) of whole cell lysates normalized to total protein from *V. cholerae* WT and Δ*ig*^{222} cultures maintaining pVector1 or pAvcD^{6xHis}. Black triangles correspond to AvcD^{6xHis} (60.6 kDa). Analysis was performed in biological triplicate and the relative signal intensity (**C**) was the determined by comparing the intensities of AvcD^{6xHIS} from paired WT and Δ*ig*^{222} lysates probed on the same blots. (**D**) An AvcD-AvcI complex formed in an AvcD concentration-dependent manner as determined by EMSA. (**E**) Trace quantities of non-specific binding of AvcD to the AvcI reverse complement (AvcI-RC) were observed. (**F)** AvcI and AvcI-RC run at essentially equivalent molecular weights on a 7 M urea denaturing PAGE. Low range ssRNA ladder (NEB).

2.5.5: AvcD is a deoxycytidylate deaminase

AvcD is a 532 amino acid polypeptide composed of two putative domains: a P-loop NTPase (PLN) domain and a DCD-like C-terminus. Interrogation of a Phyre2 [72] model of *V. cholerae* AvcD (Fig. 2.7A) revealed conserved catalytic residues in the PLN likely

involved in performing a reversible phosphotransfer of the γ-phosphate from a nucleotide triphosphate donor to a diverse group of substrates [113]. The C-terminal DCD domain contains a highly conserved zinc-dependent cytidine deaminase (CDA) active-site motif [115] ([HAE]X_{28}[PCXXC]) in the DCD domain. The constellation of residues that make up the Zn^{2+} binding pocket is composed of three critical amino acids in $AvcD^{VC}$; H382, C411, and C414. Zn^{2+} is required for the catalytic deprotonation of water by a conserved glutamate residue (E384 in $AvcD^{VC}$) for the hydrolytic deamination of a cytosine base to uridine. Ectopic expression of AvcD active site variants, in either the PLN (S52K, D162A+Q163A) or DCD domain (E384A, C411A+C414A), failed to induce *E. coli* filamentation (Fig. 2.7B) despite being abundantly expressed (Fig. 2.7C), indicating the catalytic activities originating from both the PLN and DCD domains are required for AvcD-induced filamentation.

We performed a genetic screen to identify AvD variants whose activity was no longer inhibited by AvcI by expressing a random library of *avcD* mutants in a Δ*avcD* mutant strain where *avcI* remains intact. Ectopic expression of WT *avcD* in a Δ*avcD* mutant does not induce filamentation (Fig. 2.7D) or produce small, wrinkled colonies on solid agar due to the genomic copy of *avcI* (data not shown). However, AvcD variants that are insensitive to *avcI* exhibit a small colony phenotype. Screening ~ 15,000 potential mutants, we identified five unique *avcD* mutations that encoded single amino acid substitutions (E123K, A126T, K201R, K511E, and Q514R) located in both the PLN and DCD domains that rendered AvcD insensitive to AvcI inhibition (Fig. 2.7D).

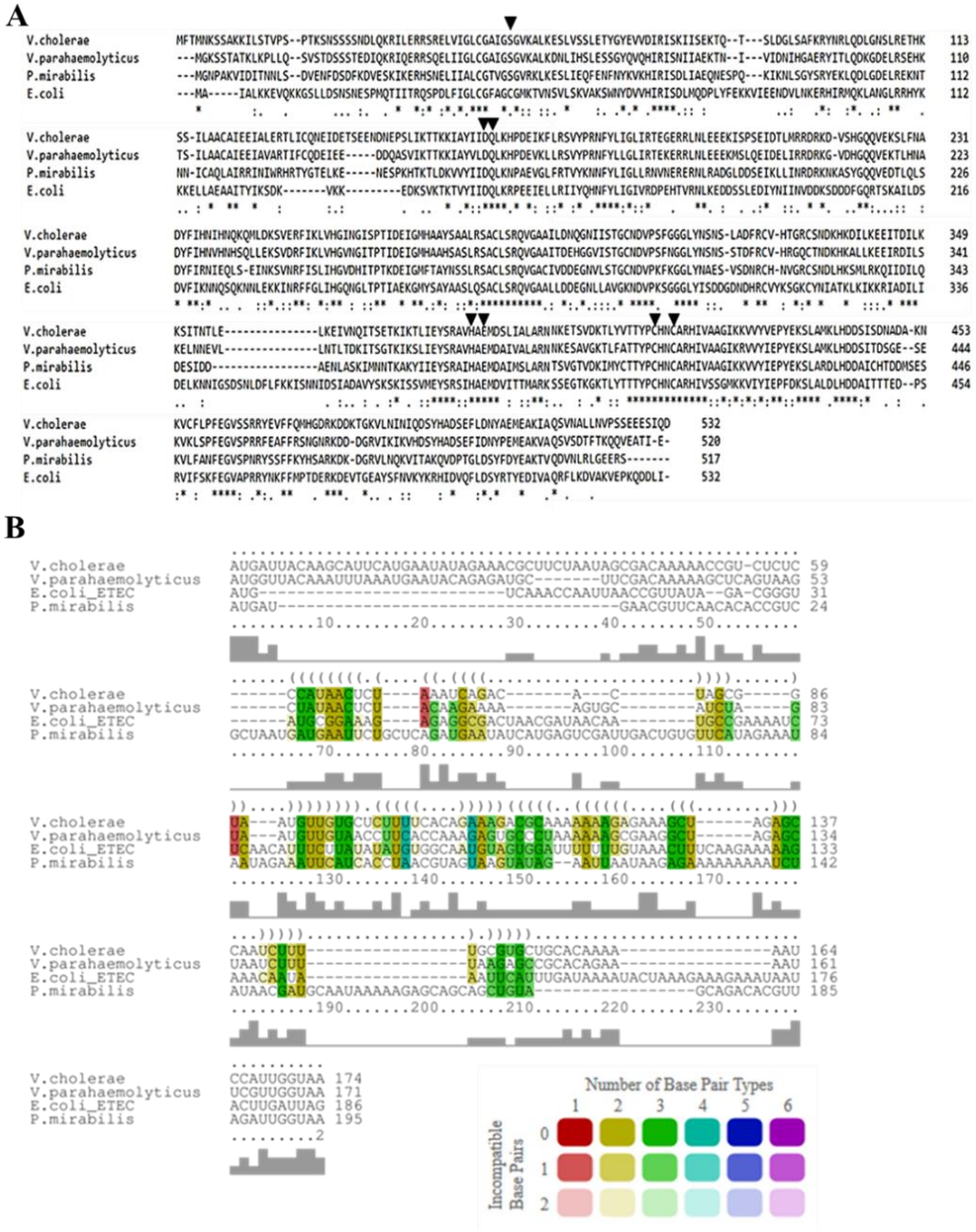


Figure 2.4. Multiple sequence alignment of homologs of AvcD and AvcI. (**A**) Amino acid alignment of the *V. cholerae* AvcD and three homologs using EMBL-EBI ClustalW [73]. "*****" indicates 100% identity, "**:**" indicates >75%, and "**.**" Indicates >50% similarity. Black triangles indicate conserved residues in *V. cholerae* AvcD targeted for site-directed mutagenesis. (**B**) Nucleotide alignment of *V. cholerae* AvcI and three homologs using LocARNA [74]. The average secondary structure is indicated in dot-bracket notation (top). Consensus identities are correlated with the height of the bars below the corresponding nucleotide. Compatible base pairs are colored according to the number of different types C-G (1), G-C (2), A-U (3), U-A (4), G-U (5) or U-G (6) of compatible base pairs in the corresponding columns. The color saturation decreases with the number of incompatible base pairs.

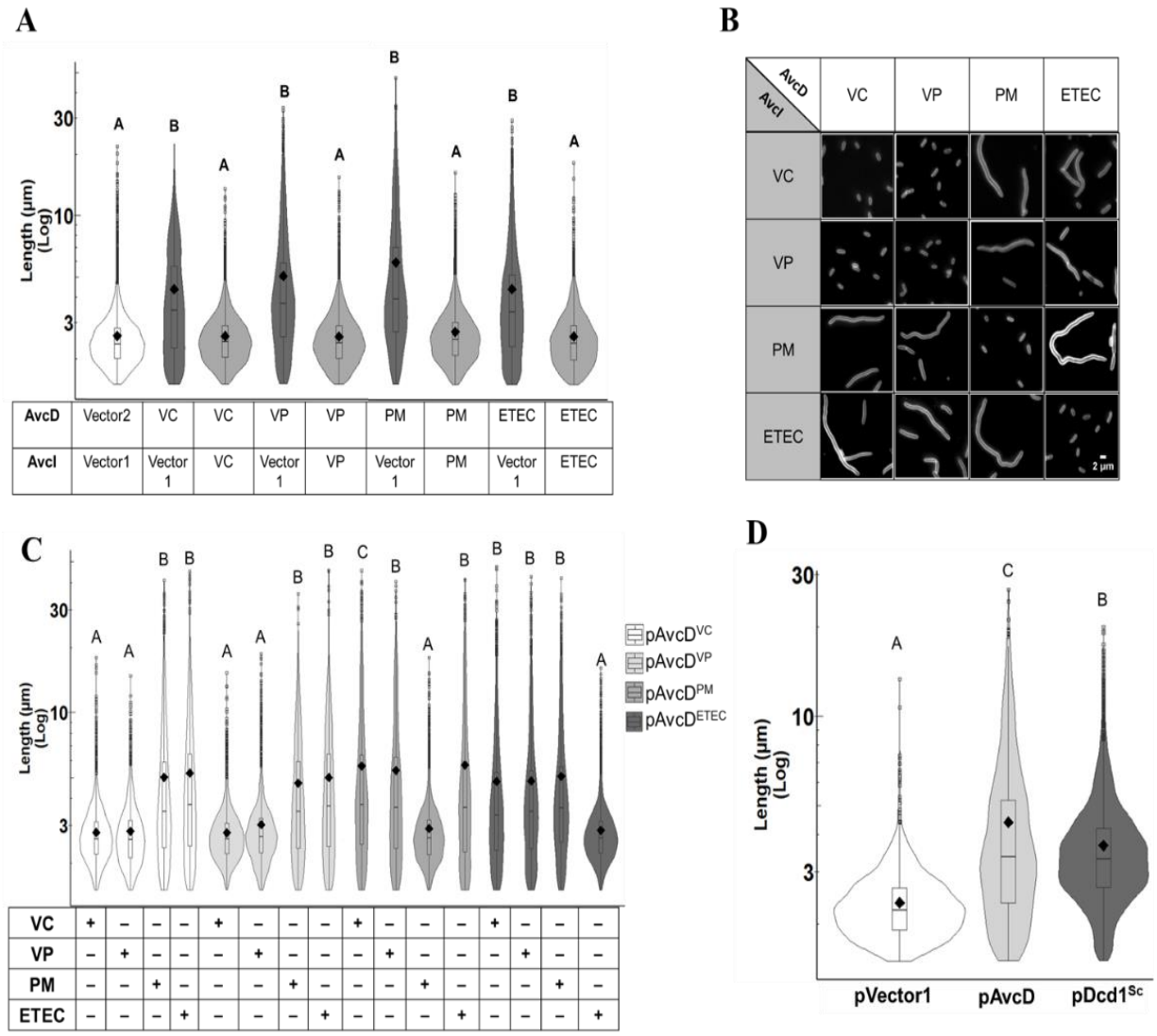


Figure 2.5. Conservation of AvcI and AvcD homologs.
(**A**) Cell length distributions of *E. coli* co-expressing P_{tac}-inducible plasmids encoding *avcD* homologs and their cognate *avcI* or vector controls. (**B**) Representative images of *E. coli* co-expressing various combinations of P_{tac}-inducible plasmids encoding homologs of *avcD* and *avcI*. Scale represents 2 µm. (**C**) Cell length distributions of *E. coli* co-expressing various combinations of P_{tac}-inducible plasmids encoding homologs of *avcD* and *avcI*. (**D**) Cell length distributions of *E. coli* expressing pAvcD, a P_{tac}-inducible plasmid encoding *dcd1* from *S. cerevisiae* (pDcd1^{Sc}), or pVector1. All cell length distributions represent ~1000-3000 cells measured per strain with summary statistics: mean (diamonds), median (horizontal black line), interquartile range (box), and data below and above the interquartile range (vertical lines). Different letters indicate significant differences at $p < 0.05$, according to Tukey's post-hoc test. VC = *Vibrio cholerae*, VP = *Vibrio parahaemolyticus*, PM = *Proteus mirabilis*, ETEC = *E. coli* ETEC.

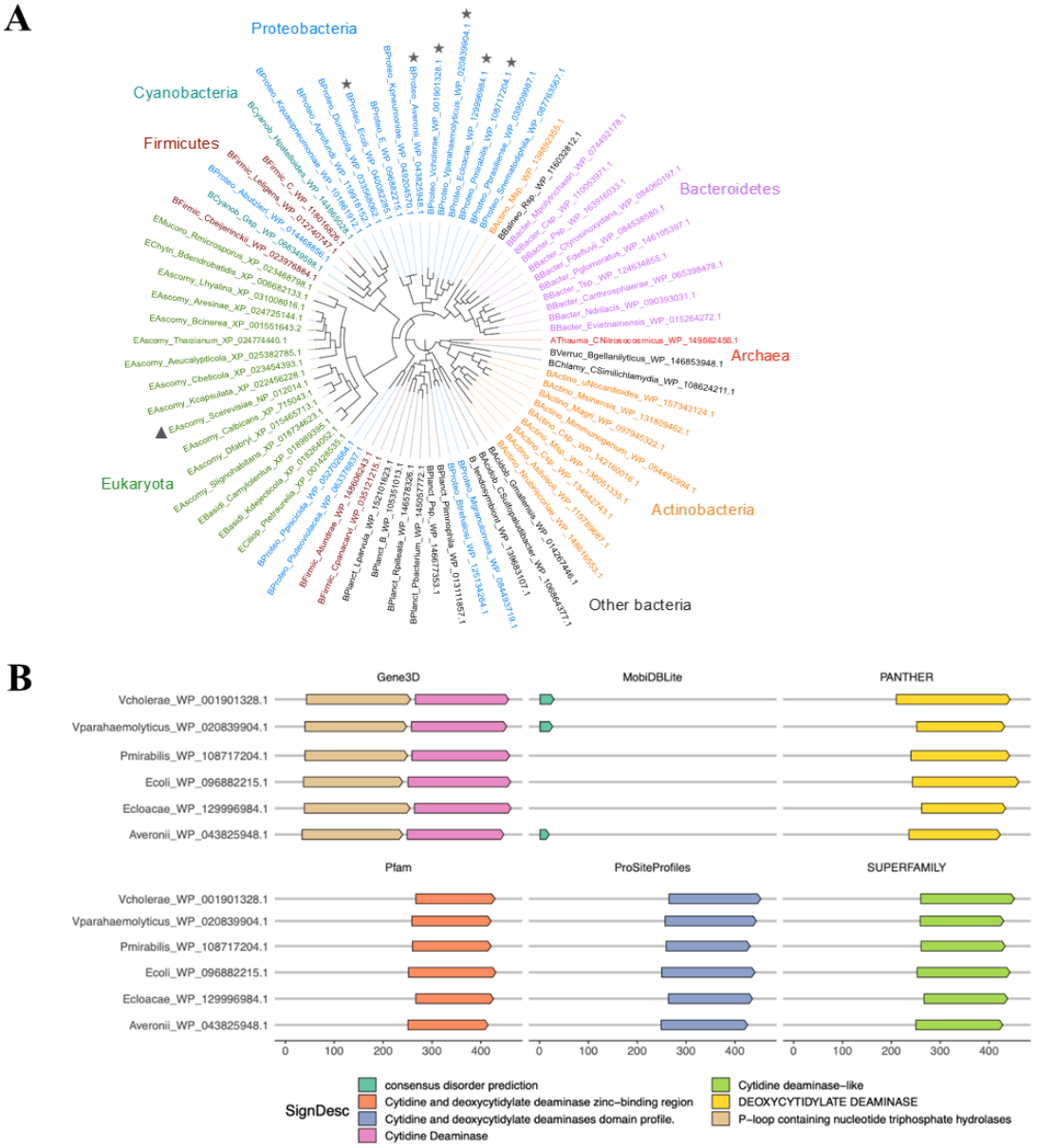


Figure 2.6. Phylogenetic analysis and domain architectures of the six AvcD query proteins.
(**A**) Phylogenetic tree of AvcD homologs from representative phyla across the tree of life. Stars indicate the six proteobacterial starting points for the homology search, as well as the eukaryotic *Saccharomyces cerevisiae dcd1* (triangle), explored in Fig. S9. (**B**) Domain architecture and secondary structure predictions for the six proteobacterial starting points (query proteins) were predicted using InterProScan [79] (Methods). Results from six main analyses are shown here for the query proteins: Gene3D (including CATH structure database), Pfam, ProSiteProfiles, PANTHER, and SUPERFAMILY protein domain profile databases, and MobiDBLite for disorder prediction. No transmembrane regions (using TMHMM) or membrane/extracellular localization were predicted for any of the proteins (using Phobius); hence not shown. Numbers (bottom) indicate the amino acid position of predicted domains and features.

Based on the Phyre2 AvcD structural model, all five residues are located on the exterior of the protein (Fig. 2.7A) suggesting they may be involved in mediating molecular interactions between AvcI and AvcD. The only mutation found within a conserved domain feature was the seemingly innocuous K201R substitution, which is modeled to lie between the two helices of the PLN LID module (Fig. 2.7A).

As the substrates of CDAs are primarily free nucleotides [116], we hypothesized that the AvcD DCD domain would also deaminate free nucleotides. Though attempts to purify active AvcD were unsuccessful, perhaps due to the absence of unknown cellular cofactors, we found soluble lysates from *E. coli* ectopically expressing AvcD specifically deaminated exogenous dCTP and dCMP substrates and no other amine-containing nucleotides tested (Fig. 2.8A). This deamination activity was not observed in the soluble lysates of *E. coli* expressing the $AvcD^{E384A}$ CDA active site variant (Fig. 2.8A). The deaminated products of dCTP and dCMP are dUTP and dUMP, respectively. To further understand the catalytic activity of AvcD we spiked 1 µM dCTP into soluble lysates collected from *E. coli* ectopically expressing either WT AvcD or a vector control and quantified the concentrations of dUTP and dUMP over 30 minutes using UPLC-MS/MS. dUTP was not detected in AvcD-expressing lysates at time 0, suggesting AvcD impacts cellular nucleotide concentrations. While the concentration of dUTP in AvcD containing lysates peaked after five minutes (min) and receded after ten min (Fig. 2.8B) the concentration of dUMP progressively increased to a final concentration of ~ 1 µM (Fig. 2.8C). No appreciable change in either dUTP or dUMP was observed in the vector control lysates for the duration of the experiment (Figs. 2.8B, 2.8C).

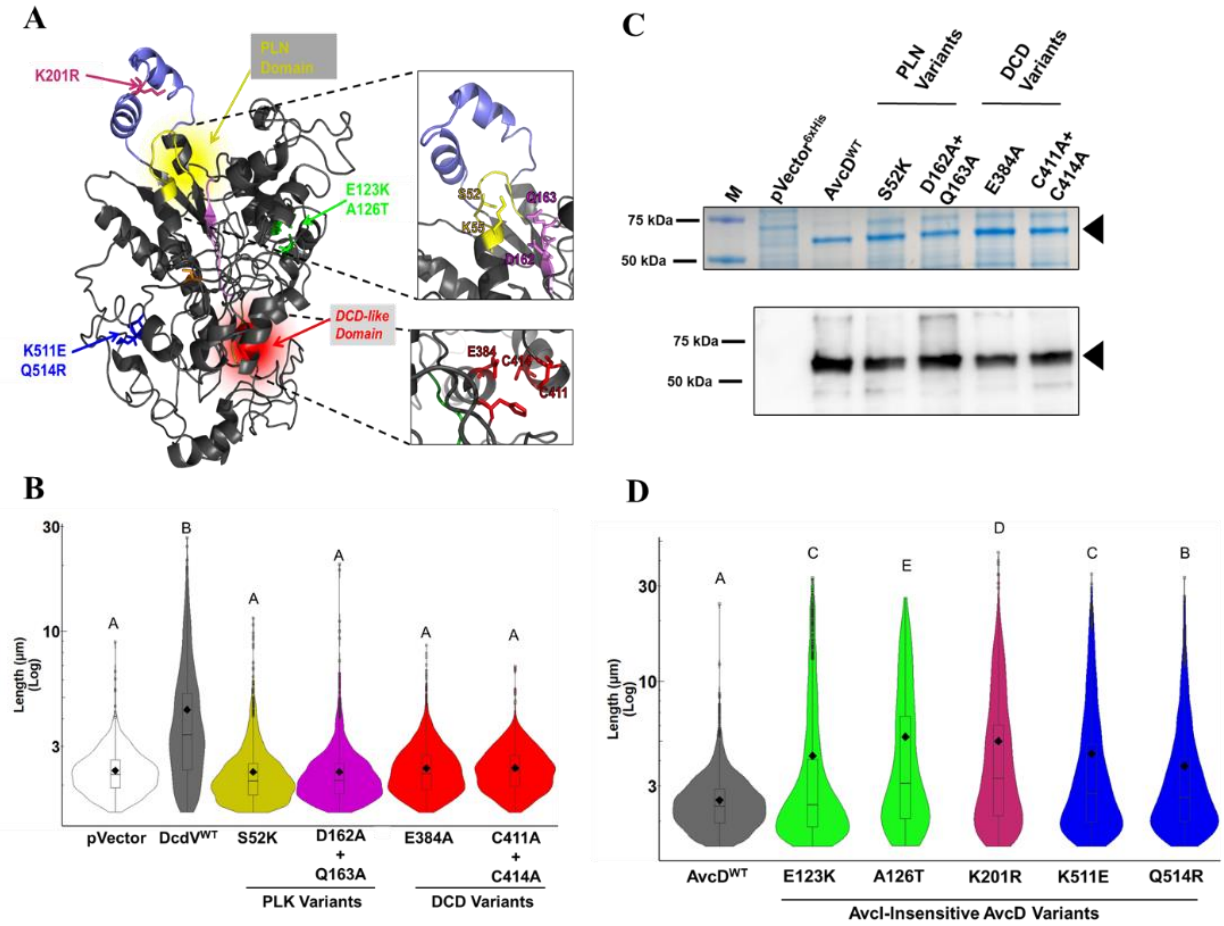


Figure 2.7. Conserved features of the N-terminal PLN domain and C-Terminal DCD domain are required for AvcD induced filamentation in *E. coli*.
(**A**) Phyre2 [72] predicted structure of AvcD from *V. cholerae* El Tor. Insets highlight conserved residues of the PLN (top) and DCD (bottom) domains selected for mutagenesis. (**B**) Cell length distributions of *E. coli* expressing pAvcD, P_{tac}-inducible plasmids encoding a variety of AvcD active-site variants, or pVector1. (**C**) Representative Coomassie stained gel (top) and anti-6x His antibody Western blot (bottom) of whole cell lysates from *E. coli* BL21(DE3) cells maintaining an empty vector ($pVector^{6xHis}$), inducible C-terminal 6x histidine tagged *avcD* (WT) or indicated *avcD* variants. Sample inputs were normalized by culture OD_{600} and resolved by SDS-PAGE. Three biological replicates of each strain were analyzed with similar results. Black triangles correspond to the predicted molecular weight of the AvcD tagged fusions (60.6 kDa). M = molecular weight marker. (**D**) Cell length distributions of Δ*avcD V. cholerae* mutant expressing the indicated AvcD variants. Distributions represent ~1700-3000 cells measured per strain and different letters indicate significant differences at $p < 0.05$, according to Tukey's post-hoc test.

The dUMP that was formed in AvcD lysates was not modified as it had the identical predicated mass of cellular dUMP. Collectively, these lysate experiments

indicate that AvcD deaminates both dCTP and dCMP substrates and cellular lysates containing AvcD ultimately funnels dCTP to dUMP, suggesting that AvcD profoundly impacts nucleotide metabolism.

We quantified the in vivo impact of AvcD activity on the intracellular concentrations of deoxynucleotide species using UPLC-MS/MS following exogenous overexpression of AvcD, $AvcD^{S52K}$, $AvcD^{E384A}$, the ETEC AvcD homolog ($AvcD^{ETEC}$), and a vector control in *E. coli*. While all strains contained similar levels of dATP, dGTP, dTTP, and dUMP, the intracellular abundances of dCTP and dCMP were significantly reduced in *E. coli* expressing AvcD and $AvcD^{ETEC}$ (Fig. 2.8). No dUTP was found following expression of AvcD or $AvcD^{ETEC}$, while dUTP was detected in the vector and two AvcD variant strains (Fig. 2.8D). Unlike the results observed with the in vitro AvcD lysates (Fig. 2.8C), no difference in dUMP was observed between any of the strains when nucleotides were extracted directly from live cells (Fig. 2.8D). We hypothesize this discrepancy is due to compensatory metabolic pathways active in live cells [117] that are lost during the preparation of cell lysates. Importantly, as we describe below, natural activation of AvcD shows increased in vivo accumulation of dUMP, analogous to our results in Fig. 2.8C.

2.5.6: AvcD induced filamentation is due to impaired genome replication

Filamentation is a phenotype often associated with thymineless death (TLD) [56] due to nucleotide starvation. A hallmark of TLD is an increased genomic origin to terminus (*ori*/*ter*) ratio resulting from repeated attempts to initiate replication from the origin that ultimately fail to reach the terminus due to a lack of dTTP substrate [118].

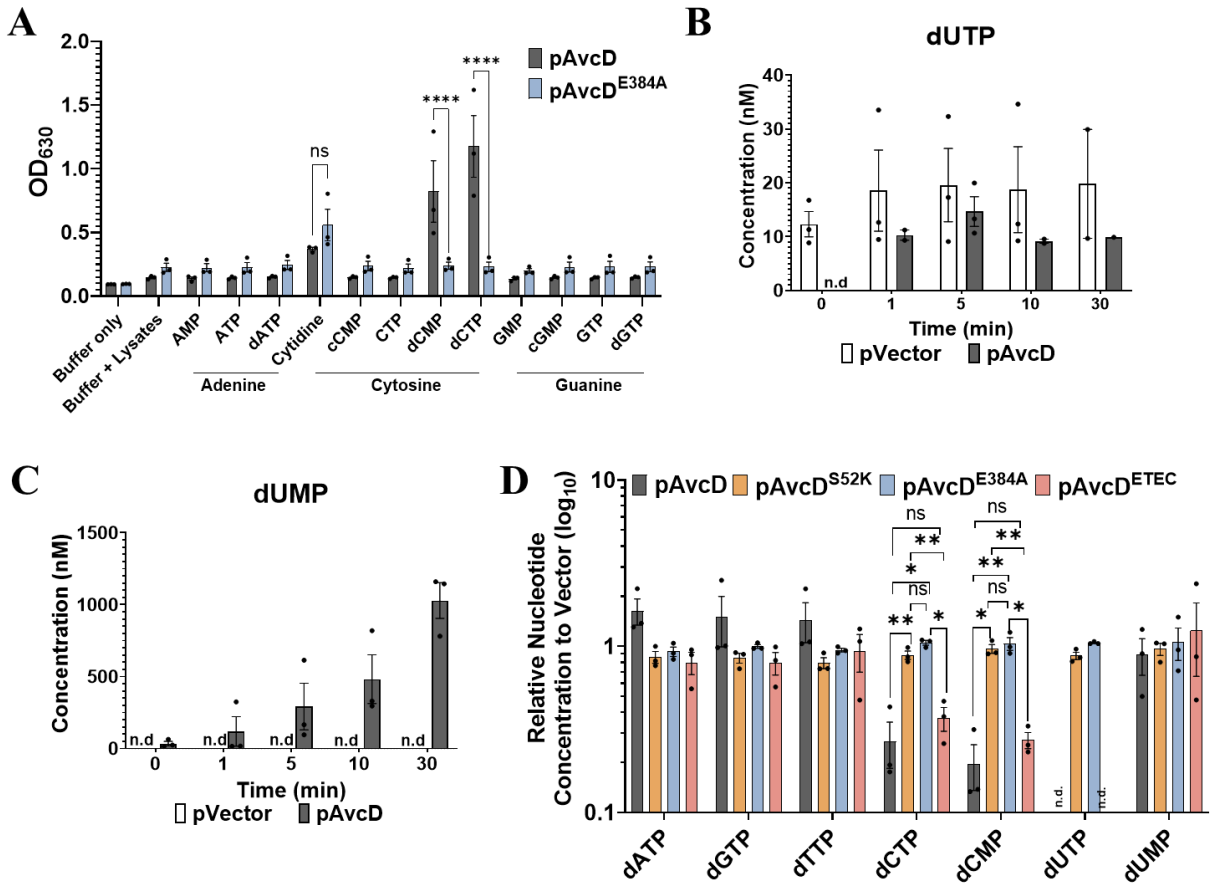


Figure 2.8. AvcD ultimately produces dUMP as final product.
(**A**) Colorimetric assay detecting the evolution of ammonia from lysates of *E. coli,* previously expressing pAvcD or pAvcD^{E384A}, incubated with 12 amine containing nucleotide substrates (37.7 mM cytidine and 7.5 mM for all other substrates) for 30 min. Data represent the mean ± SEM, *n*=3, Two-way ANOVA with Šídák's multiple-comparison test, and ns indicates not significant. Quantification of dUTP (**B**) and dUMP (**C**) using UPLC-MS/MS, in the indicated cell lysates before and after addition of 1 μM dCTP. Each bar represents mean ± SEM, *n*=3. (**D**) In vivo nucleotide concentrations of *E. coli* expressing pAvcD, AvcD active site variants (pAvcD^{S52K} pAvcD^{E384A}), or an AvcD homolog (pAvcD^{ETEC}) for 1 h measured by UPLC-MS/MS and normalized to a vector control strain. Data are graphed as mean ± SEM, *n*= 3, Two-way ANOVA with Tukey's multiple-comparison test. n.d. indicates "none detected" and ns indicates "not significant".

Hypothesizing that AvcD induced filamentation may be a consequence of replication inefficiency, analogous to TLD, we determined the Δ*ig*^{222} *V. cholerae* mutant expressing AvcD forms filaments (Fig. 2.2F, 2.9B) and has a ~3-fold higher *ori:ter* ratio than the

vector control and the AvcD variants (Fig. 2.9A), consistent with AvcD overexpression in the absence of AvcI. While the TLD-like phenotypes associated with AvcD are likely artifacts of overexpression, they provide further evidence that active AvcD alters cellular nucleotide pools. Notably, Δ*ig*^{222} *V. cholerae* does not filament without ectopic AvcD expression (Fig. 2.9B), likely due to polar effects on *avcD* expression, which is reduced relative to WT *V. cholerae* (Fig. 2.9C).

2.5.7: AvcI and AvcD constitute a toxin-antitoxin system

The genomic orientation and proximity of *avcI* to *avcD* suggest they may constitute an operon and two previous genome-wide transcriptional start site (TSS) analyses identified a common putative TSS 5' of *avcI* [119, 120]. To test if *avcD* and *avcI* are in an operon, we performed diagnostic PCR with primers located within *avcI* and *avcD* on cDNA generated from both WT and Δ*ig*^{222} RNA (Fig. 2.10A). The presence of an 839 nt product using primers spanning *avcI* to *avcD* ("ad") only amplified using the WT cDNA template confirmed that both genes are present on a shared transcript (Fig. 2.10B). Additionally, we quantified the relative abundance of *avcI* and *avcD* RNA using qRT-PCR and found the *avcI* locus was at least 18-fold more abundant than *avcD* at all growth phases (Fig. 2.10C).

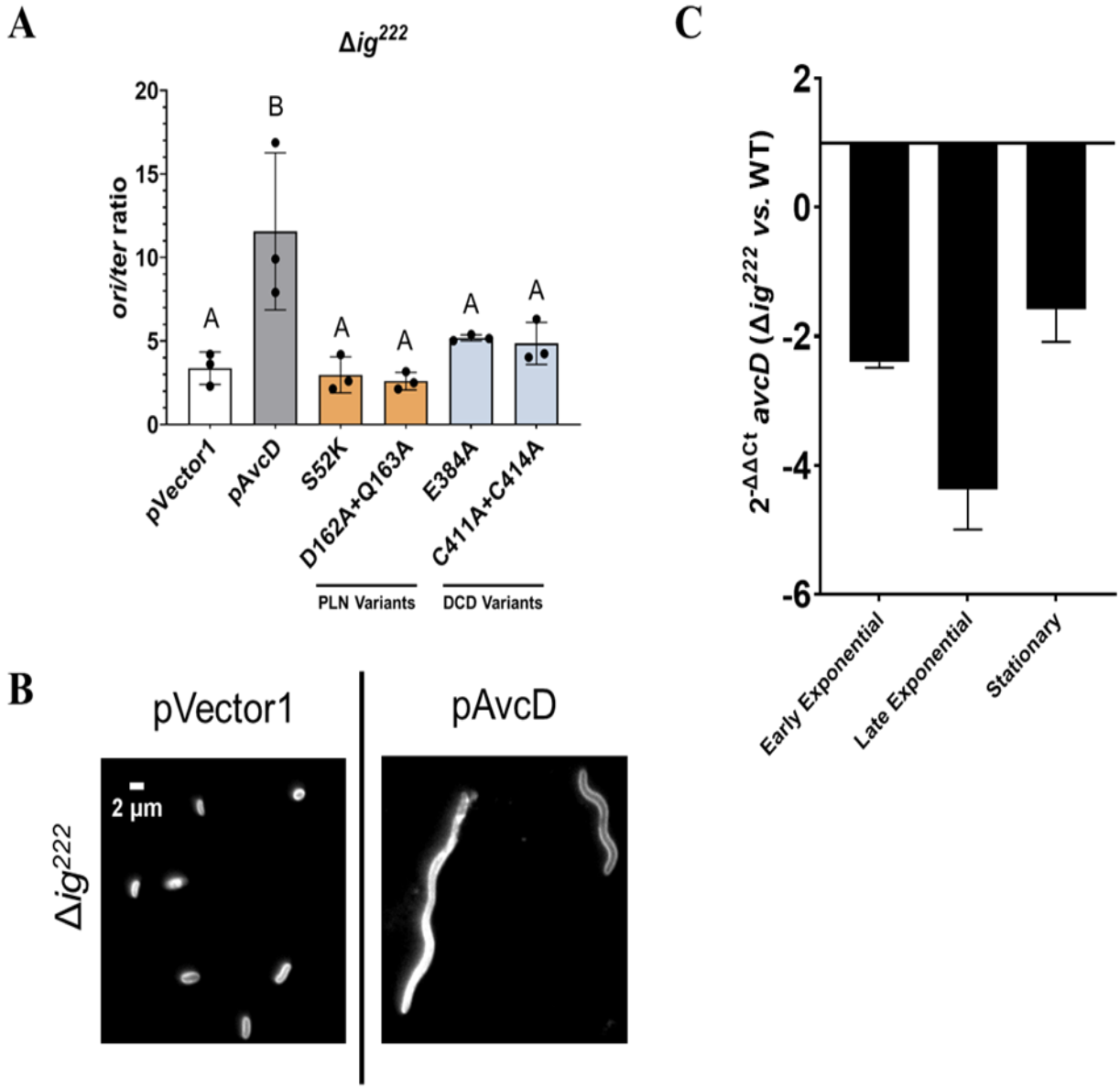


Figure 2.9. AvcD activity induces TLD-phenotype.
(**A)** *V. cholerae* mutant expressing the indicated AvcD variants. *ori/ter* ratios of Chromosome 1 in Δ*ig*^{222} *V. cholerae* strains expressing the indicated pAvcD construct and quantified using qRT-PCR. Each bar represents the mean ± SEM, *n*=3. Different letters indicate significant differences (*n*=3) at $p < 0.05$, according to Tukey's post-hoc test. (**B**) Representative images of Δ*ig*^{222} cultures maintaining an empty vector plasmid pVector1 or pAvcD. Cells were stained with FM4-64 prior to imaging. (**C**) Relative difference in *avcD* expression between Δ*ig*^{222} and WT *V. cholerae* at three different growth phases using qRT-PCR and an endogenous *gyrA* control. Data represent the mean ± SEM of three biological replicates.

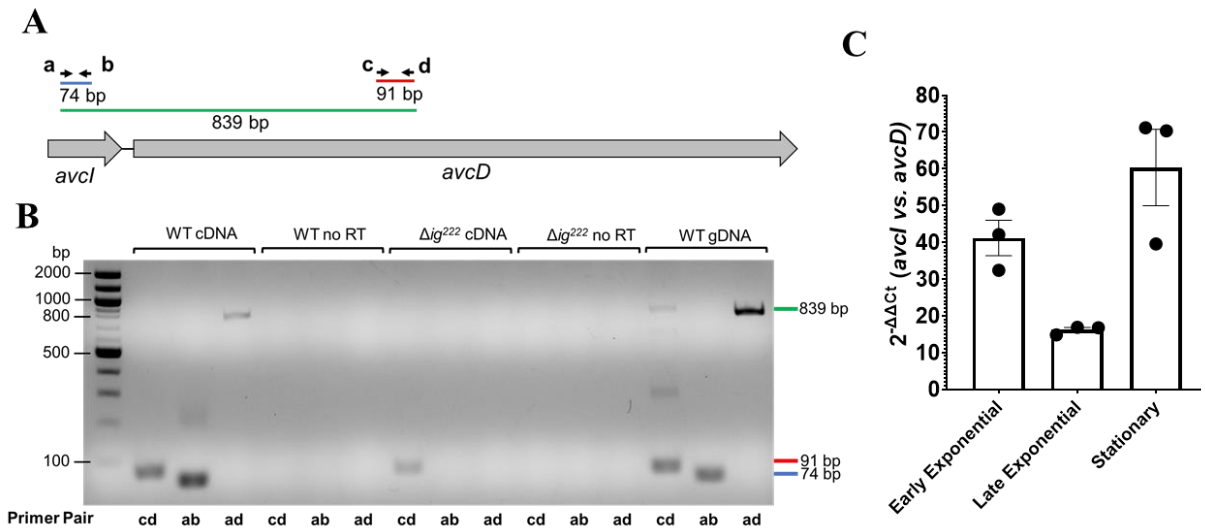


Figure 2.10. *avcI* and *avcD* are in an operon and *avcI* expression is higher than that of *avcD* in all growth phases.
(**A**) To scale genomic diagram of *avcI* and *avcD* and the primers (a, b, c, and d) used for generating diagnostic PCR products. (**B**) PCR products amplified from nucleic acid templates (above) using the indicated primer pairs (below) resolved in a 1% agarose gel. All reactions were performed in duplicate using biologically independent samples with similar results. No RT = non-reverse transcribed RNA template control. gDNA = genomic DNA control (**C**) Relative difference in transcript abundance between *avcI* and *avcD* at different growth phases in WT *V. cholerae* normalized to an endogenous *gyrA* control using qRT-PCR. Data are graphed as mean ± SEM, *n*= 3.

The co-transcription of *avcI* and *avcD* and the post-translational regulation of AvcD activity by the abundant sRNA AvcI resembles Type III Toxin-Antitoxin (TA) systems [121, 122]. In the case of the Type III TA system *toxIN,* cessation of transcription depletes the abundance of the labile sRNA antitoxin ToxI, thus liberating the ToxN endoribonuclease toxin [122]. Hypothesizing that cessation of transcription may likewise lead to activation of AvcD, we treated WT and Δ*avcD V. cholerae* with either rifampicin or spectinomycin to block transcription or translation, respectively, and measured the intracellular abundance of dCTP, dCMP, dUTP, and dUMP over time. Indeed, rifampicin treatment led to rapid and significant decreases in dCTP and dCMP and an increase in dUMP in WT *V. cholerae* but not the Δ*avcD* mutant (Figs. 2.11A-D).

Spectinomycin did not differentially alter the abundance of these nucleotides in either strain (Figs. 2.11E-H). Notably, no significant change in dUTP was observed between the strains in either condition. These data indicate that cessation of transcription, and not translation, leads to AvcD activation and demonstrate AvcI is an sRNA antitoxin.

2.5.8: cGAMP does not activate the AvcID TA system

We initiated studies of *avcD* based on our discovery that this gene frequently co-occurs in bacterial genomes with the cGAMP synthase *dncV* [24] (Fig. 2.1A), a critical member of the CBASS antiphage system [6]. We hypothesized this co-occurrence was indicative of cGAMP allosteric activation of AvcD, analogous to the activation of the CBASS-effector CapV [25]. However, co-expression of both DncV and AvcD in a Δ*capV V. cholerae*, a strain that encodes *avcI* but can no longer induce CBASS-dependent autolysis [7, 25] via CapV, did not produce filamentous cells (Fig. 2.12). This demonstrates that cGAMP does not activate AvcD. This conclusion also relies on the observation that AvcD overexpression in the double VSP islands knock out is also filamentous.

2.5.9: AvcID and homologs provide phage defense by disrupting nucleotide levels and inhibiting phage replication

Cytidine deaminases are conserved anti-viral defense mechanisms in eukaryotes [39, 47, 123] and several TA systems have been implicated in phage defense [122, 124–126]. We therefore hypothesized that *avcID* may constitute a new antiphage defense mechanism, and its association with CBASS was the result of independent

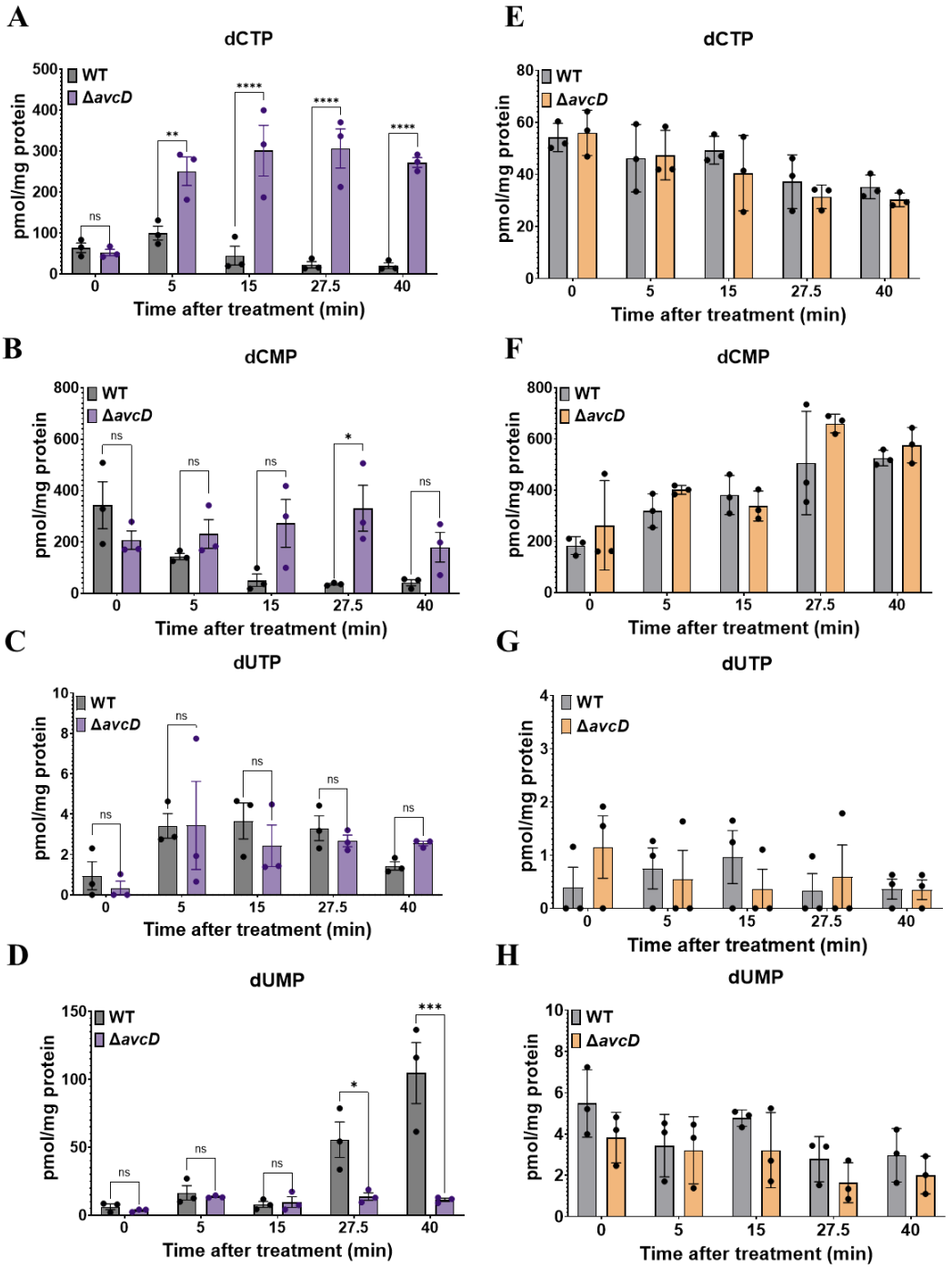


Figure 2.11 Cessation of global transcription not translation liberate AvcD enzymatic activity.
Intracellular concentration of dCTP (**A, E**), dCMP (**B, F**), dUTP (**C, G**), and dUMP (**D, H**) of WT and Δ*avcD V. cholerae* during rifampicin (purple) (250 µg/mL) or spectinomycin (orange) treatment (200 µg/mL) measured by UPLC-MS/MS and normalized to total protein. Data represent the mean ± SEM of three biological replicate cultures, Two-way ANOVA with Šídák's multiple-comparison test, and ns indicates not significant.

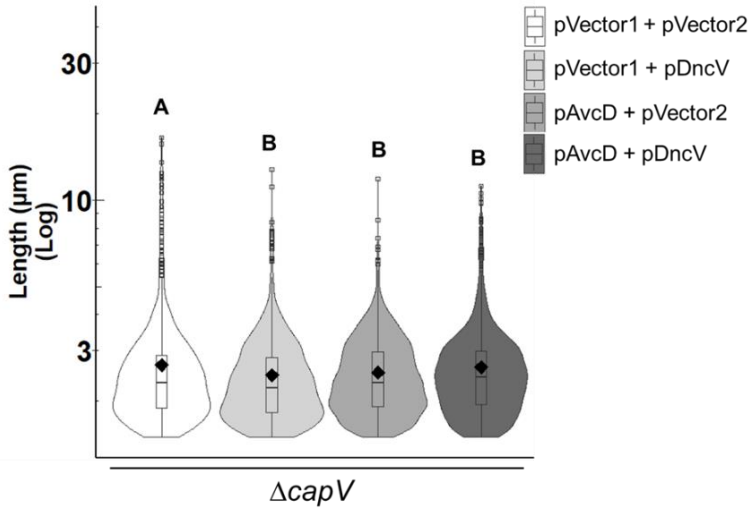


Figure 2.12. Ectopic expression of DncV and AvcD does not lead to filamentation in the Δ*capV* mutant of *V. cholerae.*
Cell length distributions measured from three biological replicates of *ΔcapV V. cholerae* cultures co-expressing either two empty vectors, pDncV and an empty vector, pAvcD and an empty vector, or pDncV and pAvcD. Distributions represent ~1200-1700 cells measured per strain. Different letters indicate significant differences at $p < 0.05$, according to Tukey's post-hoc test.

phage defense mechanisms clustering in mobile genetic elements called defense islands [7].

We showed that each *avcID* operon conferred its naïve host with resistance to at least one of ten lytic coliphage tested (Figs. 2.13A-B). As the *avcID* operons from *V. parahaemolyticus* (pAvcID^{VP}) and *E. coli* ETEC (pAvcID^{ETEC}) conferred robust protection against T3 phage, we infected *E. coli* maintaining these orthologous systems with T3 phage (MOI of 5) and measured the intracellular abundance of dCTP, dCMP, dUTP, and dUMP over time. In the presence of each *avcID* ortholog, T3 infections significantly decreased intracellular dCTP and dCMP and increased dUMP, while the dUTP level was not changed (Figs. 2.13C-F). Consistent with depletion of dCTP and dCMP decreasing phage replication, T5 phage infecting *E. coli* containing the *V. parahaemolyticus* WT *avcID* operon (pAvcID^{VP}) contained fewer phage genomes

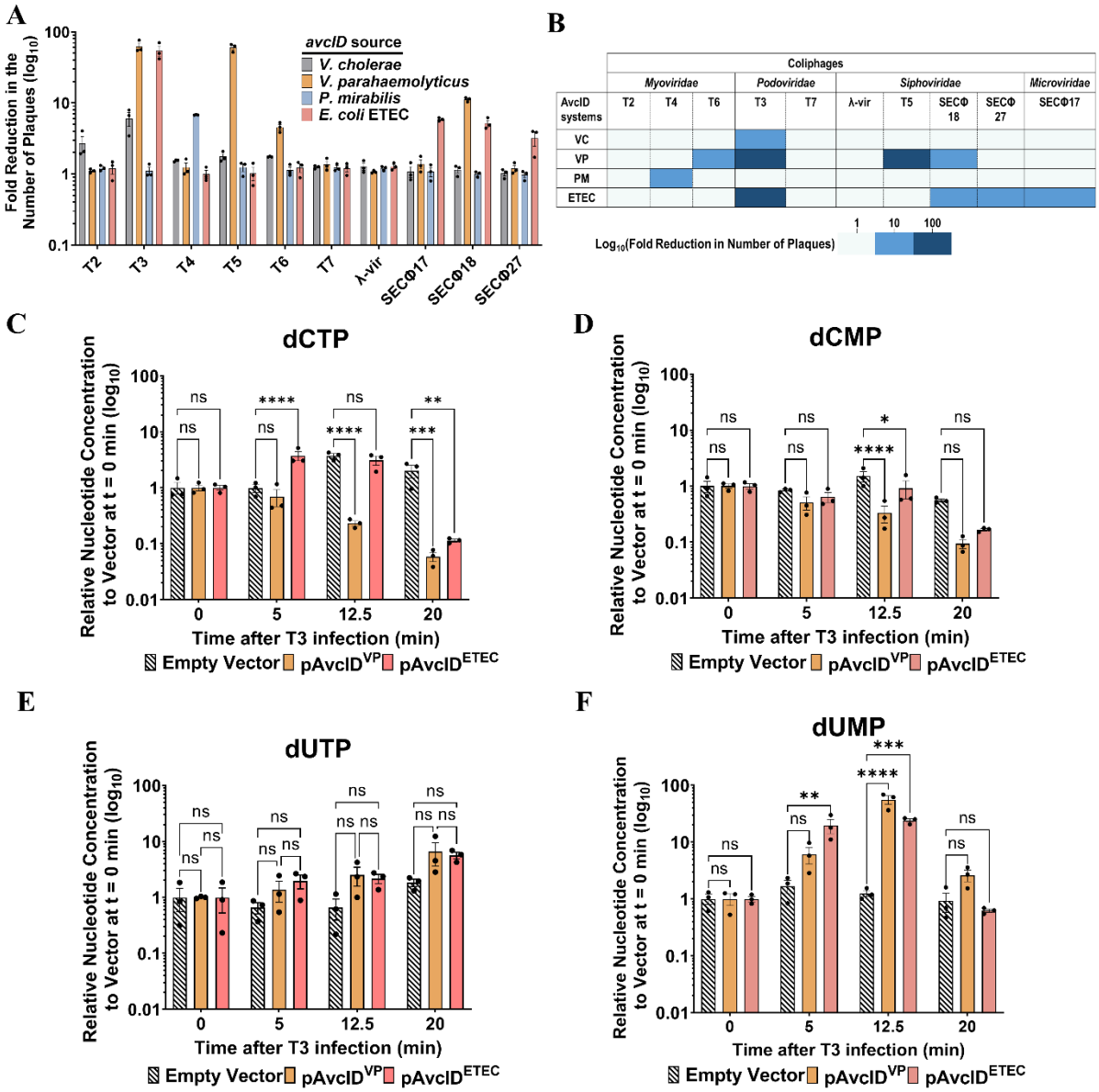


Figure 2.13. *avcID* and homologs provide phage defense against several lytic coliphages to an *E. coli* host by disrupting nucleotides level.
(**A**) Fold plaque reduction in the number of plaques conferred by four homologous *avcID* systems to a naïve *E. coli* host challenged with a panel of coliphages. Fold reduction determined by serial dilution plaque assays comparing the efficiency of plaquing on an *E. coli* host maintaining a plasmid borne *avcID* system and its native promotor against a vector control. (**B**) Summary table of (**A**). Data are the mean of the three biological replicates rounded to the nearest fold of plaque reduction relative to empty vector control. VC = *Vibrio cholerae*, VP = *Vibrio parahaemolyticus*, PM = *Proteus mirabilis*, ETEC = *E. coli* ETEC. In vivo abundance of dCTP (**C**), dCMP (**D**), dUTP (**E**), and dUMP (**F**) in an *E. coli* host carrying a vector control, pAvcID^{VP} or pAvcID^{ETEC} before and after T3 phage infection (MOI = 5). Data represent the mean ± SEM of three biological replicate cultures, Two-way ANOVA with Dunnett's post-hoc test, and ns indicates not significant.

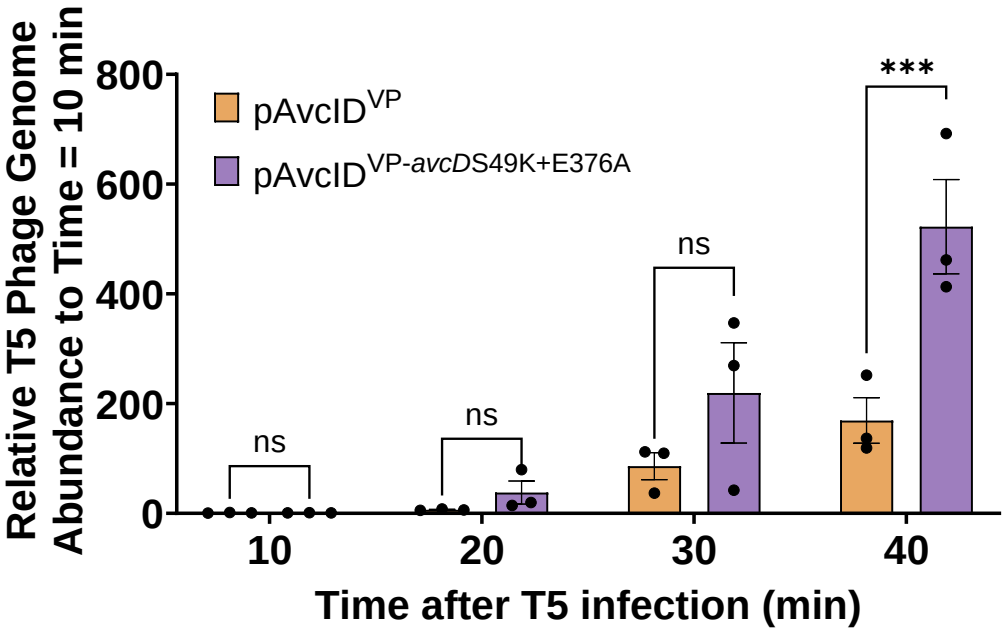


Figure 2.14. AvcID reduces phage infection replication efficiency.
Relative T5 genome abundance comparing *E. coli* expressing pAvcID^{VP} or its double point mutation variants pAvcID$^{VP\text{-}avcD\text{S49K+E376A}}$ over time. Data represents the mean ± SEM of three biological replicate cultures, Two-way ANOVA with Šídák's multiple-comparison test and ns indicates not significant.

relative to a strain containing a double active site mutant allele of *avcD*^{VP} (pAvcID$^{VP\text{-}avcD\text{S49K+E376A}}$) over the course of 40 minutes (Fig. 2.14).

2.5.10: Phage defense conferred by the AvcID system requires AvcD activity

To assess whether phage defense requires the activity of the PLN, DCD, or both domains, we generated inactive variants of AvcD derived from *V. parahaemolyticus* (pAvcID^{VP}) and assessed phage defense for T3, T5, and T6, for which the WT variant demonstrated significant protection (Fig. 2.15A). In virtually all cases, mutation of either the PLN or DCD domains, or both, abolished the AvcD-dependent protection, suggesting both domains are necessary for phage defense (Fig. 2.15A). For reasons

that are not currently understood, one exception to this finding was the E376A DCD mutant which exhibited significant, but reduced, protection against T5 phage infection, suggesting the PLN domain may have anti-phage activity on its own against certain phage.

To evaluate the dependence of changes in nucleotide levels following phage infection on the PLN and DCD domains of AvcD, *E. coli* carrying pAvcID^{VP} or the double domain active site variant pAvcID$^{VP\text{-}avcD\text{S49K+E376A}}$ were either infected with T3 or SECΦ18 or treated with a rifampicin control. In response to infection with either T3 or SECΦ18 phage, or treatment with rifampicin, pAvcID^{VP} depleted dCTP and dCMP while increasing dUMP (Figs. 2.15B-C, E). In contrast, pAvcID$^{VP\text{-}avcD\text{S49K+E376A}}$ did not significantly decrease intracellular dCTP or dCMP nor increase dUMP (Figs. 2.15B-C, E). Consistent with earlier observations (Figs. 2.11C, 2.12E) dUTP was not significantly changed in these conditions (Fig. 2.15D).

2.6: DISUCSSION

Phage defense mechanisms are often found clustered together in mobile genetic elements called defense islands [6, 9] and we speculate that the co-occurrence of AvcD and DncV in bacterial genomes is a result of their shared anti-phage activity. Uncovering the contributions to bacterial fitness of the ~36 genes encoded within the El Tor *V. cholerae* VSP-1 and 2 genomic islands may help elucidate the longevity and persistence of the seventh cholera pandemic. Our bioinformatic approach using GeneCoOccurrence accurately identified a gene network composed of the VSP-1

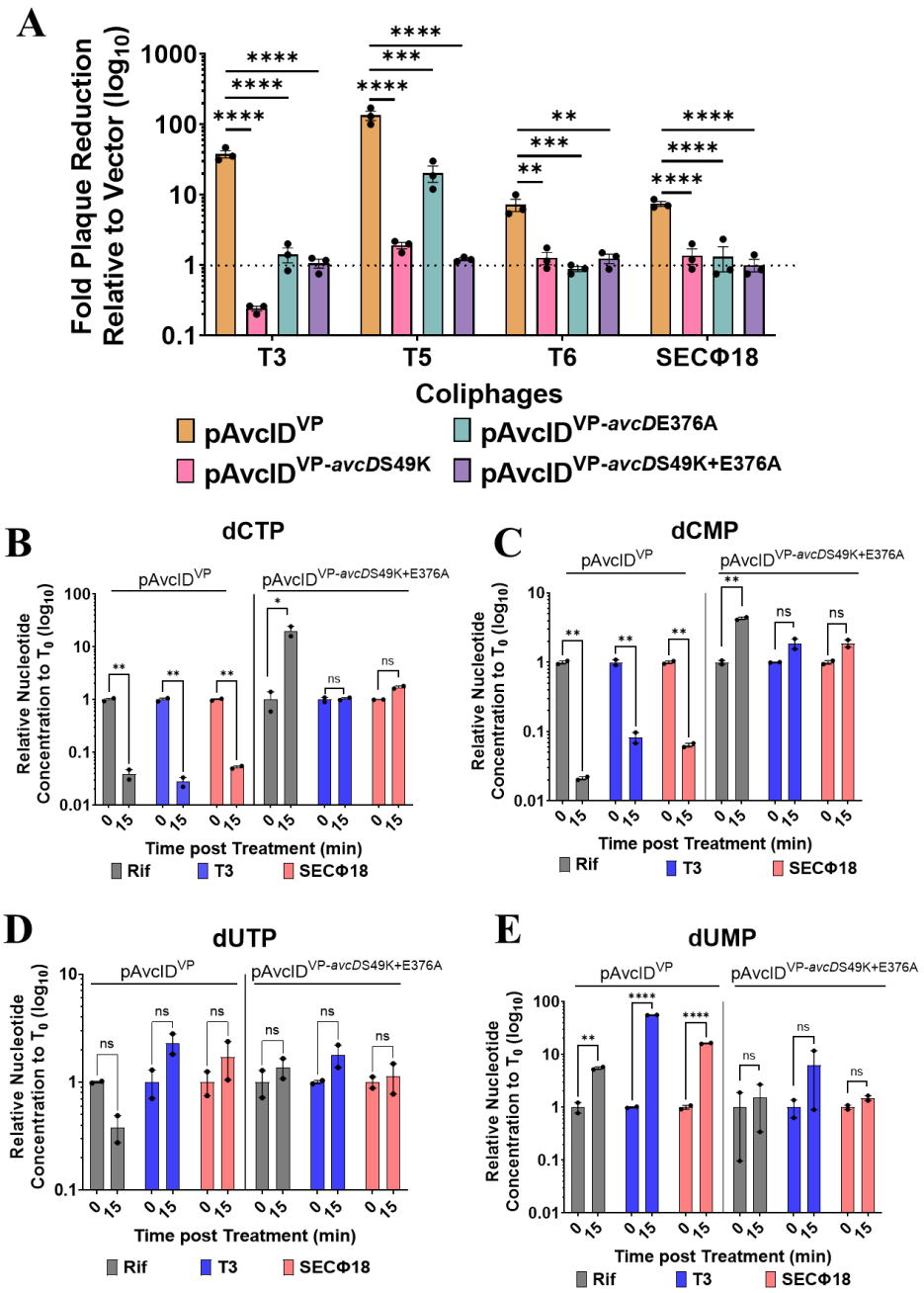


Figure 2.15. Nucleotide pool disruptions during phage infection are dependent on AvcD activity.

(**A**) Efficiency of plaquing on strains encoding WT *avcID*^{VP} from *V. parahaemolyticus* (pAvcID^{VP}) or point mutations in the PLN (S49K) and DCD (E376A) against a vector control. In vivo abundance of dCTP (**B**), dCMP (**C**), dUTP (**D**), and dUMP (**E**) of an *E. coli* host carrying vector control, the *avcID* system from *V. parahaemolyticus* with its native promoter (pAvcID^{VP}), or an inactive *avcD* mutant (pAvcID$^{VP\text{-}avcD\text{S49K+E376A}}$). Nucleotides were measured using UPLC-MS/MS, normalized to total protein. Data represent the mean ± SEM of two biological replicate cultures, Two-way ANOVA with Šídák's multiple-comparison test, and ns indicates not significant.

antiphage CBASS system (*capV-dncV-vc0180-vc0181*). Interestingly, this analysis also revealed *dncV* is frequently found in genomes with the previously uncharacterized gene *avcD.* The only function previously ascribed to *avcD* was an undefined involvement in quorumsensing (QS) controlled *V. cholerae* aggregate formation [127].

We showed that AvcD activity is post-translationally regulated by AvcI, a sRNA encoded immediately 5' of the *avcD* locus in all bacterial homologs. AvcD contains a functional DCD domain that catalyzes the deamination of deoxycytidine nucleotides and a putative PLN-like domain of unknown function. Both domains are required for AvcD to disrupt deoxynucleotide pool homeostasis, which impairs DNA replication and manifests in a filamentous cell morphology. Cell filamentation is a hallmark of TLD, observed in bacteria and eukaryotes, which arises from a sudden loss of thymine during robust cellular growth [59]. Interestingly, this phenomenon is not limited to dTTP as dGTP starvation elicits a similar response in *E. coli* and is also hypothesized to occur when other deoxynucleotide substrates become disproportionately scarce [57]. In the case of AvcD, it is conceivable the observed filamentation phenotype is a consequence of a TLD-like reduction in dCTP pools that can be termed '*cytosineless death.'* However, while AvcD activity also reduces the intracellular dC pool, overexpression of AvcD did not significantly increase the intracellular concentrations of dTTP or dUMP in vivo, suggesting a cellular compensatory pathway to combat AvcD activity is at work in growing cells. Yet, the intracellular level of dUMP increases when AvcD is activated upon transcriptional shutoff. Moreover, we speculate that the DCD and PLN domains of AvcD are responsible for this conversion of dC nucleotides to dUMP observed in the bacterial lysates, but we cannot rule out a contribution of other unknown cellular factors.

The deamination of dCTP is canonically performed by non-zinc dependent enzymes [128], making the dual substrate repertoire of dCMP and dCTP in AvcD a rare trait.

Together, our results suggests that AvcD depletes available dCMP and dCTP to protect against bacteriophage predation, and AvcD homologs are widely distributed across the tree of life. Such conservation extends to eukaryotes, and we show similar activity of the *S. cerevisia*e AvcD homolog, although the requirement of the PLN and DCD domains for activity of this enzyme remains to be tested. Likewise, Tal et. al., independently discovered many of the conclusions we describe here [129]. We speculate that AvcD becomes liberated from AvcI upon phage infection, either through cessation of transcription or degradation of AvcI (Fig. 2.16), although the molecular mechanism by which AvcD is activated remains to be formally determined. Nevertheless, in support of our model, inhibition of host transcription by infecting phage reduces the levels of ToxI sRNA antitoxin, activating the ToxN endonuclease to prevent phage production [122]. We hypothesize that the AvcD-dependent depletion of dCTP and dCMP starves infecting phage of the nucleotides necessary for genome replication. However, AvcD could impact phage infection in other ways through modification of cellular nucleotides such as disrupting phage replication or transcription and potentially decreasing phage burst size. We are currently exploring these possibilities. Purification of active AvcD can also determine if this enzyme targets other nucleotide species such as double-stranded or single-stranded DNA or RNA. In contrast to all previously described Type III TA systems [130], the AvcID system is unique in two ways: its utilization of a cytidine deaminase toxin and a RNA antitoxin that does not encode

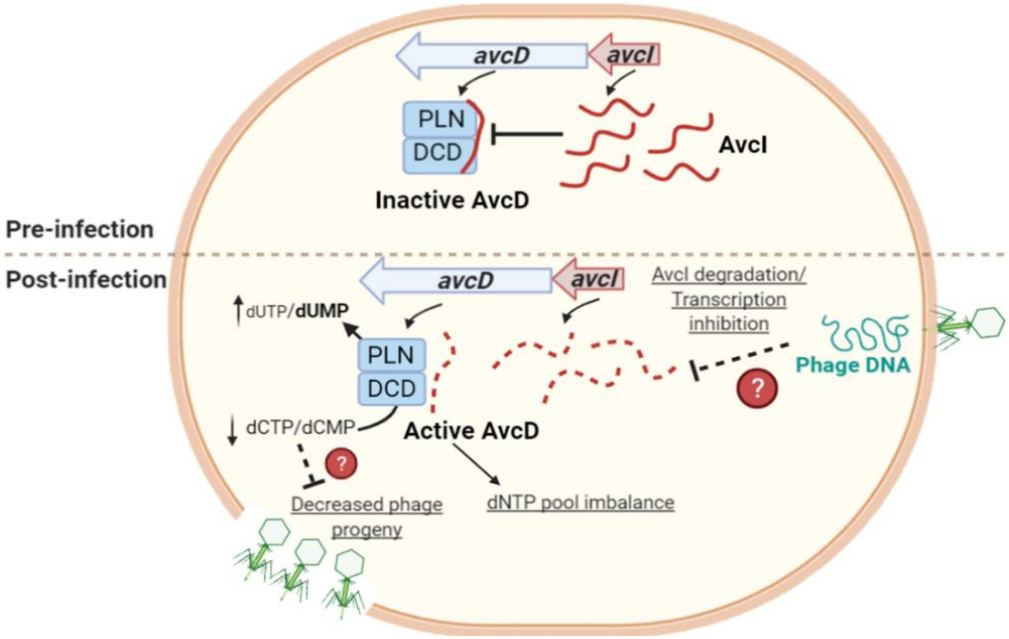


Figure 2.16. Model for AvcID-based antiphage activity in bacteria.
Top: Prior to infection, AvcD is maintained in an inactive state by the abundant sRNA, AvcI. *Bottom*: Following infection, AvcD is liberated from AvcI, likely by the cessation of global transcription or the enhanced degradation of AvcI. Active AvcD rapidly depletes available dCMP and dCTP substrates promoting the accumulation of dUMP, via deamination, which likely impairs efficient phage DNA replication and new phage virion production.

nucleotide repeats. Therefore, we propose that AvcID is a new subclass of the Type III TA family.

Exhausting deoxynucleotide pools to combat viral infection is a strategy AvcD shares with eukaryotic SAMHD1 [131–133]. Additionally, AvcD shares functional similarity with the eukaryotic anti-viral APOBEC3 [41] enzymes, which do not utilize free deoxynucleotide substrates, but rather deaminate deoxycytidines in the minus-strand DNA of retroviruses leading to viral genome instability. In a unique way, AvcD utilizes deamination of dCTP and dCMP substrates to deplete their intracellular abundance, demonstrating the biologically universal utility of draining free deoxynucleotides as an anti-viral strategy.

Chapter 3 – Transcriptional Shutoff Activates AvcID to Inhibit Phage Replication and Stability

3.1: ABSTRACT

Toxin-antitoxin (TA) systems are ubiquitous two-gene loci modules that bacteria use to regulate cellular process, such as phage defense. Here, we characterize a novel type III TA system, *avcID*, that protects *E. coli* against several bacteriophages by depleting available nucleotides. The toxin of the system (AvcD) is a deoxycytidylate deaminase that converts the dC pool to dU, increasing cellular dUMP, while the RNA antitoxin (AvcI) inhibits AvcD activity. We have shown that AvcD deaminates dC nucleotides upon phage infection yet the activation mechanism in response to phage infection is unknown. Here we show that the activation of AvcD arises from phage-induced shutoff of host transcription, leading to degradation of the labile AvcI. Interestingly, infection of phages such as T7 that are not inhibited by AvcID can also lead to the degradation of the AvcI RNA antitoxin, suggesting that transcriptional shutoff will activate the AvcID system regardless of source and depletion of AvcI is not sufficient to confer protection against phage infection. We also show that AvcID increases the abundance of defective phages that are susceptible to AvcID over the course of infection and this increase of defective phage is not reliant on the uracil-DNA glycosylase, Ung. Collectively, our data indicate that the AvcID system can regulate phage production and viability, providing further understanding of this novel TA system and its role in phage defense.

3.2: INTRODUCTION

Bacteria encounter a plethora of different conditions for which they have evolved how to respond and adapt in order to survive. These responses and adaptations can range from slowing their growth and metabolism to producing natural toxins to fend off

phage predators while protecting the host. The toxins the bacteria produce are often kept repressed until the proper conditions. One classic example of such toxic factors that are ubiquitous in bacteria and archaea is the toxin-antitoxin (TA) systems [134]. TA systems were first discovered in the 1980s and are found ubiquitously on plasmids, bacterial chromosomes, and phages [135–137]. These modules typically constitute a diverse two-gene operon that encodes a toxin and an antitoxin that neutralizes the toxin action or production [130, 138, 139]. There are currently eight types (I-VIII) of TA systems and they are classified based on the nature of the antitoxin and the mechanism by which it regulates the toxin [130, 139]. The toxins are generally proteins in all types of TA systems, except the type VIII system, in which the toxin is a small RNA (sRNA) [139, 140]. In the case of type I, III, and VIII TA systems, the antitoxins are sRNAs while the rest are small peptides [139]. Generally, antitoxins are more abundant than their cognate toxins but are more labile and readily degraded under stress conditions, allowing the toxins to exert their growth-inhibition functions [141]. It was thought that TA systems serve as a critical component of bacterial stress response systems, but this was recently disputed by a study that showed cells without TA systems have no detectable reduction in growth compared to cells that have TA systems when exposed to several classes of external stresses [142].

Though most past studies employed abiotic stressors (*i.e.* antibiotics, oxidative agents), to test the effects of type II TA systems, recent findings show that biotic stress, such as phage infection, could conceivably be the actual biological triggers of the TA systems [139, 142]. Restriction/Modifications (RMs) are known to inhibit phage infections, and they also promote plasmid maintenance when encoded in Type I TA

systems [135, 143]. Therefore, it was reasoned that TA systems may also inhibit phage infections. Indeed, so far, studies of type I-IV TA systems have demonstrated that one of their primary physiological roles is to limit phage infections [122, 124, 126, 144]. Additionally, TA modules not only are clustered and closely connected to mobile genetic elements (MGEs), but they also mediate the stabilization of MGEs by limiting gene reduction [145]. They also are highly abundant in free-living prokaryotes but not symbiotic, host-associated prokaryotes [146]. This occurrence further supports that MGEs are evolutionarily beneficial and important for the free-living prokaryotes that are constantly challenged by phages.

Vibrio cholerae is a free-living bacterium that lives in both brackish-water environments and the human gut. *V. cholerae* encodes multiple TA modules on its chromosomes. The current biotype of *V. cholerae* that is circulating and causing endemic cholera in developing countries is *V. cholerae* El Tor. El Tor contains two unique genomic islands called the Vibrio Seventh Pandemic Islands 1 and 2 (VSP-1 and -2) that are not present in the classical biotype, which was displaced by El Tor. We have recently described a novel TA module encoded on VSP-1 called AvcID [147].

The AvcID TA system encodes the AvcD toxin that deaminates dCTP and dCMP, and ultimately produces dUMP as the final product, leading to a disruption in nucleotide metabolism after phage infection [147]. AvcI is a noncoding RNA that binds to and directly inhibits the activity of AvcD. However, the mechanism by which phage induces the AvcID system remains unknown. In this work, we demonstrate that under the natural activating conditions of phage infection, the labile antitoxin sRNA AvcI is preferentially lost, allowing AvcD to deaminate dC pools. The shift in nucleotide pools produced by

AvcD inhibits the viability of the phage and the formation of new phage progeny. Though AvcID can provide *Escherichia coli* with potent defense against select lytic phage, phage resistant to AvcID can still activate AvcD activity, suggesting other dynamics of phage/host interactions are important for limiting phage replication and/or spread.

3.3: MATERIALS AND METHODS

3.3.1: Bacterial Strains, Plasmids, and Growth Conditions

The strains, plasmids, and primers used in this study are listed in Appendix 2 (Tables 1-3). Unless otherwise stated, cultures were grown in Luria-Bertani (LB) at 35°C and supplemented with ampicillin (100 µg/mL) when needed. *E. coli* BW29427, a diaminopimelic acid (DAP) auxotroph, was additionally supplemented with 300 µg/mL DAP. Plasmids were introduced into *E. coli* MG1655 or *E. coli* NR8052 through biparental conjugation using an *E. coli* BW29427 donor.

3.3.2: Phage Propagation

Coliphages were propagated on *E. coli* MG1655 grown in LB, and their titer was determined using the small drop plaque assay method, as previously described [7]. Briefly, 1 mL of overnight cultures were mixed with 50 mL of MMB agar (LB + 0.1 mM MnCl2 + 5 mM MgCl2 + 5 mM CaCl2 + 0.5% agar), tenfold serial dilutions of phages in MMB were dropped on top of them and incubated overnight at room temperature. The viral titer is expressed as plaque forming units per mL (pfu/mL).

3.3.3: Phage Infection in Liquid Culture

Overnight culture of *E. coli* carrying indicated AvcID plasmids were subcultured and grown to an OD_{600} of 0.3 and then mixed with phage at the indicated MOIs. A 150 µL aliquot of the mixtures were put into 96-well plates, and growth was measured at 2.5 min intervals with orbital shaking on a plate reader (SpectraMax M6) at 37°C for 8 hours. Data represents the mean ± SEM, *n*=3.

3.3.4: RNA Extraction for Northern Blot Following Phage Infection

RNA isolation and qRT-PCR analysis were carried out as previously described [102]. Briefly, triplicate overnight cultures of *E. coli* carrying pAvcI-AvcD-6xHis were subcultured 1:100 in LB and grown to an OD_{600} of 0.3. 1 mL of each replicate was pelleted and flash-frozen by the ethanol-dry ice slurry method. RNA was extracted using TRIzol® reagent following the manufacturer's directions (Thermo Fischer Scientific). RNA quality and quantity were determined using a NanoDrop spectrophotometer (Thermo Fischer Scientific).

3.3.5: RNA Probe Synthesis and Purification

The method for RNA probe production was modified from a previously described protocol [147]. The AvcI DNA template for *in vitro* transcription was PCR amplified from pAvcI using Q5 High-Fidelity DNA Polymerase (NEB). To incorporate the T7 promoter into the final AvcI DNA template, the forward primer included the T7 promoter sequence prior to the homologous sequence for AvcI. Additionally, the first two residues of the reverse primer were 2′-OMe modified to reduce 3′-end heterogeneity of the transcript

[108]. The PCR reaction was analyzed using a 1% agarose gel, and the band corresponding to the AvcI DNA template was excised and gel purified using Promega Gel Extraction and PCR clean up kit. AvcI RNA was synthesized by *in vitro* transcription using the T7-AvcI reverse complement DNA template and the HiScribe™ T7 High Yield RNA Synthesis Kit (NEB). Bio-11-UTP was included during the transcription reaction for Northern blot detection purposes. The transcription reactions were incubated at 37°C for 4 h. Following transcription, DNase I (NEB) was added to a final concentration of 1X per reaction and incubated at 37°C for an additional 15 min. AvcI was then purified using Monarch® RNA Cleanup Kit (NEB). Purity of product was evaluated using a TBE agarose gel. Individual aliquots of AvcI were flash frozen using liquid nitrogen and stored long-term at -80°C.

3.3.6: Northern Blot Analysis Following Phage Infection and Half-life Quantification and Analysis

1.5-2 µg total RNA was diluted 1:1 in 2x samples buffer (Invitrogen), loaded onto 7.5% TBE-Urea PAGE gels and ran for 30 min or until the front dye reached ~1 cm above the bottom of the gel at 200 V. RNA was then transferred to BrightStar™-Plus Positively Charged Nylon Membrane (Invitrogen) with a Fisherbrand™ Semidry Blotting Apparatus (Fisher Scientific) ran for 1 h at 250 mM A. RNA was then crosslinked to the membrane using the CX-2000 crosslinker compartment of UVP HybriLinker™ HL-2000 (Fisher Scientific). Each side of membrane was crosslinked at 1200 µjoules twice and dried at 50°C for at least 30 minutes to improve sensitivity. Membranes prepared this way can be stored at 4°C with desiccation for several months. The membranes were

pre-hybridized for at least 60 minutes at 60°C in ULTRAhyb™ Ultrasensitive Hybridization Buffer (Invitrogen) with gentle shaking. Next, the pre-hybridization buffer was removed, and hybridization buffer containing 1 nM of purified probe was added. The membrane was hybridized for 12-16 hours at 60°C with gentle shaking. Next, the membrane is rinsed twice every five minutes with 2x saline-sodium citrate (SSC) buffer, 0.1% SDS at 60°C and then twice every 15 minutes with 0.1X SSC, 0.1% SDS at 60°C. The biotin-labeled probes were detected using a Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific™) at RT. The membranes were then imaged using Amersham™ Imager 600. To determine the half-life of *avcI*, the band intensities were analyzed using the Fiji software and normalized to respective 0 min band intensity [101]. All Northern blots shown are representative of two independent biological replicates.

3.3.7: Western Blot Analysis of AvcD

Cells were collected in the same method as RNA extraction. Pellets were then resuspended at OD_{600} = 15 (~20 µL) in 2x Laemmi loading dye supplemented with 10% β-mercaptoethanol v/v, denatured for 10 min at 95°C, and centrifuged at 15k x g for 10 min. Samples were then analyzed by 4-20% SDS-PAGE gels (Mini-PROTEAN TGX Precast Protein Gels, Bio-Rad) alongside size standards (Precision Protein Plus, Bio-Rad or PageRuler™ Plus Prestained Protein Ladder, Thermo Scientific™). Gels were run at room temperature for 60 min at 120 V in 1x Tris/glycine/SDS running buffer. Proteins were transferred to nitrocellulose membranes (Optitran). The membranes were blocked using 5% skim milk and incubated with 1:5000 THETM His Tag Antibody, mAb, Mouse (GenScript) followed by 1:4000 Goat Anti-Mouse IgG Antibody (H&L) [HRP],

pAb (GenScript), treated with Pierce™ ECL Western Blotting Substrate, and imaged using an AmershamTM Imager 600. Western blots shown are representative of two independent biological replicates.

3.3.8: CFU/PFU Measurements Pre- and Post-Phage Infection

Overnight cultures were subcultured and split into two 10 mL aliquots and grown up to an OD_{600} of ~0.3. One aliquot was mixed with phage (T5, MOI = 0.1; T7, MOI = 0.01) and the other with an equal volume of LB (uninfected control). Both were grown in a shaking incubator (210 rpm) at 37°C. At each indicated timepoint, 1.5 mL of culture was then spun down. The supernatant from each tube was filter sterilized with 0.22 μM filter and transferred to a new tube, and the cell pellets were washed twice with equal volume of LB to remove unadsorbed phage. For PFU measurements, the supernatants were serially diluted in MMB medium (LB + 0.1 mM MnCl2 + 5 mM MgCl2 + 5 mM CaCl2) and 5 μL of each dilution was spotted on a lawn of bacteria seeded in MMB agar plate (MMB + 0.5% agar). PFU plates were then grown at RT overnight and plaques quantified the following day. For CFU measurements, resuspended cell pellets were then incubated at 37°C for 5-10 minutes before being serially diluted 10-fold in PBS and 5 μL of each dilution was spotted on LB plates. CFU plates were then grown at 37°C overnight and colonies were quantified the following day.

3.3.9: UPLC-MS/MS dNTPS Quantification

Deoxynucleotide concentrations were determined as previously described [111, 147] with minor modifications. Briefly, to measure the nucleotides after phage infection,

cells were grown in LB overnight at 37°C. Overnight cultures were subcultured 1:100 in LB and grown to OD_{600} of ~0.3. 3 mL of culture was collected for a time zero reading: 1.5 mL for dNTPs quantification and the 1.5 mL for total protein quantification. The cultures were then infected with phage (T7, MOI of 5), and an additional 3 mL were removed at each indicated subsequent time point. Culture aliquots were collected by centrifugation at 15k x g for 1 min. Pellets were resuspended in 200 µL of chilled extraction buffer [acetonitrile, methanol, ultra-pure water, formic acid (2:2:1:0.02, v/v/v/v)]. To normalize in vivo nucleotide samples, the other 1.5 mL aliquot pellet was centrifuged at 15,000 x g for 1 min, resuspended in 200 µL lysis buffer F (20 mM Tris·HCl, 1% SDS, pH 6.8), and denatured for 10 min at 95°C. Denatured lysates were centrifuged at 15,000 x g for 1 min to pellet cellular debris, and the supernatant was used to quantify the total protein concentration in the sample by using the DC protein assay (Bio-Rad) and a BSA standard curve [100]. The concentrations of deoxynucleotides detected by UPLC-MS/MS were then normalized to total protein in each sample.

All samples resuspended in extraction buffer were immediately incubated at -20ºC for 30 min after collection and centrifuged at 15,000 x g for 1 min. The supernatant was transferred to a new tube, dried overnight in a speed vacuum, and finally resuspended in 100 µL ultra-pure water. Experimental samples and deoxynucleotides standards [1.9, 3.9, 7.8, 15.6, 31.3, 62.5, and 125 nM of dATP (Invitrogen), dGTP (Invitrogen), dTTP, (Invitrogen), dCTP (Invitrogen), dCMP (Sigma), dUTP (Sigma), and dUMP (Sigma)] were analyzed by UPLC-MS/MS using an Acquity Ultra Performance

LC system (Waters) coupled with a Xevo TQ-S mass spectrometer (Waters) with an ESI source in negative ion mode.

3.3.10: Genomic Extraction and Quantification using qPCR

Phage genomes were extracted as previously described [148]. Briefly, phage lysates were treated with RNase A (Roche; 1 µg/mL), DNase I (NEB; 18 U), and lysozyme (Sigma-Aldrich; 1 mg/mL). Samples were incubated at 37°C for 90 min, and then the DNase was inactivated by incubating at 75°C for 10 min. The samples were then further treated with 0.1 mg/mL Proteinase K (Invitrogen) and 0.5% SDS and were incubated at 55°C for 1 h. Samples were then extracted once with phenol-chloroform: isoamyl alcohol (25:24:1) and second time with chloroform. DNA was isolated by ethanol precipitation with the addition of 0.3 M sodium acetate. DNA quality and quantity were determined using a NanoDrop spectrophotometer (Thermo Fischer Scientific).

For measuring phage genome abundance, 25 µL reactions consisted of 5 µL each 0.625 µM primers 1 and 2, 12.5 µL 2X SYBR master mix, and 2.5 µL of 2.5 ng/µL phage genomic DNA. qPCR reactions were performed in technical duplicates for biological triplicate samples. The relative abundance was calculated by comparing the C_t values of phage infected *E. coli* with AvcID to inactive AvcID* at each timepoints.

3.3.11: Transmission Electron Microscopy (TEM)

TEM experiments were performed at Michigan State University and were performed as previously described [149]. Briefly, samples were prepared on Quantifoil grids that have been applied on with 10 nm nanogold fiducial markers. After a plunge in

liquid nitrogen, samples were stored, and transferred and imaged under liquid nitrogen temperatures. The samples were imaged in a JEOL 2200-FS TEM, operating at 200 keV, using low dose conditions controlled by SerialEM (version 3.5.0-beta) with the use of an in-column Omega Energy Filter operating at a slit width of 35 eV [150]. The images were taken at 45,000X nominal magnification on the JEOL. Each sample was imaged with at least 10 fields of view.

3.4: RESULTS

3.4.1: AvcID system provides phage defense in liquid cultures

Previous studies demonstrated that AvcID systems derived from *V. cholerae, V. parahaemolyticus, and E. coli* ETEC can reduce the dC pool upon phage infection, yet their respective resistance profiles are different [147]. For instance, *V. parahaemolyticus* AvcID provides protection against T3, T5, T6, and SECΦ18 phages whereas *E. coli* ETEC AvcID provides protection against T3, SECΦ17, SECΦ18, and SECΦ27 [147]. To follow up on the previous study, we focused on the AvcID system derived from *V. parahaemolyticus* as it provides robust protection against a few well-characterized T-type coliphages.

Since the protection conferred by the AvcID system has thus far been based on efficiency of plaque assays (EOPs), we hypothesize that the AvcID system can also confer protection in liquid culture conditions [147]. To test this hypothesis, we infected *E. coli* cells harboring either the native AvcID or the inactive $AvcID^{S49K+E376A}$ ($pAvcID^{*}$) with T3, T5, T6, or SECΦ18 phage at varying multiplicities of infection (MOI) and tracked bacterial growth by OD_{600} over time. We also tested with T7 phage as a

negative control. At all the MOIs tested for T5, the OD_{600} of cultures harboring AvcID was always higher compared to cultures harboring AvcID*, indicating that cells survive longer as AvcID prevents the phage infection from spreading throughout the population. Similar trends were seen for AvcID cells infected with SECΦ18 phage (Fig. 3.1E). Contrary to the prior EOP results, the presence of AvcID shows poor protection against T3 and T6 in liquid cultures (Figs. 3.1A, C). However, like the EOP results, AvcID shows poor protection against T7 in liquid culture (Fig. 3.1D). These data suggested that AvcID provides defense against specific phage in liquid culture, but the phages that are sensitive to AvcID are different than the EOP experiments.

3.4.2: AvcD is activated by transcription shutoff of *avcID*

AvcI and AvcD form a complex in vitro, and AvcD is inhibited by the sRNA AvcI, suggesting AvcD inhibition is linked to the assembly of the complex. The activation of Type III TA systems can be achieved by two types of mechanisms from phage: active, which occurs when a phage product triggers release of the toxin; or passive, which occurs when phage-induced transcription shutoff reduces antitoxin levels, thereby releasing the toxin [121]. Due to the nature of the antitoxin being labile under stressful conditions, we hypothesized that AvcI is degraded upon phage infection. The stability of AvcI was assessed by Northern blot, using total RNA samples extracted from *E. coli* MG1655 cells carrying a vector with the full length *avcID* locus at several time points after various indicated treatments. The RNA samples were subsequently resolved by 7.5% urea-PAGE. Likewise, to determine whether AvcD protein levels changed

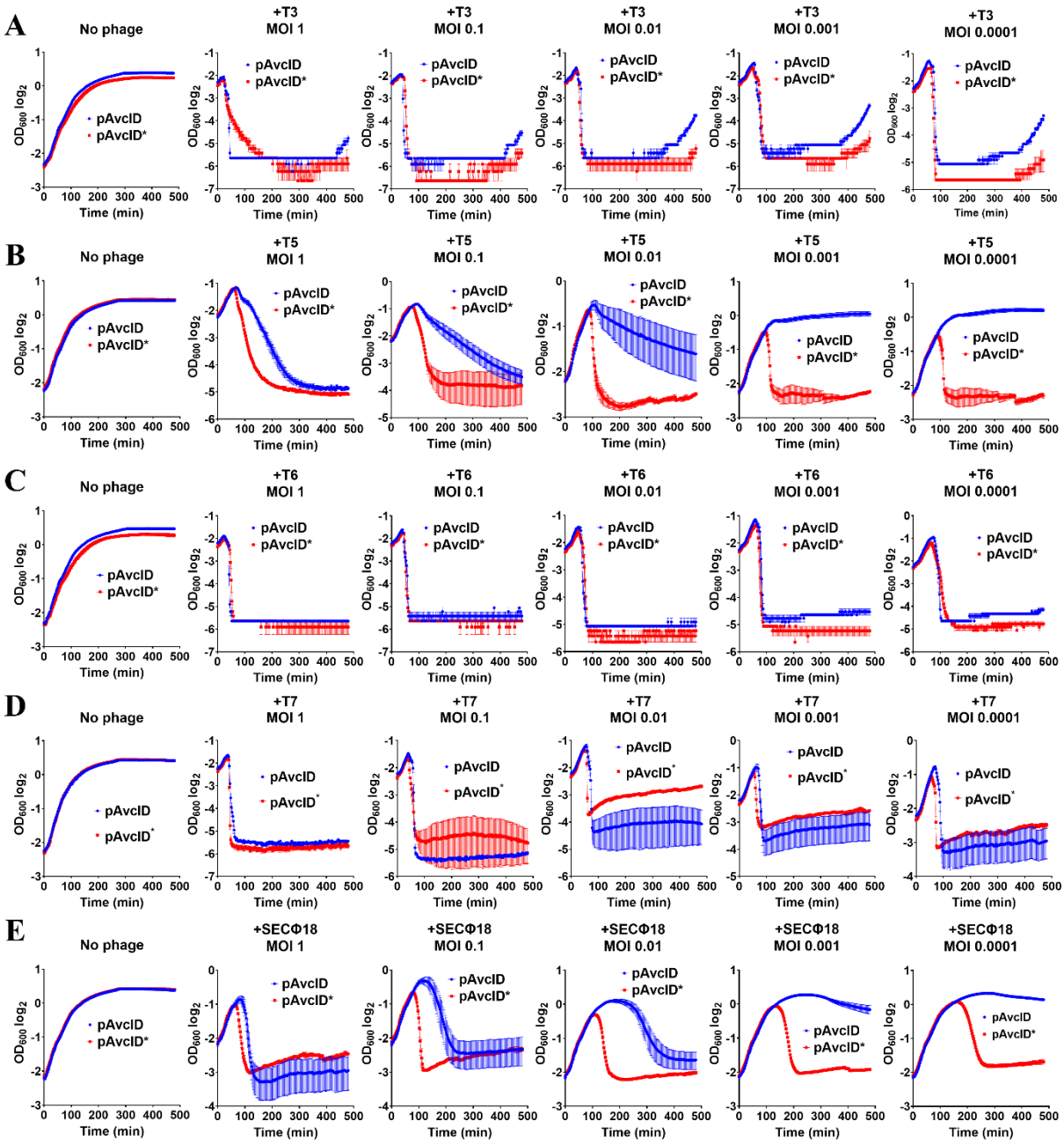


Figure 3.1. The AvcID system provides phage defense in *E. coli* in liquid culture. Growth curves for *E. coli* with active (pAvcID) or inactive (pAvcID*) AvcID system after infection with T3 (**A**), T5 (**B**), T6 (**C**), T7 (**D**), or SECΦ18 (**E**) phage at varying multiplicities of infection (MOI). Data represents the mean ± SEM of three biological replicate cultures.

concurrently with changes in *avcI* RNA levels, we tagged AvcD with a C-terminal 6xHis tag and then assessed using Western blot. Notably, the full length of *avcI* transcript is

slightly smaller than the 400 base pairs (bp) marker, which is longer than the functional unit that was previously determined (~171 bp) [147] .Additionally, the levels of AvcI decreased overtime when cells were treated with rifampicin (Fig. 3.2A). Importantly, spectinomycin, which inhibits protein synthesis instead of transcription, did not decrease *avcI* levels (Fig. 3.2B), indicating that the degradation was specific to transcriptional shutoff. The spectinomycin result is also consistent with our previous study showing that spectinomycin did not activate chromosomally encoded *avcID* in *V. cholerae* [147].

We also infected these cells with the phages T3, T5, T7, and SECΦ18. The half-life of *avcI* transcript ranges from 1.5 to 6.7 min (Fig. 3.2C-F). Notably, each phage infection results in a different *avcI* degradation rates, owning to the different phage infection processes. Concurrently, we also found that AvcD protein levels did not change significantly in any of the conditions tested. AvcID provides protection against T3, T5 and SECΦ18 phages but not T7 via EOPs, but AvcID shows more robust protection against T5 and SECΦ18 phages in liquid culture. Remarkably, the levels of *avcI* decrease in all the phage infection conditions. To determine whether AvcD is activated, we measured the intracellular abundance of dCTP, dCMP, dUTP, and dUMP before and after infecting the cells with T7 phage, as well as treated the cells with rifampicin or spectinomycin as controls. Like other phages tested, T7 infections significantly decreased intracellular dCTP and dCMP and increased dUMP, while the dUTP level was not changed, suggesting that AvcID is activated by T7 phage (Figs. 3.3A-D). Collectively, this suggests that transcriptional shutoff coupled with the fast turnover rate of *avcI* RNA leads to the release of existing AvcD from inhibition, but this is not sufficient to decrease T7 phage infection.

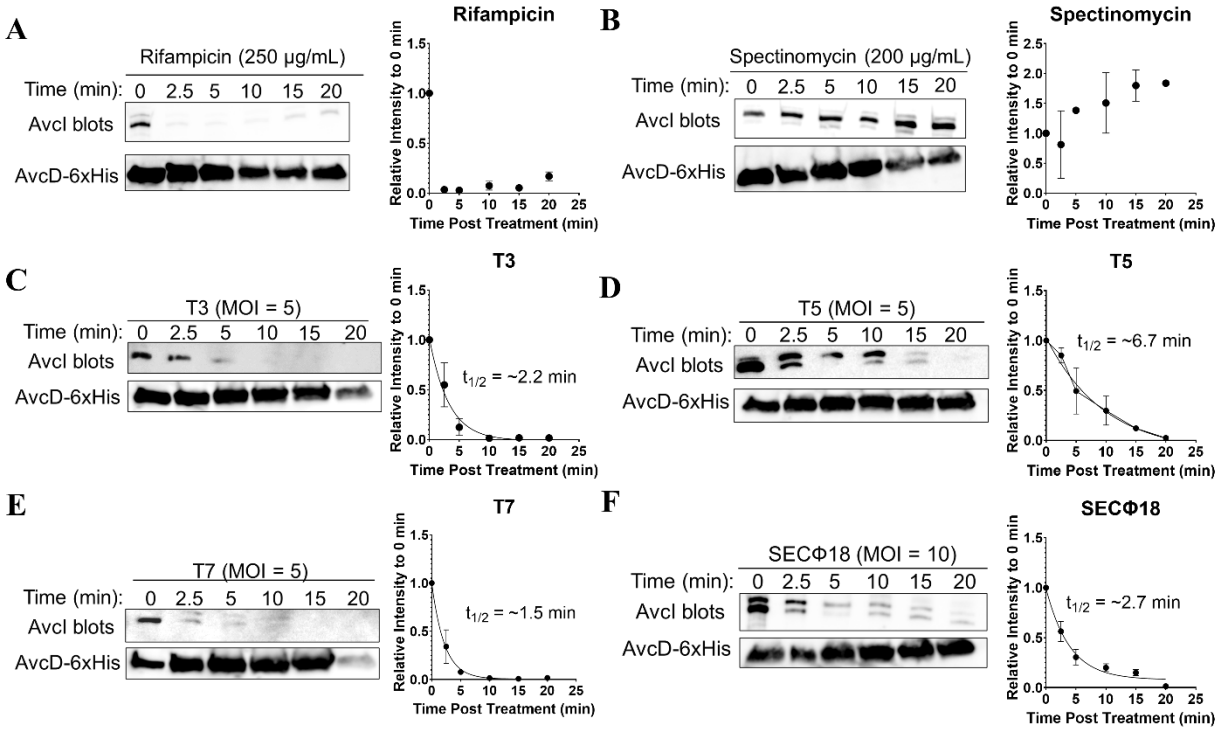


Figure 3.2. Transcriptional shutoff leads to the degradation of *avcI*.
Northern blot of *avcI* RNA using a biotinylated probe complementary to *avcI* (top) and Western blot of AvcD-6xHis using anti-6xHis antibody (bottom) during rifampicin treatment (250 µg/mL) (**A**), spectinomycin treatment (200 µg/mL) (**B**), T3 infection (**C**), T5 infection (**D**), T7 infection (**E**), and SECΦ18 infection (**F**). All the phage infection is with MOI of 5, except SECΦ18, which is with MOI of 10. The quantification of *avcI* band intensities is used to determine the half-life of *avcI*.

3.4.3: AvcID drives production of defective T5 phage

We found that AvcID provides resistance to T5 but not T7 (Figs. 2.13A, 3.1B, D), yet both phages induce the degradation of AvcI and the activation of AvcD deamination (Figs 2.15B-E, 3.2D-E, and 3.3). To further quantify the production of phage and viability of hosts over time, we performed the liquid phage infection assay on *E. coli* carrying either AvcID or inactive AvcID* and collected samples at the indicated timepoints. Phage lysates and *E. coli* cells were

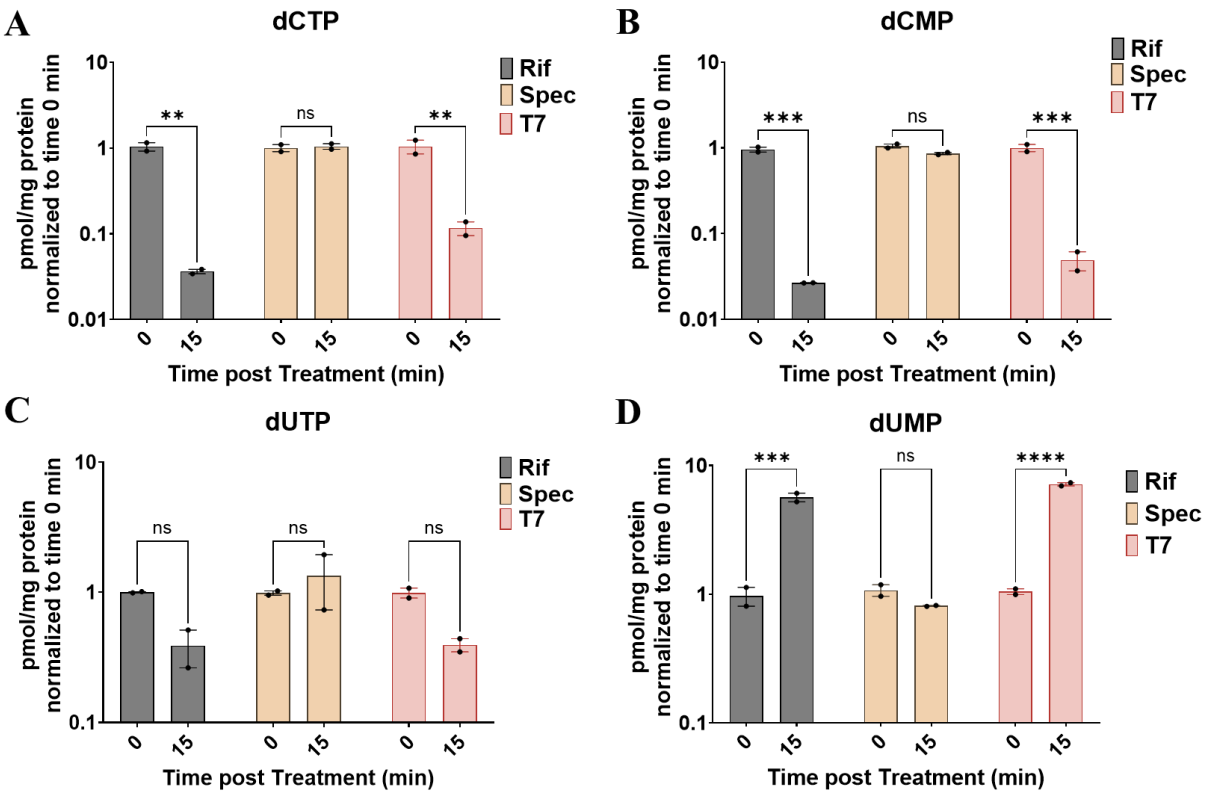


Figure 3.3. AvcD is activated by all phage studied.
In vivo abundance of dCTP (**C**), dCMP (**D**), dUTP (**E**), and dUMP (**F**) in an *E. coli* host carrying pAvcID with its native promoter before and after addition of rifampicin (250 µg/mL), spectinomycin (200µg/mL) or T7 phage (MOI = 5). Nucleotides measured using UPLC-MS/MS and normalized to total protein. Data represents the mean ± SEM of two biological replicate cultures, Two-way ANOVA with Dunnett's post-hoc test, and ns indicates not significant.

separated to measure plaque forming units (PFUs) and colony forming units (CFUs). When the cells were infected with T5, cells harboring AvcID had more viable CFUs than cells harboring inactive AvcID*, after two hours of infection (Fig. 3.4A). Indeed, the AvcID-containing cells had ~100-fold fewer PFUs than AvcID*-containing cells by 5 h of infection (Fig. 3.4B), supporting the notion that AvcID inhibits the abundance of functional T5 phages. Consistent with our liquid infection results described above, AvcID did not impact the number of CFUs and PFUs when cells were infected with T7 (Figs. 3.4E-F). Since only viable phages form plaques, we speculated

that a plaque assay could underestimate the total viral particles produced if some of those virions were defective. To determine the total viral particle abundance, we employed qPCR of phage DNA isolated from the phage particle samples used to quantify PFUs. We quantified a phage-specific gene to measure the abundance of all viral genomes. We showed that the total number of T5 phage genomes decrease over time in infected cultures of AvcID-containing cells compared to AvcID*-containing cells, but this decrease was less than that observed for PFUs between the two samples (Fig. 3.4C). We calculated the percentage of viable phages by quantifying the ratio between PFUs after infecting cultures containing either AvcID or AvcID*, and the difference of phage genome abundance after infecting AvcID or AvcID*. Using this analysis, we estimated that by 30 min only approximately 30% of T5 phage derived from cells containing AvcID are functional, and the proportion of functional phage decreases overtime, suggesting that AvcID is impairing the ability of T5 phage to form plaques (Fig. 3.4D). In contrast, virtually no difference was observed in the T7 genome abundance of phage derived from cells with AvcID or AvcID* (Fig. 3.4 G). Nearly all the T7 phages were viable even when they were from cells with AvcID, which corroborates the PFU result (Fig. 3.4H). Together this indicates that AvcID confers protection by increasing defective phage production for T5, while phage like T7 can overcome this function of AvcID through an unknown mechanism.

Consistent with our observation of defective T5 phage, transmission electron microscopy (TEM) images revealed that the T5 phage from cells containing AvcID have more phage capsids with no tails and scattered phage tails compared to T5 phage from

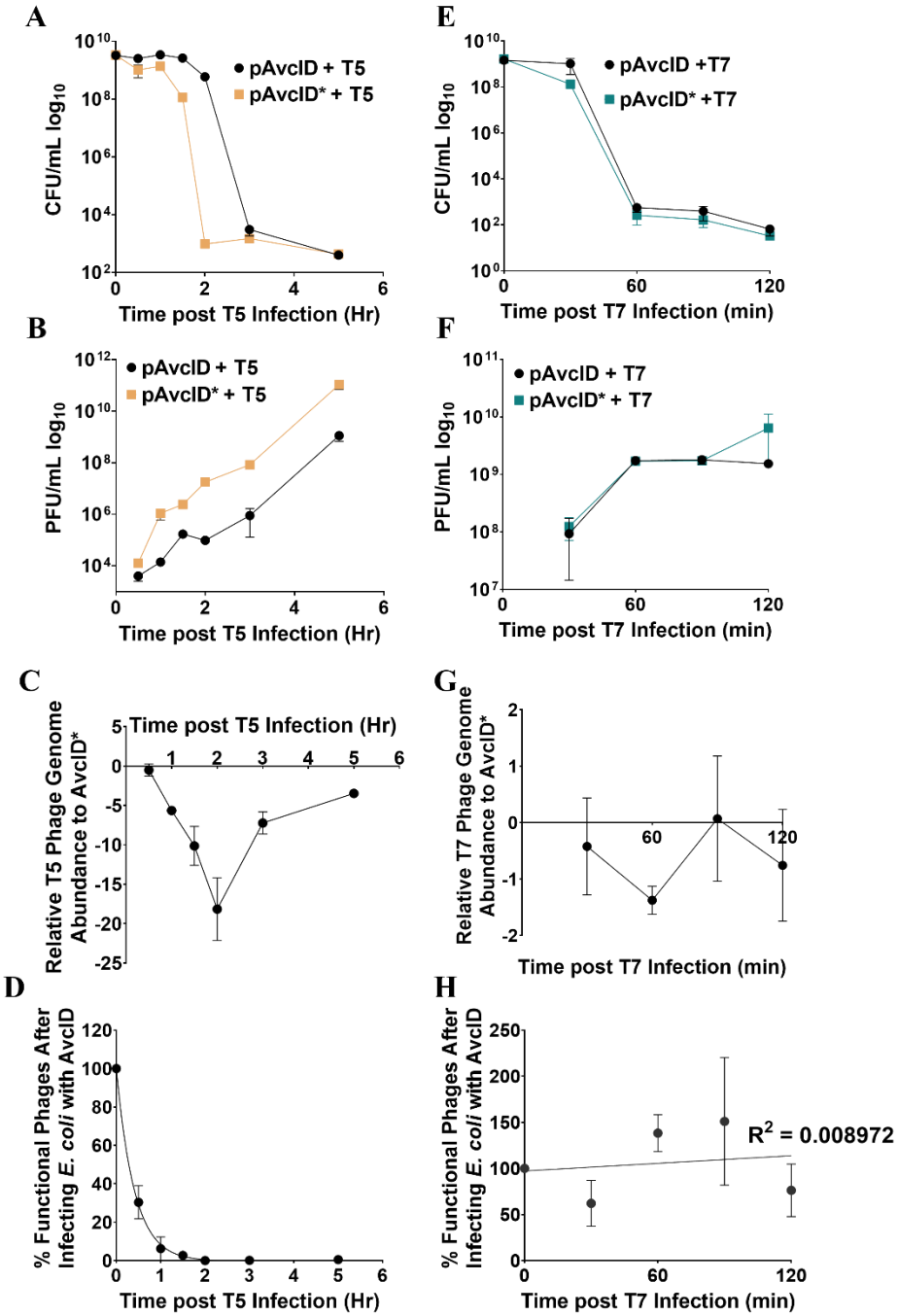


Figure 3.4. AvcID reduces the functionality of T5 but not T7 phage.
Survival of *E. coli* encoding the indicated AvcID systems as measured by CFU after infection with T5 (**A**) or T7 (**E**). PFU quantification over time in cultures with cultures of indicated AvcID systems-containing cells infected with T5 (**B**) or T7 (**F**). Relative T5 (**C**) or T7 (**G**) genome abundance comparing *E. coli* expressing pAvcID or inactive AvcID* over time. Percent viable phage after infecting cells containing AvcID with T5 (**D**) or T7 (**H**) phages. Data represents the mean ± SEM of three biological replicate cultures.

cells containing AvcID* (Fig. 3.5), yet its connection to the reduction in phage genome replication is unclear. However, the TEM results corroborate the PFU and qPCR results by confirming that the AvcID system reduces the functionality of T5 phage particles. This TEM analysis will require more samples and quantification to make this substantial claim.

3.4.4: Ung does not contribute to the AvcID antiphage defense

The dUMP nucleotide is the ultimate product produced by AvcD after deaminating dC nucleotides and it is the dominant nucleotide species after phage infection ([147] and Fig. 3.3). The increased dUMP level in the cells may lead to an increased frequency of dUMP being incorporated into the genomic DNA in place of dTMP by DNA polymerases during replication, including the phage's DNA polymerase. dUMP that has been incorporated into genomic DNA is excised by the uracil-DNA glycosylase (Ung) enzyme. An abasic site is formed after Ung releases uracil from the DNA, leading to a blockage during DNA replication. An AP endonuclease can then cleave the DNA at the abasic site, resulting in a nicked DNA strand [151, 152]. The increased dU incorporation in the genome will lead to numerous abasic sites generated by Ung. Ultimately, this could generate more nicks in DNA strands that could cause double-stranded breaks. Additionally, Ung does not distinguish between bacterial or phage genomic DNA, suggesting that AvcD and Ung could synergize to prevent phage infection. To determine whether AvcD and Ung function together to reduce phage infection, we infected *E. coli* MG1655 or Δ*ung E. coli* NR8052 carrying native AvcID or the inactive pAvcID* with T5 or SECΦ18 phage at varying MOI and tracked bacterial

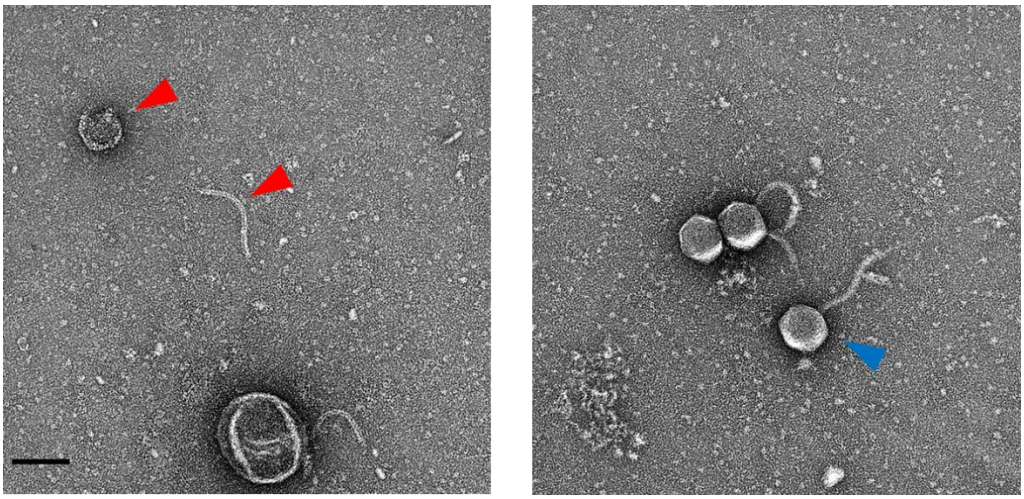


Figure 3.5. Transmission electron microscopy images showing the structure of T5 phages after infecting cells containing either AvcID (left) or AvcID* (right). The red triangle is indicating the empty capsid and the tail fiber while the blue triangle is indicating an intact phage. The scale bar represents 100 nm. *n* =1

growth by OD_{600} over time. At all the MOIs tested for T5, the OD_{600} of both strain backgrounds carrying active AvcID had a similar growth yield, suggesting that Ung is not required for AvcID to protect *E. coli* from T5 phage (Fig. 3.5A). In the absence of a functional AvcID, the absence of *ung* makes the cells slightly more susceptible to phage infection (Fig. 3.5A). When infected with SECΦ18 at an MOI of 1, the OD_{600} of the Δ*ung* mutant *E. coli* carrying AvcID dropped dramatically compared to MG1655, comparable to the Δ*ung* mutant with non-functional AvcID (Fig. 3.5B). Yet, this difference was reversed at lower MOIs, as the Δ*ung* mutant exhibited more growth than MG1655 at MOIs of 0.1 or lower. Nevertheless, at all the MOIs tested for SECΦ18, the OD_{600} of Δ*ung* strain carrying inactive AvcID* decreased significantly the most (Fig. 3.5B). Together, the data suggest that Ung alone can protect against phage infection of T5 and SECΦ18, but it is not required for AvcID to protect against phage infection.

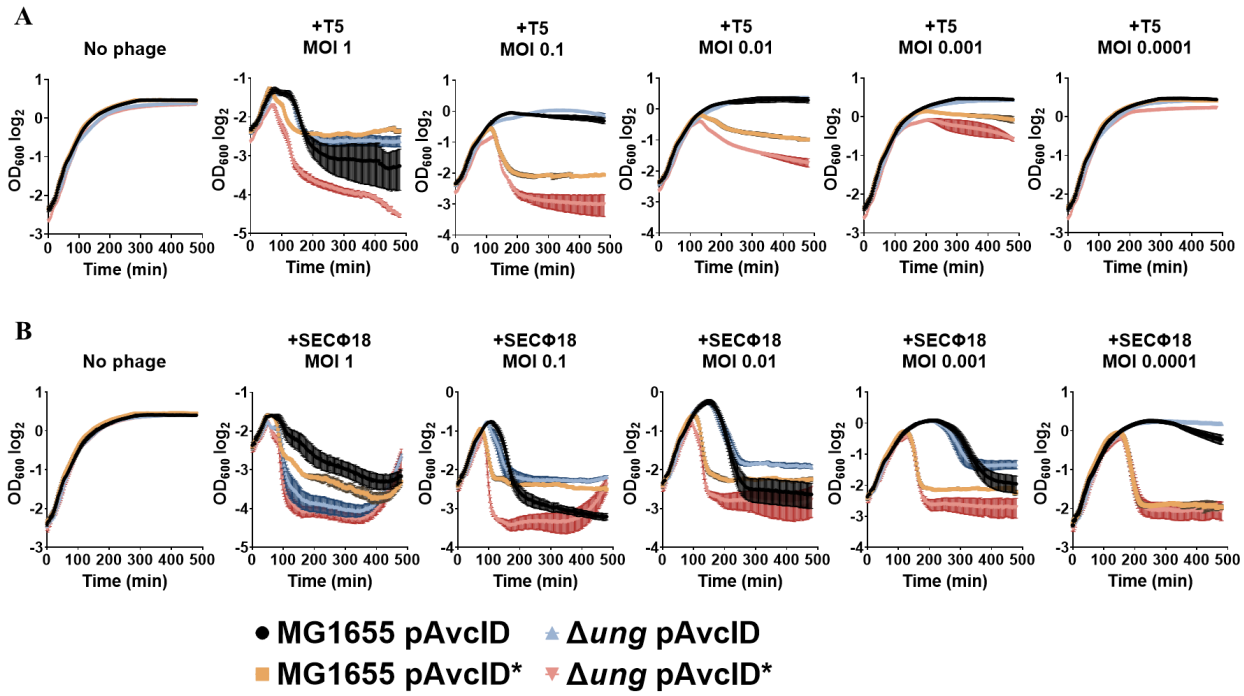


Figure 3.6. Ung and AvcID do not function together to provide phage protection. Growth curves for *E. coli* MG1655 or Δ*ung* mutant with active (pAvcID) or inactive (pAvcID*) AvcID system after infection with T5 (**A**) or SECΦ18 (B) phage at varying multiplicities of infection (MOI). Data represents the mean ± SEM of three biological replicate cultures.

3.5: DISCUSSION

Phage predation drives constant evolutionary pressure that shapes the diversity and fitness of bacteria. Increasing evidence suggests that bacteria are equipped with multiple antiphage defense systems to protect themselves from phage threat. The underlying mechanisms of certain antiphage defenses are well-characterized, such as RMs, which utilize DNA modifications to distinguish host and foreign DNA. On the contrary, the mechanism of activation of the cyclic nucleotide-based systems (*i.e.,* CBASS) in response to phage infection is generally not understood. This study reveals the mechanism of how the AvcID TA system is activated in response to phage infection and its impact on the phage's functionality. We also begin to explore the molecular

mechanisms by which AvcID specifically protects against certain phage like T5 while it is ineffective against others like T7.

Cessation of transcription is a hallmark of infection by many phages [153–155]. Our results demonstrate that transcriptional shutoff leads to the degradation of *avcI* and thus releasing the activity of AvcD to deaminate dC nucleotides. This mode of activation is consistent with other TA systems, such as the ToxIN system [121, 122]. *avcI* transcripts are produced at high rates and abundance compared to that of *avcD* [147] due to the fact there is a rho-independent transcriptional terminator between *avcI* and *avcD*, which is a characteristic of Type III TA systems [156]. Hence, the Northern blots suggest that the precursor length of the *avcI* transcript is around 400 bp. Additionally, the rho-independent transcriptional terminator allows infrequent read-through of *avcD*, causing an excess of antitoxin over toxin. This may explain why the full-length transcript of *avcID* on Northern blots. The precursor of *avcI* transcripts was shown to have a half-life of 1.5-6.7 min, which is within the normal range for an mRNA in *E. coli* [157]. However, this half-life may not fully represent the true biological half-life in their respective host cells. Furthermore, the half-life range of *avcI* transcripts may also be explained by the timescale of the lytic cycle and the mode of infection of the phage.

Though AvcID did not provide protection against T7, *avcI* transcripts were degraded post T7 infection, implying that *avcI* levels decrease when cells undergo transcriptional inhibition regardless of the cause. However, the activation of AvcD does not have any detrimental effect on the viability of T7, in contrast to T5, implying that T7 may have evolved to disregard the detrimental effect inflicted by AvcID. TEM images reveal most of the T5 phages are defective when AvcD is active. The disruption of

replication could have impacted the timing of lysis, DNA packaging, and the multiplication rate of T5 virions before disseminating to the neighboring cells, thus accounting for the decreasing number of functional phages. The capsids lacking a tail fiber in the TEM images could still have phage DNA, though that is still under investigation.

The growth of several well-known phages is inhibited when their DNA contains dUMP and Ung is present in the host cells. For instance, T5 encodes its own dUTPase for reducing the dUTP level such that dUMP is limited in its genome. Therefore, T5 can infect cells lacking Ung more readily, which is consistent with our results. However, the presence of the AvcID system prevents this T5 infection even in the absence of Ung, indicating that dUMP incorporation may not be the cause for the phage replication defect but instead it is the depletion of the dC pool that is responsible for the reduction in functionality of T5 phage particles. This depletion in the dC pool has no effect on T7 viability even though its G/C content is approximately 50%, and we are investigating possible mechanisms for this inherent difference in phages [158].

Similar to the AvcID system, Rotem Sorek's group has demonstrated bacterial dGTPases protect against phage infection by dephosphorylating dGTP to dG to inhibit phage DNA replication and that this system is also activated upon phage-induced transcriptional shutoff [129]. It is, however, unclear whether the dGTPase system is a TA system. While other types of TA systems have been demonstrated to have antiphage properties, whether they are activated in a similar mechanism as the Type III systems is unclear. Recently, the DarTG type II TA system was shown to provide phage defense by ADP-ribosylating phage DNA to disrupt DNA replication [159]. ParST,

another type II TA system, exerts its effect via modification of cellular target Prs, which is involved in nucleotide biosynthesis, though the ParST system has not been demonstrated to be involved in phage defense. The mechanism of AvcID bears a resemblance to both DarTG and ParST but is distinct from both in terms of the mechanism for toxin activation and different types of TA system.

Prior work on Type III TA systems suggests they are associated with abortive infections (Abi) due to cell death [121]. However, overexpression of AvcD does not lead to cell death but does impair genome replication, and this can be rescued by removing inducer or co-overexpressing *avcI* in trans [147]. Given that *avcI* is degraded, subsequently releasing existing AvcD to deaminate dC pools upon phage infection, we propose that protection conferred by AvcD is not through Abi. Instead, it is the phage infection that leads to cell death, with AvcD depleting available nucleotides for phage to utilize and thus decreasing the production of new functional virions in order to protect uninfected cells in the population.

Chapter 4 – Concluding Remarks

4.1: Conclusions and Significance

Antiphage defense systems are ubiquitous and often associated with MGEs in diverse microorganisms, including the human pathogen *V. cholerae*. Accumulating evidence from various antiphage defense systems suggests the importance of the bacterial immune system and its equivalent role in the eukaryotic immune system. Following the correology study developed by our collaborators in the Eva Top group, we found that AvcD is a deoxycytidylate deaminase and is post-translationally inhibited by a sRNA called AvcI. AvcD is also widely conserved in prokaryotes and eukaryotes, yet we only found AvcI homologs in prokaryotes. We also demonstrated AvcID is a novel subclass of the Type III TA system in that the toxin is not an endoribonuclease but a deoxycytidine deaminase. Phage-induced transcriptional shutoff leads to the activation of AvcID by degrading labile AvcI antitoxin, thus releasing AvcD to deplete nucleotides. Importantly, AvcID provides protection against phage infection by inhibiting phage replication and reducing the viability of the phage. Furthermore, Ung does not function together with AvcID, despite the increased level of dUMP after phage infection. In summary, we have demonstrated that AvcID is a newly described antiphage defense system.

Altering pools of available nucleotides to consequently inhibit phage replication and transcription has been shown to provide phage defense in other systems. For example, prokaryotic viperins protect against T7 phage infection by producing modified ribonucleotides that ultimately inhibit phage polymerase-dependent transcription [160]. Recently discovered by the Sorek group, the dGTPase system dephosphorylates dGTP to dG and leads to inhibition of phage replication [129]. In addition, the DarTG TA

system ADP-ribosylates phage DNA, eventually causing phage replication to stall [159]. This suggests that manipulating nucleotide pools is a conserved function of some TA systems and antiphage defense mechanisms.

Through evolution, both bacteria and phage will acquire resistance to combat one another. When we infected *E. coli* expressing different *avcID* homologs with lytic phages, we saw different *avcID* homologs confer protection against different phages. This could be possibly due to: 1) the naïve heterologous host *E. coli* lacks necessary regulatory components for these *avcID* homologs; or 2) the phage could have evolved mechanisms to counter this system. It is unclear why *avcID* derived from *P. mirabilis* confers protection against T4 phage in *E. coli*. It might be due to competition for resources since T-even phages are known to possess enzymes that can modify deoxycytosine-containing bases to evade bacterial RMs (Figure 4.1) [161]. This may explain the diversity of nucleotide-depletion strategies utilized by bacteria in order to fend off phage's evolved counter defense. We also speculate that while the AvcID and other nucleotide-depleting antiphage systems are activated amidst phage infection, these systema provide an opportunity to synergize with other phage defense systems, such as CBASS or RMs, to further target invading phages [8, 162].

Numerous antiphage systems have been uncovered through close-proximity associations within defense islands, while others were unearthed through studies of individual mechanisms in one or a few similar species of bacteria [7, 9, 121, 147, 163, 164]. Fast forward to today, two major computational pipelines, PADLOC and DefenseFinder, have been developed to detect and categorize antiphage defense systems in a given genome [165, 166]. Not only do these bioinformatic tools serve as a

comprehensive resource to investigate antiphage defense systems but they also highlight the importance of studying bacterial immune systems. Furthermore, the emergence of multi-drug resistant strains of bacteria has led to renewed interests in phage therapy [167–169]. Therefore, exploring these antiphage defense systems will not only serve as vital tools for molecular biology, but also potentially influence the clinical phage therapeutic setting.

	Coliphages									
	Myoviridae			*Podoviridae*		*Siphoviridae*				*Microviridae*
	T2	T4	T6	T3	T7	λ-vir	T5	SECΦ18	SECΦ27	SECΦ17
AvcID systems	VC	VC, PM	VC, VP	VC, VP, ETEC	None	None	VC, VP	VP, ETEC	ETEC	ETEC
Genome										
size	163.8 kb	168.9 kb	168.7 kb	38.2 kb	39.9 kb	48.5 kb	121.8 kb	44.7 kb	51.5 kb	5.5 kb
type	linear	linear	linear	linear	linear	linear	linear	linear	linear	circular
GC %	35.3%	35.3%	35.2%	49.9%	48.4%	49.8%	39.3%	54.4%	44.7%	45.9%
Proteins involve in nucleotides/ host defense pathway	• dCTP pyrophosphatase • dCMP deaminase • dCMP 5-hydroxymethyltransferase	• dCTP pyrophosphatase • dCMP deaminase • dCMP 5-hydroxymethyltransferase • dCMP hydroxymethylase • Thymidylate synthase	• dCMP deaminase • dCMP 5-hydroxymethyltransferase • dCMP hydroxymethylase • Thymidylate synthase		• Inhibitor of TA system (4.5) • Inhibitor of dGTPase (1.2)		• dUTPase • Probable thymidylate synthase • Ung-binding protein to cleave dUMP containing DNA (T5.015)			

Figure 4.1. Summary of AvcID homolog phage resistant profiles and genomic characteristic of listed coliphages. VC = *Vibrio cholerae*, VP = *Vibrio parahaemolyticus*, PM = *Proteus mirabilis*, ETEC = *E. coli* ETEC.

This dissertation has focused on understanding the mechanisms and functions of a novel TA system and has demonstrated its role in antiphage defense. Combining several interdisciplinary approaches as described herein can turn out to be a powerful means to recognize the significance of these widely conserved antiphage system. The work described here has led to considerable advances in several fields, including TA

systems, antiphage defense, and the role of genomic islands in *V. cholerae*, and laid the fundamental groundwork for many future research directions. Not only do these results build on our current understanding of the mechanisms and targets of the AvcID TA system, but they also raise further questions. For example, are there persistent cells within the population after phage infection; what is the complex structure of AvcD and AvcI; and what is influencing the phage specificity targeting by different AvcID homologs? In the following sections, I discuss these questions in more detail.

4.2: Future Directions

4.2.1 Single Cell Analysis of *avcID* Expression During Phage Infection

To further understand the studies of the AvcID system, analysis of the spatial and temporal heterogeneity of *avcID* expressing cells would provide insight into how cells are protecting themselves from phage infection. The data in Chapters 2 and 3 that are associated with phage infection experiments are often presented on a population level, but not a single-cell level. Furthermore, only a minimum fraction of cells that escape phage infection due to statistical variation can be calculated by using the Poisson distribution e^{-MOI}. According to the Poisson distribution, if the MOI is 1, then the probability of a cell receiving no phage and remaining uninfected is at least 0.37 (e^{-1}) or 37% of the cells remain uninfected. In other words, at most ~63% of the cells are infected at an MOI of 1. Even this is an overestimation of the fraction of cells infected since some of the viruses never actually infect a cell [170, 171]. Utilizing a florescent promoter-reporter of *avcID* and automated confocal fluorescence microscopy using Florescence in situ hybridization (FISH) to track the position of the phage throughout the

phage infection process, one can begin to explore the positioning and timing of gene expression of the cells that are infected compared to the cells that are not. Additionally, these gene expression patterns can be correlated with droplet-based single cell RNA sequencing techniques that can uncover the structure of the microbial communities in real time [172, 173]. As an alternative to looking at gene expression, I will tag hemagglutinin (HA) affinity tag on the C-terminus of AvcD using fluorescent-conjugated antibody to delve into the spatial-temporal dynamics of AvcD activity when encountering phage. Determining the phage infection process on a single cell level by combining information on gene expression and spatial and temporal dynamics of the cells would provide an important understanding into the bacteria-phage interaction paradigm.

4.2.2 Mechanism of Post-translational Inhibition of AvcD by AvcI

The fluorescence microscopy data and single-cell image analysis suggest that AvcI inhibits AvcD post-translationally. The in vitro EMSA data also suggest that AvcI and AvcD form a complex. While the RpfR control protein does not interact with AvcI, AvcD still forms an in vitro complex with the reverse complement of AvcI (RC-AvcI). However, the interaction of AvcD with AvcI is much stronger than the interaction with RC-AvcI. In order to show direct specificity, I propose employing another small RNA, such as Qrr sRNAs that are involved in quorum sensing, as a negative control for AvcD binding specificity [174]. Additionally, the stoichiometry and cooperativity of AvcI and AvcD remains unclear before and during phage infection. The stoichiometry of the ToxIN TA system involves assembly itself into a self-closing triangular 3 ToxN:3ToxI complex before the complex dissociates after phage infection [121]. Thus, I propose co-

crystallizing AvcD with AvcI to maintain a stabilized and natural protein-RNA complex formation inside the cell. The purification of AvcD is being carried out by our current postdoc Dr. Micah Ferrell. Micah has successfully purified AvcD with a maltose-binding protein (MBP) affinity tag and demonstrated activity of purified AvcD. However, the activity dramatically reduced after the cleavage of the MBP tag, suggesting instability of the protein. The purification requires an optimized protein purification protocol by possibly co-expressing AvcI to enhance the stability of the protein as shown for other toxin-antitoxin complex [175]. The purified proteins would undergo a deamination assay assessment. The hypothesis is that AvcI would bind to AvcD after purification, and it would show no deamination activity. On the other hand, another reaction would include a RNase-treated sample that will reduce the level of AvcI inhibitor and thus release the AvcD deamination activity. Assuming the co-purification is successful, the complex will be subjected to protein crystallization processes and the structural model would provide insight into the complex formation. The structural model would also determine the binding interaction sites between AvcI and AvcD. Interestingly, in Chapter 2, we uncovered several AvcD variants that cannot be inhibited by AvcI, and those substitutions occur on the exterior of the protein. We hypothesized that those substitution sites could be the interaction sites for the complex formation. Likewise, with high resolution, the structural model would inform us of the binding motif of AvcI since there seems to be low consensus in term of nucleotide identity among AvcI homologs. Furthermore, we can apply this to other AvcID homologs and further characterize this interaction. Based on the cross-species inhibition studies, *V. cholerae* AvcI can inhibit *V.*

parahaemolyticus AvcD in trans and vice versa. We can also apply the co-crystallization experiment and determine the heterocomplex of $AvcD_{VP}$:$AvcI_{VC}$.

4.2.3 Mechanism of Phage Specificity

The finding that different AvcID homologs have different phage resistance profiles is interesting yet warrants further investigation on the mechanism of how this occurs. The EOP results demonstrated that AvcID derived from *V. parahaemolyticus* provides protection against T3, T5, T6, and SECΦ18 phages, yet in liquid culture condition, it provides robust protection against solely T5 and SECΦ18 phages. While the EOPs are measuring PFUs and the liquid infection is measuring growth of the host, the cells have varied physiological and gene expression in both conditions. Nevertheless, T5 and SECΦ18 phages are sensitive to cells with *avcID.* To further understand the specificity of phage-resistance profiles from these *avcID* homologs, I propose to perform a phage evolution experiment involving passaging through an *E. coli* host with AvcID several times until a phage mutant that shows increased resistance compared to the ancestor phage. The Sorek group has identified mutation in the phage DNA polymerase when the phages have encountered and eventually escaped the dCTP deaminase (AvcD homolog) phage defense. However, the mechanism by which this escaped mutant phage overcomes the antiphage system is unknown. Therefore, I hypothesize that our mutant phage would have mutation(s) in the DNA polymerase gene and other genes involved in nucleotide metabolism pathways. Once I have isolated the mutant, I will titer the mutant phage using the EOP and liquid culture infection methods by comparing *E. coli* with or without *avcID*. The mutant and the wild-

type phages will subsequently be sequenced using Illumina whole-genome sequencing. The sequences will then be aligned to the reference genome and wild-type phage, and the identification of mutations will be mapped to the wild-type phage. To accelerate selection of mutant phage, I alternatively propose performing random mutagenesis of coliphages using the chemically accelerated viral evolution (CAVE) technique, in which chemical mutagens are utilized to introduce single-nucleotide mutations into the phage genome [176]. Once the library of mutant phage is generated, I will generate targeted single phage mutants as previously described [176]. Likewise, I will assess the mutant phages using EOP and liquid culture infection methods. Interestingly, T7 phage is one of the phages that was sensitive to the dCTP deaminase studied by the Sorek group, yet none of our AvcID homologs show protection against it. We utilized similar methods to study phage defense in which we cloned the *avcID* with its native promoter on a plasmid and then conjugated into *E. coli* MG1655. A fundamental difference between our and Sorek's experiments is that the origin of replication of our vector backbone is RSF1010 while theirs is p15A. The copy number of the plasmid depends on the origin of replication, and the Sorek group demonstrated that the copy number of plasmid influences the outcome of the phage resistance profile [7]. Therefore, I will re-clone tested *avcID* homologs into a backbone with p15A *ori* and compare the phage resistance profile.

APPENDICES

APPENDIX 1

Potential Second Messenger pGpG Activating Type III Secretion System in *Erwinia amylovora*

Appendix 1.1: PREFACE

The work presented in this chapter is a collaboration between me, Dr. Roshni Kharadi, and Asan Turdiev. Dr. Kharadi generated the *Erwinia amylovora* mutants, and measured *hrpS* gene expression level in vivo as well as intracellular c-di-GMP and pGpG levels. Additionally, Dr. Kharadi performed the virulence pear assay. Mr. Turdiev performed the DRaCALA and pGpG hydrolysis experiments.

Appendix 1.2: ABSTRACT

The Gram-negative plant pathogen *Erwinia amylovora*, the causative agent of fire blight, reciprocally regulates chronic biofilm formation and the Hrp type III secretion system (T3SS) during systemic plant infection. This regulation is partially modulated by cyclic di-GMP (c-di-GMP), a ubiquitous bacterial second messenger, but the mechanism by which c-di-GMP regulates the transition is unknown. C-di-GMP is synthesized by diguanylate cyclase enzymes in response to specific environmental cues, and it is degraded by two different phosphodiesterase (PDE) enzymes, an EAL domain that generates pGpG and an HD-GYP domain that generates two GMPs. However, *E. amylovora* only encodes EAL proteins suggesting c-di-GMP is degraded to pGpG. A signaling role for pGpG has not yet been described. In addition, the regulation of the T3SS is partially controlled by the transcription factor HrpS, a σ^{54}-dependent enhancer binding protein (EBP). As c-di-GMP is known to associate and regulate EBPs in other bacteria, we therefore initially hypothesized that HrpS might be the regulatory protein at which c-di-GMP exerts its effects. Our preliminary results determined that high intracellular concentrations of pGpG, generated by expressing an EAL PDE,

correlated with HrpS induction of its own transcription. Alternatively, expression of an HD-GYP PDE, which similarly reduced c-di-GMP but did not generate pGpG, did not induce *hrpS*. We further demonstrate that HrpS does not bind to c-di-GMP or pGpG but instead binds to GTP, the precursor to synthesize c-di-GMP. Our current model is that high concentrations of pGpG are generated during dispersal from biofilms due to the breakdown of c-di-GMP. pGpG functions as a novel signal to regulate the GTP pool, which in turn induces *hrpS* to activate T3SS transcription. Future experiments are testing this model by recapitulating both HrpS regulation and GTP modulation by pGpG *in vivo.* Many well-known pathogens only encode EAL PDE with no HD-GYPs, suggesting such a signaling role for pGpG could be widespread.

Appendix 1.3: INTRODUCTION

The Gram-negative plant pathogen *Erwinia amylovora* is the causative agent of fire blight, causing huge economic losses in commercial apple and pear production worldwide [177]. In 2000, Michigan's agricultural economy lost approximately $42 million due to a fire blight outbreak [178]. *E. amylovora* infection can lead to systemic disease manifestations, including flower necrosis, tissue cankers, and bacterial ooze [179]. The bacterial ooze serves as an inoculum for other susceptible hosts as it can be spread by rain or insects, rendering it difficult to control [180]. Currently, most fire blight control methods are preventive, such as quarantine, pruning and/or removal of diseased plant parts, as well as use of biological and chemical agents such as antibiotics and pesticides [181, 182]. However, effective control methods are still lacking and breeding resistant cultivars is time consuming. The occurrence of antibiotic-resistant *E.*

amylovora isolates against the current antibiotics to treat fire blight, such as streptomycin, is escalating [183].

E. amylovora invades through natural openings (*e.g.* nectarthodes of flowers), and wounds in a susceptible host and can spread systemically via the vascular system [184]. Successful infection of a susceptible host plant depends on two major virulence determinants: hypersensitive response and pathogenicity- type III secretion system (*hrp*-T3SS) and the production of the exopolysaccharide (EPS) amylovoran [181]. The roles of *hrp*-T3SS and amylovoran are to translocate bacterial effectors proteins into the host plant cell cytosol, and to facilitate the structural formation of biofilm, respectively [185, 186]. Once invaded, *E. amylovora* utilizes the *hrp*- T3SS to repress the host immune system [187]. Following the *hrp*-T3SS acute infection state, *E. amylovora* cells secrete amylovoran to form biofilm. Concomitantly, this disrupts the water flow and the physical integrity of the plant vessels, allowing the bacteria to continue to disseminate throughout the host via vascular tissues and cortical parenchyma where they repeatedly use the *hrp*-T3SS, ultimately causing systemic necrotic lesions [186, 188].

The transition between these two phenotypes is orchestrated by the ubiquitous bacterial second messenger bis-(3',5')-cyclic dimeric guanosine monophosphate (c-di-GMP). In response to changes in the environment, diguanylate cyclase (DGC) and phosphodiesterase (PDE) enzymes modulate the intracellular concentration of c-di-GMP by synthesizing and degrading c-di-GMP, respectively. DGCs encode a conserved GGDEF domain, which synthesizes c-di-GMP from two GTPs. PDEs encode an EAL domain that generates 5'- phosphoguanylyl-(3'-5')-guanosine pGpG, or an HD-GYP domain that generates two GMPs (Fig. 5.1). Additionally, HD-GYPs can also degrade

pGpG produced by EALs [189, 190]. Some bacteria also utilize oligoribonucleases, such as Orn, to complete c-di-GMP signaling by hydrolyzing pGpG to yield GMPs (Fig. 4.1) [191–193]. In *E. amylovora,* elevated concentrations of c-di-GMP are associated with increased expression of amylovoran which drives the formation of biofilms found in chronic infections, whereas reduced levels of c-di-GMP induce *hrp*-T3SS expression which is responsible for acute infection [186, 194]. The transition between these two phenotypes is orchestrated by the ubiquitous bacterial second messenger bis-(3',5')-cyclic dimeric guanosine monophosphate (c-di-GMP). In response to changes in the environment, diguanylate cyclase (DGC) and phosphodiesterase (PDE) enzymes modulate the intracellular concentration of c-di-GMP by synthesizing and degrading c-di-GMP, respectively. DGCs encode a conserved GGDEF domain, which synthesizes c-di-GMP from two GTPs. PDEs encode an EAL domain that generates 5'-phosphoguanylyl-(3'-5')-guanosine (pGpG), or an HD-GYP domain that generates two GMPs (FIG 5.1). Additionally, HD-GYPs can also degrade pGpG produced by EALs [189, 190]. Some bacteria also utilize oligoribonucleases, such as Orn, to complete c-di-GMP signaling by hydrolyzing pGpG to yield GMPs (Fig. 5.1) [191–193]. In *E. amylovora,* elevated concentrations of c-di-GMP are associated with increased expression of amylovoran which drives the formation of biofilms found in chronic infections, whereas reduced levels of c-di-GMP induce *hrp*-T3SS expression which is responsible for acute infection [186, 194].

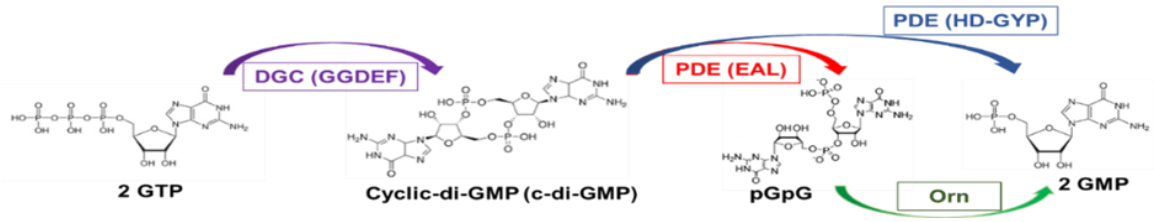


Figure 5.1 Schematic representation of c-di-GMP synthesis and degradation. The modulation of c-di-GMP levels is achieved by a diguanylate cyclase (DGC) or a phosphodiesterase (PDE). C-di-GMP is synthesized from two GTP molecules by domains of GGDEF and hydrolyzed to two different products, pGpG and GMP, by two different domains of PDE, EAL and HD-GYP, respectively. Oligoribonuclease Orn directly hydrolyzes pGpG to GMP.

Currently, c-di-GMP regulation is mediated by its interaction with effector proteins, riboswitches, and transcription factors, such as enhancer binding proteins (EBPs). The interaction of c-di-GMP with these effectors influence regulatory processes at multiple levels: transcriptional, post-transcriptional, and post-translational [195]. Recent studies have shown that c-di-GMP modulates the virulence factors in *E. amylovora* and other plant pathogens in vivo [186, 194, 196]. The mechanism by which second messengers regulate the transition between biofilm and *hrp*-T3SS remains poorly understood and has not been recapitulated in vitro. Currently, c-di-GMP regulation is mediated by its interaction with effector proteins, riboswitches, and transcription factors, such as enhancer binding proteins (EBPs). The interaction of c-di-GMP with these effectors influence regulatory processes at multiple levels: transcriptional, post-transcriptional, and post-translational [195]. Recent studies have shown that c-di-GMP modulates the virulence factors in *E. amylovora* and other plant pathogens in vivo [186, 194, 196]. The mechanism by which second messengers regulate the transition between biofilm and *hrp*-T3SS remains poorly understood and has not been recapitulated in vitro.

To infect its hosts, *E. amylovora* uses a set of clustered genes called *hrp,* which is located on an apparent pathogenicity island [177, 182]. The expression of *E. amylovora hrp*-T3SS is encoded by HrpL, an extracytoplasmic function (ECF) sigma factor. *hrpL* is partially controlled at the transcriptional level by HrpS, which belongs to the NtrC family of σ^{54}-dependent bacterial enhancer binding protein (bEBP) [185, 197–199]. HrpS autoregulates its own expression through activation of the *hrpL* promoter (P*hrpL*) and generation of a polycistronic mRNA for *hrpLXY*, and HrpY activates *hrpS* by binding to its promoter [185, 200]. Moreover, the Rcs phosphorelay is also a *hrpS* activator [185, 201]. HrpS protein consists of three conserved domains: a N-terminal receiver domain (Rec), a central AAA+ (ATPases associated with various cellular activities) involved in ATP hydrolysis and σ^{54}-interaction, and a C-terminal helix-turn-helix DNA binding domain [202, 203]. The Rec domain of bEBPs is usually around 30 amino acids long, but the N-terminal domain of HrpS is unusual in that it has a truncated Rec domain (Fig. 5.2) [201]. Several members of the bEBP family of transcription factors have been demonstrated to respond to c-di-GMP, including FleQ in *Pseudomonas aeruginosa,* FlrA and VpsR in *Vibrio cholerae*, and Clp in *Lysobacter* and *Xanthomonas* [196, 204–206]. For instance, high intracellular concentrations of c-di-GMP inhibit FlrA and promote VpsR activities to down-regulate and up-regulate the expression of genes involved in flagella and biofilm formation, respectively, and vice versa [205, 206]. C-di-GMP also regulates secretion systems in other bacteria by associating with different effectors [196]. The contribution of c-di-GMP for the induction of biofilm formation and repression of virulence and motility is becoming evident [186, 194]. However, the regulation of second messengers on HrpS activating the *hrp* system

is not clearly understood. Experiments showed that the high intracellular level of pGpG leads to the increased expression level of *hrpS*. In addition, although HrpS does not interact with pGpG but it does interact with and hydrolyze GTP. Pear infection data however demonstrated that cells with high levels of pGpG is required for virulence in *E. amylovora*. Together, the work presented here adds to the complexity of *E. amylovora hrp*-T3SS with potential second messenger pGpG as important virulence factor.

Figure 5.2 Alignment of EBPs. Alignment shows that HrpS lacks the N-terminal Rec domain relative to other EBPs. EA (*E. amylovora*); VC (*V. cholerae*); AA (*Aquifex aeolicus*); PA (*Pseudomonas aeruginosa*).

Appendix 1.4: MATERIALS AND METHODS

A1.4.1: DNA manipulation and growth conditions

The strains, plasmids, and primers used in this study are listed in Appendix 2 (Tables 1-3). Chromosomal deletion mutants in *E. amylovora* WT Ea1189 (accession no. FN666575) were constructed using the lambda red recombinase protocol described previously [207, 208]. All *Erwinia amylovora* and *Escherichia coli* strains used were grown in Luria-Bertani (LB) broth and agar medium at 28°C and 37°C, respectively. Media were amended with the following antibiotics as needed: ampicillin (100 µg/mL), chloramphenicol (10 µg/mL), kanamycin (100 µg/mL).

A1.4.2: RNA purification and qRT-PCR

Strains were grown overnight and then washed and resuspended in *hrp*-inducing minimal medium supplemented with antibiotics and 1.0 mM IPTG for 6 hours at 28°C [194]. RNA was extracted using Direct-zol RNA kit (Zymo) and cDNA synthesis was carried out using High-Capacity cDNA reverse Transcription Kit (Applied Biosystems™). qRT-PCR (Applied Biosystems™) was used to quantify gene expression for *hrpS* with *recA* as endogenous control. This assay was repeated three times with three technical replicates in each of the biological replicates.

A1.4.3: Nucleotide quantification using UPLC/MS-MS

Ultra-performance liquid chromatography coupled with tandem mass spectrometry (UPLC-MS/MS) was used to quantify intracellular c-di-GMP, pGpG, and GTP levels [186, 192]. Overnight cultures were washed and resuspended in *hrp*-inducing minimal medium supplemented with the appropriate antibiotics and 1.0 mM IPTG for 6 hours at 28°C. After normalizing the OD_{600} of all the strains, 5 mL of culture was pelleted and the nucleotides were extracted with 200 µL extraction buffer (40% methanol, 40% acetonitrile, 0.1N formic acid) and incubated for 1 h at -20°C. The cells were then pelleted, and supernatant was vacuum dried and resuspended in 200 µL of ultrapure water. C-di-GMP and pGpG were quantified separately from each of the final resuspensions using Quatro Premier XE™ (Waters). GTP was quantified using TQ-S mass spectrometer (Waters) with a parameter previously described [147]. The conditions of the MRM transitions were as follows [cone voltage (V), collision energy (eV)]: GTP, 522> 159, (15, 34).

A1.4.4: Protein Purification of HrpS

pHrpS-6xHis was transformed into *E. coli* BL21 (DE3) and grown at 37°C until OD_{600} of 0.5 and induced with 500 μM IPTG for 4 hours at 37°C or overnight at 18°C. The cells were then pelleted at 7k x g for 20 min. Cell pellets were resuspended in Buffer A (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 2 mM β-mercapoethanol, and 20% glycerol) and lysed by sonication using a Branson 450 Digital Sonifier (20% amplitude, 20 sec total, 2.5 sec on, 2.5 sec off). Crude lysates were centrifuged at 15k x g for 20 min at 4°C. Insoluble pellets were recovered as previously described [209]. Briefly, insoluble pellets were washed three times with Buffer B (20 mM Tris-HCl pH 8.0, 300 mM NaCl, and 2% SDS) and sonicated again until it became clear. Excess SDS was removed at 4°C for 3 hour or overnight. After centrifugation at 15k x g for 10 min at 4°C, KCl was added to the supernatant at the final concentration of 400 mM. SDS-KCl insoluble crystal that formed overnight at 4°C. were separate by centrifuging the samples at 15k x g at 4°C, and the remaining supernatant was subjected to Ni^{2+} resin purification. The column was washed with three column volume of Buffer C (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole). Proteins were eluted with a final concentration of 500 mM imidazole. The eluted proteins were further subjected to dialysis with 10kd-cut off dialysis tube to remove excessive imidazole. The purity of the protein was assessed using SDS-PAGE gel stained with Coomassie brilliant blue dye.

A1.4.5: NTPase Activity Assay

NTPase activities were measured as previously described using BIOMOL Green reagent (Enzo Life Sciences) as directed [210]. The reaction mixture contained 1 μM

protein, 5 mM ATP or GTP, and 5 mM $MgCl_2$, 100 mM HEPES pH 8.5, 65 mM NaCl, 5% glycerol. Reactions were incubated at 37°C for 0, 5, 10, 15, and 20 min and assayed for the release of inorganic phosphate. At each time point, a five µL aliquot was removed from the reaction and diluted 1:50 in the prepared 1.5 mL tube containing 245 µL HEPES/NaCl/Glycerol (HNG) buffer and immediately frozen in the dry ice/ethanol slurry. The amount of phosphate released was determined by comparing sample absorbance at OD_{650} with those of a phosphate standard curve. Data reported were from three separate samples of the same purified proteins assayed in duplicate.

A1.4.6: Electrophoretic Mobility Shift Assay (EMSA)

EMSA reactions were carried out by incubating purified HrpS-6xHis with 5'-FAM-labeled probes (FAM-*hrpL*). 20 µL reaction mixtures consisted of 1X binding buffer (50 ng/µL Poly(dI-dC) (Sigma), 0.5 mM $MgCl_2$, 0.1% IGEPAL CA-630, 0.05 mg/mL BSA and 5% glycerol), 2.5 nM probe, and different concentrations of protein [185]. When indicated, GTP was added at a final concentration of 5 mM; otherwise, an equal volume of water was added. A 100x molar excess competitor was added when indicated. All components except the labeled probe were mixed at room temperature and incubated for 10 min. FAM-*hrpL* was then added and the reaction mixtures were incubated for 30 min at 30°C. Reaction mixtures were loaded on a prerun 6% polyacrylamide Tris-borate-EDTA (TBE) gel, and electrophoresis was carried for 90 min at 90V at 4°C. Fluorescent band migration was detected using Typhoon FLA 9000 imager (GE Healthcare Life Sciences).

A1.4.7: DRaCALA in vitro binding assay and nucleotide hydrolysis

DRaCALA was carried as previously described with slight modifications [192] Proteins were mixed in 1x Buffer D (100 mM NaCl, 10 mM Tris-HCl pH 8.0, 5 mM $CaCl_2$) with 66 pM ^{32}P-radiolabeled nucleotide and allowed to incubate for 10 min at room temperature. The mixture was then applied to nitrocellulose sheets, dried, and imaged using a FLA7100 Fujifilm Life Science PhosphorImager. The fraction bound was quantified using Fujifilm Multi Gauge software v3.0.

For the hydrolysis experiment, proteins were mixed in 1X Buffer E (100 mM NaCl, 10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$). Before adding the protein, 1 µL aliquot was taken out, resuspended in 5 µL 0.2 M EDTA and then boiled for 5 min. After adding the protein, the reaction mixture was incubated at room temperature. 1 µL was taken out at each subsequent timepoints and resuspended in 5 µL 0.2 EDTA and boiled for 5 min. Boiled samples were ran on TLC to assess nucleotide hydrolysis.

A1.4.8: Virulence Assay

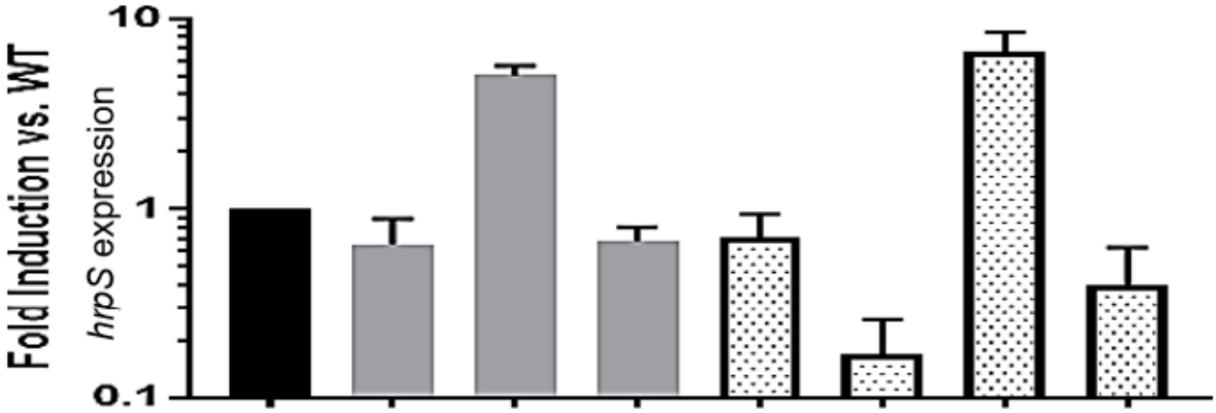
Virulence assays were conducted on immature pear fruit (*Pyrus communis* cv. Bartlett), as previously described [211]. Briefly, overnight cultures were harvested by centrifugation, and were resuspended in 0.5X sterile phosphate buffered saline (PBS). Bacteria were then inoculated on stab-wounded immature pears at a concentration of 10^4 CFU/mL, followed by an incubation at 28°C. Data were collected from immature pears in the form of necrotic lesion diameters. This experiment was repeated at least twice, with a minimum of three technical replicates per strain in each experiment.

Appendix 1.5: RESULTS

A1.5.1: Increased pGpG induces the *hrpS* expression

C-di-GMP signaling systems are regulated in a complex network. For instance, *V. cholerae* encodes over 60 DGCs and PDEs [212], making it difficult to define the function of each enzymes. To date, the functional role of c-di-GMP has been characterized, but the role of pGpG remains enigmatic. However, genetic manipulation in *E. amylovora* readily supports investigation into c-di-GMP regulatory pathways because it only encodes five putative genes with GGDEF domains (*edc* genes), three putative genes with EAL domains (*pde* genes), and no putative genes with an HD-GYP domain. *E. amylovora* thus cannot degrade c-di-GMP to two GMPs directly but instead is likely to rely on Orn to resolve pGpG [192]. The simple genetic model of *E. amylovora* is therefore a good system to study the potential signaling role of pGpG as a second messenger, as virtually all c-di-GMP must ultimately pass through this intermediate step during the completion of c-di-GMP dependent signaling. To explore the role c-di-GMP and pGpG have on the expression of *hrp*-T3SS, we measured the expression of *hrpS* by RT-qPCR as a proxy for *hrp*-T3SS expression. Using lambda-red recombination, our collaborators at Sundin lab generated mutants that were defective in degrading c-di-GMP (all three EAL-encoded PDEs knocked out (Δ3)) and synthesizing c-di-GMP (all genes encoding DGCs and PDEs (Δ8)) in *E. amylovora*. Simultaneously, the intracellular concentrations of c-di-GMP and pGpG were measured by LC-MS/MS (Fig. 5.3). In the Δ3 strain, lacking all three EAL-encoded PDEs, increased levels of c-di-GMP were detected with no detectable intracellular pGpG (Fig. 5.3). *hrpS* expression was similar to wild type (WT) in this background. Notably, when expressing an EAL on a

plasmid in Δ3 (Δ3/EAL), we observed an increased intracellular concentration of pGpG and a ~5-fold increase in *hrpS* expression when compared to WT. Conversely, when expressing a HD-GYP on a plasmid in the same background (Δ3/HD-GYP) not only did pGpG not increase but *hrpS* expression also did not change. Likewise, knockouts of all genes encoding DGCs and PDEs (Δ8) had no measurable levels of c-di-GMP or pGpG, with *hrpS* expression similar to that of WT. Expressing a DGC in Δ8 (Δ8/DGC) generated c-di-GMP but no pGpG and led to a decrease in hrpS expression. However, co- expressing an EAL and a DGC in Δ8 strain (Δ8/DGC/EAL) increased intracellular concentration of pGpG and led to an increase in *hrpS* expression compared to Δ8/DGC and WT. In contrast, co-expressing a HD-GYP and DGC in Δ8 (Δ8/DGC/HD-GYP) showed reduced c-di-GMP yet *hrpS* expression resembled the WT. Collectively, the activation of *hrpS* expression is specific to increased intracellular pGpG, owing to expressing HD-GYP—yielding GMP from c-di-GMP—did not increase *hrpS* expression compared to WT.



Strain	WT	Δ3	Δ3	Δ3	Δ8	Δ8	Δ8	Δ8
Plasmid	-	-	EAL	HDGYP	-	DGC	DGC/ EAL	DGC/ HDGYP
C-di-GMP (nM)	265.14	3657.3	183.02	324.82	ND	909.81	38.9	539.3
pGpG (nM)	11.21	ND	244.83	ND	ND	ND	214.91	ND

Figure 5.3. *hrpS* mRNA abundance responds to different levels of nucleotides in vivo. In the indicated strains, some of which contained expression plasmids for a DGC, EAL, and/or HDGYP enzyme: Δ3 is deletions of all PDEs; Δ8 is deletion of all PDEs and DGCs. (--) sign indicates absence of the plasmid. ND designates not detectable. The values represent the mean concentrations of c di GMP and pGpG across all the strains (n=3 biological replicates).

A1.5.2: HrpS does not bind to pGpG or c-di-GMP but instead binds to GTP

The expression of *hrpS* is activated by increased intracellular pGpG; therefore, I hypothesized that HrpS is interacting with pGpG to autoregulate itself. To determine if HrpS binds to pGpG, I purified HrpS-6xHis and subjected it to differential radial capillary action of ligand assay (DRaCALA), which was performed by collaborators at the University of Maryland [213]. HrpS (12.5 µM) was incubated with 66 pM of various ^{32}P-labled nucleotides— c-di-GMP, pGpG, and GTP—independently, and binding fraction was measured (Fig. 5.4A-C). We observed significant binding of HrpS to GTP but not to c-di-GMP or pGpG. To test the specificity of the interaction between HrpS and GTP, we added unlabeled excess indicated nucleotide competitors—GTP, GDP, GMP, and

ATP—to each reaction. As a result, HrpS binds to GTP/GDP with higher specificity compared to other unlabeled competitors (Fig. 5.4D). Next, we performed a pGpG hydrolysis assay to determine if HrpS is hydrolyzing the nucleotide rather than binding to it. Compared to the positive control Orn, which hydrolyzes pGpG to GMP, HrpS neither binds to pGpG nor hydrolyzes it (Fig. 5.4E). Furthermore, we performed a NTPase activity assay to determine if HrpS is hydrolyzing GTP specifically. Compared to the negative control BSA, which has no NTPase hydrolysis activity, HrpS binds to GTP and hydrolyzes GTP (Fig. 5.4 F). In summary, HrpS does not interact with pGpG but instead bind and hydrolyze GTP.

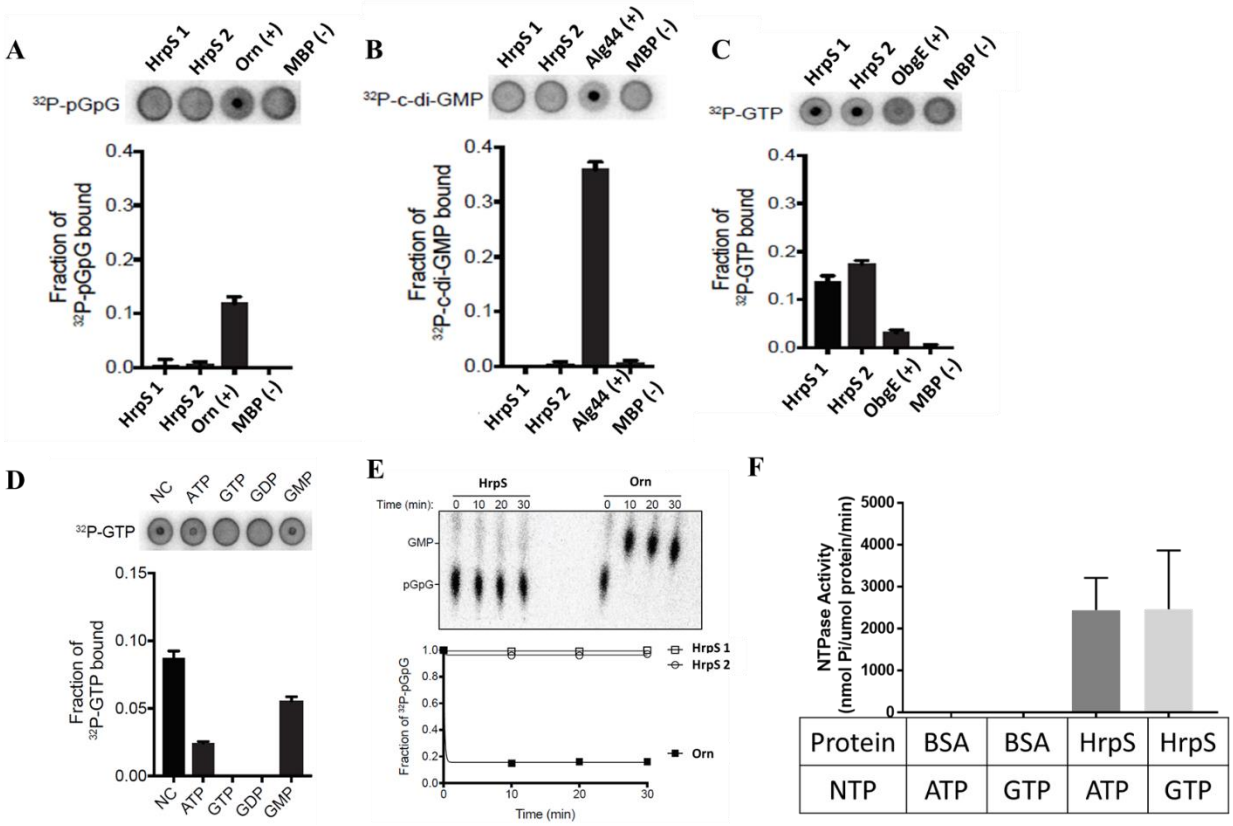


Figure 5.4. Detection of specific protein ligand interactions by DRaCALA. Graphs of fraction bound for each sample (**A**) ^{32}P-pGpG, (**B**) ^{32}P-c-di-GMP, (**C**) ^{32}P-GTP. (**D**) Competition of ^{32}P-GTP binding to HrpS by a variety of excess unlabeled nucleotides. (**E**) Kinetics of pGpG hydrolysis by HrpS and Orn (positive control). HrpS 1 and HrpS 2 come from two different protein preps; Orn, Alg44, and ObgE are positive controls for

Figure 5.4 (cont'd) binding to pGpG, c-di-GMP, and GTP, respectively; and MBP is the negative controls for binding all three nucleotides. Bars indicate means with standard deviations (n=3 biological replicates). (**F**) Nucleotide hydrolysis assay demonstrates that HrpS can hydrolyze ATP and GTP, while Bovine serum albumin (BSA) serves as a negative control. Bars indicate means with standard deviations (n=3 biological replicates).

A1.5.3: Full-length HrpS binds to *hrpL* and its own promoter

Previous studies have demonstrated in vitro binding of HrpS to its target promoters using Electrophoretic Mobility Shift Assays (EMSAs) [198, 202]. However, prior studies utilized only the DNA-binding domain of HrpS to perform EMSAs. HrpS and other classic EBPs utilize GAFTGA motif to interact in response to σ^{54} to activate transcription. In vitro analyses demonstrate that HrpS-activated promoters, such as P_{hrpL}, contain the highly conserved -24 and -12 consensus sequences utilized by σ54-RNA polymerase (RNAP) [198, 202]. To determine if the purified full-length HrpS can bind to its target promoters, we performed EMSA by incubating the *hrpL* promoter probe with various concentration of HrpS protein. Full-length HrpS binds to *hrpL* promoter in a dose-dependent manner (Fig. 5.5). The addition of an excess unlabeled probe composed of HrpS binding site was able to outcompete HrpS binding to the labeled *hrpL* probe. Conversely, when the unlabeled probe lacks the binding site, it was no longer able to abrogate the HrpS-*hrpL* band migration (Fig. 5.5). Intriguingly, we found that HrpS can bind to its own promoter through an unclear mechanism, owning to its promoter consists of σ70 consensus sequence and lacking a *hrpS* binding motif as in the P_{hrpL} (TGCAA-N4-TTGCA) [198]. Moreover, some EBPs are in their inactive dimeric state when bound to their target DNA and are then activated upon nucleotide binding and hydrolysis. Therefore, presumably, HrpS is in its dimeric state when bound to the

probe and requires nucleotide triphosphate to activate its activity (Fig. 5.5). Notably, the addition of GTP causes a higher molecular band shift, suggesting that the proteins are forming higher oligomers to interact with the DNA while bound to GTP.

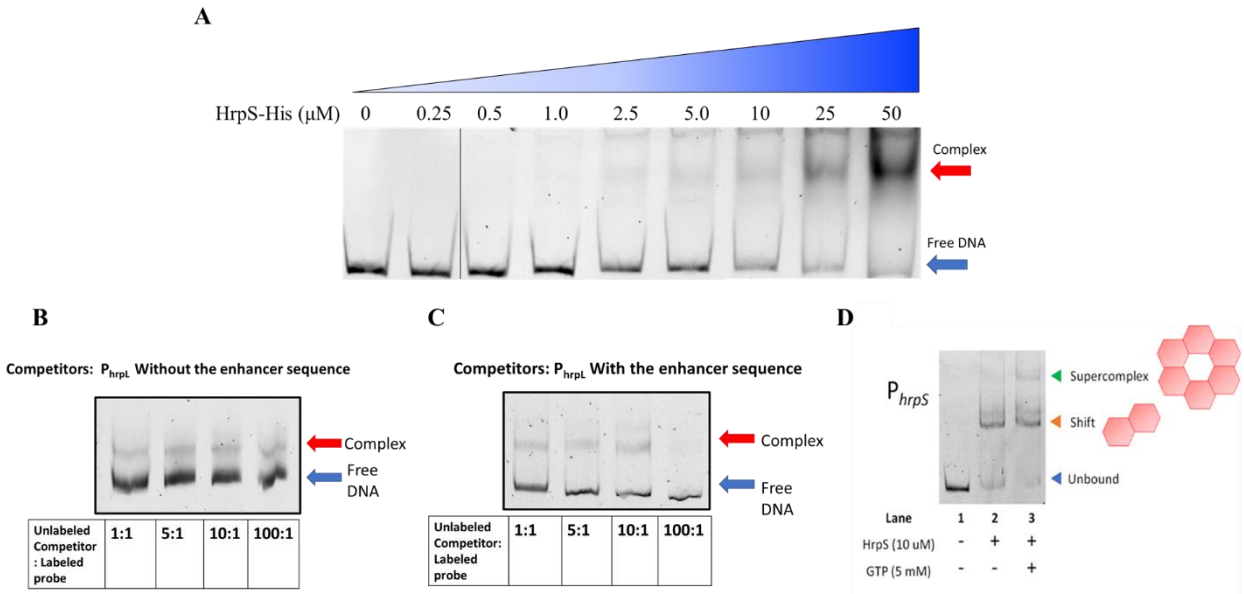


Figure 5.5. Full-length HrpS binds to *hrpL* and GTP enhances HrpS interaction with its promoter in vitro. (**A**) HrpS-6xHis were incubated with a FAM-labeled probe corresponding to the *hrpL* promoter. Unlabeled probe without HrpS binding sites (**B**) or with binding sites (**C**) competitor was added at a different increasing molar excess relative to the labeled probe in reactions with 10 µM HrpS-6xHis. GTP enhances HrpS interaction with the P_{hrpS} by forming higher oligomers in vitro. Lane 1 only contains 361-bp FAM-labeled probe of P_{hrpS}. Lane 2 contains HrpS-HIS incubated with the FAM-labeled probe. Lane 3 is the same reaction conditions as Lane 2 except 5 mM GTP was added to the binding reactions (indicated by the + sign), which creates a higher molecular shift.

A1.5.4: Increased pGpG does not lead to increased GTP to activate *hrpS* expression

The alarmones (p)ppGpp, a stringent response molecule, mediates intracellular GTP levels in Gram-positive and -negative bacteria [214–216]. Since (p)ppGpp and pGpG are both linearized guanine-based nucleotides, we hypothesize that pGpG could

also modulate the intracellular concentration of GTP to induce hrpS expression, and that the resulting HrpS change in activity prompts Hrp-T3SS expression. To determine whether the changes of pGpG correlate with intracellular GTP, we measured intracellular concentrations of GTP in the previously described strains. We found that there is no clear correlation between increasing intracellular pGpG level with increasing GTP or c-di-GMP (Fig. 5.6). In contrast, Δ3 expressing an EAL has lower GTP concentrations compared to WT and Δ3 expressing an HD-GYP. Likewise, Δ8 expressing a DGC and an EAL has lower GTP levels compared to Δ8 either expressing just a DGC or a DGC and an HDGYP, yet it has higher levels than Δ8 (Fig. 5.6). This indicates that changes in pGpG level have no effect on the level of GTP to regulate the expression of *hrpS*.

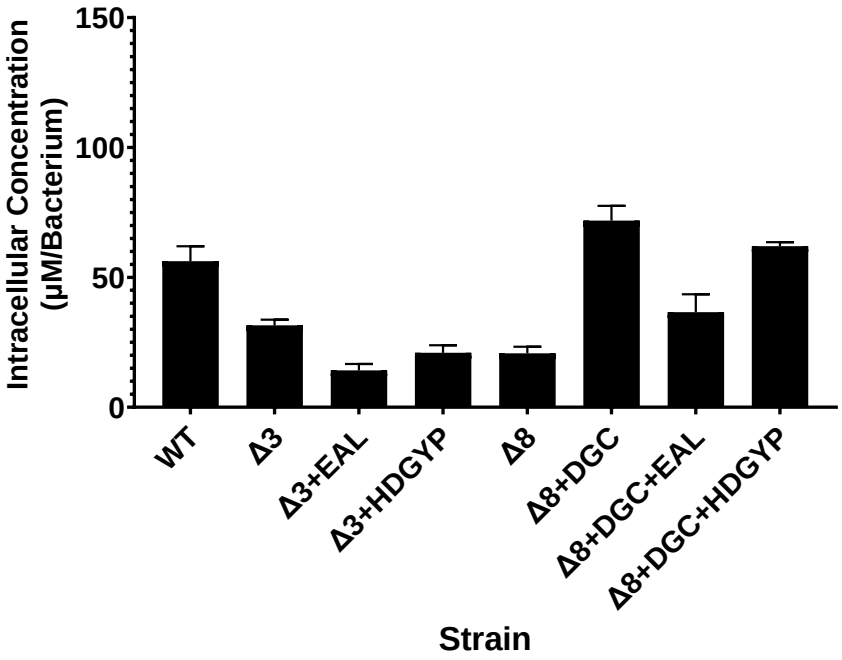


Figure 5.6. Increased pGpG level does not lead to increased GTP level. In the indicated strains, some of which contained expression plasmids for a DGC, EAL, and/or HDGYP enzyme. Δ3 is deletions of all PDEs; Δ8 is deletion of all PDEs and DGCs. (--) sign indicates absence of the plasmid. The values represent the mean concentrations of c di GMP and pGpG across all the strains (n=3 biological replicates).

A1.5.5: Increased pGpG level demonstrates higher virulency

The increased *hrpS* gene expression caused by high levesl of pGpG will upregulate the expression of *hrp*-T3SS pathogenicity island. To assess the effect of pGpG on pathogenesis, our collaborator at the Sundin lab performed an in vivo infection of various strains of *E. amylovora* in immature pears. Infection with the WT strain forms a robust lesion whereas overexpression of a DGC, increasing c-di-GMP, inhibits virulence as previous demonstrated (cite Edmunds/Castiblanco paper) (Fig. 5.7). Decreasing c-di-GMP in the DGC expressing strain only partially restores virulence relative to the WT; however, expressing an EAL in this strain fully complements WT virulence (Fig. 5.7). This result further suggests that pGpG is one of the essential factors for activating/maintaining *hrpS*-T3SS in *E. amylovora* pathogenesis.

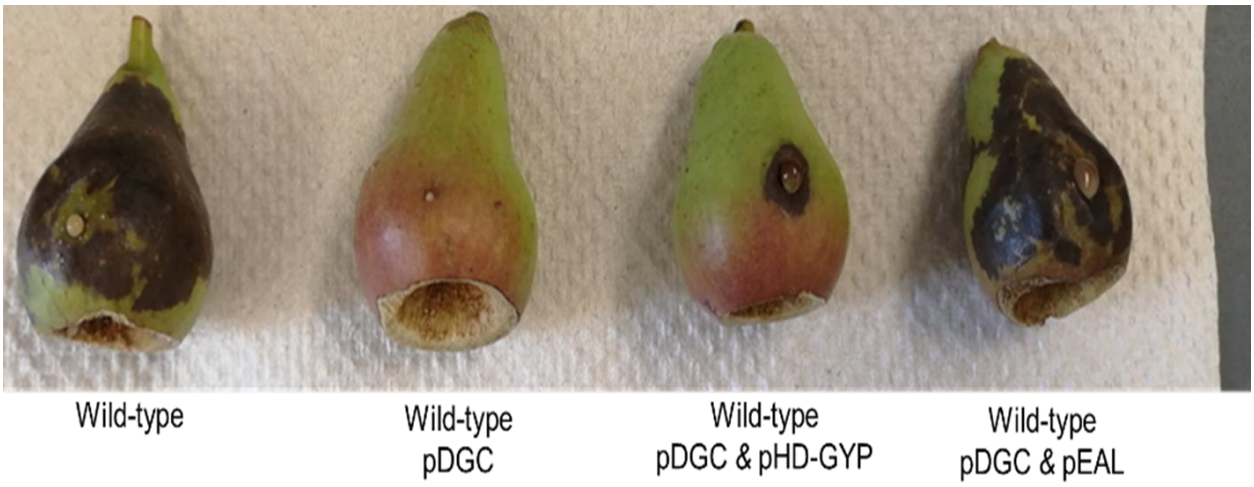


Figure 5.7. Virulence tests on immature pear fruits. Symptoms caused by *E. amylovora* WT, WT containing DGC on plasmid, WT containing DGC and HD-GYP on plasmid, and WT containing DGC and EAL on plasmid. Symptoms were recorded and photos were recorded at 4 days post-inoculation.

A1.6: DISCUSSION

We report the first evidence for a signaling role of pGpG, the breakdown product of c-di-GMP in bacteria, outside of modulation of c-di-GMP concentrations. The elevated level of *hrpS* expression correlates with the increased intracellular level of pGpG, yet HrpS does not interact with pGpG but with GTP. GTP increases the activity of HrpS to bind to its promoter in via an unknown mechanism. Nevertheless, the increasing level of pGpG does not lead to increased GTP for HrpS to utilize. Furthermore, the presence of pGpG is associated with virulence in a pear infection model. However, the effector for pGpG and its role in activating *hrpS* expression, the main regulator of *hrp*-T3SS in *E. amylovora* is still under investigation.

To date, out of 1,131 complete bacterial genomes available in GenBank, 210 (~19%) have both DGCs and EALs but not HD-GYP (Fig. 5.8) [217]. Numerous bacteria, including mammalian pathogens, such as *Escherichia coli*, *Mycobacterium tuberculosis*, *Salmonella typhimurium*, *Shigella flexneri*, and *Yersinia pestis*, also do not encode HD-GYP domain PDEs (Fig. 5.8), suggesting that pGpG may function as a signaling molecule across diverse phylogenetic backgrounds. Furthermore, a complete understanding of the second messenger regulation that contributes to the colonization and infection of *E. amylovora* may provide insight into the pathogenesis of other bacteria lacking PDE with an HD-GYP domain. Demonstrating the potential signaling role of pGpG would allow us to further understand and expand the second messenger-based signaling systems.

Actinobacteria:
Acidimicrobium ferrooxidans DSM 10331
Actinoplanes sp. SE50/110
Actinosynnema mirum DSM 43827
Amycolatopsis mediterranei U32
Arthrobacter aurescens TC1
Arthrobacter chlorophenolicus A6
Arthrobacter phenanthrenivorans Sphe3
Arthrobacter sp. FB24
Cellulomonas fimi ATCC 484
Cellulomonas [Cellvibrio] gilvus ATCC13127
Cellulomonas flavigena DSM 20109
Frankia sp. Ccl3
Frankia sp. EAN1pec
Frankia sp. EuI1c
Frankia symbiont of Datisca glomerata
Geodermatophilus obscurus DSM 43160
Gordonia bronchialis DSM 43247
Micromonospora aurantiaca ATCC 27029
Micromonospora sp. L5
Mycobacterium africanum GM041182
Mycobacterium bovis AF2122/97
Mycobacterium canettii CIPT 140010059
Mycobacterium leprae TN
Mycobacterium smegmatis str. MC2_155
Mycobacterium sp. JDM601
Mycobacterium tuberculosis H37Rv
Nocardia farcinica IFM 10152
Rhodococcus erythropolis PR4
Saccharomonospora viridis DSM 43017
Salinispora arenicola CNS-205
Salinispora tropica CNB-440
Slackia heliotrinireducens DSM 20476
Stackebrandtia nassauensis DSM 44728
Streptomyces cattleya NRRL_8057
Streptomyces griseus subsp. griseus NBRC 13350
Streptomyces scabiei 87.22
Tsukamurella paurometabola DSM 20162
Verrucosispora maris AB-18-032

Aquificae:
Hydrogenobacter thermophilus TK-6
Hydrogenobaculum sp. Y04AAS1
Sulfurihydrogenibium azorense Az-Fu1
Sulfurihydrogenobium sp. YO3AOP1
Thermocrinis albus DSM 14484

Bacteroidetes:
Salinibacter ruber DSM 13855

Cyanobacteria:
Acaryochloris marina MBIC11017
Synechococcus sp. PCC 7002

Firmicutes:
Bacillus amyloliquefaciens FZB42
Bacillus atrophaeus 1942
Bacillus cereus ATCC 14579
Bacillus cytotoxicus NVH 391-98
Bacillus licheniformis ATCC 14580
Bacillus mageterium DSM 319
Bacillus pumilus SAFR-032
Bacillus subtilis subsp. subtilis str. 168
Bacillus weihenstephanensis KBAB4
Clostridium difficile 630
Lactobacillus acidophilus NCFM
Lactobacillus brevis ATCC 367
Lactobacillus planatrum WCFS1
Lactobacillus reuteri DSM 20016
Lactobacillus salivarius UCC118
Lactococcus garvieae ATCC49165
Leuconostoc citreum KM20
Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293
Listeria innocua Clip11262
Listeria ivanovii subsp. ivanovii PAM 55
Listeria monocytogenes EGD-e
Listeria seeligeri serovar 1/2b str. SLCC3954
Listeria welshimeri serovar 6b str. SLCC5334
Oenococcus oeni PSU-1
Pediococcus pentosaceus ATCC 25745
Weissella koreensis KACC 15510

Alpha-proteobacteria:
Acidiphilium cryptum JF-5
Acidiphilium multivorum AIU301
Agrobacterium radiobacter K84
Beijerinckia indica ATCC 9039
Brevundimonas subvibrioides ATCC 15264
Brucella abortus biovar 1 str. 9-941
Brucella canis ATCC 23365
Brucella melitensis bv. 1 str. 16M
Brucella microti CCM 4915
Brucella ovis ATCC 25840
Brucella pinnipedialis B2/94
Brucella suis 1330
Candidatus Liberibacter asiaticus str. psy62
Candidatus Liberibacter solanacearum Clso- ZC1
Caulobacter crescentus CB15
Caulobacter segnis ATCC 21756
Caulobacter sp. K31
Chelativorans sp. BNC1
Dinoroseobacter shibae DFL12
Erythrobacter litoralis HTCC2594
Gluconacetobacter diazotrophicus PA1_5
Gluconacetobacter xylinus NBRC 3288
Hirschia baltica ATCC 49814
Hyphomicrobium denitrificans ATCC 51888
Hyphomonas neptinulum ATCC 15444
Jannaschia sp. CCS1
Ketogulonicigenium vaulgare Y25
Ketogulonicigenium vulgarum WSH-001
Maricaulis maris MCS10
Mesorhizobium ciceri biovar biserrula WSM1271
Methylocella silvestris BL2
Neorickettsia risticii str. Illinois
Neorickettsia sennetsu str. Miyayama
Orientia tsutsugamushi Boryong
Paracoccus dentrificans PD1222
Parvibaculum lavamentrivorans DS-1
Parvularcula bermudensis HTCC2503
Phenylobacterium zucineum HLK1
Pseudovibrio sp. FO-BEG1
Rhodobacter capsulatus SB 1003
Rhodobacter sphaeroides 2.4.1
Rhodomicrobium vannielii ATCC 17100
Rickettsia africae ESF-5
Rickettsia akarai str. Hartford
Rickettsia bellii RML369-C
Rickettsia canadensis str. McKiel
Rickettsia conorii str. Malish 7
Rickettsia felis URRWXCal2
Rickettsia heilongjiangensis 054
Rickettsia japonica YH
Rickettsia massiliae MTU5
Rickettsia prowazekii str. Madrid E
Rickettsia rickettsii str. Iowa
Rickettsia slovaca 13-B
Rickettsia typhi str. Wilmington
Roseobacter denitrificans Och 114
Roseobacter litoralis Och 149
Ruegeria pomeroyi DSS-3
Ruegeria sp. TM1040
Sinorhizobium fredii NGR234
Sinorhizobium medicae WSM419
Sinorhizobium meliloti 1021
Sphingobium japonicum UT26S
Sphingopyxis alaskensis RB2256
Starkeya novella DSM 506
Xanthobacter autotrophicus Py2
Zymonomas mobilis subsp. Mobilis ZM4

Beta-proteobacteria:
Bordetella petrii DSM 12804
Burkholderia multivorans ATCC 17616
Burkholderia rhizoxinica HKI 454
Burkholderia xenovorans LB400
Nitrosomonas europaea ATCC 19718
Polaromonas naphthalenivorans CJ2
Polaromonas sp. JS666
Ramlibacter tataouinensis TTB310
Variovorax paradoxus S110

Gamma-proteobacteria:
Acidithiobacillus caldus SM-1
Acinetobacter baumannii
Acinetobacter baylyi ADP1
Acinetobacter pittii PHEA-2
Citrobacter koseri ATCC BAA-895
Cronobacter sakazakii ATCC 35469
Dichelobacter nodosus VCS1703A
Dickeya dadantii Ech703
Dickeya zeae Ech1591
Edwardsiella ictaluri 93-146
Enterobacter asburiae LF7a
Erwinia amylovora CFBP 1430
Erwinia pyrifoliae Ep1/96
Erwinia sp. Ejp617
Erwnia tasmaniensis Et1/99
Escherichia coli K12 substr. MG 1655
Escherichia coli O157:H7
Escherichia fergusonii ATCC 35469
Legionella pneumophila subsp. Pneumophila str. Philadelphia 1
Nitrosococcus halphilus Nc4
Pseudomonas entomophila L48
Salmonella bongori NCTC 12419
Salmonella enteric serovar Paratyphi
Salmonella enterica serovar Typhi str. CT18
Salmonella typhimurium LT2
Serratia plymuthica As9
Serratia proteamaculans
Serratia sp. AS12
Shigella boydii Sb227
Shigella dysenterioe Sd197
Shigella flexneri 2a str. 301
Shigella sonnei Ss046
Sodalis glossinidius str. 'morisitans'
Xenorhabdus nematophila ATCC 19061
Yersinia pestis CO92

Epsilon-proteobacteria:
Campylobacter concisus 13826
Camplylobacter curvus 525.92
Campylobacter fetus subsp. Fetus 82-40
Camplylobacter hamnis ATCC BAA-381
Nitratiruptor sp. SB155-2
Sulfurovum sp. NBC37-1

Figure 5.8 Phylogenetic list of bacteria lacks HD-GYP domains. The listed bacteria encode c-di-GMP-associated enzymes (DGCs, PDEs) but not HD-GYP. *Adapted from http://ncbi.nlm.nih.gov/Complete_Genomes/c-di-GMP.html*.

Cyclic-di-nucleotide signaling molecules are ubiquitous and important in bacterial physiology and regulatory functions. Completion of this proposed work would illuminate the molecular mechanism supporting the hypothesis that pGpG functions as a novel signaling molecule that affects gene regulation in *E. amylovora* and potentially other well-characterized pathogens. Little known about a possible function for pGpG other than that it is the substrate for Orn degradation to yield GMP in *Pseudomonas aeruginosa* [191] and excess pGpG level feedback inhibits EAL PDE activity to regulate

c-di-GMP level [218, 219] . Uncovering its potential role would, however, further expand the second messenger-based signaling system due to its prevalence in other bacteria that encode GGDEF and EAL enzymes but no HD-GYP enzymes. We predict that the function of pGpG may be conserved across different bacterial phyla since there are other non-pathogenic bacteria that also do not encode HD-GYP enzymes. This study will develop novel paradigm of pGpG-mediated gene regulation that apply to other branches of phylogenetic tree.

APPENDIX 2

Strains, Plasmids, and Oligonucleotides

Table 1. Bacterial Strain and Phage Names and Descriptions

Strains	Name	Description	Chapter	Reference
E. coli				
DH10b		*F-mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139Δ(ara, leu)7697 galU galK λrpsL nupG*	2	ThermoFisher Scientific
BW29427		*RP4-2(TetSkan1360::FRT)*, *thr B1004*, *lacZ58*(M15), *ΔdapA1341::[erm pir^{+}]*, *rpsL*(strR), *thi-*, *hsdS-*, *pro-*	2, 3	Lab Stock
BL21(DE3)		*F- ompT hsdSB(rB -mB +) gal dcm (DE3)*	2, A1	Lab Stock
MG1655		*F- lambda- ilvG- rfb-50 rph-1*	2, 3	Lab Stock
NR8052	Δ*ung*	Δ(*pro-lac*) *thi ara trpE9777 ung-1*	3	Gift from K. Yu
078:H11 H10407	ETEC	Wild type	2	[71]
E. amylovora				
Ea1189	WT	Wild type	A1	[220]
Δ*pdeABC*	Δ3	Ea1189 Δ*eam_2228 (pdeA)*, Δ*eam_3311* (*pdeB*), and Δ*eam_3381* (*pdeC*)	A1	Appendix 1
Δ*dgcABCDE* Δ*pdeABC*	Δ8	Ea1189 Δ*pdeABC* + Δ*eam_0335* (*dgcA*), Δ*eam_0564* (*dgcB*), Δ*eam_1504* (*dgcC*), Δ*eam_2180* (*dgcD*), Δ*eam_2435* (*dgcE*),	A1	Appendix 1
V. cholerae				
C6706str2	WT or VC	Wild type O1 El Tor; Sm^{R}	2	[60]
CR01	ΔVSP-1	O1 El Tor ΔVSP-1	2	Chapter 2
CR02	ΔVSP-2	O1 El Tor ΔVSP-2	2	Chapter 2
CR03	ΔVSP-1/2	O1 El Tor ΔVSP-1/2	2	Chapter 2
BYH206	Δ*ig^{222}*	O1 El Tor Δ*ig^{222} between vc0175-vc0176 position in N16961 chromosome I [177,230-177,008]*	2	Chapter 2

Table 1. (cont'd)

BYH207	Δ*vc0176*	O1 El Tor Δ*vc0176*	2	Chapter 2
GS05	Δ*avcD*	O1 El Tor Δ*avcD*	2	Chapter 2
WLN5105	Δ*capV*	O1 El Tor Δ*capV*	2	[25]
V. parahaemolyticus				
O1:Kuk str. FDA_R31	VP	Wild type	2, 3	[69]
P. mirabilis				
AR379	PM	Wild type	2	[70]
S. cerevisiae				
yMK839	Sc	*MATa leu2-3 trp1 ura3-52*	2	[114]
Phages				
T2	T2	Wild type	2	ATCC
T3	T3	Wild type	2, 3	ATCC
T4	T4	Wild type	2	ATCC
T5	T5	Wild type	2, 3	ATCC
T6	T6	Wild type	2, 3	ATCC
T7	T7	Wild type	2, 3	ATCC
$\lambda_{virulent}$	λ_{vir}	Wild type	2	Gift from M. Laub
SECΦ17	SECΦ17	Wild type	2	Gift from M. Laub
SECΦ18	SECΦ18	Wild type	2, 3	Gift from M. Laub
SECΦ27	SECΦ27	Wild type	2	Gift from M. Laub

Table 2. Plasmid Names and Descriptions

Plasmids	Name	Relevant characteristics	Chapter	Reference
pEVS141	pVector1	pEVS143 without P_{tac}; Km^r	2	[64]
pEVS143		Broad-host range P_{tac} overexpression vector; Km^r	2, A1	[62]
pMMB67EH	pVector2	Broad-host range P_{tac} overexpression vector; Amp^r	2	[63]
pKAS32		Suicide vector for mutant construction, Amp^r	2	[61]
pLAFR	pLAFR	pLAFR; Tet^r	2	Gift from B. Bassler
pCCD7	pCCD7	pLAFR::*VSP-1*; Tet^r	2	[25]
pET28b	$pVector^{6xHis}$	T7 promoter; Km^r	2, A1	Novagen
pBRP353	pDncV	pMMB67EH::*dncV*; Amp^r	2	[25]
pCMW204	pAvcD	pEVS143::*avcD;* Km^r	2	Chapter 2
pGBS87	pAvcD/$pAvcD^{VC}$	pMMB67EH::*avcD*; Amp^r	2	Chapter 2
pGBS65	$pAvcD^{6xHis}$	pET28b*::avcD-6xHis C-term*; Km^r (*only* in *E. coli* BL21(DE3))	2	Chapter 2
pGBS98	$pAvcD^{6xHis}$	pEVS143::*avcD-6xHis C-term;* Km^r (*only* in *V. cholerae*)	2	Chapter 2
$pAvcD^{4-532}$		pET28b::$avcD^{4-532}$*-6xHis N-term;* Km^r	2	Chapter 2
pGBS71	$pAvcD^{E384A}$	pEVS143::*avcD-E384A;* Km^r	2	Chapter 2
pGBS82	$pAvcD^{E384A}$	pET28b::*avcD-E384A-6xHis C-term*; Km^r (*only* for in vitro and Western blot)	2	Chapter 2
pGBS81	$pAvcD^{C411A+C414A}$	pEVS143::*avcD-C411A+C414A*; Km^r	2	Chapter 2

Table 2 (cont'd)

pGBS75	pAvcD$^{C411A+C414A}$	pET28b::*avcD-C411A+C414A-6xHis C-term*; Km^r	2	Chapter 2
pGBS103	pAvcD^{S52K}	pEVS143::*avcD-S52K;* Km^r	2	Chapter 2
pGBS114	pAvcD^{S52K}	pET28b::*avcD-S52K-6xHis C-term*; Km^r	2	Chapter 2
pGBS106	pAvcD$^{D162A+Q163A}$	pEVS143::*avcD-D162A+Q163A;* Km^r	2	Chapter 2
pGBS116	pAvcD$^{D162A+Q163A}$	pET28b::*avcD-D162A+Q163A-6xHis C-term*; Km^r	2	Chapter 2
pGBS80	pIg^{222}	pEVS143::*Ig^{222}, (position in N16961 chromosome I [177,230-177,008]);* Km^r	2	Chapter 2
pGBS108	pIg$^{222\text{-}STOP}$	pEVS143::*ig^{222}*-1C>T, 2T>A; Km^r	2	Chapter 2
pGBS110	pAvcI	pEVS143::*avcI* (*position in N16961 chromosome I [177,181-177,008]);* Km^r	2	Chapter 2
pAW01	pAvcI$^{RBS\text{-}less}$	pEVS143:*avcI* without RBS; Km^r	2	Chapter 2
pGBS111	pAvcI^{STOP}	pEVS143::*avcI*-1A>T, 2T>A, 3G>A; Km^r	2	Chapter 2
pGBS118	pAvcI^{InteriorSTOP}	pEVS143::*avcI*-17A>T, 18T>A, 19G>A; Km^r	2	Chapter 2
pBYH49	pAvcI$^{49\text{-}186}$	pEVS143::*Ig^{222}* truncation (49-186 NT); Km^r	2	Chapter 2
pBYH52	pAvcI$^{49\text{-}204}$	pEVS143::*Ig^{222}* truncation (49-204 NT); Km^r	2	Chapter 2
pBYH53	pAvcI$^{49\text{-}214}$	pEVS143::*Ig^{222}* truncation (49-214 NT); Km^r	2	Chapter 2

Table 2 (cont'd)

pBYH54	pAvcI$^{49\text{-}218}$	pEVS143::*Ig*^{222} truncation (49-218 NT); Km^{r}	2	Chapter 2
pBYH55	pAvcI$^{66\text{-}222}$	pEVS143::*Ig*^{222} truncation (66-222 NT); Km^{r}	2	Chapter 2
pBYH56	pAvcI$^{86\text{-}222}$	pEVS143::*Ig*^{222} truncation (86-222 NT); Km^{r}	2	Chapter 2
pBYH57	pAvcI$^{123\text{-}222}$	pEVS143::*Ig*^{222} truncation (123-222 NT); Km^{r}	2	Chapter 2
pGBS120	pAvcD^{E123K}	pEVS143::*avcD*-E123K; Km^{r}	2	Chapter 2
pGBS131	pAvcD^{A126T}	pEVS143::*avcD*-A126T; Km^{r}	2	Chapter 2
pGBS128	pAvcD^{K201R}	pEVS143::*avcD*-K201R; Km^{r}	2	Chapter 2
pGBS129	pAvcD^{K511R}	pEVS143::*avcD*-K511R; Km^{r}	2	Chapter 2
pGBS130	pAvcD^{Q514R}	pEVS143::*avcD*-Q514R; Km^{r}	2	Chapter 2
pGBS124	pAvcD^{ETEC}	pEVS143::*avcD* from *Escherichia coli* O78:H11 H10407 (ETEC); Km^{r} (*only* for mass spec experiment)	2	Chapter 2
pGBS125	pAvcI^{ETEC}	pEVS143::*avcI* from *E. coli* O78:H11 H10407 (ETEC); Km^{r}	2	Chapter 2
pGBS126	pAvcD^{ETEC}	pMMB67EH::*avcD* from ETEC; Amp^{r}	2	Chapter 2
pAW07	pAvcI^{VP}	pEVS143::*avcI* from *V. parahaemolyticus* O1:Kuk str. FDA_R31; Km^{r}	2	Chapter 2
pAW06	pAvcD^{VP}	pMMB67EH::*avcD* from *V. parahaemolyticus* O1:Kuk str. FDA_R31; Amp^{r}	2	T Chapter 2

Table 2 (cont'd)

pAW02	pAvcI^{PM}	pEVS143::*avcI* from *P. mirabilis* AR379; Km^r	2	Chapter 2
pAW04	pAvcD^{PM}	pMMB67EH::*avcD* from *P. mirabilis* AR379; Amp^r	2	Chapter 2
pBRP15		pMMB67EH without P_{tac}; Amp^r	2	Chapter 2
pBYH64		pBRP15::*avcI-avcD* operon with its upstream intergenic region position in *V. cholerae* N16961 [177,759-176,932]; Amp^r	2	Chapter 2
pBYH67	pAvcID (Ch. 3)	pBRP15:: *avcI-avcD* operon with its upstream intergenic region position in *V. parahaemolyticus* O1:Kuk str. FDA_R31 (CP006004) [468,152-466,174]; Amp^r	2, 3	Chapter 2
pBYH65		pBRP15:: *avcI-avcD* operon with its upstream intergenic region position in *P. mirabilis* AR379 (NZ_CP029133), [3,698,504-3,700,828]; Amp^r	2	Chapter 2
pBYH63		pBRP15:: *avcI-avcD* operon with its upstream intergenic region position in ETEC (NC_017723.1), [2,280-4,414]; Amp^r	2	Chapter 2

Table 2 (cont'd)

pBYH69		pEVS143::*dcd1* from *Saccharomyces cerevisiae*; Km^r	2	Chapter 2
pBYH81	$pAvcID^{VP-avcDS49K}$	pBYH67::*avcD*-S47K; Amp^r	2	Chapter 2
pBYH82	$pAvcID^{VP-avcDE376K}$	pBYH6::*avcD*-E376K; Amp^r	2	Chapter 2
pBYH83	$pAvcID^{VP-avcDS49K+E376K}$ (Ch. 2); pAvcID* (Ch. 3)	pBYH67::*avcD*-S47K+E376K; Amp^r	2, 3	Chapter 2, 3
pBYH84		pBYH67::*avcD*-6xHis; Amp^r	3	Chapter 3
pCRR01		Deletion construct for ΔVSP-1, Amp^r	2	Chapter 2
pCRR02		Deletion construct for ΔVSP-2, Amp^r	2	Chapter 2
pBYH36		Deletion construct for Δig^{222}, Amp^r	2	Chapter 2
pBYH37		Deletion construct for *Δvc0176,* Amp^r	2	Chapter 2
pGBS88		Deletion construct for Δ*avcD,* Amp^r	2	Chapter 2
pBYH2	pHrpS-6xHis	pET28b*::hrpS-6xHis C-term*; Km^r (*only* in *E. coli* BL21(DE3))	A1	Appendix 1
pWL001	pEAL	pEVS143::*vc1086*; Km^r	A1	Lab Stock
pWL002	pHDGYP/pGYP	pEVS143::*vca0681*; Km^r	A1	Lab stock
pLFC11	pDGC	pACYCduet-1::*edcC* and *edcE* with their native promoter; Cm^r	A1	[186]

Table 3. Oligonucleotides Used in This Study

Name	Primer use	Sequence	Chapter	Reference
Vector Construction				
CMW3009	*avcD* F[1] EcoRI + RBS[3] (pEVS143-AvcD)	GGAAACAGCCTCGACAGGCCTAGGAGGAAGCTAAATTGTTTACAATGAATAAGTCCTCCG	2	Chapter 2
CMW3010	*avcD* R[2] BamHI (pEVS143-AvcD)	CATAAAGCTTGCTCAATCAATCACCGGATCCTAGTCTTGGATGCTCTCTTC	2	Chapter 2
CMW3025	*avcD* F EcoRI + RBS (pMMB67EH-AvcD)	ATTTCACACAGGAAACAGAGGAGCTAAGGAAGCTAAATTGTTTACAATGAATAAGTCCTC	2	Chapter 2
CMW3026	*avcD* R BamHI (pMMB67EH-AvcD)	CCTGCAGGTCGACTCTAGAGCTAGTCTTGGATGCTCTC	2	Chapter 2
CMW3066	*avcD*+6His R BamHI (pEVS143-AvcD-His_6)	GCTTGCTCAATCAATCACCGTTAGTGGTGGTGGTGGTGCTCGATGTCTTGG	2	Chapter 2
CMW3079	*Ig*[222] F EcoRI + RBS (pEVS143-Ig^{222})	CAGCCTCGACAGGCCTAGGAGGAGCTAAGGAAGCTAAACTGTTCGCAAATCATACTTTAG	2	Chapter 2
CMW3080	*Ig*[222] R BamHI (pEVS143-Ig^{222}, pEVS143-AvcI & pEVS143-AvcI 3' end truncations and interior stop codon)	GCTTGCTCAATCAATCACCGTTACCAATGGATTTTTTGTG	2	Chapter 2
CMW3081	*Ig*[222-STOP] F EcoRI + RBS (pEVS143-$Ig^{222\text{-}STOP}$)	CAGCCTCGACAGGCCTAGGAGGAGCTAAGGAAGCTAAATAGTTCGCAAATCATAGTTTAG	2	Chapter 2
CMW3093	*avcD* F NcoI (pET28b-AvcD-His_6)	AACTTTAAGAAGGAGATATACATGTTTACAATGAATAAGTCCTCCGC	2	Chapter 2

Table 3 (cont'd)

CMW3094	*avcD* R XhoI (pET28b-AvcD-His_6)	CTCAGTGGTGGTGGTG GTGGTGCTCGATGTCTT GGATGCTCTCTTCTTCA CTCGATGG	2	Chapter 2
CMW3102	*avcI* F EcoRI + RBS (pEVS143-AvcI & pEVS143-AvcI 5' end truncations)	CTCGACAGGCCTAGGA GGAGCTAAGGAAGCTA AAATGATTACAAGCATT CATGAATATAG	2	Chapter 2
CMW3103	*avcI* F EcoRI + RBS (pEVS143-$AvcI^{STOP}$)	CTCGACAGGCCTAGGA GGAGCTAAGGAAGCTA AATAAATTACAAGCATT CATGAATATAG	2	Chapter 2
CMW3128	$avcI^{49-186}$ F EcoRI + RBS (pEVS143-$AvcI^{49-186}$)	ACAGCCTCGACAGGCC TAGGAGGAGCTAAGGA AGCTAAAATGATTACAA G	2	Chapter 2
CMW2129	$avcI^{49-186}$ R BamHI (pEVS143-$AvcI^{49-186}$)	GCTTGCTCAATCAATCA CCGGGCTCTAGCTTTCT CTTTTTTTGCGTCTTTC	2	Chapter 2
CMW3162	$avcD^{ETEC}$ F EcoRI + RBS (pEVS143-$AvcD^{ETEC}$)	ACAGCCTCGACAGGCC TAGGAGGAGCTAAGGA AGCTAAAATGGCTATAG CTTTGAAAAAG	2	Chapter 2
CMW3163	$avcD^{ETEC}$ R BamHI (pEVS143-$AvcD^{ETEC}$)	GCTTGCTCAATCAATCA CCGTTAAATCAAGTCAT CTTGTTTTG	2	Chapter 2
CMW3164	$avcD^{ETEC}$ F EcoRI + RBS (pMMB67EH-$AvcD^{ETEC}$)	AATTTCACACAGGAAAC AGAGGAGCTAAGGAAG CTAAAATGGCTATAGCT TTGAAAAAGG	2	Chapter 2
CMW3165	$avcD^{ETEC}$ F BamHI (pMMB67EH-$AvcD^{ETEC}$)	CCTGCAGGTCGACTCT AGAGTTAAATCAAGTCA TCTTGTTTTGG	2	Chapter 2
CMW3166	$avcI^{ETEC}$ F EcoRI + RBS (pEVS143-$AvcI^{ETEC}$)	ACAGCCTCGACAGGCC TAGGAGGAGCTAAGGA AGCTAAAATGTCAAACC AATTAACCG	2	Chapter 2

Table 3 (cont'd)

CMW3167	*avcI*^{ETEC} F BamHI (pEVS143-AvcI^{ETEC})	GCTTGCTCAATCAATCACCGCTAATCAAGTATTATTTCTTTCTTTAGTATTTTATC	2	Chapter 2
CMW3180	*avcI*^{VP} F EcoRI + RBS (pEVS143-AvcI^{VP})	ACAGCCTCGACAGGCCTAGGAGGAGCTAAGGAAGCTAAAATGGTTACAAATTTAAATG	2	Chapter 2
CMW3181	*avcI*^{VP} R BamHI (pEVS143-AvcI^{VP})	GCTTGCTCAATCAATCACCGTTACCAACGAATTTTCTGTGCGGCTCTTAAAAG	2	Chapter 2
CMW3184	*avcD*^{VP} F EcoRI + RBS (pMMB67EH-AvcD^{VP})	CAATTTCACACAGGAAACAGAGGAGCTAAGGAAGCTAAAATGGGAAAATCCTCTA	2	Chapter 2
CMW3185	*avcD*^{VP} R BamHI (pMMB67EH-AvcD^{VP})	CCTGCAGGTCGACTCTAGAGTTATTCAATAGTGGCTTCTACTTGTTGCTTTGTGAATG	2	Chapter 2
CMW3189	*avcI* F EcoRI (pEVS143-AvcI)	ACAGCCTCGACAGGCCTAGGATGATTACAAGCATTCATGAATATAGAAACGCTTC	2	Chapter 2
CMW3192	*avcI*^{PM} F EcoRI + RBS (pEVS143-AvcI^{PM})	ACAGCCTCGACAGGCCTAGGAGGAGCTAAGGAAGCTAAAATGAACGTTCAAC	2	Chapter 2
CMW3193	*avcI*^{PM} R BamHI (pEVS143-AvcI^{PM})	GCTTGCTCAATCAATCACCGTTACCAATCTAACGTGTCTGCTACAGCTGC	2	Chapter 2
CMW3196	*avcD*^{VP} F EcoRI + RBS (pMMB67EH-AvcD^{PM})	CAATTTCACACAGGAAACAGAGGAGCTAAGGAAGCTAAAATGGGTAATCC	2	Chapter 2
CMW3197	*avcD*^{VP} R BamHI (pMMB67EH-AvcD^{PM})	CCTGCAGGTCGACTCTAGAGTTAACTTCTCTCTTCACCTAAACGAAGATTTAC	2	Chapter 2
CMW3200	*avcI*$^{49-204}$ R BamHI (pEVS143-AvcI$^{49-204}$)	GCTTGCTCAATCAATCACCGTGCAGCACGCAAAAGATTG	2	Chapter 2

Table 3 (cont'd)

CMW3201	*avcI*$^{49-214}$ R BamHI (pEVS143-AvcI$^{49-214}$)	GCTTGCTCAATCAATCACCGGGATTTTTTGTGCAGCAC	2	Chapter 2
CMW3202	*avcI*$^{49-218}$ R BamHI (pEVS143-AvcI$^{49-218}$)	GCTTGCTCAATCAATCACCGCAATGGATTTTTTGTGCAGCACGCAAAAGA	2	Chapter 2
CMW3203	*avcI*$^{66-222}$ F EcoRI + RBS (pEVS143-AvcI$^{66-222}$)	ACAGCCTCGACAGGCCTAGGAGGAGCTAAGGAAGCTAAAGAATATAGAAACG	2	Chapter 2
CMW3204	*avcI*$^{86-222}$F EcoRI + RBS (pEVS143-AvcI$^{86-222}$)	ACAGCCTCGACAGGCCTAGGAGGAGCTAAGGAAGCTAAAATAGCGACAAAAAC	2	Chapter 2
CMW3205	*avcI*$^{123-222}$ F EcoRI + RBS (pEVS143-AvcI$^{123-222}$)	ACAGCCTCGACAGGCCTAGGAGGAGCTAAGGAAGCTAAAAGACACTAGCG	2	Chapter 2
CMW3306	*avcI-avcD* w/ upstream intergenic regions from *V. cholerae* F BamHI (pAvcI-AvcD^{VC})	CGGGAAACCTGTCGTGCCAGCTAGTCTTGGATGCTCTC	2	Chapter 2
CMW3307	*avcI-avcD* w/ upstream intergenic regions from *V. cholerae* R EcoRI (pAvcI-AvcD^{VC})	CCTGCAGGTCGACTCTAGAGATAGAGACACTATATTTAGTGTTTAATTAAC	2	Chapter 2
CMW3308	*avcI-avcD* w/ upstream intergenic regions from ETEC F BamHI (pAvcI-AvcD^{ETEC})	CGGGAAACCTGTCGTGCCAGTTAAATCAAGTCATCTTGTTTTGGTTC	2	Chapter 2

Table 3 (cont'd)

CMW3309	*avcI-avcD* w/ upstream intergenic regions from ETEC R EcoRI (pAvcI-AvcD^{ETEC})	CCTGCAGGTCGACTCTAGAGAGGCTCCGCTGAGAAAAAATTC	2	Chapter 2
CMW3310	*avcI-avcD* w/ upstream intergenic regions from *P. mirabilis* F BamHI (pAvcI-AvcD^{PM})	CGGGAAACCTGTCGTGCCAGTTAACTTCTCTCTTCACCTAAAC	2	Chapter 2
CMW3311	*avcI-avcD* w/ upstream intergenic regions from *P. mirabilis* R EcoRI (pAvcI-AvcD^{PM})	CCTGCAGGTCGACTCTAGAGTGCTTTAACTCCTAAAGG	2	Chapter 2
CMW3312	*avcI-avcD* w/ upstream intergenic regions from *V. parahaemolyticus* F BamHI (pAvcI-AvcD^{VP})	CGGGAAACCTGTCGTGCCAGTTATTCAATAGTGGCTTCTAC	2, 3	Chapter 2,3
CMW3313	*avcI-avcD* w/ upstream intergenic regions from *V. parahaemolyticus* R EcoRI (pAvcI-AvcD^{VP})	TGCCTGCAGGTCGACTCTAGAGTCACTTTGCTGATTTAAGCAGAT	2, 3	Chapter 2,3
CMW3335	*dcd1^{Sc}* F EcoRI (pEVS143-Dcd1)	ACAGCCTCGACAGGCCTAGGAGGAGCTAAGGAAGCTAAAATGTTAATTGGTGTAAG	2	Chapter 2
CMW3336	*dcd1^{Sc}* R BamHI (pEVS143-Dcd1)	GCTTGCTCAATCAATCACCGTTAAATCATCACAATTCTTGGTTC	2	Chapter 2

Table 3 (cont'd)

EWAvcDFwd	*avcD*$^{4\text{-}532}$ F NdeI (pAvcD$^{4\text{-}532}$) For protein purification	GTGCCGCGCGGCAGCCATATGAATAAGTCCTCCGCAAA	2	Chapter 2
EWAvcDRev	*avcD*$^{4\text{-}532}$ R XhoI (pAvcD$^{4\text{-}532}$) For protein purification	TGGTGGTGGTGGTGGTGCTTAGTCTTGGATGCTCTCTTCTT	2	Chapter 2
CMW3504	*hrpS* F NcoI (pET28b::*hrpS*-6xHis)	ACTTTAAGAAGGAGATATACATGAACATCAGGAATAGTGAAC	A1	Appendix 1
CMW3505	*hrpS* R XhoI (pET28b::*hrpS*-6xHis)	AGTGGTGGTGGTGGTGGTGCTCGAGCTGAGCAATAACC	A1	Appendix 1
Site-directed Mutagenesis				
CMW3011	*avcD* (E384A) F (pEVS143-AvcD^{E384A} & pET28b-AvcD^{E384A})	CAAGAGCGGTTCATGCTGCAATGGATTCTCTTATAGC	2	Chapter 2
CMW3012	*avcD* (E384A) R (pEVS143-AvcD^{E384A} & pET28b-AvcD^{E384A})	GCTATAAGAGAATCCATTGCAGCATGAACCGCTCTTG	2	Chapter 2
CMW3013	*avcD* (C411A + C414A) F (pEVS143-AvcD$^{C411A+C414A}$)	TATATGTTACGACATATCCGGCTCACAACGCTGCGCGACACATCGTTGCTG	2	Chapter 2
CMW3014	*avcD* (C411A + C414A) R (pEVS143-AvcD$^{C411A+C414A}$)	CAGCAACGATGTGTCGCGCAGCGTTGTGAGCCGGATATGTCGTAACATATA	2	Chapter 2
CMW3021	*avcD* (K55A) F (pEVS143-AvcD^{K55A})	GCTATTGGCTCTGGTGTAGCGGCATTAAAAGAGAGTTTAGTTAGTTCTCTTGAGACATAT	2	Chapter 2
CMW3022	*avcD* (K55A) R (pEVS143-AvcD^{K55A})	ATATGTCTCAAGAGAACTAACTAAACTCTCTTTTAATGCCGCTACACCAGAGCCAATAGC	2	Chapter 2

Table 3 (cont'd)

CMW3104	*avcD* (D162A + Q163A) F (pEVS143-AvcD$^{D162A+Q163A}$)	CGCATACATCATCGCG GCGTTAAAGCACCCTG ATGAAATCAAATTCC	2	Chapter 2
CMW3105	*avcD* (D162A + Q163A) R (pEVS143-AvcD$^{Q162A+Q163A}$)	GGAATTTGATTTCATCA GGGTGCTTTAACGCCG CGATGATGTATGCG	2	Chapter 2
CMW3110	*avcD* (S52K) F (pEVS143-AvcD^{S52K})	CCTCTGTGGGGCTATT GGCAAAGGTGTAAAGG CATTAAAAGAGAG	2	Chapter 2
CMW3111	*avcD* (S52K) R (pEVS143-AvcD^{S52K})	CTCTCTTTTAATGCCTTT ACACCTTTGCCAATAGC CCCACAGAGG	2	Chapter 2
CMW3112	*avcD* (S52P) F (pEVS143-AvcD^{S52P})	CCTCTGTGGGGCTATT GGCCCGGGTGTAAAGG CATTAAAAGAGAG	2	Chapter 2
CMW3113	*avcD* (S52P) R (pEVS143-AvcD^{S52P})	CTCTCTTTTAATGCCTTT ACACCCGGGCCAATAG CCCCACAGAGG	2	Chapter 2
CMW3114	*avcD* (S52W) F (pEVS143-AvcD^{S52W})	CCTCTGTGGGGCTATT GGCTGGGGTGTAAAGG CATTAAAAGAGAG	2	Chapter 2
CMW3115	*avcD* (S52K) R (pEVS143-AvcD^{S52W})	CTCTCTTTTAATGCCTTT ACACCCCAGCCAATAG CCCCACAGAGG	2	Chapter 2
CMW3118	*avcI* (interior alternative frame stop) F (pEVS143-AvcI17A>T, 18T>A, 19G>A)	AAGGAAGCTAAAATGAT TACAAGCATTCTAAAAT ATAGAAACGCTTCTAAT AGCG	2	Chapter 2
CMW3119	*avcI* (interior alternative frame stop) R (pEVS143-AvcI17A>T, 18T>A, 19G>A)	CGCTATTAGAAGCGTTT CTATATTTTAGAATGCTT GTAATCATTTTAGCTTC CTT	2	Chapter 2
CMW3124	*avcD* (E123K) F (pEVS143-AvcD^{E123K})	GCAGCCTGTGCTATCAA AGAAATTGCGCTGG	2	Chapter 2

Table 3 (cont'd)

CMW3125	*avcD* (E123K) R (pEVS143-AvcD^{E123K})	CCAGCGCAATTTCTTTGATAGCACAGGCTGC	2	Chapter 2
CMW3172	*avcD* (A126T) F (pEVS143-AvcD^{A126T})	GCTATCGAAGAAATTACGCTGGAAAGAACATTAATCTGTC	2	Chapter 2
Cmw3173	*avcD* (A126T) R (pEVS143-AvcD^{A126T})	GACAGATTAATGTTCTTTCCAGCGTAATTTCTTCGATAGC	2	Chapter 2
CMW3448	*avcD^{VP}* (S49K) F	ATTGGTCTTTGTGGAGCTATAGGCAAGGGTGTGAAAGCACTAAAAGATAAC	2, 3	Chapter 2, 3
CMW3449	*avcD^{VP}* (S49K) R	GTTATCTTTTAGTGCTTTCACACCCTTGCCTATAGCTCCACAAAGACCAAT	2 ,3	Chapter 2, 3
CMW3450	*avcD^{VP}* (E376A) F	GAGAGCTGTACACGCAGCAATGGATGCCATTGTTG	2, 3	Chapter 2, 3
CMW3451	*avcD^{VP}* (E376A) R	CAACAATGGCATCCATTGCTGCGTGTACAGCTCTC	2, 3	Chapter 2, 3
CMW3456	*avcD^{VP}*-6xHis knock-in F	GTAGAAGCCACTATTGAACACCACCACCACCACCACTAACTGGCACGACAGGTTTC	3	Chapter 3
CMW3457	*avcD^{VP}*-6xHis knock-in R	GAAACCTGTCGTGCCAGTTAGTGGTGGTGGTGGTGGTGTTCAATAGTGGCTTCTAC	3	Chapter 3
Gene Deletion				
CMW2794	ΔVSP-2 up[4] F; CR02 & CR03	GTGGAATTCCCGGGAGAGCTCGGCTTGTTCACTATCGTAATAATGC	2	Chapter 2
CMW2795	ΔVSP-2 up R; CR02 & CR03	GGAGGGGCCACCACTGGGAGGGCACCAGATTC	2	Chapter 2
CMW2796	ΔVSP-2 down[5] F; CR02 & CR03	GCCCTCCCAGTGGTGGCCCCTCCCAGGT	2	Chapter 2
CMW2797	ΔVSP-2 down R; CR02 & CR03	AGCTATAGTTCTAGAGGTACGGGCATTAAGGTGGTGGAAACCG	2	Chapter 2
CMW2814	ΔVSP-1 up F; CR01 & CR03	GTGGAATTCCCGGGAGAGCTGGCTTTACTGTTATTCGC	2	Chapter 2

Table 3 (cont'd)

CMW2815	ΔVSP-1 up R; CR01 & CR03	TACCATGTAGTAGCGGTATCGAGATTCC	2	Chapter 2
CMW2816	ΔVSP-1 down F; CR01 & CR03	GATACCGCTACTACATGGTAACGAACTCTTC	2	Chapter 2
CMW2817	ΔVSP-1 down R; CR01 & CR03	AGCTATAGTTCTAGAGGTACCGCTAAGTTTGTGGATGC	2	Chapter 2
CMW2970	Δ*vc0176* up F; BYH207	ATAACAATTTGTGGAATTCCCGGGAGAGCTGGGAATCGAATATTGAGAG	2	Chapter 2
CMW2971	Δ*vc0176* up R; BYH207	ATATAGTGTCTCTATTTATGGCTCATAATCTTGAAG	2	Chapter 2
CMW2972	Δ*vc0176* down F; BYH207	GATTATGAGCCATAAATAGAGACACTATATTTAGTGTTTAATTAAC	2	Chapter 2
CMW2973	Δ*vc0176* down R; BYH207	TGCGCATGCTAGCTATAGTTCTAGAGGTACTATGAAACTTATTTCTATACTCTCAG	2	Chapter 2
CMW3067	Δ*avcD* up F; GS05	GTGGAATTCCCGGGAGAGCTACTATATTTAGTGTTTAATTAACAAAAAAC	2	Chapter 2
CMW3068	Δ*avcD* up R; GS05	CAGACTAAAGCCTGAAATTATGAAACTTATTTCTATAC	2	Chapter 2
CMW3069	Δ*avcD* down F; GS05	TAATTTCAGGCTTTAGTCTGGAAAATTCACTTTTC	2	Chapter 2
CMW3070	Δ*avcD* down R; GS05	AGCTATAGTTCTAGAGGTACACATGGAGCATGATCAGG	2	Chapter 2
CMW3071	Δ*Ig*^{222} up F; BYH206	ATAACAATTTGTGGAATTCCCGGGAGAGCTTCTCAAAGAAGCACGTAAAAAAG	2	Chapter 2
CMW3072	Δ*Ig*^{222} up R; BYH206	CAAGAATTAACGTGGTAAAGTGCGCACATTCTAC	2	Chapter 2
CMW3073	Δ*Ig*^{222} down F; BYH206	AATGTGCGCACTTTACCACGTTAATTCTTGATTAGC	2	Chapter 2
CMW3074	Δ*Ig*^{222} down R; BYH206	TGCGCATGCTAGCTATAGTTCTAGAGGTACTCATTTTCTTCTGAGGTTTC	2	Chapter 2

Table 3 (cont'd)

qPCR				
CMW2926	*gyrA* F	TGGCCAGCCAGAGATCAAG	2	Chapter 2
CMW2927	*gyrA* R	ACCCGCAGCGGTACGA	2	Chapter 2
CMW3206	*avcD* F	TCGACCAGTTAAAGCACCCT	2	Chapter 2
CMW3207	*avcD* R	CCTTCTGTACGGATCAAGCCA	2	Chapter 2
CMW3208	*avcI* F	GTGAATGGATATTTCGGTGGA	2	Chapter 2
CMW3209	*avcI* R	TTGTCGCTATTAGAAGCGTT	2	Chapter 2
CMW3288	*oriI* F	CAGGTGAACCAGCAAAATCGA	2	[103]
CMW3289	*oriI* R	TGGTATTGAAGCTCAATGCGG	2	[103]
CMW3290	*terI* F	TTCAAGCTGAGGCGGATTTG	2	[103]
CMW3291	*terI* R	GCTCATTGGCTTCTTGTGCTT	2	[103]
CMW3502	*orf124*_T5 F	AGGTGCTAGCAACCACTGAC	2, 3	Chapter 2, 3
CMW3503	*orf124*_T5 R	CGTCCGATTTCGACGGTTTG	2, 3	Chapter 2, 3
CMW3506	*p52*_T7_F	CAGAACTCATGGCAAGCACG	3	Chapter 3
CMW3507	*p52*_T7_R	TAAAGCCCTCCGCTTGGTTT	3	Chapter 3
Electrophoretic Mobility Shift Assay (EMSA)				
CMW2779	FW HrpLpr	**FAM-**CCTGGCGAACCTTCAATGATGAGAGCAGTTGTCATTGTGT	A1	Appendix 1
CMW2780	RV HrpLpr	TGGCTTGCTCCGTTACTAAATCAGGTGATGCCTTAGCGGC	A1	Appendix 1
CMW2784	FW HrpSpr	**FAM-**GAATGCTCTTATATTTGTCTCTCGC	A1	Appendix 1
CMW2785	RV HrpSpr	AAAAAATTACCCCTGCCCTATCCAT	A1	Appendix 1
CMW3510	*hrpL* binding site FW	ATATGCAACTTATTGCAAATTTTGGCG	A1	Appendix 1

Table 3 (cont'd)

CMW3511	Upstream hrpL_pr w/o BS_RV	ATGTCGCTGGCGACTTT	A1	Appendix 1
In vitro Transcription Synthesis				
EJW002	*avcI*^{VC} RNA F	GACCATGATTACGCCATAATACGACTCACTATAGGGATGATTACAAGCATTCATG	2	Chapter 2
EJW003	*avcI*^{VC} RNA R	[mU][mU]ACCAATGGATTTTTTGTGC	2	Chapter 2
EJW016	*avcI*^{VC}-RC RNA F	GACCATGATTACGCCATAATACGACTCACTATAGGGTTACCAATGGATTTTTTG	2	Chapter 2
EJW017	*avcI*^{VC}-RC RNA R	[mA][mU]GATTACAAGCATTCATG	2	Chapter 2
CMW3454	*avcI*^{VP} RNA probe F	GACCATGATTACGCCATAATACGACTCACTATAGGGTTACCAACGAATTTTCTG	3	Chapter 3
CMW3455	*avcI*^{VP} RNA probe R	[mA][mU]GGTTACAAATTTAAATG	3	Chapter 3

[1]F = Forward
[2]R= Reverse
[3]RBS= Ribosomal Binding Site
[4]Up= Amplifies Upstream Fragment
[5]Down= Amplifies Downstream Fragment

Table 4. Maximum Conservation of Homologs from Different Phylogenetic Lineages

AvcD homologs summary table Lineages and percentage similarities of AvcD homologs containing both DCD and PLN domains		
DomArch.Gene3D	Lineage	Max%Positive
PLN+DCD	Bacteria>Proteobacteria	100.00
PLN+DCD	Bacteria>Bacteroidetes	58.80
PLN+DCD	Bacteria>Balneolaeota	56.02
PLN+DCD	Bacteria>Actinobacteria	55.63
PLN+DCD	Archaea>Thaumarchaeota	53.65
PLN+DCD	Bacteria>Firmicutes	52.27
PLN+DCD	Bacteria>Planctomycetes	52.27
PLN+DCD	Bacteria	51.88
NABP+PLN+DCD	Bacteria>Proteobacteria	51.09
PLN+DCD	Bacteria>Acidobacteria	49.70
PLN+DCD	Bacteria>Verrucomicrobia	48.69
PLN+DCD	Bacteria>Chlamydiae	45.25
PLN+DCD+NABP	Bacteria>Proteobacteria	42.48
PLN+DCD+NABP+NABP	Bacteria>Proteobacteria	39.85
PLN+PLN+DCD	Bacteria>Proteobacteria	36.75
PLN+DCD	Bacteria>Cyanobacteria	34.63
PLN+DCD	Eukaryota>Ascomycota	27.63
PLN+DCD	Eukaryota>Ciliophora	27.57
PLN+DCD	Eukaryota>Basidiomycota	25.10
PLN+DCD	Eukaryota>Chytridiomycota	23.77
PLN+DCD	Eukaryota>Mucoromycota	22.63
PLN+DCD	Eukaryota>Apicomplexa	19.96
PLN+DCD	Eukaryota>Streptophyta	19.39
PLN+Znf_CCHC+DCD	Eukaryota>Ascomycota	16.92

Abbreviations. PLN, P-loop containing nucleotide triphosphate hydrolases; DCD, Cytidine Deaminase domain 2; NABP, Nucleic acid-binding proteins; Znf_CCHC, Zinc finger CCHC-type

Key
(% pos)
100
15

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Haudiquet M, Buffet A, Rendueles O, Rocha EPC. 2021. Interplay between the cell envelope and mobile genetic elements shapes gene flow in populations of the nosocomial pathogen *Klebsiella pneumoniae.* PLOS Biol 19:e3001276.

2. Rodríguez-Beltrán J, Sørum V, Toll-Riera M, de la Vega C, Peña-Miller R, Millán ÁS. 2020. Genetic dominance governs the evolution and spread of mobile genetic elements in bacteria. Proc Natl Acad Sci U S A 117:15755–15762.

3. Rankin DJ, Rocha EPC, Brown SP. 2011. What traits are carried on mobile genetic elements, and why? Heredity (Edinb) 106:1–10.

4. Touchon M, Moura de Sousa JA, Rocha EP. 2017. Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. Curr Opin Microbiol 38:66–73.

5. McKitterick AC, Seed KD. 2018. Anti-phage islands force their target phage to directly mediate island excision and spread. Nat Commun 9:2348.

6. Makarova KS, Wolf YI, Snir S, Koonin E V. 2011. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. J Bacteriol 193:6039–6056.

7. Cohen D, Melamed S, Millman A, Shulman G, Oppenheimer-Shaanan Y, Kacen A, Doron S, Amitai G, Sorek R. 2019. Cyclic GMP–AMP signalling protects bacteria against viral infection. Nature 574:691–695.

8. Bernheim A, Sorek R. 2020. The pan-immune system of bacteria: antiviral defence as a community resource. Nat Rev Microbiol 18:113–119.

9. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R. 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science (80-) 359:eaar4120.

10. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406:477–83.

11. Rabaan AA. 2018. Cholera: an overview with reference to the Yemen epidemic. Front Med https://doi.org/10.1007/s11684-018-0631-2.

12. Zhang P, Li F, Liang W, Li J, Kan B, Wang D. 2014. The seventh pandemic *Vibrio cholerae* O1 El Tor isolate in China has undergone genetic shifts. J Clin Microbiol 52:964–7.

13. Pradhan S, Mallick SK, Chowdhury R. 2013. Classical Biotype Is Converted to the Viable Non-Culturable State when Cultured with the El Tor Biotype. PLoS One 8:e53504.

14. Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF, Mekalanos JJ. 2002. Comparative genomic analysis of *Vibrio cholerae*: Genes that correlate with cholera endemic and pandemic disease. Proc Natl Acad Sci 99:1556–1561.

15. Faruque SM, Mekalanos JJ. 2003. Pathogenicity islands and phages in *Vibrio cholerae* evolution. Trends Microbiol 11:505–510.

16. Hu D, Liu B, Feng L, Ding P, Guo X, Wang M, Cao B, Reeves PR, Wang L. 2016. Origins of the current seventh cholera pandemic. Proc Natl Acad Sci 113:E7730–E7739.

17. Murphy RA, Boyd EF. 2008. Three Pathogenicity Islands of *Vibrio cholerae* Can Excise from the Chromosome and Form Circular Intermediates. J Bacteriol 190:636–647.

18. Hacker J, Carniel E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep 2:376–81.

19. Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H. 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol Microbiol 23:1089–97.

20. Hacker J, Kaper JB. 2000. Pathogenicity islands and the evolution of microbes. Annu Rev Microbiol 54:641–79.

21. Cabezón E, Ripoll-Rozada J, Peña A, de la Cruz F, Arechaga I. 2015. Towards an integrated model of bacterial conjugation. FEMS Microbiol Rev 39:81–95.

22. De La Cruz F, Frost LS, Meyer RJ, Zechner EL. 2010. Conjugative DNA metabolism in Gram-negative bacteria. FEMS Microbiol Rev 34:18–40.

23. Beyhan S, Tischler AD, Camilli A, Yildiz FH. 2006. Differences in Gene Expression between the Classical and El Tor Biotypes of *Vibrio cholerae* O1. Infect Immun 74:3633–3642.

24. Davies BW, Bogard RW, Young TS, Mekalanos JJ. 2012. Coordinated Regulation of Accessory Genetic Elements Produces Cyclic Di-Nucleotides for *V. cholerae* Virulence. Cell 149:358–370.

25. Severin GB, Ramliden MS, Hawver LA, Wang K, Pell ME, Kieninger A-K, Khataokar A, O'Hara BJ, Behrmann L V., Neiditch MB, Benning C, Waters CM, Ng W-L. 2018. Direct activation of a phospholipase by cyclic GMP-AMP in El Tor *Vibrio cholerae*. Proc Natl Acad Sci 115:E6048–E6055.

26. Kazi MI, Conrado AR, Mey AR, Payne SM, Davies BW. 2016. ToxR Antagonizes H-NS Regulation of Horizontally Acquired Genes to Drive Host Colonization. PLoS Pathog 12:e1005570.

27. Zhu D, Wang L, Shang G, Liu X, Zhu J, Lu D, Wang L, Kan B, Zhang J-R, Xiang Y. 2014. Structural biochemistry of a *Vibrio cholerae* dinucleotide cyclase reveals cyclase activity regulation by folates. Mol Cell 55:931–937.

28. Murphy SG, Alvarez L, Adams MC, Liu S, Chappie JS, Cava F, Dörr T. 2019. Endopeptidase Regulation as a Novel Function of the Zur-Dependent Zinc Starvation Response. MBio 10.

29. Murphy SG, Johnson BA, Ledoux CM, Dörr T. 2021. *Vibrio cholerae*'s mysterious Seventh Pandemic island (VSP-II) encodes novel Zur-regulated zinc starvation genes involved in chemotaxis and cell congregation. PLoS Genet 17:e1009624.

30. Chibani-Chennoufi S, Bruttin A, Dillmann ML, Brüssow H. 2004. Phage-host interaction: An ecological perspective. J Bacteriol 186:3677–3686.

31. Weitz JS, Hartman H, Levin SA. 2005. Coevolutionary arms races between bacteria and bacteriophage. Proc Natl Acad Sci U S A 102:9535–40.

32. Stern A, Sorek R. 2011. The phage-host arms-race: Shaping the evolution of microbes. Bioessays 33:43.

33. Shamim Hasan Zahid M, Nashir Udden SM, Faruque ASG, Calderwood SB, Mekalanos JJ, Faruque SM. 2008. Effect of phage on the infectivity of *Vibrio cholerae* and emergence of genetic variants. Infect Immun 76:5266–5273.

34. Zahid MSH, Waise Z, Kamruzzaman M, Ghosh AN, Balakrish Nair G, Khairul

Bashar SAM, Mekalanos JJ, Faruque SM. 2011. An experimental study of phage mediated bactericidal selection & emergence of the El Tor *Vibrio cholerae*. Indian J Med Res 133:218.

35. Lowey B, Whiteley AT, Keszei AFA, Morehouse BR, Mathews IT, Antine SP, Cabrera VJ, Kashin D, Niemann P, Jain M, Schwede F, Mekalanos JJ, Shao S, Lee ASY, Kranzusch PJ. 2020. CBASS Immunity Uses CARF-Related Effectors to Sense 3′–5′- and 2′–5′-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. Cell 182:38-49.e17.

36. O'Donovan GA, Neuhard J. 1970. Pyrimidine metabolism in microorganisms. Bacteriol Rev 34:278–343.

37. Weiss B. 2007. The Deoxycytidine Pathway for Thymidylate Synthesis in *Escherichia coli*. J Bacteriol 189:7922–7926.

38. Weiss B, Wang L. 1994. De novo synthesis of thymidylate via deoxycytidine in dcd (dCTP deaminase) mutants of *Escherichia coli*. J Bacteriol 176:2194–2199.

39. Krishnan A, Iyer LM, Holland SJ, Boehm T, Aravind L. 2018. Diversification of AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread roles in immunity. Proc Natl Acad Sci 115:E3201–E3210.

40. Chiu Y-L, Greene WC. 2008. The APOBEC3 Cytidine Deaminases: An Innate Defensive Network Opposing Exogenous Retroviruses and Endogenous Retroelements. Annu Rev Immunol 26:317–353.

41. Stavrou S, Ross SR. 2015. APOBEC3 Proteins in Viral Immunity. J Immunol 195:4565–70.

42. Sharma S, Patnaik SK, Thomas Taggart R, Kannisto ED, Enriquez SM, Gollnick P, Baysal BE. 2015. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat Commun 6:6881.

43. Thompson FL, Iida T, Swings J. 2004. Biodiversity of vibrios. Microbiol Mol Biol Rev 68:403–31, table of contents.

44. Son MS, Taylor RK. 2011. Genetic Screens and Biochemical Assays to Characterize *Vibrio cholerae* O1 Biotypes: Classical and El Tor. Curr Protoc Microbiol 22A:6A.2.1-6A.2.17.

45. O'Shea YA, Finnan S, Reen FJ, Morrissey JP, O'Gara F, Boyd EF. 2004. The Vibrio seventh pandemic island-II is a 26.9 kb genomic island present in *Vibrio*

cholerae El Tor and O139 serogroup isolates that shows homology to a 43.4 kb genomic island in V. vulnificus. Microbiology 150:4053–4063.

46. Nusrin S, Gil AI, Bhuiyan NA, Safa A, Asakura M, Lanata CF, Hall E, Miranda H, Huapaya B, Vargas G. C, Luna MA, Sack DA, Yamasaki S, Nair GB. 2009. Peruvian *Vibrio cholerae* O1 El Tor strains possess a distinct region in the Vibrio seventh pandemic island-II that differentiates them from the prototype seventh pandemic El Tor strains. J Med Microbiol 58:342–354.

47. Lada AG, Frahm Krick C, Kozmin SG, Mayorov VI, Karpova TS, Rogozin IB, Pavlov YI. 2011. Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. Biochem 76:131–146.

48. Kreuzer KN, Brister JR. 2010. Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. Virol J 7:358.

49. You L, Suthers PF, Yin J. 2002. Effects of *Escherichia coli* Physiology on Growth of Phage T7 In Vivo and In Silico. J Bacteriol 184:1888–1894.

50. Endy D, Kong D, Yin J. 1997. Intracellular kinetics of a growing virus: A genetically structured simulation for bacteriophage T7. Biotechnol Bioeng 55:375–389.

51. Zhang Y, Maley F, Maley GF, Duncan G, Dunigan DD, Van Etten JL. 2007. Chloroviruses encode a bifunctional dCMP-dCTP deaminase that produces two key intermediates in dTTP formation. J Virol 81:7662–71.

52. Graziani S, Xia Y, Gurnon JR, Van Etten JL, Leduc D, Skouloubris S, Myllykallio H, Liebl U. 2004. Functional Analysis of FAD-dependent Thymidylate Synthase ThyX from *Paramecium bursaria* Chlorella Virus-1. J Biol Chem 279:54340–54347.

53. Zhang Y, Moriyama H, Homma K, Van Etten JL. 2005. Chlorella Virus-Encoded Deoxyuridine Triphosphatases Exhibit Different Temperature Optima. J Virol 79:9945–9953.

54. Chiu CS, Ruettinger T, Flanegan JB, Greenberg GR. 1977. Role of deoxycytidylate deaminase in deoxyribonucleotide synthesis in bacteriophage T4 DNA replication. J Biol Chem 252:8603–8608.

55. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W. 2003. Bacteriophage T4 Genome. Microbiol Mol Biol Rev 67:86–156.

56. Guzmán EC, Martín CM. 2015. Thymineless death, at the origin. Front Microbiol 6:499.

57. Itsko M, Schaaper RM. 2014. dGTP starvation in *Escherichia coli* provides new insights into the thymineless-death phenomenon. PLoS Genet 10:e1004310.

58. Kuong KJ, Kuzminov A. 2010. Stalled replication fork repair and misrepair during thymineless death in *Escherichia coli.* Genes to Cells 15:619–634.

59. Ahmad SI, Kirk SH, Eisenstark A. 1998. Thymine metabolism and thymineless death in prokaryotes and eukaryotes. Annu Rev Microbiol 52:591–625.

60. Helene Thelin K, Taylor RK. 1996. Toxin-coregulated pilus, but not mannose-sensitive hemagglutinin, is required for colonization by *Vibrio cholerae* O1 El Tor biotype and O139 strains. Infect Immun 64:2853–2856.

61. Skorupski K, Taylor RK. 1996. Positive selection vectors for allelic exchange. Gene 169:47–52.

62. Bose JL, Rosenberg CS, Stabb E V. 2008. Effects of luxCDABEG induction in *Vibrio fischeri*: Enhancement of symbiotic colonization and conditional attenuation of growth in culture. Arch Microbiol 190:169–183.

63. Fürste JP, Pansegrau W, Frank R, Blöcker H, Scholz P, Bagdasarian M, Lanka E. 1986. Molecular cloning of the plasmid RP4 primase region in a multi-host-range tacP expression vector. Gene 48:119–131.

64. Dunn AK, Millikan DS, Adin DM, Bose JL, Stabb E V. 2006. New rfp- and pES213-derived tools for analyzing symbiotic *Vibrio fischeri* reveal patterns of infection and lux expression in situ. Appl Environ Microbiol 72:802–10.

65. Edelheit O, Hanukoglu A, Hanukoglu I. 2009. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. BMC Biotechnol 9:61.

66. Kim P-J, Price ND. 2011. Genetic co-occurrence network across sequenced microbes. PLoS Comput Biol 7:e1002340.

67. Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30.

68. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. 2019. New approach for understanding genome variations in KEGG. Nucleic Acids Res 47:D590–

D595.

69. Lüdeke CHM, Kong N, Weimer BC, Fischer M, Jones JL. 2015. Complete Genome Sequences of a Clinical Isolate and an Environmental Isolate of *Vibrio parahaemolyticus*. Genome Announc 3:e00216-15.

70. Lutgring JD, Machado M-J, Benahmed FH, Conville P, Shawar RM, Patel J, Brown AC. 2018. FDA-CDC Antimicrobial Resistance Isolate Bank: a Publicly Available Resource To Support Research, Development, and Regulatory Requirements. J Clin Microbiol 56.

71. Evans DJ, Evans DG. 1973. Three characteristics associated with enterotoxigenic *Escherichia coli* isolated from man. Infect Immun 8:322–8.

72. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–58.

73. Madeira F, Madhusoodanan N, Lee J, Tivey ARN, Lopez R. 2019. Using EMBL-EBI Services via Web Interface and Programmatically via Web Services. Curr Protoc Bioinforma 66:e74.

74. Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R. 2012. LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. RNA 18:900–14.

75. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

76. Tatusova T, Ciufo S, Federhen S, Fedorov B, McVeigh R, O'Neill K, Tolstoy I, Zaslavsky L. 2015. Update on RefSeq microbial genomes resources. Nucleic Acids Res 43:D599–D605.

77. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733-45.

78. Federhen S. 2012. The NCBI Taxonomy database. Nucleic Acids Res 40:D136–

D143.

79. Soding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–W248.

80. Nielsen H. 2017. Predicting Secretory Proteins with SignalP. Methods Mol Biol 1611:59–73.

81. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11 Edited by F. Cohen. J Mol Biol 305:567–580.

82. Käll L, Krogh A, Sonnhammer ELL. 2004. A Combined Transmembrane Topology and Signal Peptide Prediction Method. J Mol Biol 338:1027–1036.

83. Käll L, Krogh A, Sonnhammer ELL. 2007. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res 35:W429-32.

84. Mistry J, Finn R. 2007. Pfam: a domain-centric method for analyzing proteins and proteomes. Methods Mol Biol 396:43–58.

85. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.

86. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Xenarios I, Yeh L-S, Young S-Y, Mitchell AL. 2017. InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res 45:D190–D199.

87. Wilkins D. 2020. gggenes. available: https://github.com/wilkox/gggenes.

88. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the Tidyverse. J Open Source Softw 4:1686.

89. Müller K. 2017. here: A Simpler Way to Find Your Files. 1.0.1. available: https://cran.r-project.org/web/packages/here/index.html.

90. Lassmann T. 2019. Kalign 3: multiple sequence alignment of large data sets. Bioinformatics https://doi.org/10.1093/bioinformatics/btz795.

91. R Core Team. 2019. R: A language and environment for statistical computing. R Found Stat Comput. available: https://www.r-project.org/.

92. RStudio Team. 2020. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA. available: https://www.rstudio.com/.

93. Burke JT, Chen SZ, Sosinski LM, Johnston JB, Ravi J. 2022. MolEvolvR: A web-app for characterizing proteins using molecular evolution and phylogeny. bioRxiv 2022.02.18.461833.

94. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–7.

95. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113.

96. Bonatesta E, Horejs-Kainrath C, Bodenhofer U. 2020. Multiple Sequence Alignment. 1.1.2. available: https://www.bioinf.jku.at/software/msa/.

97. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490.

98. Rambaut A. 2009. FigTree. availabe: https://tree.bio.ed.ac.uk/software/figtree/.

99. Paradis E, Blomberg S, Bolker B, Brown J, Claramunt S, Claude J, Cuong HS, Desper R, Didier G, Durand B, Dutheil J, Ewing R, Gascuel O, Guillerme T, Heibl C, Ives A, Jone B, Krah F, Lawson D, Lefort V, Legendre P, Lemon J, Louvel G, Marcon E, McCloskey R, Nylander J, Opgen-Rhein R, Popescu A-A, Royer-Carenzi M, Schliep K, Strimmer K, de Vienne D. 2020. ape: Analyses of Phylogenetics and Evolution. 5.4–1. available: https://ape-package.ird.fr/.

100. Fernandez NL, Hsueh BY, Nhu NTQ, Franklin JL, Dufour YS, Waters CM. 2020. *Vibrio cholerae* adapts to sessile and motile lifestyles by cyclic di-GMP regulation of cell shape. Proc Natl Acad Sci 117:29046–29054.

101. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein

V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. Nat Methods 9:676–682.

102. Fernandez NL, Waters CM. 2019. Cyclic di-GMP Increases Catalase Production and Hydrogen Peroxide Tolerance in *Vibrio cholerae*. Appl Environ Microbiol 85.

103. Dryselius R, Izutsu K, Honda T, Iida T. 2008. Differential replication dynamics for large and small Vibrio chromosomes affect gene dosage, expression and location. BMC Genomics 9:559.

104. Koirala D, Lewicka A, Koldobskaya Y, Huang H, Piccirilli JA. 2020. Synthetic Antibody Binding to a Preorganized RNA Domain of Hepatitis C Virus Internal Ribosome Entry Site Inhibits Translation. ACS Chem Biol 15:205–216.

105. Piccirilli JA, Koldobskaya Y. 2011. Crystal structure of an RNA polymerase ribozyme in complex with an antibody fragment. Philos Trans R Soc Lond B Biol Sci 366:2918–28.

106. Shelke SA, Shao Y, Laski A, Koirala D, Weissman BP, Fuller JR, Tan X, Constantin TP, Waggoner AS, Bruchez MP, Armitage BA, Piccirilli JA. 2018. Structural basis for activation of fluorogenic dyes by an RNA aptamer lacking a G-quadruplex motif. Nat Commun 9:4542.

107. Koirala D, Shao Y, Koldobskaya Y, Fuller JR, Watkins AM, Shelke SA, Pilipenko E V., Das R, Rice PA, Piccirilli JA. 2019. A conserved RNA structural motif for organizing topology within picornaviral internal ribosome entry sites. Nat Commun 10:3629.

108. Kao C, Rüdisser S, Zheng M. 2001. A simple and efficient method to transcribe RNAs with reduced 3’ heterogeneity. Methods 23:201–5.

109. Dong H, Liu Y, Zu X, Li N, Li F, Zhang D. 2015. An enzymatic assay for high-throughput screening of cytidine-producing microbial strains. PLoS One 10:e0121612.

110. Ngo TT, Phan APH, Yam CF, Lenhoff HM. 1982. Interference in determination of ammonia with the hypochlorite-alkaline phenol method of Berthelot. Anal Chem 54:46–49.

111. Massie JP, Reynolds EL, Koestler BJ, Cong J-PJ-P, Agostoni M, Waters CM. 2012. Quantification of high-specificity cyclic diguanylate signaling. Proc Natl Acad Sci 109:12746–12751.

112. Waldron EJ, Snyder D, Fernandez NL, Sileo E, Inoyama D, Freundlich JS, Waters CM, Cooper VS, Neiditch MB. 2019. Structural basis of DSF recognition by its receptor RpfR and its regulatory interaction with the DSF synthase RpfF. PLOS Biol 17:e3000123.

113. Leipe DD, Koonin E V., Aravind L. 2003. Evolution and classification of P-loop kinases and related proteins. J Mol Biol 333:781–815.

114. Handee W, Li X, Hall KW, Deng X, Li P, Benning C, Williams BL, Kuo M-H. 2016. An Energy-Independent Pro-longevity Function of Triacylglycerol in Yeast. PLOS Genet 12:e1005878.

115. Boussardon C, Avon A, Kindgren P, Bond CS, Challenor M, Lurin C, Small I. 2014. The cytidine deaminase signature HxE(x) n CxxC of DYW1 binds zinc and is necessary for RNA editing of ndhD-1. New Phytol 203:1090–1095.

116. Iyer LM, Zhang D, Rogozin IB, Aravind L. 2011. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. Nucleic Acids Res 39:9473–97.

117. Vértessy BG, Tóth J. 2009. Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases. Acc Chem Res 42:97–106.

118. Khan SR, Kuzminov A. 2019. Thymineless Death in *Escherichia coli* Is Unaffected by Chromosomal Replication Complexity. J Bacteriol 201.

119. Papenfort K, Förstner KU, Cong J-P, Sharma CM, Bassler BL. 2015. Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. Proc Natl Acad Sci 112:E766–E775.

120. Krin E, Pierlé SA, Sismeiro O, Jagla B, Dillies M-AA, Varet H, Irazoki O, Campoy S, Rouy Z, Cruveiller S, Médigue C, Coppée J-YY, Mazel D. 2018. Expansion of the SOS regulon of *Vibrio cholerae* through extensive transcriptome analysis and experimental validation. BMC Genomics 19:1–18.

121. Short FL, Akusobi C, Broadhurst WR, Salmond GPC. 2018. The bacterial Type III toxin-antitoxin system, ToxIN, is a dynamic protein-RNA complex with stability-dependent antiviral abortive infection activity. Sci Rep 8:1013.

122. Guegler CK, Laub MT. 2021. Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection. Mol Cell 81:2361-2373.e9.

123. Conticello SG, Thomas CJF, Petersen-Mahrt SK, Neuberger MS. 2005. Evolution

of the AID/APOBEC Family of Polynucleotide (Deoxy)cytidine Deaminases. Mol Biol Evol 22:367–377.

124. Dy RL, Przybilski R, Semeijn K, Salmond GPCC, Fineran PC. 2014. A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. Nucleic Acids Res 42:4590–605.

125. Koga M, Otsuka Y, Lemire S, Yonesaki T. 2011. *Escherichia coli* rnlA and rnlB compose a novel toxin-antitoxin system. Genetics 187:123–130.

126. Pecota DC, Wood TK. 1996. Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. J Bacteriol 178:2044–2050.

127. Jemielita M, Wingreen NS, Bassler BL. 2018. Quorum sensing controls *Vibrio cholerae* multicellular aggregate formation. Elife 7:e42057.

128. Johansson E, Fanø M, Bynck JH, Neuhard J, Larsen S, Sigurskjold BW, Christensen U, Willemoës M. 2005. Structures of dCTP Deaminase from *Escherichia coli* with Bound Substrate and Product. J Biol Chem 280:3051–3059.

129. Tal N, Millman A, Stokar-Avihail A, Fedorenko T, Leavitt A, Melamed S, Yirmiya E, Avraham C, Amitai G, Sorek R. 2021. Antiviral defense via nucleotide depletion in bacteria. bioRxiv 2021.04.26.441389.

130. Harms A, Brodersen DE, Mitarai N, Gerdes K. 2018. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. Mol Cell 70:768–784.

131. Ayinde D, Casartelli N, Schwartz O. 2012. Restricting HIV the SAMHD1 way: Through nucleotide starvation. Nat Rev Microbiol 10:675–680.

132. Shepard C, Xu J, Holler J, Kim DH, Mansky LM, Schinazi RF, Kim B. 2019. Effect of induced dNTP pool imbalance on HIV-1 reverse transcription in macrophages. Retrovirology 16:29.

133. Oo A, Kim DH, Schinazi RF, Kim B. 2020. Viral protein X reduces the incorporation of mutagenic noncanonical rNTPs during lentivirus reverse transcription in macrophages. J Biol Chem 295:657–666.

134. Page R, Peti W. 2016. Toxin-antitoxin systems in bacterial growth arrest and persistence. Nat Chem Biol 2016 124 12:208–214.

135. Ogura T, Hiraga S. 1983. Mini-F plasmid genes that couple host cell division to plasmid proliferation. Proc Natl Acad Sci 80:4784–4788.

136. Gotfredsen M, Gerdes K. 1998. The Escherichia coli relBE genes belong to a new toxin–antitoxin gene family. Mol Microbiol 29:1065–1076.

137. Lehnherr H, Maguin E, Jafri S, Yarmolinsky MB. 1993. Plasmid Addiction Genes of Bacteriophage P1: doc, which Causes Cell Death on Curing of Prophage, and phd, which Prevents Host Death when Prophage is Retained. J Mol Biol 233:414–428.

138. Guglielmini J, Van Melderen L. 2011. Bacterial toxin-antitoxin systems: Translation inhibitors everywhere. Mob Genet Elements 1:283–290.

139. Song S, Wood TK. 2020. A Primary Physiological Role of Toxin/Antitoxin Systems Is Phage Inhibition. Front Microbiol 11:1895.

140. Choi JS, Kim W, Suk S, Park H, Bak G, Yoon J, Lee Y. 2018. The small RNA, SdsR, acts as a novel type of toxin in *Escherichia coli.* RNA Biol 15:1319–1335.

141. Peltier J, Hamiot A, Garneau JR, Boudry P, Maikova A, Hajnsdorf E, Fortier L-C, Dupuy B, Soutourina O. 2020. Type I toxin-antitoxin systems contribute to the maintenance of mobile genetic elements in *Clostridioides difficile*. Commun Biol 3:718.

142. LeRoux M, Culviner PH, Liu YJ, Littlehale ML, Laub MT. 2020. Stress Can Induce Transcription of Toxin-Antitoxin Systems without Activating Toxin. Mol Cell 79:280-292.e8.

143. Enikeeva FN, Severinov K V., Gelfand MS. 2010. Restriction-modification systems and bacteriophage invasion: who wins? J Theor Biol 266:550.

144. Hazan R, Engelberg-Kulka H. 2004. *Escherichia coli* mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. Mol Genet Genomics 272:227–234.

145. Szekeres S, Dauti M, Wilde C, Mazel D, Rowe-Magnus DA. 2007. Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. Mol Microbiol 63:1588–605.

146. Pandey DP, Gerdes K. 2005. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. Nucleic Acids Res 33:966–976.

147. Severin GB, Hsueh BY, Elg CA, Dover JA, Rhoades CR, Wessel AJ, Ridenhour BJ, Top EM, Ravi J, Parent KN, Waters CM. 2021. A Broadly Conserved Deoxycytidine Deaminase Protects Bacteria from Phage Infection. bioRxiv

2021.03.31.437871.

148. Dover JA, Burmeister AR, Molineux IJ, Parent KN. 2016. Evolved populations of *Shigella flexneri* phage Sf6 acquire large deletions, altered genomic architecture, and faster life cycles. Genome Biol Evol 8:2827–2840.

149. Schrad JR, Abrahão JS, Cortines JR, Parent KN. 2020. Structural and Proteomic Characterization of the Initiation of Giant Virus Infection. Cell 181:1046-1061.e6.

150. Mastronarde DN. 2005. Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol 152:36–51.

151. Lindahl T, Ljungquist S, Siegert W, Nyberg B, Sperens B. 1977. DNA N-glycosidases: properties of uracil-DNA glycosidase from *Escherichia coli*. J Biol Chem 252:3286–94.

152. Lindahl T. 1990. Repair of intrinsic DNA lesions. Mutat Res 238:305–11.

153. Hinton DM. 2010. Transcriptional control in the prereplicative phase of T4 development. Virol J 7:289.

154. Hesselbach BA, Nakada D. 1977. “Host shutoff” function of bacteriophage T7: involvement of T7 gene 2 and gene 0.7 in the inactivation of *Escherichia coli* RNA polymerase. J Virol 24:736–45.

155. Davison J. 2015. Pre-early functions of bacteriophage T5 and its relatives. Bacteriophage 5:e1086500.

156. Brantl S, Jahn N. 2015. sRNAs in bacterial type I and type III toxin-antitoxin systems. FEMS Microbiol Rev 39:413–427.

157. Bernstein JA, Khodursky AB, Lin P-H, Lin-Chao S, Cohen SN. 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci U S A 99:9697–702.

158. De Paepe M, Taddei F. 2006. Viruses’ life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. PLoS Biol 4:e193.

159. LeRoux M, Srikant S, Littlehale MH, Teodoro G, Doron S, Badiee M, Leung AKL, Sorek R, Laub MT. 2021. The DarTG toxin-antitoxin system provides phage defense by ADP-ribosylating viral DNA. bioRxiv 2021.09.27.462013.

160. Bernheim A, Millman A, Ofir G, Meitav G, Avraham C, Shomar H, Rosenberg MM, Tal N, Melamed S, Amitai G, Sorek R. 2020. Prokaryotic viperins produce diverse antiviral molecules. Nature https://doi.org/10.1038/s41586-020-2762-2.

161. Weigele P, Raleigh EA. 2016. Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. Chem Rev 116:12655–12687.

162. Westra ER, Swarts DC, Staals RHJ, Jore MM, Brouns SJJ, van der Oost J. 2012. The CRISPRs, They Are A-Changin': How Prokaryotes Generate Adaptive Immunity. Annu Rev Genet 46:311–339.

163. Hays SG, Seed KD. 2020. Dominant *Vibrio cholerae* phage exhibits lysis inhibition sensitive to disruption by a defensive phage satellite. Elife 9.

164. Millman A, Bernheim A, Stokar-Avihail A, Fedorenko T, Voichek M, Leavitt A, Oppenheimer-Shaanan Y, Sorek R. 2020. Bacterial Retrons Function In Anti-Phage Defense. Cell 183:1551-1561.e12.

165. Payne LJ, Todeschini TC, Wu Y, Perry BJ, Ronson CW, Fineran PC, Nobrega FL, Jackson SA. 2021. Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. Nucleic Acids Res 49:10868–10878.

166. Tesson F, Hervé A, Touchon M, d'Humières C, Cury J, Bernheim A. 2021. Systematic and quantitative view of the antiviral arsenal of prokaryotes. bioRxiv 2021.09.02.458658.

167. Yen M, Cairns LS, Camilli A. 2017. A cocktail of three virulent bacteriophages prevents *Vibrio cholerae* infection in animal models. Nat Commun 8:14187.

168. Lin DM, Koskella B, Lin HC. 2017. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. World J Gastrointest Pharmacol Ther 8:162–173.

169. Cisek AA, Dąbrowska I, Gregorczyk KP, Wyżewski Z. 2017. Phage Therapy in Bacterial Infections Treatment: One Hundred Years After the Discovery of Bacteriophages. Curr Microbiol 74:277–283.

170. Shabram P, Aguilar-Cordova E. 2000. Multiplicity of infection/multiplicity of confusion. Mol Ther 2:420–1.

171. Ellis EL, Delbrück M. 1939. The Growth of Bacteriophage. J Gen Physiol 22:365–84.

172. Brennan MA, Rosenthal AZ. 2021. Single-Cell RNA Sequencing Elucidates the Structure and Organization of Microbial Communities. Front Microbiol 12:713128.

173. McNulty R, Sritharan D, Liu S, Hormoz S, Rosenthal AZ. 2021. Droplet-based single cell RNA sequencing of bacteria identifies known and previously unseen cellular states. bioRxiv 2021.03.10.434868.

174. Shao Y, Bassler BL. 2012. Quorum-sensing non-coding small RNAs use unique pairing regions to differentially control mRNA targets. Mol Microbiol 83:599–611.

175. Bøggild A, Sofos N, Andersen KR, Feddersen A, Easter AD, Passmore LA, Brodersen DE. 2012. The crystal structure of the intact *E. coli* RelBE toxin-antitoxin complex provides the structural basis for conditional cooperativity. Structure 20:1641–8.

176. Favor AH, Llanos CD, Youngblut MD, Bardales JA. 2020. Optimizing bacteriophage engineering through an accelerated evolution platform. Sci Reports 2020 101 10:1–10.

177. Smits THM, Rezzonico F, Kamber T, Blom J, Goesmann A, Frey JE, Duffy B. 2010. Complete Genome Sequence of the Fire Blight Pathogen *Erwinia amylovora* CFBP 1430 and Comparison to Other Erwinia spp. Mol Plant-Microbe Interact 23:384–393.

178. Norelli JL, Jones AL, Aldwinckle HS. 2003. Fire Blight Management in the Twenty-first Century: Using New Technologies that Enhance Host Resistance in Apple. Plant Dis 87:756–765.

179. Oh C-S, Beer S V. 2005. Molecular genetics of *Erwinia amylovora* involved in the development of fire blight. FEMS Microbiol Lett 253:185–92.

180. Billing E. 1974. The Effect of Temperature on the Growth of the Fireblight Pathogen, *Erwinia amylovora*. J Appl Bacteriol 37:643–648.

181. Khan MA, Zhao Y, Korban SS. 2012. Molecular Mechanisms of Pathogenesis and Resistance to the Bacterial Pathogen *Erwinia amylovora*, Causal Agent of Fire Blight Disease in Rosaceae. Plant Mol Biol Report 30:247–260.

182. Borruso L, Salomone-Stagni M, Polsinelli I, Schmitt AO, Benini S. 2017. Conservation of *Erwinia amylovora* pathogenicity-relevant genes among Erwinia genomes. Arch Microbiol 199:1335–1344.

183. Sundin GW, Wang N. 2018. Antibiotic Resistance in Plant-Pathogenic Bacteria.

Annu Rev Phytopathol 56:161–180.

184. Piqué N, Miñana-Galbis D, Merino S, Tomás J. 2015. Virulence Factors of *Erwinia amylovora*: A Review. Int J Mol Sci 16:12836–12854.

185. Lee JH, Zhao Y. 2018. Integration of multiple stimuli-sensing systems to regulate HrpS and type III secretion system in *Erwinia amylovora*. Mol Genet Genomics 293:187–196.

186. Edmunds AC, Castiblanco LF, Sundin GW, Waters CM. 2013. Cyclic Di-GMP Modulates the Disease Progression of *Erwinia amylovora*. J Bacteriol 195:2155–2165.

187. Pester D, Milčevičová R, Schaffer J, Wilhelm E, Blümel S. 2012. *Erwinia amylovora* Expresses Fast and Simultaneously hrp/dsp Virulence Genes during Flower Infection on Apple Trees. PLoS One 7:e32583.

188. Koczan JM, McGrath MJ, Zhao Y, Sundin GW. 2009. Contribution of Erwinia amylovora Exopolysaccharides Amylovoran and Levan to Biofilm Formation: Implications in Pathogenicity. Phytopathology 99:1237–1244.

189. Romling U, Galperin MY, Gomelsky M. 2013. Cyclic di-GMP: the First 25 Years of a Universal Bacterial Second Messenger. Microbiol Mol Biol Rev 77:1–52.

190. Ryjenkov DA, Tarutina M, Moskvin O V., Gomelsky M. 2005. Cyclic Diguanylate Is a Ubiquitous Signaling Molecule in Bacteria: Insights into Biochemistry of the GGDEF Protein Domain. J Bacteriol 187:1792–1798.

191. Chen G, Zhao Q, Zhu F, Chen R, Jin Y, Liu C, Pan X, Jin S, Wu W, Cheng Z. 2016. Oligoribonuclease is required for the type III secretion system and pathogenesis of *Pseudomonas aeruginosa*. Microbiol Res 188–189:90–96.

192. Orr MW, Donaldson GP, Severin GB, Wang J, Sintim HO, Waters CM, Lee VT. 2015. Oligoribonuclease is the primary degradative enzyme for pGpG in *Pseudomonas aeruginosa* that is required for cyclic-di-GMP turnover. Proc Natl Acad Sci U S A 112:E5048-57.

193. Orr MW, Weiss CA, Severin GB, Turdiev H, Kim S-K, Turdiev A, Liu K, Tu BP, Waters CM, Winkler WC, Lee VT. 2018. A Subset of Exoribonucleases Serve as Degradative Enzymes for pGpG in c-di-GMP Signaling. J Bacteriol 200.

194. Kharadi RR, Castiblanco LF, Waters CM, Sundin GW. 2018. Phosphodiesterase Genes Regulate Amylovoran Production, Biofilm Formation, and Virulence in

Erwinia amylovora. Appl Environ Microbiol 85.

195. Hengge R. 2010. Cyclic-di-GMP reaches out into the bacterial RNA world. Sci Signal 3:pe44.

196. López-Baena FJ, Vinardell JM, Medina C. 2019. Regulation of Protein Secretion Systems Mediated by Cyclic Diguanylate in Plant-Interacting Bacteria. Front Microbiol 10:1289.

197. Wei ZM, Beer S V. 1995. hrpL activates *Erwinia amylovora* hrp gene transcription and is a member of the ECF subfamily of sigma factors. J Bacteriol 177:6201–10.

198. Lee JH, Sundin GW, Zhao Y. 2016. Identification of the HrpS binding site in the hrpL promoter and effect of the RpoN binding site of HrpS on the regulation of the type III secretion system in *Erwinia amylovora.* Mol Plant Pathol 17:691–702.

199. Ramos LS, Lehman BL, Sinn JP, Pfeufer EE, Halbrendt NO, McNellis TW. 2013. The fire blight pathogen *Erwinia amylovora* requires the rpoN gene for pathogenicity in apple. Mol Plant Pathol 14:838–843.

200. Zhao Y, Wang D, Nakka S, Sundin GW, Korban SS. 2009. Systems level analysis of two-component signal transduction systems in *Erwinia amylovora*: role in virulence, regulation of amylovoran biosynthesis and swarming motility. BMC Genomics 10:245.

201. Wei Z, Kim JF, Beer S V. 2000. Regulation of hrp Genes and Type III Protein Secretion in *Erwinia amylovora* by HrpX/HrpY, a Novel Two-Component System, and HrpS. Mol Plant-Microbe Interact 13:1251–1262.

202. Wang J, Shao X, Zhang Y, Zhu Y, Yang P, Yuan J, Wang T, Yin C, Wang W, Chen S, Liang H, Deng X. 2018. HrpS Is a Global Regulator on Type III Secretion System (T3SS) and Non-T3SS Genes in P*seudomonas savastanoi* pv. phaseolicola. Mol Plant-Microbe Interact 31:1232–1243.

203. Bush M, Dixon R. 2012. The Role of Bacterial Enhancer Binding Proteins as Specialized Activators of 54-Dependent Transcription. Microbiol Mol Biol Rev 76:497–529.

204. Baraquet C, Murakami K, Parsek MR, Harwood CS. 2012. The FleQ protein from *Pseudomonas aeruginosa* functions as both a repressor and an activator to control gene expression from the pel operon promoter in response to c-di-GMP. Nucleic Acids Res 40:7207–18.

205. Srivastava D, Hsieh M-L, Khataokar A, Neiditch MB, Waters CM. 2013. Cyclic di-GMP inhibits *Vibrio cholerae* motility by repressing induction of transcription and inducing extracellular polysaccharide production. Mol Microbiol 90:1262–76.

206. Hsieh M-L, Hinton DM, Waters CM. 2018. VpsR and cyclic di-GMP together drive transcription initiation to activate biofilm formation in *Vibrio cholerae*. Nucleic Acids Res 46:8876–8887.

207. Zhao Y, Sundin GW, Wang D. 2009. Construction and analysis of pathogenicity island deletion mutants of *Erwinia amylovora*. Can J Microbiol 55:457–64.

208. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–5.

209. He C, Ohnishi K. 2017. Efficient renaturation of inclusion body proteins denatured by SDS. Biochem Biophys Res Commun 490:1250–1253.

210. Rule CS, Patrick M, Sandkvist M. 2016. Measuring In Vitro ATPase Activity for Enzymatic Characterization. J Vis Exp 2016.

211. Zhao Y, He S-Y, Sundin GW. 2006. The *Erwinia amylovora* avrRpt2EA gene contributes to virulence on pear and AvrRpt2EA is recognized by Arabidopsis RPS2 when expressed in *Pseudomonas syringae*. Mol Plant Microbe Interact 19:644–54.

212. Conner JG, Zamorano-Sánchez D, Park JH, Sondermann H, Yildiz FH. 2017. The ins and outs of cyclic di-GMP signaling in *Vibrio cholerae*. Curr Opin Microbiol 36:20–29.

213. Orr MW, Lee VT. 2017. Differential Radial Capillary Action of Ligand Assay (DRaCALA) for High-Throughput Detection of Protein–Metabolite Interactions in Bacteria, p. 25–41. *In* Methods in molecular biology (Clifton, N.J.). Methods Mol Biol.

214. Kriel A, Bittner AN, Kim SH, Liu K, Tehranchi AK, Zou WY, Rendon S, Chen R, Tu BP, Wang JD. 2012. Direct regulation of GTP homeostasis by (p)ppGpp: a critical component of viability and stress resistance. Mol Cell 48:231–41.

215. Kriel A, Brinsmade SR, Tse JL, Tehranchi AK, Bittner AN, Sonenshein AL, Wang JD. 2014. GTP dysregulation in *Bacillus subtilis* cells lacking (p)ppGpp results in phenotypic amino acid auxotrophy and failure to adapt to nutrient downshift and regulate biosynthesis genes. J Bacteriol 196:189–201.

216. Gaca AO, Colomer-Winter C, Lemos JA. 2015. Many means to a common end: the intricacies of (p)ppGpp metabolism and its control of bacterial homeostasis. J Bacteriol 197:1146–56.

217. Chou S-H, Galperin MY. 2016. Diversity of Cyclic Di-GMP-Binding Proteins and Mechanisms. J Bacteriol 198:32–46.

218. Lacey MM, Partridge JD, Green J. 2010. *Escherichia coli* K-12 YfgF is an anaerobic cyclic di-GMP phosphodiesterase with roles in cell surface remodelling and the oxidative stress response. Microbiology 156:2873–2886.

219. Cohen D, Mechold U, Nevenzal H, Yarmiyhu Y, Randall TE, Bay DC, Rich JD, Parsek MR, Kaever V, Harrison JJ, Banin E. 2015. Oligoribonuclease is a central feature of cyclic diguanylate signaling in *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A 112:11359–11364.

220. Burse A, Weingart H, Ullrich MS. 2004. The phytoalexin-inducible multidrug efflux pump AcrAB contributes to virulence in the fire blight pathogen, *Erwinia amylovora.* Mol Plant Microbe Interact 17:43–54.