CRISPR REGULATION AND RNA BIOGENESIS

IN STAPHYLOCOCCUS EPIDERMIDIS

by

FORREST CHARLES WALKER

ASMA HATOUM-ASLAN, COMMITTEE CHAIR
JULIE OLSON
PATRICK FRANTOM

A THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in the
Department of Biological Sciences
in the Graduate School of The
University of Alabama

TUSCALOOSA, ALABAMA

2017
ABSTRACT

CRISPR-Cas is a prokaryotic adaptive immune system that recognizes and degrades specific nucleic acid invaders, producing a “memory” of past infections that allows for heritable immunity. CRISPR-Cas systems are widespread among prokaryotes, including many human pathogens such as Staphylococcus aureus and S. epidermidis. These species can cause severe diseases, increasingly so with rising proportions of antibiotic-resistant staphylococcal infections being reported, including the ‘superbug’ MRSA. Novel therapies are required to deal with the growing antibiotic resistance crisis and prevent a post-antibiotic era. These therapies may include using novel bactericidal agents, such as bacteriophages, as well as strategies to combat the transfer of antibiotic resistance between these pathogens. CRISPR-Cas stands poised as a major potential barrier to the implementation of bacteriophage therapy, as these systems can adapt and make whole populations of bacteria immune to these treatments. Additionally, CRISPR-Cas is known to block horizontal gene transfer of antibiotic resistance plasmids between staphylococcal strains, and is thus crucial for limiting the spread of antibiotic resistance. Understanding how this system functions will aid the development of novel ways to modulate these systems to allow for successful bacteriophage therapies and prevent the spread of antibiotic resistance. We have investigated the mechanistic roles of the cas genes in S. epidermidis CRISPR-Cas immunity and identified non-Cas nucleases that may also be involved in this immunity, linking CRISPR to important physiological processes such as RNA degradation and recycling. Additionally, we have started characterizing transcriptional regulation of CRISPR-Cas in this important human
pathogen, identifying environmental cues that the bacteria use to ‘decide’ when to permit or deny entrance of nucleic acid invaders. Finally, we are beginning to probe the ways in which bacteriophages suppress the bacterial immune response to permit infection in the face of CRISPR immunity
**LIST OF ABBREVIATIONS AND SYMBOLS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CNS</strong></td>
<td>Coagulase-Negative Staphylococci</td>
</tr>
<tr>
<td><strong>MGE</strong></td>
<td>Mobile genetic element: nucleic acids that are exchanged between cells, typically plasmids or bacteriophages</td>
</tr>
<tr>
<td><strong>MRSA</strong></td>
<td>Methicillin-resistant <em>Staphylococcus aureus</em></td>
</tr>
<tr>
<td><strong>HGT</strong></td>
<td>Horizontal gene transfer: the spread of mobile genetic elements between organisms via several processes including transduction, transformation, and conjugation</td>
</tr>
<tr>
<td><strong>CRISPR</strong></td>
<td>Clustered, regularly interspaced short palindromic repeats: genomic loci that encode for CRISPR RNAs which specify the target of CRISPR-Cas immunity</td>
</tr>
<tr>
<td><strong>Cas</strong></td>
<td>CRISPR-Associated: refers to the genes responsible for carrying out CRISPR RNA-mediated interference of mobile genetic elements found in the genome with the CRISPR repeat arrays</td>
</tr>
<tr>
<td><strong>Cas10-Csm</strong></td>
<td>The Cas complex found in <em>S. epidermidis</em> RP62a and another Type III-A CRISPR-Cas systems, characterized by the presence of Cas10, composed of a total of five Cas proteins</td>
</tr>
<tr>
<td><strong>crRNA</strong></td>
<td>CRISPR RNAs: short RNAs transcribed from the CRISPR locus which specify the target of CRISPR-Cas immunity via base-pairing</td>
</tr>
<tr>
<td><strong>bp</strong></td>
<td>Base-pairs</td>
</tr>
<tr>
<td><strong>nt</strong></td>
<td>Nucleotides</td>
</tr>
<tr>
<td><strong>crRNA</strong></td>
<td>CRISPR RNAs: short RNAs transcribed from the CRISPR locus which specify the target of CRISPR-Cas immunity via base-pairing</td>
</tr>
<tr>
<td><strong>bp</strong></td>
<td>Base-pairs</td>
</tr>
<tr>
<td><strong>TSB</strong></td>
<td>Tryptic Soy Broth</td>
</tr>
<tr>
<td><strong>BHI</strong></td>
<td>Brain Heart Infusion</td>
</tr>
</tbody>
</table>
\textit{spc1} The 35-bp region of the CRISPR locus that encodes for a crRNA that targets the nickase gene found on conjugative staphylococcal plasmids

\textit{spc2} The 35-bp region of the CRISPR locus that encodes for a crRNA that targets the staphylococcal bacteriophage CNPH82

\textit{PAGE} Polyacrylamide Gel Electrophoresis

\textit{cDNA} Complementary DNA: DNA synthesized from an RNA template

\textit{Aca} Anti-CRISPR associated protein/gene

\textit{Acr} Anti-CRISPR protein/gene
ACKNOWLEDGMENTS

I would like to thank all my family and friends who stuck by me during the last two years as well as my wonderful boyfriend, Joe, as I pursued my master’s degree and put up with me talking incessantly about CRISPR. I would especially like to thank my advisor, Asma Hatoum-Aslan, for graciously taking me on as a master’s student in her first year as a professor here. I, of course, could not have completed any of this work without her wonderful guidance and support, and I am truly indebted to her for her mentorship. I would also like to thank my undergraduate assistant, Emma Culbertson, for the hours she has put into this project and her invaluable aid. Additionally, I need to acknowledge all my lab mates who have always been there to lend a helping hand when needed, especially Katie Cater and Lucy Chou-Zheng. Finally, I am grateful to the members of my committee, Julie Olson and Patrick Frantom, for their support and for agreeing to help me through this process. Without any one of these individuals, I could not have made it to this point.
CONTENTS

ABSTRACT ......................................................................................................................... ii

LIST OF ABBREVIATIONS AND SYMBOLS ................................................................. iv

ACKNOWLEDGMENTS ................................................................................................... vi

LIST OF TABLES ............................................................................................................... x

LIST OF FIGURES ......................................................................................................... xi

CHAPTER 1 INTRODUCTION ......................................................................................... 1

1.1 *Staphylococcus epidermidis* Pathogenicity .............................................................. 1

1.2 Horizontal Gene Transfer and *S. epidermidis* Virulence ......................................... 3

1.3 Bacteriophage Therapy ............................................................................................. 6

1.4 Anti-Phage Defense Mechanisms ................................................................................ 7

1.5 CRISPR-Cas in *S. epidermidis* RP62a ................................................................... 9

CHAPTER 2 crRNA MATURATION .................................................................................... 12

2.1 INTRODUCTION ........................................................................................................ 12

2.2 METHODOLOGY ....................................................................................................... 13

2.2.1 Bacterial Growth Conditions ................................................................................ 13

2.2.2 Plasmid Construction and Cloning ....................................................................... 14

2.2.3 CRISPR Functionality Assays .............................................................................. 14

2.2.4 RNA Extraction and Biotin Pulldowns ................................................................. 15

2.2.5 Double-Filter Protein-Nucleic Acid Binding Assays ............................................ 16
2.2.6 Statistical Analysis.................................................................17
2.3 RESULTS AND DISCUSSION.........................................................17
  2.3.1 Identification of Putative Active Site Residues .......................17
  2.3.2 Anti-Plasmid Immunity of pCRISPR-Cas Mutants ..................20
  2.3.3 crRNA Size Distribution of pCRISPR-Cas Mutants ...............26
  2.3.4 Anti-Bacteriophage Immunity of pCRISPR-Cas Mutants ..........29
  2.3.5 Biochemical Characterization of Csm5 ................................31
  2.3.6 Roles of Non-Cas Nucleases in Cas10-Csm Maturation ..........34
2.4 FUTURE DIRECTIONS ..................................................................35

CHAPTER 3 REGULATION OF CRISPR-CAS IN S. EPIDERMIDIS ..........37
  3.1 INTRODUCTION ........................................................................36
  3.2 METHODOLOGY .......................................................................38
    3.2.1 Bacterial Growth Conditions ...........................................39
    3.2.2 Plasmid Construction and Cloning ...................................39
    3.2.3 GFP Reporter Expression Assays ......................................40
    3.2.4 RNA Extraction and Primer Extension ..............................41
    3.2.5 Statistical Analysis ...........................................................42
  3.3 RESULTS AND DISCUSSION .....................................................42
    3.3.1 Identification and Characterization of crRNA Promoters ......42
    3.3.2 Screening for Environmental Regulators of crRNA Expression ..48
  3.4 FUTURE DIRECTIONS ...............................................................56

CHAPTER 4 INHIBITION OF CRISPR-CAS BY BACTERIOPHAGES ............57
  4.1 INTRODUCTION ........................................................................57
4.2 METHODOLOGY .................................................................................................................. 59
  4.2.1 Preparation of Phage High-Titer Lysates ....................................................................... 59
  4.2.2 Phage Genomic DNA Extraction .................................................................................. 59
  4.2.3 Statistical Analysis ........................................................................................................ 60

4.3 RESULTS AND DISCUSSION .......................................................................................... 60
  4.3.1 Identification of Putative Anti-CRISPR Proteins in Staphylococcal Phages ............... 60
  4.3.2 Characterization of Phages Resistant to Cas10-Csm Interference ................................ 64
  4.3.3 Identification of Homologues of Putative Staphylococcal Anti-CRISPR Genes ....... 71

4.4 FUTURE DIRECTIONS ....................................................................................................... 72

CHAPTER 5 CONCLUSIONS ..................................................................................................... 74

REFERENCES ............................................................................................................................. 76

APPENDIX ................................................................................................................................ 89
LIST OF TABLES

2.1 Conserved charged residues in Csm2, Csm3, and Csm5 are required for optimal anti-plasmid immunity in a heterologous host ................................................................. 23

2.2 A conserved glutamate in Csm5 is required for efficient anti-plasmid immunity in a native host .................................................................................................................. 25

3.1 Four promoters are predicted to be present within Regions A and B of the crRNA leader. ...45

3.2 List of environmental conditions examined as potential regulators of CRISPR-Cas and summarized results ............................................................................................................. 49
LIST OF FIGURES

1.1 The CRISPR-Cas system detects and interferes with invasive genetic elements in three steps. .................................................................................................................................................. 10

2.1 Identification of conserved residues in the CRISPR-associated proteins Csm2, Csm3, and Csm5 ........................................................................................................................................... 20

2.2 Conserved charged residues in Csm2, Csm3, and Csm5 are essential in a heterologous background for CRISPR-Cas anti-plasmid immunity ................................................................. 22

2.3 A conserved, charged glutamate in Csm5 is required for efficient anti-plasmid immunity in a native host ........................................................................................................................... 24

2.4 Mutations in conserved, charged residues of Csm3 and Csm5 result in altered crRNA profiles in a heterologous and native host ........................................................................................................ 27

2.5 Complex formation is essential, but wild-type crRNA sizes are not required, for Cas10-Csm anti-bacteriophage immunity ................................................................................................ 30

2.6 Csm5 binds to crRNAs with an affinity in the nanomolar range in a MgCl₂-enhanced manner .................................................................................................................................................. 34

3.1 Two large regions of the crRNA leader are required for wild-type crRNA expression levels in a GFP reporter assay .................................................................................................................. 44

3.2 Multiple putative, well-conserved promoters are present within the crRNA leader region .... 46

3.3 Two promoters are involved in wild-type expression of S. epidermidis RP62a crRNAs ...... 48

3.4 NaCl represses crRNA expression ............................................................................................ 52

3.5 Zn²⁺ represses the expression of crRNAs in a dose-dependent manner. ................................. 54

4.1 Putative acr and aca genes are found within the lysis region of the genomes of three bacteriophages that infect S. epidermidis RP62a ................................................................................. 61
4.2 The putative aca and acr genes of ISP, K7, and MP16 are conserved at the amino acid level, though the putative acrlaca in ISP are divergent from those of K7 and MP16..................63

4.3 The regions downstream of the putative acrlaca locus is well-conserved between ISP, K7, and MP16 but the region upstream is only well-conserved between K7 and MP16 ......64

4.4 Native Cas10-Csm expression is unable to effectively protect against infection by MP16 and K7, but can protect against ISP...............................................................66

4.5 The evasion of Cas10-Csm immunity by K7 and MP16 is not eliminated by overexpression of cas genes..................................................................................70
CHAPTER 1
INTRODUCTION

1.1 *Staphylococcus epidermidis* Pathogenicity

*Staphylococcus epidermidis* is a Gram positive bacterium typically seen as a common human commensal organism, found ubiquitously on the skin (Otto, 2009). While most humans carry *S. epidermidis* on their skin without any negative effects, it is also recognized as a major opportunistic human pathogen. Coagulase-negative staphylococci (CNS), a general grouping which includes all staphylococci except the most pathogenic species, *S. aureus*, are the most frequent infectious agent causing bloodstream infections related to indwelling medical devices, causing approximately one third of nosocomial bloodstream infections, with mortality in 21% of those cases (Edmond, 1999). *S. epidermidis* is the main infectious member of the CNS, causing up to 79% of both nosocomial and non-hospital associated bloodstream infections by CNS (Pfaller et al., 1999). CNS, including *S. epidermidis*, are also a prominent pathogen in infants, accounting for up to 40% of neonatal bacteremia cases (Anday & Talbot, 1985). Other infections caused by CNS range from colonization of catheters to native or prosthetic heart valve-associated endocarditis, the latter of which has 24% mortality (Chu et al., 2008, 2009). *S. epidermidis* is an opportunistic human pathogen of growing concern causing numerous diseases with considerable mortality in humans.

Although *S. epidermidis* typically causes less severe infections than its more virulent sister species, *S. aureus*, it nonetheless possesses its own suite of virulence factors allowing it to
establish and maintain infection. One major factor that greatly increases the number of infections by \textit{S. epidermidis} is its ubiquitous colonization of humans (Otto, 2009). \textit{S. epidermidis} has been seen to be the most common CNS in diverse areas of the human body ranging from the arm pits to the groin (Cavanagh et al., 2016). This widespread colonization allows \textit{S. epidermidis} access into the blood, where many of its infections occur, if the skin is penetrated in any location. Additionally, its prevalent carriage by humans, allows for potential infection of any individual as well as the spread of different strains between hosts from skin-to-skin contact (Han, Yang, & Park, 2016; Massey, Horsburgh, Lina, Höök, & Recker, 2006). As with any bacterial pathogen in the era of modern medicine, one of the major virulence factors for \textit{S. epidermidis} is antibiotic resistance, allowing persistent colonization in the face of treatment. Resistance to penicillins is common within \textit{Staphylococcus} species, including \textit{S. epidermidis}, with methicillin resistance found in 70-89\% of nosocomial infections (Diekema et al., 2001; Division of Healthcare Quality Promotion, National Center for Infectious Diseases, 2004). This widespread resistance to methicillin within staphylococci is predominantly due to the presence of a genomic island known as the SCC\textit{mec} cassette (J. Liu et al., 2016). These cassettes come in a variety of types that differ based on exact gene content and organization, but are unified in that they carry the \textit{meca} gene, conferring resistance to methicillin and other related antibiotics, as well as site-specific recombinases that allow the integration of this mobile genetic element (MGE). This same cassette, in addition to causing methicillin-resistant \textit{S. epidermidis}, is also responsible for the appearance of the ‘superbug’ methicillin-resistant \textit{S. aureus} (MRSA) (Teodoro, Mattos, Cavalcante, Pereira, & dos Santos, 2012). The best-characterized virulence factor possessed by \textit{S. epidermidis} is its ability to form biofilms, allowing for its infection of indwelling medical devices, which are the most common sites of \textit{S. epidermidis} infections (Arciola, Campoccia,
There are a variety of macromolecules from which *S. epidermidis* forms biofilms, including polysaccharides, proteins, and extracellular DNA (Büttner, Mack, & Rohde, 2015). Biofilms allow the population of cells to enter into a more hardy state, increasing their ability to evade antibiotics, bacteriophages, and the human immune system (Otto, 2008). Many clinical isolates of *S. epidermidis* have also been shown to produce members of a toxin family, the phenol-soluble modulins, which can be found on certain subtypes of the aforementioned SCCmec cassette and thus spread easily between *Staphylococcus* species (Peschel & Otto, 2013; Queck et al., 2009). Additionally, there are very rare examples of *S. epidermidis* isolates which possess toxins, including the toxic shock syndrome toxin and enterotoxins, that are more typically associated with *S. aureus*, allowing those specific isolates to become much more virulent than their toxin-free relatives (Crass & Bergdoll, 1986; Madhusoodanan et al., 2011; Podkowik et al., 2016). Primarily due to its ubiquity, ability to form biofilms, antibiotic resistance, and readily-spread toxins and other virulence factors from other, more pathogenic, staphylococci, *S. epidermidis* has developed significant pathogenicity, leading to numerous infections annually.

1.2 Horizontal Gene Transfer and *S. epidermidis* Virulence

The one thread unifying the virulence factors possessed by *S. epidermidis* is horizontal gene transfer (HGT). HGT is a mechanism by which bacteria and archaea share mobile genetic elements (MGEs), such as plasmids or transposons, with other organisms outside of their descendants. There are diverse ways in which prokaryotes can perform HGT, but the three main mechanisms are conjugation, bacteriophage transduction, and transformation (García-Aljaró et al., 2017). Conjugation is a means of direct transfer of MGEs, particularly conjugative plasmids.
that often carry antibiotic-resistance genes, between a donor and recipient cell via a multiprotein secretion complex, the type IV secretion system (Goessweiner-Mohr, Arends, Keller, & Grohmann, 2017). Transduction is a mechanism of HGT carried out by bacteriophages (Penadé S, Chen, Quiles-Puchalt, Carpena, & Novick, 2015). During the packaging of phage DNA into the capsid, mispackaging events can occur wherein bacterial DNA, such as plasmids, are mistakenly packaged into the capsid. These mispackaged phages are still capable of being released from the initial cell and injecting their DNA into a new host cell, introducing the mispackaged bacterial DNA into the new host, spreading that MGE to a new organism.

Transformation is the final canonical mechanism of HGT wherein some bacterial species are capable of directly acquiring DNA from the extracellular environment (I. Chen & Dubnau, 2004). These three mechanisms of HGT have together shaped the evolution of bacteria, including virulence, across the kingdom, including within the staphylococci.

Staphylococci are very proficient at conjugation both within and between species (Lindsay, 2014; Marraffini & Sontheimer, 2008). Conjugative staphylococcal plasmids carrying resistance against antibiotics such as mupirocin, gentamycin, and kanamycin are well-characterized, with at least a dozen multiresistance plasmids from staphylococci having been sequenced (M. A. Liu, Kwong, Jensen, Brzoska, & Firth, 2013). In addition to plasmids, other mobile genetic elements such as genomic pathogenicity islands have long been observed to be shared between *S. aureus* and *S. epidermidis*, including the aforementioned SCCmec cassettes which carry methicillin resistance and staphylococcal toxins (Méric et al., 2015). In addition to antibiotic resistance and toxins, other key *S. epidermidis* virulence factors are found in MGEs. The ica operon, responsible for polysaccharide-based biofilm formation, is notably regulated by and strongly associated with IS256, an insertion sequence commonly found in clinical *S.*
epidermidis isolates (Kozitskaya et al., 2004). This is a clear example of horizontal transfer between different staphylococcal species increasing the virulence of S. epidermidis. Bap, a common protein involved in proteinaceous biofilm formation, has also been proposed to be carried within a transposon and has likely been recently shared between diverse staphylococcal species including S. epidermidis and S. aureus (Tormo, Knecht, Götz, Lasa, & Penadés, 2005). It seems that HGT of virulence factors including biofilm formation and antibiotic resistance has greatly contributed to the current pathogenicity of S. epidermidis, through conjugation or phage transduction.

S. aureus bacteriophages have been shown to carry many staphylococcal virulence factors, including toxins such as staphylokinase, various staphylococcal enterotoxins, and exfoliative toxin A, which are involved in diverse S. aureus infections (Deghorain & Van Melderen, 2012; Malachowa & Deleo, 2010). S. aureus phages are known to facilitate transfer of additional virulence determinants such as pathogenicity islands and antibiotic resistance plasmids, and it is conceivable that S. epidermidis is capable of similar transduction between different S. epidermidis strains. Additionally, S. epidermidis and other CNS may act as reservoirs for pathogenicity genes to be spread to other, already pathogenic species, increasing their pathogenicity (Otto, 2013). Strains of CNS may linger on human or animal skin, carrying virulence determinants such as antibiotic resistance, but not being cleared from the body as they, on their own, do not pose a threat (Han et al., 2016). However, if they pass their genes into more pathogenic species such as S. aureus and further increase their virulence, this could result in the spread of methicillin resistance and other virulence factors into S. aureus, as has been proposed to occur with bap and antibiotic resistance, above. This is particularly likely considering that many staphylococcal phages, such as ISP or K, are polyvalent, attacking various S. aureus and
CNS species and may facilitate transduction between these species (O’Flaherty et al., 2005; Vandersteegen et al., 2011). Horizontal gene transfer has the potential to transform this human commensal into a devastating pathogen and may also allow *S. epidermidis* to increase the pathogenicity of other strains which it encounters.

### 1.3 Bacteriophage Therapy

Considering the possibly lethal effects of this organism and its growing antibiotic resistance, significant research has gone into the potential of bacteriophage therapy as a means of combating infections of *S. epidermidis*, as with many other pathogens. Bacteriophage therapy refers to the use of bacteriophages as antimicrobial agents to treat bacterial infections, and the concept has existed for over a century, since the very early days of bacteriophage research (Wittebole, De Roock, & Opal, 2014). Bacteriophages infect and kill very specific species of bacteria without disrupting any others, preventing the widespread devastation of the commensal bacteria in the body that is seen with treatments such as antibiotics which can, in turn, lead to the development of further disease. Because of their efficacy and specificity, implementation of and research into bacteriophage therapy, has been popular in some areas of the world, particularly some of the former soviet republics such as Poland and Georgia, having been abandoned in much of the world after the discovery of antibiotics (Anna Cisek, Iwona Da, Karolina Paulina Gregorczyk, & Zbigniew Wy, 2016). There have been several papers published about bacteriophage therapy as a means of combating infections with *S. epidermidis* and related organisms, particularly as it applies to biofilm dispersal (Gutiérrez et al., 2015; Takemura-Uchiyama et al., 2013). Additional work has been done characterizing the administration of staphylococcal bacteriophages *in vivo* in the context of an established infection, showing that
bacteriophage therapy has the potential to ameliorate staphylococcal disease in a mouse model (Matsuzaki et al., 2003). Staphylococcal bacteriophages have also been proposed to control the incidence of *S. aureus* infections from food, predominantly in dairy products (Bueno, García, Martínez, & Rodríguez, 2012; García, Martínez, Obeso, & Rodríguez, 2008). In addition to using whole staphylococcal phages as antimicrobial agents, individual phage-derived lysins are capable of killing *S. aureus* in vitro and in vivo and *S. epidermidis* in vitro and may represent another avenue for phage therapy, clinically or in food production (Cater et al., 2017; Fischetti, 2010; Rashel et al., 2007). While there have been advancements in recent years in developing staphylococcal bacteriophages capable of being used clinically, there are still significant hurdles to overcome before it can be used clinically, both due to legal/safety considerations as well as significant molecular hurdles that could prevent efficient treatment using bacteriophage.

1.4 Anti-Phage Defense Mechanisms

One potential barrier to the widespread implementation of bacteriophages as therapeutics is the prevalence of bacterial defense systems. These systems prevent the uptake of mobile genetic elements (such as plasmids, transposons, and bacteriophages), including restriction modification and abortive infection systems that are found within the staphylococci and many other prokaryotes (Belkebir & Azeddoug, 2012; Depardieu et al., 2016). Restriction modification is the prototypical example of a mechanism by which prokaryotes are able to escape from bacteriophage infection (Vasu & Nagaraja, 2013). Restriction modification systems are composed of at least two enzymes: one that acts as a restriction enzyme and another that acts as a methyltransferase. The restriction enzyme recognizes short DNA sequences and endonucleolytically cleaves the recognized DNA if the DNA is unmethylated (Ershova, Rusinov,
Spirin, Karyagina, & Alexeevski, 2015). The methyltransferase acts to protect the bacterial genome from recognition and degradation by methylating the DNA. Phage or plasmid DNA which enters the cell will likely not be methylated and will thus be recognized and destroyed by the restriction-modification system, prevent infection or HGT (Ershova et al., 2015). Abortive infection systems are a second class of anti-phage defense systems that act by inducing death of the bacterial host cell in addition to inhibiting replication of the infecting phage (Samson, Magadán, Sabri, & Moineau, 2013). This ‘altruistic’ sacrifice of one infected cell prevents the multiplication and spread of the phage progeny to other cells in the population, preventing widespread cell death. Restriction modification and abortive infection systems are both able to prevent infection by bacteriophages, but do not have a high level of adaptability, as they only inhibit a subset of phages which happen to be recognized.

On the other hand, the CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR associated proteins) system has recently been recognized as a powerful, robust, and adaptable immune system that develops an immune memory of past infections (van der Oost, Westra, Jackson, & Wiedenheft, 2014a). CRISPR-Cas is found in 40% of sequenced bacteria and 90% of archaea and is the only known adaptive prokaryotic immune system (Makarova et al., 2015). There are six distinct types of CRISPR-Cas systems (I-VI), classified on the basis of their protein content (Koonin, Makarova, & Zhang, 2017). These systems are linked by their ability to specifically recognize and degrade invasive genetic elements that bear significant nucleotide similarity to nucleic acid invaders that the bacterium’s ancestors have previously encountered. *S. epidermidis* RP62a is a strain isolated from a patient with an intravascular catheter-associated case of sepsis which, along with many other Staphylococcal clinical isolates, carries a Type III-A CRISPR-Cas system (Cao et al., 2016; Christensen,
Simpson, Bisno, & Beachey, 1982; Li et al., 2016; Marraffini & Sontheimer, 2008). This system recognizes specific RNA sequences, leading to degradation of complementary RNA and DNA by activating the Cas nucleases (Cas10, Csm3, and Csm6) (Jiang, Samai, & Marraffini, 2016; Samai et al., 2015).

1.5 CRISPR-Cas in *S. epidermidis* RP62a

CRISPR interference in *S. epidermidis* RP62a and other Type III-A CRISPR-Cas systems occurs in three steps (Figure 1.1). The first step is adaptation, in which the bacteria develop a memory of a previously-seen mobile genetic element (Figure 1.1A). This occurs when a mobile genetic element is degraded by various processes, such as restriction modification, within a cell and short (~35 bp) fragments (spacers) of the invasive nucleic acid are integrated into the CRISPR locus within the bacterial genome (Sternberg, Richter, Charpentier, & Qimron, 2016). Adaptation has not been re-created experimentally in *S. epidermidis* RP62a or any other Type III-A systems, but must have occurred at some point in the past, as there are three spacers encoding immunity against nucleic acid invaders in the CRISPR-Cas locus. In the second step of the CRISPR pathway, crRNA biogenesis, spacers are transcribed and processed (Figure 1.1B). Each transcribed spacer (pre-crRNA) is nucleolytically processed in two steps. The first step is called primary processing, in which Cas6 cleaves the pre-crRNAs into intermediate crRNAs of 71-nt each that are then cleaved again in a process called maturation, resulting in the production of mature crRNAs of predominantly 31, 37, and 43-nt length (Hatoum-Aslan, Maniv, & Marraffini, 2011; Hatoum-Aslan, Samai, Maniv, Jiang, & Marraffini, 2013). These mature crRNAs are incorporated into a protein complex, known as the Cas10-Csm complex, composed of five of the nine Cas (CRISPR-Associated) proteins, Cas10, Csm2, Csm3, Csm4, and Csm5.
This ribonucleoprotein complex forms the active surveillance complex within the cell. In the final step of the CRISPR pathway, interference, this surveillance complex degrades specific nucleic acid invaders (Figure 1.1C). Surveillance complexes within the cell recognize RNA sequences bearing high complementarity to the crRNAs through cognate base pairing, activating cleavage of the RNA and nearby DNA sequences, destroying the invader (Jiang et al., 2016; Samai et al., 2015). CRISPR-Cas systems are unique among anti-bacteriophage defense systems in that they have a very high specificity for nucleic acid sequence, requiring significant homology within the ~35 bp spacer region for this cleavage to occur, with mutations in as few as 3 nucleotides capable of ablating interference (Maniv, Jiang, Bikard, & Marraffini, 2016). It is also noteworthy for its efficiency in interfering with both bacteriophages and plasmids, decreasing efficiency of conjugation by at least 100-fold and often completely preventing successful plaque formation by bacteriophages (Jiang et al., 2016; Maniv et al., 2016; Walker, Chou-Zheng, Dunkle, & Hatoum-Aslan, 2017).

Figure 1.1
The CRISPR-Cas system detects and interferes with invasive genetic elements in three steps. In adaptation (A), an invasive MGE is recognized and degraded by cellular nucleases, and short fragments of the nucleic acid is integrated into the genome by Cas1/2, forming spacers (colored diamonds). In crRNA biogenesis (B), these spacers are transcribed into pre-crRNAs which are
nucleolytically processed into 71-nt long intermediate crRNAs, which are further processed via maturation into mature crRNAs. Mature crRNAs combine with the Cas proteins to form the Cas10-Csm complex. In interference (C), this complex recognizes and degrades nucleic acid invaders with sequences complementary to one or more of the spacers, preventing infection.

CRISPR-Cas is one of the most powerful characterized prokaryotic immune systems due to its versatility and efficiency. Thus, in many genera, including *Staphylococcus*, it stands as one of the key barriers to rampant dissemination of antibiotic resistance and other virulence factors via horizontal gene transfer. Similarly, it is also a significant barrier to the efficacy of bacteriophage therapy, as it allows prokaryotes to develop immunity to treatment and pass this immunity on to its progeny indefinitely (Nilsson, 2014). The *S. epidermidis* RP62a system naturally possesses spacers that target conjugative plasmids bearing antibiotic resistance as well as transducing phages, allowing exploration of multiple routes of horizontal gene transfer in this important human pathogen (Maniv et al., 2016; Marraffini & Sontheimer, 2008). Therefore, this system provides a well-characterized model system for exploring how the staphylococci control horizontal gene transfer and allows for exploration of how this system may prevent the spread of virulence in a pathogen. This information from *S. epidermidis* may also be applicable to its more-pathogenic sister species, *S. aureus*, many clinical strains of which have been sequenced and shown to possess similar Type III-A CRISPR-Cas systems (Cao et al., 2016). Understanding this system will allow us to understand the ways in which bacterial populations and species spread virulence factors, becoming pathogenic, and allow for the development potential methods of controlling the spread of virulence factors such as antibiotic resistance in the future as well as improving methods of clinical or industrial applications of bacteriophages.
2.1 Introduction

For our knowledge of CRISPR-Cas to potentially be used to improve human health, it is necessary to understand how it functions on the molecular level. This may allow development of therapies that could interfere with or aid CRISPR-Cas functionality to modulate the spread of mobile genetic elements or increase the efficiency of bacteriophage therapy. Within *S. epidermidis*, a significant amount is known about the third step of CRISPR interference while less is known about the crucial step of crRNA biogenesis. Although the first step of crRNA processing, primary processing, is well characterized on the molecular level and appears to be conserved between many Type I and III systems, maturation, the last cleavage event that produces the fully mature crRNAs, is less well understood (Carte, Wang, Li, Terns, & Terns, 2008; Reimann et al., 2017). It is known that one of the Cas proteins, Csm3, is responsible for determining the length of the mature crRNAs, but does not perform the cleavage itself, and that several Cas proteins, Csm2, Csm3, and Csm5, are essential for maturation (Hatoum-Aslan et al., 2011, 2013). Csm3 polymerizes along the length of the intermediate crRNA, protecting the RNA from degradation by an unknown nuclease, with different numbers of Csm3 proteins polymerized along the crRNA leading to the different sizes of mature crRNAs observed. The nuclease that is responsible for maturation is not known in our system, though it has been hypothesized to be a non-Cas nuclease as is seen in Type II CRISPR-Cas systems
Exploring this relatively uncharacterized step of CRISPR interference will allow us to develop a more well-rounded understanding of CRISPR functionality and expand our knowledge of bacterial immunity as well as potentially lead to therapeutic opportunities in the future.

In this work, we attempted to identify the maturation nuclease of the *S. epidermidis* RP62a Type III-A CRISPR-Cas system. We used alanine-scanning mutations to explore the role of predicted catalytic residues within four *cas* genes and identify amino acids that are essential for proper maturation and/or CRISPR-Cas immunity. We further investigated the molecular properties of two Cas proteins, Csm3 and Csm5, in more detail, discerning potential roles for their residues in complex stability, nucleic acid binding, and protein-protein interactions. We also identify non-Cas nucleases which may be involved in the functionality of our system in maturation or in other steps of CRISPR interference.

2.2 Methodology

2.2.1 Bacterial Growth Conditions

*S. aureus* RN4220 (Kreiswirth et al., 1983) was grown in tryptic soy broth (TSB) medium. *S. epidermidis* strains RP62a and LM1680 were grown in brain-heart infusion (BHI) medium. *E. coli* strains DH5α and BL21 (DE3) codon plus cells (EMD Millipore) were grown in Luria Bertani (LB) broth. All strains were grown at 37 °C unless stated otherwise. Media were supplemented with antibiotics as needed: 15 μg/mL neomycin (for selection of *S. epidermidis*), 10 μg/mL chloramphenicol (for selection of pC194 and pCRISPR-Cas and its derivatives), 5 μg/mL mupirocin (for selection of pG0400), 50 μg/mL kanamycin (for selection of pET28b-
derived plasmids), and/or 30 μg/mL chloramphenicol (for selection of *E. coli* BL21 (DE3) codon plus cells). Media was additionally supplemented with 5 mM CaCl$_2$ when bacteriophages were used.

2.2.2 Plasmid Construction and Cloning

pCRISPR-Cas derivatives were constructed from pCRISPR-Cas (Hatoum-Aslan et al., 2013) using either inverse PCR or Gibson assembly (Gibson et al., 2009). Constructs were first electroporated into *S. aureus* RN4220 and intended mutations were confirmed using PCR and sequencing. If necessary, one confirmed isolate of each construct was purified from RN4220 using the EZNA Plasmid Mini Kit (Omega) and electroporated into *S. epidermidis* LM1680 (Hatoum-Aslan et al., 2013).

pET28b-His10Smt3 derivatives were constructed from amplified pET28b plasmid and PCR products from RP62a genomic DNA and assembled using Gibson assembly. Constructs were transformed into *E. coli* DH5α cells and plasmids were purified from at least two different isolates using the EZNA Mini Prep Kit and were confirmed with sequencing. At least one confirmed plasmid for each construct was then transformed into *E. coli* BL21 (DE3) codon plus cells for protein purification.

2.2.3 CRISPR Functionality Assays

Conjugation assays were carried out using a filter-mating method. Single colonies of *S. aureus* RN4220 or *S. epidermidis* LM1680 carrying pCRISPR-Cas or its derivatives (recipients) were inoculated into 2 mL cultures and grown until OD$_{600}$ of a 1:10 dilution was $\geq 0.2$, as was *S. aureus* RN4220 carrying pG0400 (donors). Donors and recipients were combined in a 4:1 ratio
and allowed to filter mate on a 0.45 μM nitrocellulose filter overnight. Growth on the filter was resuspended in media and serially diluted to 10⁻⁷ before being plated onto plates selecting for both pCRISPR-Cas and pG0400 (transconjugants: chloramphenicol and mupirocin, as well as neomycin when carried out in *S. epidermidis*) or just pCRISPR-Cas (recipients: chloramphenicol as well as neomycin when carried out in *S. epidermidis*) to allow for enumeration. Conjugation efficiency was calculated as the ratio of transconjugants to recipients.

Phage infection assays were carried out by combining, in a total volume of 200 μL, a 1:100 dilution of overnight culture of *S. epidermidis* LM1680 carrying pCRISPR-Cas or its derivatives and an MOI (multiplicity of infection, ratio of phage particles to bacterial cells) of 10 of phage CNPX (Maniv et al., 2016) in BHI broth. Cultures were grown for 15 hours in a 96-well plate in a SpectraMax M2e, with OD600 measured every 15 minutes. Data shown are the average of three technical replicates, ± one standard deviation.

2.2.4 RNA Extraction and Biotin Pulldowns

RNA was extracted from *S. aureus* RN4220 or *S. epidermidis* LM1680 cells carrying pCRISPR-Cas or its derivatives using TRIzol. 10 mL of media was inoculated with a 1:20 dilution of overnight cultures of the indicated strains and grown to mid-log (OD600~1). The cells were centrifuged (5000 rpm, 5 min), washed twice in 1 mL TSM (Tris-Sucrose-Magnesium) buffer (50 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 500 mM sucrose), digested in 500 μL TSM supplemented with 20 μg/mL lysostaphin for 20 minutes at 37 °C, and pelleted. The pellet was thoroughly resuspended in 750 μL TRIzol and centrifuged (4°C, 15 minutes, 12,000 x g). The aqueous layer was taken into a fresh tube and RNA was precipitated by the addition of 500 μL of 100% isopropanol and incubation at room temperature for 10 minutes. The RNA was centrifuged
(4°C, 10 minutes, 12,000 x g) and washed in 75% ethanol. RNA was precipitated in a final centrifugation (4°C, 5 minutes, 12,000 x g) and air-dried before being resuspended in nuclease-free water.

crRNAs were purified from the whole RNA extract using a biotinylated DNA probe with a sequence complementary to spacer 1 as described previously (Hatoum-Aslan et al., 2011). 50 μg of RNA was incubated with 1 μL of 20 μM probe and the RNA was denatured at 70 °C for 1 minute before annealing at room temperature for 45 minutes. 15 μL of magnetic streptavidin-coated beads (Sera-Mag™ Magnetic Streptavidin Coated particles, GE Healthcare) were added to the tube and allowed to anneal for 1 hour at room temperature. The beads and bound RNA were washed 3 times with 15 μL of annealing buffer (5 mM Tris-HCl (pH 8.3), 75 mM KCl, 1 mM EDTA (pH 8.0)) to remove unbound RNAs. The beads were resuspended in 30 μL of annealing buffer and heated for 2 minutes at 98 °C to dissociate the streptavidin and biotin. The supernatant was 5’ end-labeled with γ[32P]ATP using T4 PNK (NEB). Labeled RNAs were run on a denaturing polyacrylamide gel and visualized on a storage phosphor screen using a Typhoon FLA 7000 (GE).

2.2.5 Double-Filter Protein-Nucleic Acid Binding Assays

Double-filter assays were carried out to measure protein binding affinities for nucleic acids (Tanaka & Schwer, 2005; Wong & Lohman, 1993). 0.2 μm nitrocellulose membranes were soaked in 0.5 M NaOH for 10 minutes before being rinsed three times in dH2O. 0.45 μm nylon membranes were soaked in 0.1 M EDTA (pH 8.0) for 10 minutes, soaked three times for ten minutes each in 1.0 M NaCl, rinsed briefly in 0.5 M NaOH, and rinsed three times in dH2O. Membranes were soaked at 4 °C for at least 1 hour in binding buffer (40 mM Tris-HCl (pH 7.5),
2 mM DTT, 1 mM EDTA (pH 8.0), 10% glycerol). The membranes were assembled in a 96-well dot-blot apparatus with the nylon membrane beneath the nitrocellulose membrane. Trace amounts of 5' radiolabeled RNA probe was combined with the indicated amounts of Csm5 and incubated at room temperature for 10 minutes. The wells of the dot-blot apparatus were washed in 200 μL of cold binding buffer and the sample was applied. After pulling the sample through, the wells were washed twice in 100 μL of cold binding buffer. Filters were air-dried, exposed to a storage phosphor screen, and imaged using a Typhoon FLA 7000 phosphorimager. The fraction of bound RNA was calculated with the formula (intensity of nitrocellulose signal) / (intensity of nitrocellulose signal + intensity of nylon signal) using ImageQuant (GE).

2.2.6 Statistical Analysis

All statistical analyses (t-tests and standard deviations) were calculated using Microsoft Excel. P-values were calculated using a 2-tailed Student’s t-test.

2.3 Results and Discussion

2.3.1 Identification of Putative Active Site Residues

The identification of the nuclease(s) responsible for CRISPR RNA maturation was begun by examining the cas genes themselves. Primary processing, the first step of crRNA nucleolytic processing, is known in this system to be carried out by one of the Cas proteins, Cas6 (Carte et al., 2008; Hatoum-Aslan et al., 2011), and several of the Cas proteins (Csm3, Csm4, and Csm5) are predicted to be nucleases with as-of-yet unknown function (Makarova et al., 2011), so it is possible that one or more of the Cas proteins carries out maturation. Additionally, three of the
Cas proteins (Csm2, Csm3, and Csm5) are known to be essential for maturation, so one or more of them may be directly involved in the process as the nuclease (Hatoum-Aslan et al., 2011).

For these reasons, multiple sequence alignments of each of these three proteins were performed with ten homologues from diverse prokaryotic species, two of which (Csm3 and Csm5) are predicted to be nucleases (Figure 2.1). From this alignment, many residues that are highly conserved across the diverse species were identified. Acidic residues that may carry out metal-dependent nuclease activity as well as basic residues that may be involved in protein-protein or protein-nucleic acid interactions were of particular interest, so the focus was on highly conserved lysine (K), arginine (R), glutamate (E), and aspartate (D) residues. From our multiple sequence alignment, seventeen highly conserved, charged amino acids that had not previously been investigated were identified as candidates for mutagenesis to assess their role in crRNA maturation and CRISPR-Cas functionality. All seventeen of these mutations were separately introduced into a plasmid, pCRISPR-Cas, which carries all nine Cas genes with the N-terminus of csm2 tagged with a 6-His tag to allow for Cas10-Csm complex purification, as well as the entire CRISPR repeat-spacer array and the leader from S. epidermidis RP62a. These plasmids were introduced to S. aureus RN4220, a heterologous host in which the S. epidermidis CRISPR-Cas system is known to function (Samai et al., 2015), for initial screening of defects in both CRISPR-Cas functionality and crRNA maturation.
A) Csm2

1 10 20 30
--- --- --- ---
DYAKTKSGK -- TIDETFHEVYKSNKNVKDR

40 50 60 70 80
KGKEKQYLENGLTASKLLEILMEQVNRLYTLAIFNSNEQDQNEEFCIDEFL

90 100 110 120 130
KIKFGEAL-REKSVDEFKTLMPITDEVIKESKFFPFDYCKYFEAL

VAAYKAKYEE-

B) Csm3

1 10 20 30
--- --- --- ---
MYSLVSGTLEVVTHGKSEMGESSMGILARSEPVVRDL

40 50 60 70 80
KTLKLEESSIAGKKNKLTVKAHFGLKKQPS -- HNQDDERVMLFGSS

90 100 110 120
EKSS --- IQRAALQSTAFFSEKTFEH -- MONDIAYTTKFPENTIK

130 140 150
LTAVMFKIERVTRGRLDVEFPIVYVV --- ESEQVDETENIKAIHL

170 180 200 210
KEVHVEEGCGTEGRCPEKDNIETVGEY -- DSTMILKE

C) Csm5

1 10 20 30 40
--- --- --- --- ---
-TIKNEYEVKIK- GIISSGQVKQDLYDEPNSKTMVINCNKLV

50 60 70 80 90
KFKE -- KLLLTYQONFLAYET -- KNPENGLKDYIADQVQKSEWEAVF

100 110 120 130 140
SSKMYNQGKYGNTREKFENDLHLMVQGQNKLPEGSSKSMKLKKT

SKVN -- EKNKD- ---

150 160 170 180
YSKKEKSEKPIESNLAIQKIDINSA -- KSMPLY
Identification of conserved residues in the CRISPR-associated proteins Csm2 (A), Csm3 (B), and Csm5 (C). Each alignment was performed using PRALINE multiple sequence alignment (J Heringa, 1999; Jaap Heringa, 2002; Simossis & Heringa, 2005) using 10 homologs of the indicated S. epidermidis RP62a Cas proteins, identified via BLASTp of each protein against the NCBI non-redundant protein sequence database. Residues are colored by PRALINE based on their conservation as indicated below each alignment, with 0 being not conserved and 10 being highly conserved. Amino acids are labeled above each row based on the position of that amino acid within the indicated protein of S. epidermidis and each 10th amino acid is also underlined. Below each row, (*) indicates an amino acid that was mutated to alanine in this study and (^) indicates residues mutated in previous studies.

2.3.2 Anti-Plasmid Immunity of pCRISPR-Cas Mutants

All seventeen of the mutant pCRISPR-Cas plasmids were screened for defects in CRISPR-based anti-plasmid immunity using conjugation assays. The first spacer of the RP62a CRISPR system, spc1, bears 100% nucleotide identity to the nickase gene found on staphylococcal conjugative plasmids, such as pG0400, allowing the native Cas10-Csm complex to interfere with their conjugation (Marraffini & Sontheimer, 2008). To assess the level of anti-plasmid immunity bestowed by pCRISPR-Cas, S. aureus RN4220 donor cells carrying pG0400 are combined with recipient cells carrying pCRISPR-Cas, one of its derivatives, or a negative control that just carries pC194 (Ehrlich, 1977), the chloramphenicol resistance plasmid on which pCRISPR-Cas is built, which lacks any CRISPR components. The donor is allowed to exchange DNA with the recipient overnight via filter mating and the resulting growth is plated on selective media, allowing for enumeration of the total number of cells carrying pCRISPR-Cas or its
derivative (recipients) as well as the number of cells which successfully took up pG0400 (transconjugants), which carries mupirocin resistance as a selectable marker. The ratio of transconjugants to recipients is calculated, giving the efficiency of conjugation. When a functional CRISPR-Cas system is present, conjugation will not occur due to degradation of pG0400, resulting in a low conjugation efficiency. If CRISPR-Cas is either not present or non-functional, conjugation will be allowed to occur, resulting in a higher conjugation efficiency.

When all the mutant pCRISPR-Cas derivatives were screened in this manner, several were observed to have either partial or total defects in anti-plasmid immunity (Figure 2.2, Table 2.1). One of the four Csm2 mutations, three of five Csm3 mutations, and four of eight Csm5 mutations appeared to have lowered anti-plasmid immunity (relative to the wild-type) in this heterologous host (indicated with * in Table 2.1). All other mutations were seen to have no noticeable defects in anti-plasmid immunity.
Figure 2.2
Conserved charged residues in Csm2 (A), Csm3 (B), and Csm5 (C) are essential in a heterologous background for CRISPR-Cas anti-plasmid immunity. Conjugation efficiency assays were performed using *S. aureus* RN4220 strains carrying pCRISPR-Cas (WT), pC194 (the negative control, NC), or mutant derivatives of pCRISPR-Cas, with the mutations identified along the x-axis. Grey bars represent the counts of recipients, white bars the counts of transconjugants. Results shown are the average of three technical replicates, error bars represent one standard deviation.
<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Conjugation Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCRISPR-Cas (WT)</td>
<td>6.75 ($\pm$ 9.34) x 10^{-7}</td>
</tr>
<tr>
<td>pC194 (NC)</td>
<td>3.55 ($\pm$ 4.10) x 10^{-4}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm2 K47A, R49A) *</td>
<td>4.21 ($\pm$ 1.15) x 10^{-5}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm2 E92A)</td>
<td>4.17 ($\pm$ 1.44) x 10^{-7}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm2 E128A)</td>
<td>2.22 ($\pm$ 3.85) x 10^{-7}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm2 D141A)</td>
<td>6.13 ($\pm$ 3.69) x 10^{-7}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm3 K4A)</td>
<td>8.96 ($\pm$ 9.57) x 10^{-7}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm3 K52A, K54A, R56A) *</td>
<td>1.96 ($\pm$ 0.99) x 10^{-4}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm3 E140A)</td>
<td>2.18 ($\pm$ 0.83) x 10^{-7}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm3 D179A) *</td>
<td>1.20 ($\pm$ 1.39) x 10^{-4}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm5 D28A)</td>
<td>1.70 ($\pm$ 1.20) x 10^{-6}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm5 D32A) *</td>
<td>1.92 ($\pm$ 2.67) x 10^{-4}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm5 D115A)</td>
<td>1.19 ($\pm$ 1.64) x 10^{-6}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm5 D122A) *</td>
<td>1.05 ($\pm$ 0.47) x 10^{-4}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm5 D153A)</td>
<td>2.25 ($\pm$ 0.90) x 10^{-7}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm5 D162A) *</td>
<td>1.16 ($\pm$ 0.32) x 10^{-4}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm5 D178A)</td>
<td>1.24 ($\pm$ 2.02) x 10^{-6}</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm5 E191A) *</td>
<td>8.26 ($\pm$ 2.38) x 10^{-6}</td>
</tr>
</tbody>
</table>

Table 2.1
Conserved charged residues in Csm2, Csm3, and Csm5 are required for optimal anti-plasmid immunity in a heterologous host. Shown are the calculation efficiencies (the ratio of transconjugants to recipients) from the data shown in Figure 2.2, ± 1 standard deviation. Mutants which appeared to have a decrease in efficiency of anti-plasmid CRISPR immunity which were further studied in *S. epidermidis* LM1680 are indicated with *.

Seven of the eight mutants which were shown to be defective in anti-plasmid interference in *S. aureus* RN4220 were transformed into *S. epidermidis* LM1680, a derivative of RP62a with a significant genome deletion that removes the entire CRISPR-Cas locus and a large number of other genes encompassing ~11% of the RP62a genome (Hatoum-Aslan et al., 2013). Csm3 K52A, K54A, R56A was not transformed into *S. epidermidis* LM1680 as it was found that this triple mutation prevented the complex from assembling, which is likely why maturation and anti-plasmid immunity were defective (Walker et al., 2017). *S. epidermidis* LM1680, due to its lack of CRISPR-Cas, has no anti-plasmid immunity on its own, providing a host with a more similar background to *S. epidermidis* RP62a than that which is found in *S. aureus* RN4220 but which is still CRISPR-Cas negative. This allows one to answer the question if any defects seen in anti-plasmid immunity are also seen in a native host and are thus truly caused by the mutation and not
an interplay between the mutation and the non-native host background. The additional screen in
*S. epidermidis* LM1680 was not carried out on all created mutants due to the difficulty in
transforming large plasmids into *S. epidermidis*, which is not readily competent (Monk, Shah,
Xu, Tan, & Foster, 2012). These seven mutants were screened via the same conjugation assay in
the *S. epidermidis* LM1680 background, and of the eight, only one mutation (Csm5 E191A,
indicated with *) was shown to significantly decrease the efficacy of anti-plasmid immunity in *S.
epidermidis* LM1680 as compared to the wild-type pCRISPR-Cas (Figure 2.3, Table 2.2). Csm3
K4A was also screened in the same manner because, while it does not have abrogated anti-
plasmid immunity in *S. aureus* RN4220 (Figure 2.2), it does have a non-wild type crRNA profile
(Figure 2.4), but it was similarly found to have no significant defect in anti-plasmid immunity.

![Figure 2.3](image)

A conserved, charged glutamate in Csm5 is required for efficient anti-plasmid immunity in a
native host. Conjugations efficiency assays were performed as in Figure 2.2 but in the
background of *S. epidermidis* LM1680. * indicates p-value (t-test) < 0.05.
A conserved glutamate in Csm5 is required for efficient anti-plasmid immunity in a native host. Shown are the calculation efficiencies (the ratio of transconjugants to recipients) from the data shown in Figure 2.3, ± 1 standard deviation. The mutant which has a significant (p-value < 0.05) decrease in efficiency of anti-plasmid CRISPR immunity is indicated with *. P-values were calculated using a t-test, relative to pCRISPR-Cas (WT).

Interestingly, although eight distinct mutations across the three Cas proteins investigated in this mutagenesis analysis abrogated anti-plasmid immunity without disrupting Cas10-Csm complex formation in a heterologous host, only one of those mutants displayed a similar phenotype in a native background. This could indicate that non-Cas components, which may differ between *S. aureus* and *S. epidermidis*, could be essential for proper CRISPR-Cas immunity. Several non-Cas proteins have been shown to be essential for CRISPR immunity in other CRISPR systems, such as RNaseIII, which performs crRNA processing in Type II CRISPR-Cas systems, and PNPase, an exonuclease required for interference in a unique CRISPR system in *Listeria* (Deltcheva et al., 2011; Sesto et al., 2014). It is possible that one or more non-Cas proteins that are required for CRISPR-Cas functionality in the presence of these mutants are absent or sufficiently divergent in *S. aureus*, preventing successful interference in the non-native host. Alternatively, the intracellular environment, in terms of cofactors, chaperones, or other molecules, may differ enough between the two species that mutants that can be tolerated in one species result in dysfunction in another species with a distinct environment. Csm5 E191A was the only mutation in any of the *cas* genes which disrupted anti-plasmid immunity without

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Conjugation Efficiency</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCRISPR-Cas (WT)</td>
<td>1.34 (± 0.53) × 10^{-6}</td>
<td></td>
</tr>
<tr>
<td>pC194 (NC)</td>
<td>5.94 (± 3.59) × 10^{-4}</td>
<td>0.104</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm2 K47A, R49A)</td>
<td>2.00 (± 1.76) × 10^{-7}</td>
<td>0.053</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm3 K4A)</td>
<td>8.94 (± 9.57) × 10^{-6}</td>
<td>0.303</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm3 R141A)</td>
<td>1.23 (± 1.17) × 10^{-6}</td>
<td>0.897</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm3 D179A)</td>
<td>8.24 (± 6.67) × 10^{-7}</td>
<td>0.357</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm5 D32A)</td>
<td>7.00 (± 5.20) × 10^{-7}</td>
<td>0.209</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm5 D122A)</td>
<td>3.39 (± 2.53) × 10^{-7}</td>
<td>0.063</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm5 D162A)</td>
<td>8.94 (± 5.97) × 10^{-7}</td>
<td>0.389</td>
</tr>
<tr>
<td>pCRISPR-Cas (Csm5 E191A) *</td>
<td>1.40 (± 0.46) × 10^{-5}</td>
<td>0.039</td>
</tr>
</tbody>
</table>

*Table 2.2*
disrupting complex formation in both backgrounds (Walker et al., 2017). This may indicate that this residue is involved in an important step in CRISPR-Cas functionality, potentially maturation and/or interference itself. Interestingly, the functional defect in seen in Csm5 E191A does not appear to be a complete lack of interference, as conjugation efficiency for this mutant (1.4 x 10^{-5}) seems to be intermediate between the efficiency for wild-type pCRISPR-Cas and the negative control (1.3 x 10^{-6} and 5.9 x 10^{-4}). This may indicate that this amino acid is advantageous, but not absolutely required, for CRISPR-Cas anti-plasmid immunity. The hypothetical roles that this residue may play in CRISPR interference, as well as further investigations of phenotypes associated with these mutations, are discussed in detail in upcoming sections.

2.3.3 crRNA Size Distribution of pCRISPR-Cas Mutants

In addition to functional screens of the pCRISPR-Cas mutants, the ability of each mutant to generate proper, mature crRNAs was assayed using a biotin pulldown of spc1 crRNAs from strains carrying each pCRISPR-Cas mutant. Typically, the most prominent mature crRNA observed is 43-nt long, with additional, less-prominent species at 37 and 31-nt in length (Hatoum-Aslan et al., 2013). Using a complementary biotinylated oligonucleotide, spc1 crRNAs were captured from whole RNA extracts of *S. aureus* RN4220 bearing pCRISPR-Cas or its derivatives, the crRNAs were labeled, and then run via denaturing PAGE for visualization. Of the seventeen mutants that were profiled in *S. aureus* RN4200, six had crRNA profiles that were significantly different from the wild-type pCRISPR-Cas (indicated with * in Figure 2.4). These six mutants were additionally transformed into *S. epidermidis* LM1680 for profiling of the crRNAs in a native host. These six mutants maintained their distorted crRNA size distributions
in an *S. epidermidis* background, with their size distributions falling into one of three general categories:

1) crRNA maturation was totally abrogated, with little/no mature crRNAs being visible, similar to a knockout of their respective genes (Csm3 K52A, K54A, R56A (in *S. aureus*); Csm5 E191A)

2) crRNA sizes were shifted downward, with either the 37 and 31-nt species becoming much more prominent (Csm3 K4A; Csm3 D179A)

3) Mature crRNAs were selectively processed to primarily the 37-nt species, with little 43 or 31-nt crRNAs (Csm3 R141A; Csm5 D162A)

---

*Figure 2.4*
Mutations in conserved, charged residues of Csm3 and Csm5 result in altered crRNA profiles in a heterologous and native host. Pulldowns of crRNAs using a biotinylated oligo complementary to *spc1* crRNA were performed on RNA extracts of *S. aureus* RN4220 or *S. epidermidis* LM1680 (*) strains bearing pCRISPR-Cas or its indicated mutant. Labeled are the intermediate crRNA, 71-nt long, and the three main mature crRNAs, 43, 37, and 31-nt in length.
A total of nine pCRISPR-Cas mutants, including those transformed due to altered crRNA size distribution or impaired interference in *S. aureus*, were transformed into and further tested in *S. epidermidis* LM1680. While there may be a correlation between abrogated anti-plasmid immunity and defective crRNA maturation in a heterologous host (since six of the ten mutants shown to have a defect in anti-plasmid immunity in *S. aureus* also had defects in crRNA maturation), there does not appear to be a strong correlation in the native background as, although the disrupted crRNA size distribution remained the same, anti-plasmid immunity was seen to be entirely functional in all but one mutant. This one mutant, Csm5 E191A, was the only one in its crRNA profile group, those with complete or almost complete abrogation of crRNA maturation whose entire functionality could be fully assessed. The Csm3 K52A, K54A, R56A mutation, which gave a similar crRNA profile, was shown to just prevent the Cas10-Csm complex from forming and was not pursued further (Walker et al., 2017). As the only *cas* gene mutant which eliminated proper crRNA maturation without interfering with Cas10-Csm complex formation also resulted in defective anti-plasmid immunity, this may indicate that maturation is essential for maximum CRISPR-Cas interference, but further mutants which cause a similar defect in maturation would be necessary to confirm this hypothesis. Although there is some evidence that specific disruptions of crRNA maturation may eliminate CRISPR anti-plasmid immunity, the relationship is not clear at this point. Further work will need to be done to identify the nuclease(s) responsible for this activity and eliminate maturation entirely before firm conclusions can be drawn.
2.3.4 Anti-Bacteriophage Immunity of pCRISPR-Cas Mutants

Although our primary functional screen based on its reproducibility and its easily quantifiable nature is the conjugation assay measuring anti-plasmid immunity, it is not the only way to measure CRISPR-Cas activity. In general, CRISPR-Cas interferes with the uptake of two distinct nucleic acid invaders: 1) mobile genetic elements such as plasmids or transposons and 2) bacteriophages (van der Oost, Westra, Jackson, & Wiedenheft, 2014b). The details of interference of each distinct class of invader by the *S. epidermidis* Cas10-Csm complex is known to differ substantially. One example of these differences is that the protein Csm6 is required for anti-plasmid immunity while it is dispensable in some cases of anti-bacteriophage immunity (Hatoum-Aslan, Maniv, Samai, & Marraffini, 2014; Jiang et al., 2016). To test whether our mutants which disrupt crRNA sizes may disrupt only one or both facets of CRISPR immunity, any mutants that were shown to have disrupted crRNA profiles and/or abrogated anti-plasmid immunity in *S. aureus* RN4220 were additionally screened for functional defects in anti-bacteriophage interference. The second spacer of the *S. epidermidis* RP62a CRISPR-Cas locus, spc2, bears 100% complementarity to phage CNPX and has been shown to be able to completely prevent plaque formation by this phage when it is present along with the *cas* genes (Maniv et al., 2016). This spacer is present on pCRISPR-Cas and its derivatives, allowing anti-bacteriophage immunity to be assayed with this phage. This was accomplished in liquid culture with OD600 measured every fifteen minutes in an overnight co-incubation of phage CNPX with strains of *S. epidermidis* LM1680 bearing pCRISPR-Cas or its derivatives.

In this assay, the results were somewhat different from those seen for anti-plasmid immunity (Figure 2.5). All but one of the tested mutants behaved exactly as in anti-plasmid immunity, with anti-bacteriophage immunity entirely functional. Interestingly, Csm5 E191A,
which had significantly decreased anti-plasmid immunity, had no decrease in anti-bacteriophage immunity, as it was able to grow exactly as wild-type pCRISPR-Cas in the presence of phage CNPX. It appears that, somewhat consistent with previous results concerning the necessity of Csm6 and Csm3 in anti-plasmid immunity vs. anti-phage immunity, that anti-phage immunity is much more robust and resilient in the face of cas gene mutations. The only mutation that disrupted anti-phage immunity was Csm3 K52A, K54A, R56A, which makes sense since the Cas10-Csm complex does not form with this mutation and no interference can occur without complex formation (Hatoum-Aslan et al., 2013; Walker et al., 2017). This highlights the complexity of Cas10-Csm interference as another example of a discrepancy between anti-plasmid and anti-bacteriophage immunity and may indicate that wild-type crRNA sizes are less crucial in anti-phage immunity than anti-plasmid immunity, although a potential mechanism for this difference is unknown.

**Figure 2.5**
Complex formation is essential, but wild-type crRNA sizes are not required, for Cas10-Csm anti-bacteriophage immunity. *S. epidermidis* LM1680 bearing pCRISPR-Cas or an indicated mutant
was infected with bacteriophage CNPX at an MOI of 10 and growth was measured by taking OD 600 nm measurements at 15 min intervals for 15 hours.

2.3.5 Biochemical Characterization of Csm5

Based on data presented herein and in previous works, we hypothesized that Csm5 was the enzyme responsible for crRNA maturation for the following reasons: 1) Csm5 is essential for crRNA maturation (Hatoum-Aslan et al., 2014); 2) deletion of csm5 eliminated only Csm5 itself from the complex (Hatoum-Aslan et al., 2014); 3) Csm5 is predicted to have nuclease activity (Makarova et al., 2011); 4) Csm5, based on structures from other Type III-A and III-B CRISPR-Cas complexes, is located at the 3’ end of the crRNA, which is where the maturation cleavage event is hypothesized to occur (Rouillon et al., 2013; Spilman et al., 2013); and 5) a mutation in Csm5, E191A, is the only one identified in this work which abrogates crRNA maturation without also eliminating Cas10-Csm complex formation (Figure 2.4, Walker et al., 2017).

While Csm3 is similar to Csm5 in many of these facets, there are additional important caveats in regards to hypothesizing that Csm3 is the maturation nuclease. While Csm3 is essential for maturation, its elimination completely prevents any of the complex from forming (Hatoum-Aslan et al., 2014), so it may just be that the complex is destabilized in its absence as opposed to Csm3 playing any direct role in maturation. Additionally, Csm3 is thought to compose the backbone of the complex, polymerized along the length of the crRNA, and is thus located in a less advantageous position for access to the 3’ end of the crRNA than Csm5 (Rouillon et al., 2013; Spilman et al., 2013). Finally, mutations in Csm3 resulted in changes in crRNA sizes, not total abrogation of maturation as with Csm5 E191A. The exception to this is Csm3 K52A, K54A, R56A, but this, again, just destabilizes the complex and its role in maturation cannot be differentiated from this disruption of complex formation. The other
residues identified here as involved in proper crRNA size distribution, K4 and D179 within Csm3, are also proposed to play a structural role in promoting crRNA maturation. Csm3 is known to be responsible for the lengths of crRNAs, as it polymerizes along the length of the intermediate crRNA and the number of Csm3 subunits on the crRNA dictate the length of crRNA seen in that specific complex (Hatoum-Aslan et al., 2013). Csm3 K4 and D179 are predicted to form an inter-protein interaction between adjacent Csm3 subunits, with their mutation eliminating that interaction, preventing Csm3 from polymerizing as efficiently, resulting in shorter crRNAs. This is supported by in silico evidence, which shows, based on structures of homologous proteins, that the two residues are predicted to form a salt bridge together (Walker et al., 2017). In addition, in vitro evidence shows these mutant Csm3 proteins, when purified from E. coli, are no longer able to polymerize (Walker et al., 2017). Similarly, Csm2 was not investigated further for many of these reasons, including its less advantageous predicted structural positioning within the complex and lack of mutants that abrogate proper crRNA maturation (Figure 2.4, Rouillon et al., 2013; Spilman et al., 2013).

For these reasons, the role that Csm5 may play in maturation was further investigated. We hypothesized that Csm5 is either the nuclease that directly performs the cleavage responsible for maturation or, alternatively, that it recruits a distinct nuclease to the 3’ end of the crRNA to facilitate maturation. To investigate these possibilities, Csm5 was cloned into E. coli and purified for biochemical characterization.

The first assays performed were simply combining, in vitro, Csm5 with radiolabeled RNA substrates similar to crRNAs in the presence of various metals. If Csm5 is an RNase, the RNAs should be degraded and the size of the RNAs, when visualized with denaturing PAGE, should be visibly smaller. However, in the conditions tested, no degradation of RNAs was seen
in the presence of Csm5 (Walker et al., 2017). This is consistent with previous work which had been unable to detect degradation of crRNAs by the purified *S. epidermidis* Cas10-Csm complex (Hatoum-Aslan et al., 2013), but inconsistent with Csm5 being the maturation nuclease.

To begin to investigate the second possibility, that Csm5 acts to recruit another cellular nuclease to the crRNA and that second nuclease is responsible for maturation, the ability of Csm5 to bind to crRNAs was characterized. If Csm5 is recruiting another protein to the crRNA, it may be binding the crRNA and acting essentially as a bridge between the RNA and the nuclease. This binding of Csm5 to RNAs was probed using a double-filter assay. Csm5 was incubated with radio-labeled RNAs and then passed through two filters. The top filter was nitrocellulose and captured the protein and any RNAs bound to the protein. The bottom filter was nylon and captured any additional RNAs that were not bound to protein and passed through the top filter. Comparing the amount of RNA bound to each filter allows one to determine what percentage of RNA was bound to Csm5 at differing protein concentrations, allowing for calculation of binding affinity. Our results showed that Csm5 binds to RNAs with an affinity in the nanomolar range (Kd= 100.6 nM) in the presence of MgCl₂, with weaker binding when EDTA was added (Kd= 201.9 nM) (Figure 2.6). This indicates that Csm5 may, in the context of the Cas10-Csm complex, bind directly to the crRNA and facilitate recruitment of an additional nuclease to perform maturation.
2.3.6 Roles of Non-Cas Nucleases in Cas10-Csm Maturation

Mass spectrometry analysis was performed on the Cas10-Csm complex purified from *S. epidermidis* LM1680 to identify any cellular nucleases that co-purify with the complex. We were particularly interested in 3’-5’ exonucleases, as we hypothesized that the maturation nuclease acts exonucleotically in that direction. From these data, we identified three nucleases that fit this description: polyribonucleotide phosphorylase (PNPase), RNaseR, and Cbf1 (Walker et al.,
Further *in vitro* assays in which these enzymes were purified from *E. coli* and combined with Csm5 and RNAs demonstrated that Csm5 selectively inhibits or activates nucleolytic degradation of the RNAs by these proteins. RNaseR and Cbf1 were both entirely inhibited from degrading the RNA, likely due to Csm5 tightly sitting on the RNA, consistent with our above findings that Csm5 binds RNA with high affinity. PNPase, however, had its nucleolytic activity activated in the presence of Csm5, indicating that Csm5 may functionally interact with PNPase, stimulating its degradation of RNAs. Additionally, further *in vitro* evidence from 2-dimensional gel electrophoresis and affinity pull-down assays demonstrated a physical interaction between Csm5 and PNPase, but not Csm3 and PNPase (Walker et al., 2017). It is conceivable that the one mutation which disrupted crRNA maturation and anti-plasmid immunity in *S. epidermidis* RP62a, Csm5 E191A, may have abrogated the interaction between Csm5 and PNPase. This could explain why pCRISPR-Cas with Csm5 E191A has a defect, if PNPase is essential for CRISPR-Cas functionality and/or crRNA maturation. However, Csm5 E191A was not able to be purified from *E. coli* and there is no biochemical evidence for this connection.

2.4 Future Directions

Although it appears that Csm5 interacts physically and functionally with PNPase, it is unknown exactly how this may impact Cas10-Csm functionality. The Cas10-Csm complex degrades both RNA and DNA targets (Jiang et al., 2016; Samai et al., 2015), so any degradation by PNPase may not be involved in maturation, but rather target elimination, and further work will need to be done to elucidate any role(s) PNPase may perform in relation to CRISPR. This question is further complicated because PNPase is a nonspecific exonuclease, able to degrade RNA and DNA equally, as well as a nonspecific polymerase, randomly adding nucleotides to the
3’ end of polynucleotide substrates (Briani, Carzaniga, & Dehò, 2016). PNPase could be involved in degradation of RNA targets, degradation of DNA targets, maturation of crRNAs, or a combination of the three. Further work, including knocking out PNPase and characterizing the resulting mutant and its phenotypes in crRNA maturation, anti-plasmid immunity, and anti-phage immunity, will need to be performed to dissect the exact role that this protein plays in Cas10-Csm immunity.
CHAPTER 3
REGULATION OF CRISPR-CAS IN *S. EPIDERMIDIS*

3.1 Introduction

Regulation of CRISPR-Cas systems is one of the most important areas of research for potential therapeutic applications of CRISPR-Cas knowledge. One of the largest threats to modern medicine is the growing antibiotic resistance crisis, and CRISPR stands as a potential barrier to the spread of antibiotic resistance through its ability to prevent HGT. Understanding what signals activate or repress CRISPR expression will allow understanding of when and where prokaryotes are sharing these mobile genetic elements with each other and allow us to develop strategies to combat this spread. Additionally, knowledge about regulating CRISPR expression and functionality, particularly in terms of repression, may allow the development of more efficient means of bacteriophage therapy for clearing infections by allowing the inactivation of a major defense system that could prevent effective treatment using bacteriophages.

Thus far, nothing is known about transcriptional or translational control of the Cas10-Csm system in *S. epidermidis* RP62a, but regulation of CRISPR systems in other prokaryotes has been much better characterized (Patterson, Yevstigneyeva, & Fineran, 2017). Perhaps the best-characterized regulatory network for any CRISPR system is that of the Type I-E system in *E. coli*. In this system, a complex network of proteins including H-NS (histone-like nucleoid structuring protein), LeuO, and CRP (cAMP receptor protein), respond to cellular metabolite levels, specifically concentrations of glucose and amino acids, and compete for overlapping...
binding sites surrounding the cas operon, with the cas genes expressed only in high glucose concentrations (Pul et al., 2010; Westra et al., 2010; Yang, Chen, Huang, Huang, & Tseng, 2014). A very similar regulatory network has also been shown to control expression of CRISPR in Salmonella enterica, with LRP (Leucine Responsive regulatory Protein) also playing a key role in basal inactivation of CRISPR-Cas (Hernández-Lucas et al., 2008; Medina-Aparicio et al., 2011). In contrast to what is seen in E. coli, in Pectobacterium atrosepticum and Thermus thermophilus, high levels of glucose actually repress their Type I-F CRISPR-Cas systems through the homologues of CRP in these organisms (Patterson, Chang, Taylor, & Fineran, 2015; Shinkai et al., 2007). Metabolite concentrations are well-characterized regulators of CRISPR-Cas expression across species and CRISPR types and likely play roles in regulating CRISPR immunity in S. epidermidis.

Another wide-spread, well-characterized activator of CRISPR-Cas expression is cell envelope stress, particularly phage infection. In order to mount a robust response to bacteriophage infection, the most immediately-threatening incursion to which CRISPR immunity responds, it makes sense that the cell would increase expression of immune genes when infection is detected, and this has been observed in several different organisms. Although the exact mechanism is not known, phage infection is known to induce CRISPR-Cas expression in Streptococcus thermophilus, Thermus thermophilus, Sulfolobus islandicus, and Sulfolobus solfataricus (Agari et al., 2010; Fusco et al., 2015; León-Sobrino, Kot, & Garrett, 2016; Quax et al., 2013; Young et al., 2012). Intriguingly, in S. islandicus there are two distinct CRISPR-Cas systems, and phage infection activates one locus while repressing the other in a unique contradictory example of phage-responsive regulation of CRISPR-Cas (León-Sobrino et al., 2016). Additionally, cell-envelope stress-responsive two-component regulators VicK/R and
BaeS/R have been characterized to be involved in regulating CRISPR-Cas expression in *Streptococcus mutans* and *E. coli*, respectively, providing a direct link between cell-envelope stress and CRISPR-Cas regulation (Perez-Rodrigues et al., 2011; Serbanescu et al., 2015).

Various metabolic and phage-related stresses are known to regulate CRISPR-Cas systems in other organisms, yet nothing is still known about regulation of CRISPR in *S. epidermidis*. We sought to start filling in this gap in our knowledge about this system to provide an understanding of how this organism responds to bacteriophage infection and how it regulates horizontal gene transfer. This included three main goals: 1) characterizing the crRNA promoter(s) to understand how the system is transcribed, 2) identifying small-molecule stressors that regulate crRNA expression, and 3) identifying putative transcription factors which may play a role in regulating CRISPR immunity. Additional findings led to exploration of bacteriophage-encoded genes that may play non-transcriptional, non-translational roles in regulating CRISPR-Cas immunity.

3.2 Methodology

3.2.1 Bacterial Growth Conditions

Bacterial strains were grown in the same media and antibiotics as described in 2.2.1, except where noted.

3.2.2 Plasmid Construction and Cloning

Derivatives of pCRISPR-GFP and pSpacer were constructed using inverse PCR and passaged through *S. aureus* RN4220 to *S. epidermidis* RP62a as described in 2.2.2. pET28b derivatives for protein purification in *E. coli* were also constructed as described in 2.2.2.

39
Derivatives of pGG-BsaI and pCasSA were constructed using Golden Gate assembly as described previously (W. Chen, Zhang, Yeo, Bae, & Ji, 2017). Recovery sequences for phage genome editing were introduced into derivatives of pCasSA using Gibson assembly.

3.2.3 GFP Reporter Expression Assays

A SpectraMax i3x (Molecular Devices) was used to measure the GFP fluorescence of S. epidermidis RP62a carrying pCRISPR-GFP or its derivatives. Overnight cultures of RP62a carrying pCRISPR-GFP or its derivatives were diluted 1:100 into BHI with appropriate antibiotics and any additional compounds as indicated in the text. Triplicate wells of a 96-well plate (Grenier) were inoculated with 200 μL of this mixture. Plates were incubated with shaking at 37 °C overnight with automated OD600 and GFP fluorescence measurements taken every 15 minutes using the SpectraSoft 7 software (Molecular Devices). Data were analyzed using Microsoft Excel to determine the fluorescence relative to OD600 at the transition from exponential to stationary phase growth. For the purposes of normalizing the growth stage of all treatments for accurate comparison, a novel formula was developed to determine the time point at which growth transitioned from the exponential to stationary phase. OD600, with the BHI blank subtracted, had its first derivative calculated relative to time (in minutes). Exponential phase was defined as the period starting at the first time point at which the first derivative of OD600 relative to time in minutes was greater than or equal to $1 \times 10^{-3}$ and ending at the first time point at which the first derivative was less than $1 \times 10^{-3}$. The first measurement at which the first derivative was less than $1.00 \times 10^{-3}$ was deemed the transition point from exponential to stationary phase and was used for calculations. The relative fluorescence of each well at the transition point was calculated by dividing GFP fluorescence by OD600. OD600 of a well
containing sterile BHI was subtracted from the OD600 of each well at the transition point. GFP fluorescence of each well was divided by the blanked OD600, and the three technical replicates of each plasmid or condition were averaged. These values were then blanked by subtracting the relative fluorescence of the negative control, pC194, the plasmid on which pCRISPR-GFP is based, from the relative fluorescence at the transition point of each condition. For assays determining the effects of environmental conditions on crRNA expression, the relative fluorescence values of pCRISPR-GFP at each condition were blanked against the relative fluorescence of pC194 at that same condition.

3.2.4 RNA Extraction and Primer Extension

RNA extraction was performed as described in 2.2.3 with minor modifications. S. epidermidis RP62a cultures were inoculated, 1:100 from an overnight culture, in BHI and supplemented with the indicated concentrations of either sodium chloride or neomycin and zinc chloride, and grown at 37°C until OD600~1. Extractions were further performed as described in 2.2.3.

Primer extension analyses were performed to quantify the relative expression levels of crRNAs. 5 μg of whole RNA extracts were combined with trace amounts of radio-labeled primers complementary to 1) the spc1 crRNA and 2) the 5S rRNA in annealing buffer. RNAs were melted for 1 minute at 98 °C and then allowed to anneal at 48 °C for 45 minutes, then diluted to 30 μL with reverse transcriptase synthesis buffer (50 mM Tris-HCl (pH 8.3), 75 mM KCl, 4.5 mM MgCl₂, 15 mM DTT, 0.6 mM dNTPs, 5 U/μL M-MLV Reverse Transcriptase (NEB)). Extension was then performed at 37 °C for 30 minutes before being precipitated with 1/10 volume sodium acetate (pH 5.2), 3 volumes 100% ethanol, and 0.5 μL 100 mg/mL tRNA at
-80 °C. Pellets were air-dried and resuspended in 3.5 μL 40 μg/mL RNaseA for 3 minutes and dissolved in formamide loading buffer for visualization via denaturing PAGE as described in 2.2.3.

3.2.5 Statistical Analysis

Statistical analyses were performed as in 2.2.6.

3.3 Results and Discussion

3.3.1 Identification and Characterization of crRNA Promoters

To begin developing any hypotheses about CRISPR regulation by cis or trans regulatory factors, one needs to know how CRISPR is normally transcribed, starting with identification of any promoters driving CRISPR expression. This work was begun by using a GFP reporter for crRNA expression, called pCRISPR-GFP, as a model of CRISPR-Cas expression because crRNAs are essential for CRISPR functionality and may function as a proxy to reveal overall regulation of the entire CRISPR-Cas system. Additionally, most characterized regulators have been described in their roles of regulating cas expression, with fewer works investigating crRNA expression, so this will further the understanding of a less-studied aspect of CRISPR regulation (Patterson et al., 2017). The reporter consists of the leader region of the S. epidermidis RP62a CRISPR locus, 800-bp upstream of the first direct repeat of the locus, followed by a Shine-Dalgarno sequence (from the csm3 gene, which has a Shine-Dalgarno sequence identical to the consensus, AGGAGG, introduced to allow this RNA-encoding gene to express GFP), and a staphylococcal codon-optimized GFP gene. This plasmid was constructed, transformed into S.
epidermidis RP62a, and used as the base plasmid for genetic manipulation to identify the crRNA promoter(s) present in *S. epidermidis*.

After confirming that pCRISPR-GFP can indeed fluoresce much higher than the negative control, a mutational analysis was performed to begin searching for the crRNA promoter(s). Starting just upstream of the Shine-Dalgarno sequence (with a 2-bp buffer between the Shine-Dalgarno and the first deletion), a total of eleven regions of 10-40 bp in length were deleted from pCRISPR-GFP, with the resulting mutant plasmids transformed into *S. epidermidis* RP62a and assayed for their fluorescence. Fluorescence was normalized against OD600 at the end of exponential phase and graphed (Figure 3.1). The deletions of nt 191-170 and 42-24 each cause a large increase in fluorescence, up to ~40% for the deletion of 42-24. This could indicate that one or both regions are important for maintaining wild-type levels of crRNA expression and may contain binding sites for one or more transcriptional repressors. However, this initial probe was attempting to identify the promoter(s) of crRNA expression, so the potential regulators in these regions have not yet been investigated further, but are of interest for future explorations. One would expect that a mutation in a promoter would result in a significant decrease in basal fluorescence, and that was the object of this initial probe. From these data, two specific regions (marked A and B in Figure 3.1) stand out as having decreased fluorescence. While region A has a greater decrease (~80%) than region B (~20%), both have consistent decreases in fluorescence over a large (60 or 68 bp, respectively) area, which one would expect from a promoter, which covers at least 40-bp of DNA. These areas became the subjects of more intense investigation.
Figure 3.1
Two large regions of the crRNA leader are required for wild-type crRNA expression levels in a GFP reporter assay. 10-40 bp regions of the crRNA leader on pCRISPR-GFP were deleted and fluorescence was measured at the end of the logarithmic phase and normalized against OD600. The indicated deleted nucleotides are measured relative to the GFP Shine-Dalgarno sequence, with the first position of the Shine-Dalgarno equaling +1.

After identifying regions A and B, both regions were searched for potential promoter sequences. The phiSITE PromoterHunter prokaryotic promoter prediction tool was used to identify putative promoters within each region (Klucar, Stano, & Hajduk, 2009). From this search, three putative promoters were identified within Region A and just one was identified within Region B (Figure 3.2, Table 3.1). To further narrow down which promoters were most likely to be active, the conservation of these sequences within staphylococcal Type III-A crRNA promoters was determined by first identifying Type III-A CRISPR-Cas systems in diverse staphylococcal species. BLASTp was used to identify homologs of *S. epidermidis* RP62a, the characteristic gene for Type III-A systems, in all published, fully aligned staphylococcal genomes in the NCBI database. Of the 12 genomes identified, 11 contained crRNA leader regions with enough similarity to be aligned to the *S. epidermidis* RP62a crRNA leader using
Clustal Omega (McWilliam et al., 2013). Using that multiple sequence alignment, a sequence logo of Region A was generated using the WebLogo server to identify which predicted promoters were well conserved (Crooks, Hon, Chandonia, & Brenner, 2004). None of the three Region A promoters were perfectly conserved (Figure 3.2B). The proposed -35 region for promoter A.1 was not very well conserved, although the first two nucleotides (TA) were 100% conserved and are the most well-conserved nucleotides in the -35 region. Additionally, the -10 region for promoter A.1 was 100% conserved, supporting its potential role in crRNA expression. Promoter A.2 had a 100% conserved -35 region, but its -10 sequence was not very well conserved across the staphylococcal species. Promoter A.3 had a fairly well conserved -10 region, but its -35 was not very well conserved. The single Region B promoter (B.1) was very well conserved, 100% across the -35 region and almost 100% across the -10 element, with only the A in position 5 not 100% conserved. From this, 3 putative promoters were identified and we chose to further investigate promoter A.1 of Region A (primarily due to its higher score from phiSITE (see Table 3.1)) as well as the lone promoter within Region B.

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Score (-35)</th>
<th>Score (-10)</th>
<th>Spacing (nt)</th>
<th>Energy Score</th>
<th>Final Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1</td>
<td>1.97</td>
<td>3.50</td>
<td>18</td>
<td>4.24</td>
<td>3.25</td>
</tr>
<tr>
<td>A.2</td>
<td>0.31</td>
<td>2.01</td>
<td>17</td>
<td>4.24</td>
<td>3.01</td>
</tr>
<tr>
<td>A.3</td>
<td>-0.21</td>
<td>1.24</td>
<td>15</td>
<td>4.24</td>
<td>2.90</td>
</tr>
<tr>
<td>B.1</td>
<td>4.36</td>
<td>2.88</td>
<td>17</td>
<td>4.73</td>
<td>3.21</td>
</tr>
</tbody>
</table>

*Table 3.1*

Four promoters are predicted to be present within Regions A and B of the crRNA leader. Promoters were identified using the default settings of phiSITE and the scores given by phiSITE for each promoter region, as well as the energy score and final score of the entire promoter, are shown. Spacing indicates the number of nucleotides between the predicted -35 and -10 regions of each promoter.
Figure 3.2
Multiple putative, well-conserved promoters are present within the crRNA leader region. (A) Three promoters were identified by phiSITE within Region A and one within Region B, with the -35, -10, and +1 sites underlined. (B) At least one promoter present in each Region is well-
conserved among staphylococci, with the -35 and -10 of each promoter identified in A indicated with black bars.

Now that the target promoters were identified, a single mutation was introduced within each -10 region to examine the importance of these putative -10 regions in pCRISPR-GFP fluorescence. The A in the second position of each promoter was mutated to a T using inverse PCR on pCRISPR-GFP. As this location is very well conserved in -10 regions, one would expect that, if these are active -10 regions, that the mutation would reduce or completely knock out fluorescence of the reporter. That is what resulted, with the mutated A.1 promoter causing a fluorescence decrease of 80% compared to wild-type and the mutated B.1 promoter causing an approximately 20% drop in fluorescence (Figure 3.3), roughly equal to the decrease seen for the deletions within Region A and B, respectively (Figure 3.1). The two mutations were combined on a single plasmid to determine if the effects were additive, which would indicate that both promoters are active, with each contributing a portion of the fluorescence seen from wild-type pCRISPR-GFP. When the double mutant was tested, it effectively eliminated fluorescence from pCRISPR-GFP, with only ~2% of the wild-type fluorescence detected (Figure 3.3). These observations support the hypothesis that crRNAs are under the control of two promoters, A.1 and B.1, responsible for about 80% and 20% of fluorescence, respectively. Further biochemical work probing RNA polymerase binding via permanganate footprinting or identification of the transcription start site using primer extension are currently underway to confirm this hypothesis.
Figure 3.3
Two promoters are involved in wild-type expression of S. epidermidis RP62a crRNAs. Complement switch mutations (A->T) were made to the -11 position of putative promoters A.1 and B.1 (see Figure 2.2) separately and together in pCRISPR-GFP and fluorescence was measuring in late-logarithmic phase and normalized against OD600. * indicates p-value < 0.01; *** indicates p-value < 0.0001; **** indicates p-value < 0.00001.

3.3.2 Screening for Environmental Regulators of crRNA Expression

Knowing that many environmental factors including oxidative stress (Strand et al., 2010), glucose (Patterson et al., 2015; Yang et al., 2014), phage infection (Agari et al., 2010; Young et al., 2012), and DNA damaging agents (Klaiman, Steinfels-Kohn, & Kaufmann, 2014) are known to regulate CRISPR-Cas expression, we continued our exploration of CRISPR-Cas regulation using our pCRISPR-GFP reporter plasmid to identify environmental factors that impact crRNA expression. We screened through many factors, including antibiotics (penicillin, tetracycline, mupirocin, etc.), phage infection, and DNA-damaging agents (mitomycin C, hydrogen peroxide) without seeing any real change in fluorescence that could not just be explained by changes in growth (Table 3.2). However, through the screening, two environmental factors were identified that may act as repressors of crRNA expression.
<table>
<thead>
<tr>
<th>Condition</th>
<th>Results</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibiotic stress</td>
<td>No conclusive effects</td>
<td>Tested 12 antibiotics across diverse antibiotic classes: penicillin, tetracycline, mupirocin, gentamycin, kanamycin, spectinomycin, erythromycin, vancomycin, ampicillin, lysostaphin, mupirocin, and rifampicin</td>
</tr>
<tr>
<td>Bacteriophage Infection</td>
<td>No conclusive effects</td>
<td></td>
</tr>
<tr>
<td>Ethanol</td>
<td>No conclusive effects</td>
<td>Tested due to its ability to induce biofilm formation in <em>S. epidermidis</em> (Knobloch et al., 2001)</td>
</tr>
<tr>
<td>Increased carbohydrates</td>
<td>No conclusive effects</td>
<td>Tested because of previous reports of glucose regulating CRISPR-Cas in other systems (Patterson et al., 2015; Yang et al., 2014)</td>
</tr>
<tr>
<td>(glucose, sucrose)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metals</td>
<td>Zn^{2+} acts as a repressor; no other conclusive effects</td>
<td></td>
</tr>
<tr>
<td>Oxidative stress</td>
<td>No conclusive effects</td>
<td>Hydrogen peroxide was examined due to previous reports of oxidative stress regulating CRISPR-Cas (Strand et al., 2010)</td>
</tr>
<tr>
<td>Sodium chloride</td>
<td>Acts as a repressor</td>
<td></td>
</tr>
</tbody>
</table>

*Table 3.2*

List of environmental conditions examined as potential regulators of CRISPR-Cas and summarized results.

Through the two-component systems BaeS/R and VicK/R, cell envelope stress is known to regulate expression of CRISPR-Cas in *E. coli* and *S. mutans*, respectively (Perez-Rodrigues et al., 2011; Serbanescu et al., 2015), and we hypothesized that a similar effect may be seen in *S. epidermidis* RP62a under osmotic shock conditions which stress the cell envelope. Salt has also been shown to regulate CRISPR-Cas expression in *Thermoproteus tenax*, potentially through an envelope stress-related mechanism (Plagens, Tjaden, Hagemann, Randau, & Hensel, 2012). Sodium chloride was used as the stressor and expression of cells carrying pCRISPR-GFP grown in various concentrations of NaCl, up to 15% was examined. This was conducted without the
addition of neomycin and chloramphenicol, as the growth of cells was heavily impacted by the combined presence of the antibiotics and NaCl and we wanted to investigate higher concentrations of salt to probe the true range of crRNA regulation by sodium chloride. In this assay, like in assays with many antibiotics we tested, salt would kill the cells at high concentrations and there was similarly a decrease in fluorescence with increasing salt concentration (Figure 3.4A). However, unlike in previous assays, the degree to which fluorescence decreased was much greater than the decrease in growth, which is evident when fluorescence is normalized against OD600 (Figure 3.4B). An almost complete knock-down of crRNA expression was seen at 15% NaCl and around a 60-70% reduction in crRNAs at 10% NaCl.

To confirm these results, we used primer extension to measure RNA levels more directly. RNAs were extracted from mid-log *S. epidermidis* RP62a cells grown in varying concentrations of salt, again, without the presence of neomycin. crRNAs were using radiolabeled primers complementary to spc1. 5S rRNAs were also reverse transcribed and quantified to allow for normalization of crRNAs. From this, repression of the crRNAs was again seen with increasing concentrations of salt, confirming that salt repressed crRNA expression (Figure 3.4C, D). The degree of repression seen with pCRISPR-GFP was ~90% in 15% NaCl, while the repression was ~60% from the primer extension data, but the two methods of measuring crRNA abundance generally corroborate each other.
A) OD600 vs Time (min)

- 0%
- 2.50%
- 5%
- 7.50%
- 10%
- 15%

B) Relative Normalized GFP Fluorescence vs Concentration NaCl (%)

- 0%
- 2.5%
- 5%
- 7.5%
- 10%
- 15%

* and ** indicate statistical significance.
Figure 3.4
NaCl represses crRNA expression. A) Growth curve of *S. epidermidis* RP62a carrying pCRISPR-GFP in BHI supplemented with the indicated concentrations of NaCl. B) Bar graph of fluorescence of pCRISPR-GFP in the indicated concentrations of NaCl at the end of the logarithmic phase, normalized against the GFP fluorescence of a BHI blank and the OD600, relative to the normalized fluorescence of RP62a carrying pCRISPR-GFP in BHI without added NaCl. C) Representative gel showing the results of primer extension analyses on whole RNA extracts from *S. epidermidis* mid-log cells grown in the presence of the indicated amounts of NaCl, including the rRNA normalization control. D) Average of triplicate primer extension experiments, with crRNA abundance normalized against rRNA abundance at the indicated NaCl concentrations and then normalized against the crRNA:rRNA ratio in BHI with no added NaCl. Data are from three separate experiments, ± one standard deviation. * indicates p-value < 0.01; ** indicates p-value < 0.001.

In addition to the stressors mentioned above, the effects of various metals on fluorescence of pCRISPR-GFP were also tested. Iron limitation has been shown to regulate CRISPR in *Synechocystis* (Hein, Scholz, Voß, & Hess, 2013), so we hypothesized that iron or other metals may impact CRISPR expression. Additionally, at least one alternative sigma factor, σ^{54} in *Salmonella typhimurium*, is known to regulate CRISPR expression (Samuels et al., 2013), so we were also interested in exploring potential roles for alternate sigma factors in *S. epidermidis*. In *S. aureus*, manganese is known to activate the alternative sigma factor σ^{B} (Pané-Farré, Jonas, Förstner, Engelmann, & Hecker, 2006). Another known regulator of the *S. aureus* σ^{B} is salt.
Pané-Farré et al., 2006), which has been shown to regulate crRNA expression in *S. epidermidis* RP62a, furthering the interest in exploring potential roles of this alternative sigma factor in regulating CRISPR, directly or indirectly. For these reasons, the roles that metals may play in regulating CRISPR-Cas was investigated, through alternative sigma factors or otherwise. A panel of several metal salts (FeSO₄, iron (III) citrate, ZnCl₂, NiCl₂, MnCl₂, MgCl₂, CaCl₂, CoCl₂) was tested for any significant effects on pCRISPR-GFP fluorescence. Of the eight metals, only Zn²⁺ produced significant, reproducible regulation which, at a concentration of 750 μM, significantly repressed pCRISPR-GFP fluorescence up to 70-80% (Figure 3.5B). The repression seen with Zn²⁺ was nearly equal to that seen with NaCl, yet ZnCl₂ caused only minor growth defects (Figure 3.5A, B). This phenotype was confirmed using primer extension, and again, even at concentrations below 100 μM Zn²⁺, we saw a similar significant repression of crRNA expression (Figure 3.5C, D). In the 10-mL format in which RNA extractions are performed, *S. epidermidis* RP62a growth is affected by Zn²⁺ at a much lower concentration, with significant growth defects at 250 μM Zn²⁺ and no growth occurring with 500 μM Zn²⁺, and it is therefore not feasible to increase concentrations to the level used in the GFP reporter format. Even with this increased growth defect for RNA extractions, primer extension of crRNAs in BHI with up to 100 μM Zn²⁺ corroborate the repression seen in reporter assays with pCRISPR-GFP. Zn²⁺ appears to be a potent repressor of crRNA expression through an unknown mechanism.
Figure 3.5

Zn\(^{2+}\) represses the expression of crRNAs in a dose-dependent manner. A) A growth curve of *S. epidermidis* RP62a carrying pCRISPR-GFP in BHI media supplemented with the indicated concentrations of ZnCl\(_2\) (μM). B) Bar graph of fluorescence of pCRISPR-GFP in the indicated concentrations of ZnCl\(_2\) at the end of the logarithmic phase, normalized against the GFP fluorescence of a blank and the OD600, relative to the normalized fluorescence of RP62a carrying pCRISPR-GFP in BHI without added NaCl. C) Representative primer extension gel of crRNAs from *S. epidermidis* RP62a in the absence or presence of ZnCl\(_2\) (0, 10, 50, 100 μM) showing crRNAs and the rRNA control. D) Graphed results primer extension analyses, as in C, with crRNA abundance normalized against rRNA abundance at the indicated ZnCl\(_2\) concentrations and then normalized against the crRNA:rRNA ratio in BHI with no added ZnCl\(_2\) and then averaged. Data are from three separate experiments, ± one standard deviation. * indicates p-value < 0.01; ** indicates p-value < 0.001.

Interestingly, between promoters A.1 and B.1, there appears to be a binding site similar to that for PerR, a Fur-family (Ferric uptake regulator) regulator that responds to hydrogen peroxide stress (M. J. Horsburgh, Ingham, & Foster, 2001; Malcolm J. Horsburgh, Clements, Crossley, Ingham, & Foster, 2001). Even though a nearly-perfect PerR binding site is present, we were
unable to reproducibly show that peroxide regulates expression of the crRNAs. However, PerR is just one of several Fur-family regulators within *S. epidermidis* RP62a, and the Fur family proteins appear to share similar binding sequences (M. J. Horsburgh et al., 2001). One of the other Fur family proteins present in the genome of *S. epidermidis* RP62a is Zur, the zinc-uptake regulatory protein. Zur has been shown in *S. aureus* to repress its regulon in a zinc-dependent manner, similar to the repression we have seen here with Zn\(^{2+}\) (Lindsay & Foster, 2001). It is possible that Zur is the transcription factor responsible for regulating crRNA expression in response to Zn\(^{2+}\) based on the similar expression phenotype and the presence of a putative binding site between the two identified crRNA promoters. While Zur is typically presented as regulating genes responsible for zinc uptake (Gabriel, Miyagi, Gaballa, & Helmann, 2008; Lindsay & Foster, 2001; Mortensen, Rathi, Chazin, & Skaar, 2014), it also regulates a number of different genes of various functions, some unknown, in diverse organisms (Gilston et al., 2014; Pawlik et al., 2012), so it is possible that the CRISPR-Cas locus could be part of its regulon in *S. epidermidis*. Additionally, Zn\(^{2+}\) is a known structural cofactor for proper biofilm formation in *S. aureus* (Formosa-Dague, Speziale, Foster, Geoghegan, & Dufrêne, 2016). Biofilms are optimal conditions for horizontal gene transfer and are closely linked to the spread of MGEs in diverse species (Madsen, Burmølle, Hansen, & Sørensen, 2012). Conjugation in *S. aureus* has been shown to be significantly increased in biofilm as compared to planktonic cells, indicating that a large amount of horizontal transfer of factors such as antibiotic resistance likely occurs in the context of a biofilm (Savage, Chopra, & O’Neill, 2013). Therefore, it may be that organisms like *S. epidermidis* may repress CRISPR-Cas in high Zn\(^{2+}\) environments, as those conditions promote biofilm formation, where the exchange of beneficial genes is more likely than attack by bacteriophages. This could allow the organism to promote horizontal gene transfer under certain
circumstances where it is not likely to have a negative effect. A similar explanation could be
given for NaCl-related repression of CRISPR-Cas, as NaCl is known to transcriptionally induce
biofilm formation in *S. epidermidis* via a number of intermediary transcription factors (Lim,
Jana, Luong, & Lee, 2004; Rowe et al., 2016; Xu, Zou, Lee, & Ahn, 2010). Common regulators
of biofilm formation such as Zn$^{2+}$ and NaCl may act to repress CRISPR-Cas expression to
promote horizontal gene transfer in relatively ‘safe’ conditions, where CRISPR-Cas will likely
not be needed to defend against bacteriophages.

### 3.4 Future Directions

Thus far, some evidence has been gathered that is beginning to unravel how CRISPR-Cas
is transcriptionally regulated in *S. epidermidis* RP62a. Biochemical work such as permanganate
footprinting to confirm the identity of the identified crRNA promoters will be performed to
confirm the position of promoters A and B. Further biochemical and genetic work will be done
to investigate putative transcription factors that may be responsible for carrying out the
repression due to NaCl and ZnCl$_2$, including knocking out putative regulatory factors to
investigate the expression of crRNAs in a mutant background. Finally, further work will be
conducted to characterize the regulation of the *cas* genes. This includes construction of reporter
plasmids for one or more of the *cas* genes, as has been done for the crRNAs, as well as
development of direct assays of transcriptional activity, such as qPCR. Together, this work will
allow for an understanding of the regulatory network which controls this important physiological
process in *S. epidermidis* and allow us to decode the ‘thought process’ these organisms use to
determine whether or not CRISPR-Cas should be expressed.
CHAPTER 4
INHIBITION OF CRISPR-CAS BY BACTERIOPHAGES

4.1 Introduction

Although transcriptional regulation is clearly a rich source of insights regarding the ways in which prokaryotes modify the functionality of their CRISPR-Cas systems, it is not the only mechanism by which CRISPR-Cas systems are modulated. In any system involving a host being infected by a pathogen, there will inevitably be an evolutionary arms race for dominance. The host will develop a defense against the pathogen, the pathogen will then acquire a means of subverting that defense, requiring the host to provide another defense against the pathogen or be killed. This is seen in all levels of host-pathogen interactions. One example in eukaryotes is inactivation of antibodies, the primary means by which higher eukaryotic immune systems recognize pathogens and mount a response, by pathogens, such as staphylococcal protein A, which binds to and inactivates antibodies (Falugi, Kim, Missiakas, & Schneewind, 2013). This is also seen in bacteriophage-bacteria interactions, where bacteriophages such as *Staphylococcus* phage K or *E. coli* phage T4 have developed resistance to a key bacterial defense system, restriction-modification (RM), either through simple point mutations of their genomes removing RM recognition sequences or by acquisition of methylases or similar genes that modify the bacteriophage genomes, preventing recognition of the bacteriophage’s nucleic acids by RM systems (Labrie, Samson, & Moineau, 2010). It stands to reason that bacteriophage would also develop defenses against CRISPR-Cas immunity, and this has been demonstrated in several
ways. A bacteriophage genome has been sequenced and shown to contain the H-NS gene, which, in *E. coli*, acts as a repressor of CRISPR-Cas immunity (Skennerton et al., 2011). It is possible that this phage would be able to transcriptionally repress the CRISPR-Cas system of its host, allowing the phage to kill the host before a CRISPR-Cas immune response could be mounted. Additionally, mutations within the protospacer or protospacer-adjacent motifs can often result in resistance to CRISPR-Cas immunity as they may render the system’s spacers useless against that phage, allowing it to persist and successfully infect (Deveau et al., 2008).

Perhaps the most sophisticated mechanism by which bacteriophages have been demonstrated to subvert CRISPR-Cas immunity is via anti-CRISPR proteins. First described in *Pseudomonas aeruginosa* phages, several phage-encoded proteins that prohibit CRISPR-Cas immunity have been described within diverse Proteobacteria species for Type I and II systems (Bondy-Denomy, Pawluk, Maxwell, & Davidson, 2013; Pawluk, Amrani, et al., 2016; Pawluk, Staals, et al., 2016; Pawluk, Bondy-Denomy, Cheung, Maxwell, & Davidson, 2014). Some of these proteins have been characterized at the molecular, structural level and inhibit CRISPR-Cas immunity by different mechanisms, including preventing crRNAs from binding to the target RNA and by binding to and inactivating Cas nucleases, blocking target degradation (Bondy-Denomy et al., 2015; Chowdhury et al., 2017; Pawluk, Amrani, et al., 2016). Although several examples of anti-CRISPR proteins have been identified and characterized, they are thus far limited to phages that infect bacteria in the phylum Proteobacteria and none have been shown to inhibit Type III CRISPR-Cas immunity. Recognizing this clear gap in our full understanding of bacteriophage-bacteria interactions in relation to CRISPR-Cas immunity, we endeavored to identify and characterize putative anti-CRISPR proteins encoded by Staphylococcal phages that inhibit the Type III-A CRISPR-Cas system of *S. epidermidis* RP62a.
4.2 Methodology

4.2.1 Preparation of Phage High-Titer Lysates

Two mL of BHI containing neomycin and 5 mM calcium chloride were inoculated simultaneously from the freezer stocks of the host strain, *S. epidermidis* RP62a, and freezer stocks of a phage, ISP, K7, or MP16. This mixture was allowed to grow overnight and filtered through a 0.45 μm syringe filter. 50-100 μL of this lysate was added to a 1:100 dilution of an overnight culture of *S. epidermidis* RP62a with 5 mM CaCl₂, alongside a negative control with no added phage. After ~3-4 hours, when clearing was obvious in the tubes containing phage as compared to the negative control, 1 mL of growing bacteria from the negative control was added to each tube containing a phage and the media in the negative control was replenished. The cells were allowed to grow and once per hour, another 1 mL of bacteria from the negative control was added to the tubes containing phages. This was repeated 3-4 times, the tubes containing phage were centrifuged (5,000 rpm, 5 minutes), and the supernatant was filtered through a 0.45 μm syringe filter. Titers were determined by plating serial dilutions (10⁰-10⁻⁷) on a top agar overlay of a 1:20 dilution of an overnight culture of *S. epidermidis* RP62a.

4.2.2 Phage Genomic DNA Extraction

Phage genomic DNA (gDNA) was isolated from high-titer lysates for use in PCR and sequence analysis. 200 μL of a high-titer phage lysate was combined with 200 μL of 25:24:1 phenol:chloroform:isoamyl alcohol, vortexed for 1 minute, and centrifuged (17,000 x g, 5 minutes). The aqueous phase was removed and phenol extracted as above. DNA was precipitated from the aqueous phase with 2.5 volumes 100% ethanol and 1/10 volume 3.0 M NaOAc, pH 5.2
and placed on ice for 10 minutes before being centrifuged as above. Samples were washed in 1 mL 75% EtOH and air-dried and the pellet was resuspended in 30 μL dH₂O. 1 μL was used as a template for PCR.

4.2.3 Statistical Analysis

Statistical analyses were performed as in 2.2.6.

4.3 Results and Discussion

4.3.1 Identification of Putative Anti-CRISPR Proteins in Staphylococcal Phages

Based on the current anti-CRISPR literature, there is a conserved genetic organization found in all identified anti-CRISPR genes. While the anti-CRISPR proteins themselves vary widely in size and sequence, they have thus far always been found directly upstream of a putative transcriptional regulator (called the anti-CRISPR associated protein, Aca) that is conserved and which has been the main method used to identify novel anti-CRISPR genes (Pawluk, Staals, et al., 2016). We have several staphylococcal phages in our lab, including novel staphylococcal phages isolated and characterized within the lab, and we began searching for a similar genetic architecture within our sequenced phages. We simply searched within our collection of phage genome sequences for relatively short (~50-150 amino acids) proteins with predicted transcription factor, based on the lengths of characterized Aca proteins Aca1 and Aca2 (79 and 125 amino acids, respectively) (Bondy-Denomy et al., 2013). We were particularly interested in putative aca genes found within the area surrounding the bacteriophage structural/lysis genes, where most confirmed anti-CRISPR genes have been found (Pawluk, Staals, et al., 2016). From this simple search, we identified three phages that infect S.
epidermidis RP62a that possess putative aca/acrloci: ISP (Vandersteegen et al., 2011), MP16 (not yet published), and K7 (not yet published) (Figure 4.1).

**Figure 4.1**
Putative acr and aca genes are found within the lysis region of the genomes of three bacteriophages that infect *S. epidermidis* RP62a. The two ORFs directly upstream and downstream of the predicted acr/aca are depicted and annotated. ORFs are not drawn to scale. Colors indicate the predicted function of each gene according to the NCBI Conserved Domain Database (orange=packaging, blue=lysis, grey=hypothetical genes of unknown function) or our own identification (purple=putative anti-CRISPR, red=putative anti-CRISPR associated regulator).

Each identified putative aca/acr pair is located within the lysis cassette of their respective phage as indicated by the presence of lytic genes (in blue) found immediately upstream of the putative Acr in all three phages as well as the co-localization of the acr/aca pairing on the same strand as other structural and lytic genes. Now that we had identified putative acr/aca genes, we performed a multiple sequence alignment of each set of genes with each other to examine how highly conserved they were between the three phages as well as alignment of the entire lysis cassettes to see how the synteny of the three loci compared. From the multiple sequence alignment of the putative acr and aca genes, two things immediately become apparent (Figure 4.2). First, the Aca protein sequence is better conserved than the Acr (63% vs 44% identity between the three genes), consistent with what has been seen in characterized anti-CRISPR genes where the Aca is well-conserved and used to identify novel anti-CRISPR genes (Bondy-
Additionally, the putative acr genes and the putative aca genes within K7 and MP16 are much better conserved compared to each other (100% sequence identity at the protein level and the nucleotide level) than when compared to the putative acr/aca of ISP. This is consistent with the relative conservation of the DNA sequence surrounding the acr/aca locus according to a multiple sequence alignment with progressiveMauve (Darling, Mau, & Perna, 2010). This genome alignment shows that the locus is conserved much more closely between K7 and MP16 (in green) than between either phage and ISP (in purple), especially upstream of the acr/aca locus (Figure 4.3). Phages K7 and MP16 are, in general, much more closely related to each other than to ISP, with average nucleotide identity of 97.93% (Goris et al., 2007), while neither K7 nor MP16 could even have their average nucleotide identity with ISP calculated, so this discrepancy in conservation makes sense evolutionarily.
The putative Aca and Acr proteins of ISP, K7, and MP16 are conserved at the amino acid level, though the putative Acr/Aca in ISP are divergent from those of K7 and MP16. A) PRALINE multiple-sequence alignment of the putative Acr protein sequences from ISP, K7, and MP16. B) PRALINE multiple-sequence alignment of the putative Aca sequences from ISP, K7, and MP16. Residues are colored by PRALINE based on their conservation as indicated below each alignment, with 0 being not conserved and 10 being highly conserved. Conservation scores are also indicated beneath each residue- * indicates 100% conservation.

**Figure 4.2**
Figure 4.3
The regions downstream of the putative acr/aca locus is well-conserved between ISP, K7, and MP16 but the region upstream is only well-conserved between K7 and MP16. The relative height of the bars indicates level of conservation between the phages: purple regions are conserved between all three phages and green regions are conserved only between K7 and MP16. The same ORFs highlighted in Figure 4.1 are indicated here and colored per their predicted functions as in Figure 4.1. (orange=packaging, blue=lysis, grey=hypothetical genes of unknown function, purple=putative anti-CRISPR, red=putative anti-CRISPR associated regulator).

4.3.2 Characterization of Phages Resistant to Cas10-Csm Interference

To assess whether the putative Acr proteins in these phage genomes were functional, we tested the ability of the native CRISPR-Cas system of S. epidermidis RP62a to target the three phages, ISP, K7, and MP16. If we were to introduce a spacer into a CRISPR-competent background and saw that plaquing (phage lysis) no longer occurred, then the phage is clearly sensitive to interference by our Cas10-Csm complex and it is not likely that any anti-CRISPR activity against our Type III-A system is present in that phage genome. If, however, one or more spacers are not able to prevent plaquing, indicating that the phage(s) are not impaired by CRISPR-Cas immunity, one explanation could be that the phage encodes a mechanism for evading CRISPR-Cas immunity, such as anti-CRISPR genes. To test these two possibilities, we designed three spacers against each phage genome, with the spacers regions with 100% conservation between phages K7 and MP16 (Figure 4.4A). For all three phages, one spacer targeted the polA gene and another targeted a lysis, testing interference of protospacers.
transcribed in each direction on the phage genome. Additionally, the third spacer targeting K7 and MP16 is complementary to the putative acr gene itself. The final spacer against ISP targets a hypothetical gene distal to the other two targeted loci. In all cases except for the one targeting the hypothetical ISP gene, there is no matching between the 5’ tag (the final 8 nucleotides on the 5’ end) of the crRNA and the targeted region, as complementarity in this region is known to abrogate Cas10-Csm immunity (Marraffini & Sontheimer, 2010). The spacer targeting the hypothetical ISP gene has two matches between the 5’ tag and the target, and so it is possible that, if plaquing is seen, the interference is simply being prevented by this match. The three spacers were introduced into a plasmid, pSpacer, carrying the crRNA leader followed by only a single repeat. Each plasmid was introduced into S. epidermidis LAM104, a derivative of RP62a with the CRISPR repeat/spacer array deleted (Marraffini & Sontheimer, 2008). On its own, LAM104 will produce all the Cas proteins, but no interference will occur since there are no crRNAs present. This allows us to saturate all the Cas10-Csm complexes present in the cell with our designed crRNAs without any competition from other crRNAs, which should allow for efficient interference of the phages. Serial dilutions of each phage can then be spotted onto lawns of S. epidermidis LAM104 strains carrying the various pSpacer derivatives and plaque formation will be seen only if phage infection is successful, indicating non-functional CRISPR immunity.
Figure 4.4
Native Cas10-Csm expression is unable to effectively protect against infection by MP16 and K7, but can protect against ISP. A) Diagrams of the genomes of ISP, K7, and MP16. Black arrows indicate the spacers which were tested for anti-phage interference: for K7 and MP16, 1 targets the putative anti-CRISPR and for ISP, 1 targets a hypothetical gene; for all three phages, 2 targets a lysin and 3 targets PolA. For K7 and ISP, spacers 4, 5, and 6 target the same genomic loci in the reverse direction. Green and yellow arrows are predicted ORFs larger than 50 amino acids in the reverse and forward directions, respectively. B) Quantitative results and representative images of a plaquing assay of ISP on *S. epidermidis* LAM104 bearing pSpacer derivatives encoding spacers against the indicated genes. In the images, serial dilutions are shown, from $10^0$ on the left to $10^5$ at right. C) Representative images of plaquing of K7 and MP16 on *S. epidermidis* LAM104 bearing pSpacer derivatives encoding spacers against the indicated genes. Serial dilutions are shown, from $10^0$ on the left to $10^{-7}$ at right. D) Quantification of K7 and MP16 plaquing on the assay shown in part C. Shown are averages of at least duplicate, ± one standard deviation.

For ISP, all three spacers 100% efficiently prevented the phage from lysing the bacteria, with no plaques seen (3.8B). Interestingly, K7 and MP16 continued to plaque at the same level as on the non-targeting strain in the presence of spacers targeting all three genomic loci, albeit with reduced plaque size for the three targeting spacers, particularly those targeting PolA and the putative anti-CRISPR (3.8C, D). K7 plaques on pSpacer-PolA were very small and not readily countable and is not depicted in Figure 4.4D as there was no way to be confident about the counts. However, it can be seen in Figure 4.4C that plaquing is indeed occurring at similar levels to the other strains. The *S. epidermidis* RP62a CRISPR-Cas system is known to have strand-specificity, functioning only when the crRNA is complementary to the RNA transcribed across
its target (Goldberg, Jiang, Bikard, & Marraffini, 2014). To confirm that the spacers were not simply inactive due to targeting the incorrect strand, we designed another set of three spacers complementary to K7 and MP16 which overlapped the original set, but targeting the opposite strand (Figure 4.4A). Again, we saw no reduction in counts of plaque forming units but, unlike with the first set of spacers, we no longer saw a reduction in plaque size in the presence of these spacers, supporting the hypothesis that the plaque size decrease is due to partially successful Cas10-Csm interference (Figure 4.4 C, D).

It is interesting that K7 and MP16 appear to be resistant to Cas10-Csm immunity, while ISP is not, since all three have very similar putative anti-CRISPR genes. This could indicate that there is a CRISPR evasion mechanism, but it is not caused by the gene(s) that have been identified so far, or it could indicate that ISP has either a defective or potentially non-Cas10-Csm anti-CRISPR. K7 and MP16 are much more closely related to each other than to ISP, so it is logical that those two phages possess similar immune evasion mechanisms that may not be found in ISP, especially considering the 100% identity between their putative anti-CRISPR loci (Figure 4.2). Additionally, K7 and MP16 have a much more limited host range than ISP, plaquing only on S. epidermidis, and, for K7, only S. epidermidis RP62a and its derivatives. ISP, on the other hand, is able to plaque on a range of staphylococci, including S. aureus strains (Vandersteegen et al., 2011). ISP may use its similar gene to evade CRISPR-Cas immunity of another type, such as a Type II system found in S. aureus (Ran et al., 2015), or a divergent Cas10-Csm system. In any case, further genetic and biochemical work need to be carried out. It seems that there is likely some small amount of CRISPR-Cas activity against K7 and MP16 in the case of the first set of designed spacers, explaining the reduction in plaque size seen in these strains. However, it is still
possible that, as opposed to the phages possessing anti-CRISPR activity, the tested spacers are not sufficient for providing robust anti-phage immunity.

To test this second possibility, the three spacers that target the correct strand of K7 and MP16 were introduced into pGG-BsaI, a plasmid which bears the crRNA leader region followed by a single repeat, as before, but in this case followed by two BsaI recognition sites, allowing for rapid, efficient introduction of new spacers via Golden Gate assembly (Goldberg et al., 2014). More importantly, this plasmid bears the entire suite of cas genes from S. epidermidis RP62a, allowing for overexpression of the entire CRISPR-Cas system, not just the crRNAs. If there is a genuine CRISPR inhibition mechanism being employed, the excess Cas10-Csm complexes in the cell should be able to overcome the inhibition to some degree, allowing for increased interference of the phages compared to the previous crRNA overexpression constructs. If the spacers are just intrinsically insufficient to allow for interference, there should be no decrease in titer even in this overexpression construct. We introduced these three plasmids into S. epidermidis LM1680, the RP62a derivative that entirely lacks a CRISPR-Cas system in its genome, and again plated K7 and MP16 onto an overlay of these strains. In all cases, we observed a slight decrease in titer (~2-log difference) when plated onto strains bearing the overexpression construct with the three targeting spacers as compared to the negative control, pGG-BsaI without a spacer targeting either phage (for MP16, none of the targeting strains are significantly different (p-value > 0.05), while for K7, all three are significantly decreased (p-value = 0.03 for all three) (Figure 4.5). Additionally, the plaques appeared to be more turbid than those typically seen with K7 and MP16, further indicating that these phages are being targeted by the Cas10-Csm complex to some degree, but their mechanism of CRISPR-Cas inactivation/evasion is strong enough that they are still able to plaque.
Figure 4.5
The evasion of Cas10-Csm immunity by K7 and MP16 is not eliminated by overexpression of cas genes. A) Representative plaquing assay of the indicated phages on S. epidermidis LM1680 bearing pGG-BsaI or one of its derivatives targeting the indicated genes on K7 and MP16. B) Quantification of the assay shown in A. Values are an average of a technical triplicate, ± one standard deviation. * indicates p-value < 0.05.

While we still lack any biochemical or genetic evidence of the exact nature of the CRISPR-Cas evasion or inhibition performed by K7 and MP16, we have a fair amount of preliminary evidence in vivo suggesting that at least two staphylococcal phages are capable of escape from Cas10-Csm interference. This resistance to CRISPR-Cas immunity is seemingly very powerful as it can overcome interference by the highly overexpressed Cas10-Csm complex (Figure 4.5). Additionally, we have a target gene that we suspect is the gene responsible for the observed CRISPR evasion based on in silico evidence of shared synteny with characterized anti-
CRISPR genes. Ongoing work in the lab is seeking to knockout this predicted Acr from the phage genomes using Cas9 from *Streptococcus pyogenes* (W. Chen et al., 2017). Additionally, we are attempting to introduce the putative *acr* and *aca* into a plasmid so they can be expressed outside of the context of infection by K7 and MP16, allowing examination of their ability to inhibit CRISPR-Cas immunity against other phages as well as anti-plasmid immunity.

4.3.3 Identification of Homologues of Putative Staphylococcal Anti-CRISPR Genes

Interestingly, our identified putative *acr* and *aca* genes appear widespread among sequenced bacteriophages. When the genes are searched using tBLASTn, searching against the NCBI non-redundant protein database, many staphylococcal, and non-staphylococcal, phages are among the top results (Appendix). Thirty-five distinct staphylococcal phages were identified in the tBLASTn search of the MP16/K7 *aca* and 32 from that of the MP16/K7 *acr*. Three phages were not found from the *acr* tBLASTn that were found for the *aca*: Twort, SA5, and Sb-1. Upon further inspection, a gene similar in length to the *acr* of MP16/K7 was found in the genomes of Twort and Sb-1, although the latter is annotated with an ATA start codon and these genes were found to be similar to the putative *acr* genes from other phages identified in the tBLASTn. SA5 bears a good homolog of the *aca*, but it has a 2-bp insertion near the start of what would be the Acr coding sequence, causing a frame shift mutation and the introduction of numerous nonsense mutations. In all, 35 staphylococcal phage genomes were found to carry sequences resembling both a putative *aca* and putative *acr* homologous to those identified in K7 and MP16, with the caveats expressed above in regards to Twort, SA5, and Sb-1.

Staphylococcal phages are all members of the *Caudovirales* order and divided into three families, the *Podoviridae, Siphoviridae,* and *Myoviridae* based on phage morphology and
genome sequences (Deghorain & Van Melder, 2012). There are 130 Staphylococcal phage genomes in GenBank, with 77 siphophages, 40 myophages, and 13 podophages according to their taxonomy listed in GenBank. All 39 phages (including 5 which are just point mutants of the same phage) identified from tBLASTn searches with the K7/MP16 putative aca and acr are within the Myoviridae family. The lone phage identified as a staphylococcal myophage lacking an acr/aca in the GenBank database is SA1, which, based on tBLASTn and BLAST searches with several CDS and genomic regions, does not appear to bear homology to any staphylococcal phages, rather producing hits predominantly from Salmonella and Enterobacter phages, and there are no publications which have characterized this phage. Thus, there is some doubt that this phage is truly a staphylococcal phage and is being disregarded in this analysis. Excluding this one outlier, all sequenced staphylococcal myophages carry a putative acr/aca pair and no staphylococcal podo- or siphophages bear these genes even though they comprise ~70% of sequenced staphylococcal phages. Our putative acr/aca genes appear to be widespread among one family of staphylococcal phages, but not at all found in other families, contrary to what has been seen in the Proteobacteria, where, with the exception of a single phage found in Pectobacterium phage ZF40, which is a myophage, all characterized anti-CRISPR genes are encoded either within siphophages or within bacterial genomes (Bondy-Denomy et al., 2013; Pawluk, Amrani, et al., 2016; Pawluk, Staals, et al., 2016).

4.4 Future Directions

While only preliminary results exist, the possibility of anti-CRISPR proteins against Type-III CRISPR-Cas systems is exciting and would confirm speculations that such counter-defense mechanisms are widespread among bacteria and their parasites. Additional work will be
done to knock-out these putative anti-CRISPR proteins and characterize the activities of the resulting mutant phages. Biochemical work will also be carried out to identify the mechanism by which these phages prevent Cas10-Csm anti-bacteriophage immunity. As these putative genes are widespread among staphylococcal myophages, it would also be interesting to investigate the ability of these other phages to evade immunity to strengthen the argument for these genes as the source of anti-CRISPR activity.
CHAPTER 5
CONCLUSIONS

CRISPR-Cas systems, initially regarded as a niche, but intriguing, aspect of bacterial physiology have gained increasing renown in recent years as they have been developed for diverse biotechnological and medicinal applications, including genome editing, infectious disease diagnosis, and RNA tracking (Doudna & Charpentier, 2014; Gootenberg et al., 2017; Nelles et al., 2016). In order to continue developing novel applications and to understand the effect of CRISPR-Cas on pathogenesis, it is necessary to understand all aspects of CRISPR-Cas functionality. To that end, we have investigated the crucial step of crRNA biogenesis in the Cas10-Csm complex of \textit{S. epidermidis} RP62a. Our work provides information regarding key interactions essential for CRISPR formation and begins to shed some light on the process of maturation in this system. Additionally, we have provided evidence that non-Cas nucleases may be involved in Cas10-Csm functionality, deepening the known connections between CRISPR and other cellular processes. Furthermore, our exploration of crRNA regulation, particularly the identification of Zn$^{2+}$ as a powerful regulator of crRNA expression allows us to begin to explore the question of how these organisms respond to environmental cues to ‘decide’ when to degrade or permit entrance of invasive genetic elements. This broadens our understanding of regulation of horizontal gene transfer in important human pathogens, a crucial process in our antibiotic resistance crisis. Finally, our identification of bacteriophages resistant to Cas10-Csm immunity provides insight into host-pathogen interactions on a microscopic level and may potentially aid
in the development of more efficient bacteriophage therapies by leading to the use of CRISPR-Cas-resistant therapies. The work presented herein is the foundation for further research into many aspects of CRISPR-Cas immunity which may have implications in human health in the future.
REFERENCES


76


### APPENDIX

#### A.1 tBLASTn Results of φK7/φMP16 Putative Anti-CRISPR Protein

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Subject</th>
<th>Bit-Score</th>
<th>E-Value</th>
<th>% Identity</th>
<th>% Positives</th>
<th>Subject Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>KF021268</td>
<td>Staphylococcus phage phiIBB-SEP1, complete genome</td>
<td>142.1</td>
<td>1.2209</td>
<td>7E-37</td>
<td>73</td>
<td>74</td>
</tr>
<tr>
<td>KP027447</td>
<td>Staphylococcus phage phiIPLA-C1C, complete genome</td>
<td>142.1</td>
<td>1.4262</td>
<td>1E-37</td>
<td>73</td>
<td>74</td>
</tr>
<tr>
<td>JX194239</td>
<td>Staphylococcus phage SA11, complete genome</td>
<td>59.30</td>
<td>1.4778</td>
<td>9E-08</td>
<td>31</td>
<td>45</td>
</tr>
<tr>
<td>KX532239</td>
<td>Staphylococcus phage StAP1, complete genome</td>
<td>59.30</td>
<td>1.4778</td>
<td>8E-08</td>
<td>31</td>
<td>45</td>
</tr>
<tr>
<td>KP881332</td>
<td>Staphylococcus phage Stau2, complete genome</td>
<td>59.30</td>
<td>1.4778</td>
<td>7E-08</td>
<td>31</td>
<td>45</td>
</tr>
<tr>
<td>JX846612</td>
<td>Staphylococcus phage vB_SauM_Remus, complete genome</td>
<td>59.30</td>
<td>1.4492</td>
<td>8E-08</td>
<td>31</td>
<td>45</td>
</tr>
<tr>
<td>JX846613</td>
<td>Staphylococcus phage vB_SauM_Romulus, complete genome</td>
<td>59.30</td>
<td>1.4492</td>
<td>6E-08</td>
<td>31</td>
<td>45</td>
</tr>
<tr>
<td>AY954969</td>
<td>Bacteriophage G1, complete genome</td>
<td>51.98</td>
<td>6.4063</td>
<td>2E-06</td>
<td>34</td>
<td>48</td>
</tr>
<tr>
<td>Accession</td>
<td>Organism</td>
<td>Description</td>
<td>GC content (%)</td>
<td>GC skew (M)</td>
<td>AT content (%)</td>
<td>AT skew (M)</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>-------------</td>
<td>----------------</td>
<td>-------------</td>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>JX080302</td>
<td>Staphylococcus phage 676Z, complete genome</td>
<td></td>
<td>51.98</td>
<td>78</td>
<td>6.4065</td>
<td>34</td>
</tr>
<tr>
<td>KJ206559</td>
<td>Staphylococcus phage 812, complete genome</td>
<td></td>
<td>51.98</td>
<td>78</td>
<td>6.4064</td>
<td>34</td>
</tr>
<tr>
<td>JX080301</td>
<td>Staphylococcus phage A3R, complete genome</td>
<td></td>
<td>51.98</td>
<td>78</td>
<td>6.4063</td>
<td>34</td>
</tr>
<tr>
<td>EU418428</td>
<td>Staphylococcus phage A5W, complete genome</td>
<td></td>
<td>51.98</td>
<td>78</td>
<td>6.4064</td>
<td>34</td>
</tr>
<tr>
<td>JX080303</td>
<td>Staphylococcus phage Fi200W, complete genome</td>
<td></td>
<td>51.98</td>
<td>78</td>
<td>6.4065</td>
<td>34</td>
</tr>
<tr>
<td>JQ686190</td>
<td>Staphylococcus phage GH15, complete genome</td>
<td></td>
<td>51.98</td>
<td>78</td>
<td>6.4063</td>
<td>34</td>
</tr>
<tr>
<td>KP687431</td>
<td>Staphylococcus phage IME-SA1, complete genome</td>
<td></td>
<td>51.98</td>
<td>78</td>
<td>6.4063</td>
<td>34</td>
</tr>
<tr>
<td>KR902361</td>
<td>Staphylococcus phage IME-SA118, complete genome</td>
<td></td>
<td>51.98</td>
<td>78</td>
<td>6.4063</td>
<td>34</td>
</tr>
<tr>
<td>KR908644</td>
<td>Staphylococcus phage IME-SA119, complete genome</td>
<td></td>
<td>51.98</td>
<td>78</td>
<td>6.4063</td>
<td>34</td>
</tr>
<tr>
<td>KP687432</td>
<td>Staphylococcus phage IME-SA2, complete genome</td>
<td></td>
<td>51.98</td>
<td>78</td>
<td>6.4063</td>
<td>34</td>
</tr>
<tr>
<td>FR852584</td>
<td>Staphylococcus phage ISP complete genome</td>
<td></td>
<td>51.98</td>
<td>78</td>
<td>6.4063</td>
<td>34</td>
</tr>
<tr>
<td>JX878671</td>
<td>Staphylococcus phage JD007, complete genome</td>
<td></td>
<td>51.98</td>
<td>78</td>
<td>6.4064</td>
<td>34</td>
</tr>
<tr>
<td>Accession</td>
<td>Organism</td>
<td>Source Type</td>
<td>GC%</td>
<td>MAF</td>
<td>Ens</td>
<td>Start</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
<td>------</td>
<td>------</td>
<td>----</td>
<td>-------</td>
</tr>
<tr>
<td>AY176327</td>
<td>Staphylococcus phage K, complete genome</td>
<td>51.98</td>
<td>6.4060</td>
<td>1E-06</td>
<td>34</td>
<td>48</td>
</tr>
<tr>
<td>KJ888149</td>
<td>Staphylococcus phage MCE-2014, complete genome</td>
<td>51.98</td>
<td>6.4064</td>
<td>E-06</td>
<td>34</td>
<td>48</td>
</tr>
<tr>
<td>JX080304</td>
<td>Staphylococcus phage MSA6, complete genome</td>
<td>51.98</td>
<td>6.4065</td>
<td>5E-06</td>
<td>34</td>
<td>48</td>
</tr>
<tr>
<td>KM216423</td>
<td>Staphylococcus phage P108, complete genome</td>
<td>51.98</td>
<td>6.4063</td>
<td>7E-06</td>
<td>34</td>
<td>48</td>
</tr>
<tr>
<td>JX080305</td>
<td>Staphylococcus phage P4W, complete genome</td>
<td>51.98</td>
<td>6.4065</td>
<td>3E-06</td>
<td>34</td>
<td>48</td>
</tr>
<tr>
<td>AB903967</td>
<td>Staphylococcus phage phiSA012 DNA, complete genome</td>
<td>51.98</td>
<td>6.4064</td>
<td>E-06</td>
<td>34</td>
<td>48</td>
</tr>
<tr>
<td>KY581279</td>
<td>Staphylococcus phage pSa-3, complete genome</td>
<td>51.98</td>
<td>6.4063</td>
<td>E-06</td>
<td>34</td>
<td>48</td>
</tr>
<tr>
<td>AB853330</td>
<td>Staphylococcus phage S25-3 DNA, complete genome</td>
<td>51.98</td>
<td>6.4063</td>
<td>5E-06</td>
<td>34</td>
<td>48</td>
</tr>
<tr>
<td>AB853331</td>
<td>Staphylococcus phage S25-4 DNA, complete genome</td>
<td>51.98</td>
<td>6.4061</td>
<td>5E-06</td>
<td>34</td>
<td>48</td>
</tr>
<tr>
<td>JX080300</td>
<td>Staphylococcus phage Staph1N, complete genome</td>
<td>51.98</td>
<td>6.4064</td>
<td>9E-06</td>
<td>34</td>
<td>48</td>
</tr>
<tr>
<td>KC012913</td>
<td>Staphylococcus phage Team1, complete genome</td>
<td>51.98</td>
<td>6.4063</td>
<td>8E-06</td>
<td>34</td>
<td>48</td>
</tr>
<tr>
<td>KP027446</td>
<td>Staphylococcus phage phiIPLA-RODI, complete genome</td>
<td>50.44</td>
<td>2.4247</td>
<td>7E-05</td>
<td>34</td>
<td>49</td>
</tr>
<tr>
<td>XM_009942890</td>
<td>PREDICTED: Opisthocomus hoazin FRAS1 related extracellular matrix 3 (FREM3), partial mRNA</td>
<td>34.65</td>
<td>8.2456</td>
<td>14</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>
A.2 tBLASTn Results of φK7/φMP16 Putative Anti-CRISPR Associated Protein

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Subject</th>
<th>Bit-Score</th>
<th>E-Value</th>
<th>% Identity</th>
<th>% Positives</th>
<th>Subject Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>KP027447</td>
<td>Staphylococcus phage phiIPLA-C1C, complete genome</td>
<td>149.0</td>
<td>5.0587 7E-40</td>
<td>76</td>
<td>76</td>
<td>MFMEKENKPLVLKKVREEKGYTLMRNLASELVNSYSYWEYGKKKPRTNNLQKLEKTLGVPEELEKSEWEEQDE*</td>
</tr>
<tr>
<td>KF021268</td>
<td>Staphylococcus phage phiIBB-SEP1, complete genome</td>
<td>146.3</td>
<td>5.4056 9E-39</td>
<td>75</td>
<td>75</td>
<td>MFMEKENKPLVLKKVREEKRYTLMRNLASELVNSYSYWEYGKKKPRTNNLQKLEKTLGVPEELEKSEWEEQDE*</td>
</tr>
<tr>
<td>AB853330</td>
<td>Staphylococcus phage S25-3 DNA, complete genome</td>
<td>95.90 05</td>
<td>2.2379 6E-21</td>
<td>48</td>
<td>60</td>
<td>MDEKKESKLNLQKIRVEKGHTLMRNLASELGVHYSLIYWEYGKKKPRSAKLALNTPGKELFKELEEDD</td>
</tr>
<tr>
<td>JX878671</td>
<td>Staphylococcus phage JD007, complete genome</td>
<td>95.90 05</td>
<td>2.2379 8E-21</td>
<td>48</td>
<td>60</td>
<td>MDEKKESKLNLQKIRVEKGHTLMRNLASELGVHYSLIYWEYGKKKPRSAKLALNTPGKELFKELEEDD</td>
</tr>
<tr>
<td>AB903967</td>
<td>Staphylococcus phage phiSA012 DNA, complete genome</td>
<td>95.90 05</td>
<td>2.2379 8E-21</td>
<td>48</td>
<td>60</td>
<td>MDEKKESKLNLQKIRVEKGHTLMRNLASELGVHYSLIYWEYGKKKPRSAKLALNTPGKELFKELEEDD</td>
</tr>
<tr>
<td>JX080305</td>
<td>Staphylococcus phage P4W, complete genome</td>
<td>90.12 25</td>
<td>2.8225 E-19</td>
<td>43</td>
<td>56</td>
<td>KKEKPLNLQKIRVEKGHTLMRNLASELGVHYSLIYWEYGKKKPRSDNLMLRNLALNTPGKELFKELEEDD</td>
</tr>
<tr>
<td>JX080303</td>
<td>Staphylococcus phage Fi200W, complete genome</td>
<td>90.12 25</td>
<td>2.8225 E-19</td>
<td>43</td>
<td>56</td>
<td>KKEKPLNLQKIRVEKGHTLMRNLASELGVHYSLIYWEYGKKKPRSDNLMLRNLALNTPGKELFKELEEDD</td>
</tr>
<tr>
<td>JX080302</td>
<td>Staphylococcus phage 676Z, complete genome</td>
<td>90.12 25</td>
<td>2.8225 E-19</td>
<td>43</td>
<td>56</td>
<td>KKEKPLNLQKIRVEKGHTLMRNLASELGVHYSLIYWEYGKKKPRSDNLMLRNLALNTPGKELFKELEEDD</td>
</tr>
<tr>
<td>HQ163896</td>
<td>Staphylococcus phage Sb-1, complete genome</td>
<td>88.96 69</td>
<td>6.0899 2E-19</td>
<td>43</td>
<td>55</td>
<td>KKEKPLNLQKIRVEKGHTLMRNLASELGVHYSLIYWEYGKKKPRSDNLMLRNLALNTPGKELFKELEEDD</td>
</tr>
<tr>
<td>Accession</td>
<td>Organism</td>
<td>Genome Type</td>
<td>GCContent</td>
<td>A/N</td>
<td>STR</td>
<td>Similarity</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------------</td>
<td>------------------------------</td>
<td>-----------</td>
<td>-----</td>
<td>-----</td>
<td>------------</td>
</tr>
<tr>
<td>AY176327</td>
<td>Staphylococcus phage K, complete genome</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0899</td>
<td>4E-19</td>
</tr>
<tr>
<td>AB853331</td>
<td>Staphylococcus phage S25-4 DNA, complete</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0900</td>
<td>4E-19</td>
</tr>
<tr>
<td>JX875065</td>
<td>Staphylococcus phage SA5, complete genome</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0901</td>
<td>5E-19</td>
</tr>
<tr>
<td>KY581279</td>
<td>Staphylococcus phage pSa-3, complete genome</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0901</td>
<td>7E-19</td>
</tr>
<tr>
<td>KJ206563</td>
<td>Staphylococcus phage 812 isolate host-range</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0901</td>
<td>8E-19</td>
</tr>
<tr>
<td>KJ206562</td>
<td>Staphylococcus phage 812 isolate host-range</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0901</td>
<td>8E-19</td>
</tr>
<tr>
<td>FR852584</td>
<td>Staphylococcus phage ISP complete genome</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0901</td>
<td>8E-19</td>
</tr>
<tr>
<td>AY954969</td>
<td>Bacteriophage G1, complete genome</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0901</td>
<td>9E-19</td>
</tr>
<tr>
<td>KJ206561</td>
<td>Staphylococcus phage 812 isolate host-range</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0902</td>
<td>1E-19</td>
</tr>
<tr>
<td>KJ206560</td>
<td>Staphylococcus phage 812 isolate host-range</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0902</td>
<td>1E-19</td>
</tr>
<tr>
<td>KR902361</td>
<td>Staphylococcus phage IME-SA118, complete</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0902</td>
<td>1E-19</td>
</tr>
<tr>
<td>JQ686190</td>
<td>Staphylococcus phage GH15, complete genome</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0902</td>
<td>1E-19</td>
</tr>
<tr>
<td>KP687431</td>
<td>Staphylococcus phage IME-SA1, complete</td>
<td></td>
<td>88.96</td>
<td>69</td>
<td>6.0902</td>
<td>2E-19</td>
</tr>
<tr>
<td>Accession</td>
<td>Phage Name</td>
<td>Identity</td>
<td>Length</td>
<td>GC%</td>
<td>Strain</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------</td>
<td>----------</td>
<td>--------</td>
<td>-----</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>KM216423</td>
<td>Staphylococcus phage P108, complete genome</td>
<td>88.96</td>
<td>69</td>
<td>43</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>KC012913</td>
<td>Staphylococcus phage Team1, complete genome</td>
<td>88.96</td>
<td>69</td>
<td>43</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>KP687432</td>
<td>Staphylococcus phage IME-SA2, complete genome</td>
<td>88.96</td>
<td>69</td>
<td>43</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>JX080301</td>
<td>Staphylococcus phage A3R, complete genome</td>
<td>88.96</td>
<td>69</td>
<td>43</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>KR908644</td>
<td>Staphylococcus phage IME-SA119, complete genome</td>
<td>88.96</td>
<td>69</td>
<td>43</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>KJ888149</td>
<td>Staphylococcus phage MCE-2014, complete genome</td>
<td>88.96</td>
<td>69</td>
<td>43</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>KJ206559</td>
<td>Staphylococcus phage 812, complete genome</td>
<td>88.96</td>
<td>69</td>
<td>43</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>KP027446</td>
<td>Staphylococcus phage phiPLA-RODI, complete genome</td>
<td>88.96</td>
<td>69</td>
<td>43</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>EU418428</td>
<td>Staphylococcus phage A5W, complete genome</td>
<td>88.96</td>
<td>69</td>
<td>43</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>JX080300</td>
<td>Staphylococcus phage Staph1N, complete genome</td>
<td>88.96</td>
<td>69</td>
<td>43</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>JX080304</td>
<td>Staphylococcus phage MSA6, complete genome</td>
<td>88.96</td>
<td>69</td>
<td>43</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>KF766114</td>
<td>Staphylococcus phage K, complete genome</td>
<td>88.96</td>
<td>69</td>
<td>43</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>JX846613</td>
<td>Staphylococcus phage vB_SauM_Romulus, complete genome</td>
<td>81.64</td>
<td>81</td>
<td>41</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Identity</td>
<td>Length</td>
<td>Similarity (%)</td>
<td>E-value</td>
<td>Length (nt)</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------------</td>
<td>----------</td>
<td>--------</td>
<td>----------------</td>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>JX846612</td>
<td>Staphylococcus phage vB_SauM_Remus, complete genome</td>
<td>81.64</td>
<td>81</td>
<td>7E-16</td>
<td>41</td>
<td>53</td>
</tr>
<tr>
<td>KP881332</td>
<td>Staphylococcus phage Stau2, complete genome</td>
<td>80.10</td>
<td>73</td>
<td>9E-16</td>
<td>41</td>
<td>52</td>
</tr>
<tr>
<td>KX532239</td>
<td>Staphylococcus phage StAP1, complete genome</td>
<td>80.10</td>
<td>73</td>
<td>5E-16</td>
<td>41</td>
<td>52</td>
</tr>
<tr>
<td>JX194239</td>
<td>Staphylococcus phage SA11, complete genome</td>
<td>80.10</td>
<td>73</td>
<td>7E-16</td>
<td>41</td>
<td>52</td>
</tr>
<tr>
<td>KP339049</td>
<td>Enterococcus phage EFDG1, complete genome</td>
<td>77.02</td>
<td>58</td>
<td>8E-14</td>
<td>35</td>
<td>49</td>
</tr>
<tr>
<td>AY954970</td>
<td>Staphylococcus phage Twort, complete genome</td>
<td>71.63</td>
<td>3</td>
<td>1E-13</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>AB609718</td>
<td>Enterococcus phage phiEF24C-P2 DNA, complete genome</td>
<td>70.86</td>
<td>26</td>
<td>2E-12</td>
<td>37</td>
<td>49</td>
</tr>
<tr>
<td>AP009390</td>
<td>Enterococcus phage phiEF24C DNA, complete genome</td>
<td>70.86</td>
<td>26</td>
<td>2E-12</td>
<td>37</td>
<td>49</td>
</tr>
<tr>
<td>KJ801817</td>
<td>Enterococcus phage ECP3, complete genome</td>
<td>70.86</td>
<td>26</td>
<td>3E-12</td>
<td>37</td>
<td>49</td>
</tr>
<tr>
<td>KJ668715</td>
<td>Listeria phage LMTA-34, partial genome</td>
<td>68.55</td>
<td>14</td>
<td>1E-11</td>
<td>32</td>
<td>43</td>
</tr>
<tr>
<td>DQ004855</td>
<td>Listeria bacteriophage P100, complete genome</td>
<td>68.16</td>
<td>62</td>
<td>8E-11</td>
<td>32</td>
<td>43</td>
</tr>
<tr>
<td>KJ535721</td>
<td>Listeria phage List-36, complete genome</td>
<td>68.16</td>
<td>62</td>
<td>8E-11</td>
<td>32</td>
<td>43</td>
</tr>
<tr>
<td>KJ591604</td>
<td>Listeria phage LMTA-148, complete genome</td>
<td>68.16</td>
<td>62</td>
<td>9E-11</td>
<td>32</td>
<td>43</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Identity</td>
<td>E-value</td>
<td>Score</td>
<td>Length</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------</td>
<td>----------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>KJ094033</td>
<td>Listeria phage LP-048, complete genome</td>
<td>68.16</td>
<td>1.3025</td>
<td>9E-11</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>JQ797329</td>
<td>Listeria phage vB_LmoM_AG20, complete genome</td>
<td>68.16</td>
<td>1.3025</td>
<td>9E-11</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>KM373208</td>
<td>Listeria phage WIL-1, complete genome</td>
<td>68.16</td>
<td>1.3026</td>
<td>E-11</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>KJ094029</td>
<td>Listeria phage LP-064, complete genome</td>
<td>68.16</td>
<td>1.3026</td>
<td>E-11</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>JX126918</td>
<td>Listeria phage LP-125, complete genome</td>
<td>68.16</td>
<td>1.3026</td>
<td>E-11</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>KJ094031</td>
<td>Listeria phage LP-124, complete genome</td>
<td>68.16</td>
<td>1.3026</td>
<td>E-11</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>KJ094030</td>
<td>Listeria phage LP-083-2, complete genome</td>
<td>68.16</td>
<td>1.3026</td>
<td>E-11</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>KJ586795</td>
<td>Listeria phage LMTA-94, complete genome</td>
<td>68.16</td>
<td>1.3026</td>
<td>E-11</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>KJ591605</td>
<td>Listeria phage LMTA-57, complete genome</td>
<td>68.16</td>
<td>1.3026</td>
<td>E-11</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>DQ003638</td>
<td>Listeria phage A511, complete genome</td>
<td>68.16</td>
<td>1.3026</td>
<td>E-11</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>KJ535722</td>
<td>Listeria phage LMSP-25, complete genome</td>
<td>68.16</td>
<td>1.3026</td>
<td>E-11</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>KJ586794</td>
<td>Listeria phage LMTA-34, complete genome</td>
<td>68.16</td>
<td>1.3026</td>
<td>E-11</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>KX961630</td>
<td>Bacillus phage QCM8, complete genome</td>
<td>56.99</td>
<td>1.3120</td>
<td>6E-07</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Length</td>
<td>GC-content</td>
<td>Similarity Score</td>
<td>Identity Score</td>
<td>Accessions</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>--------</td>
<td>------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>------------</td>
</tr>
<tr>
<td>KF669652</td>
<td>Bacillus phage Grass, complete genome</td>
<td>55.45</td>
<td>4.4522 5E-07</td>
<td>28</td>
<td>38</td>
<td>LRRKRLDHGYTIYSLAAKLGVNFSTI SYWETGEKFPRRQKLEELDFLVNG YRELFKD</td>
</tr>
<tr>
<td>KJ010547</td>
<td>Bacillus phage Bp8p-C, complete genome</td>
<td>55.45</td>
<td>4.6752 E-07</td>
<td>27</td>
<td>36</td>
<td>LKKFRMDSGYSINTLAALKLGVDSSTI SYWESGKRFPRRNVLQLEDFLVNSY REL</td>
</tr>
<tr>
<td>KJ010548</td>
<td>Bacillus phage Bp8p-T, complete genome</td>
<td>55.45</td>
<td>4.6752 E-07</td>
<td>27</td>
<td>36</td>
<td>LKKFRMDSGYSINTLAALKLGVDSSTI SYWESGKRFPRRNVLQLEDFLVNSY REL</td>
</tr>
<tr>
<td>AP013029</td>
<td>Bacillus phage phiNIT1 DNA, complete genome</td>
<td>55.45</td>
<td>4.7212 E-07</td>
<td>28</td>
<td>38</td>
<td>LRRKRLDHGYTIYSLAAKLGVNFSTI SYWETGEKFPRRQKLEELDFLVNG YRELFKD</td>
</tr>
<tr>
<td>AB930182</td>
<td>Bacillus phage SPG24 DNA, contig00001 sequence</td>
<td>54.68</td>
<td>7.1183 2E-07</td>
<td>28</td>
<td>37</td>
<td>LRRKRLDHGYTIYSLADKLGVDSSTI SYWETGEKFPRRQKLEELDFLVNG YRELFKD</td>
</tr>
<tr>
<td>KM051843</td>
<td>Bacillus phage Bobb, complete genome</td>
<td>54.29</td>
<td>9.3602 2E-07</td>
<td>26</td>
<td>36</td>
<td>MTSKSEHTPRLKLYRLDGGFTIYSL ADRVGVHYSSGWSRNFKPRIDK LMILEEIFKVGYREL</td>
</tr>
<tr>
<td>KU253712</td>
<td>Bacillus phage Eldridge, complete genome</td>
<td>51.21</td>
<td>1.2259 6E-05</td>
<td>27</td>
<td>39</td>
<td>LEMKNKTPKRLKRYMDAGITYTL ADRKLEEILGKHYSSISYWENGEKVLPRIGTVALEDFLVNGYRELSDLTDD</td>
</tr>
<tr>
<td>KM236245</td>
<td>Bacillus phage Mater, complete genome</td>
<td>51.21</td>
<td>1.2624 7E-05</td>
<td>31</td>
<td>43</td>
<td>MTQKSEHTPRLKLYRLDGRFTIYTL ADRVGVHYSSGWSRNFKPRIDK LMILEEIFDENYREL</td>
</tr>
<tr>
<td>KM236246</td>
<td>Bacillus phage Moonbeam, complete genome</td>
<td>50.83</td>
<td>1.8310 4E-05</td>
<td>26</td>
<td>39</td>
<td>MTSKSEHTPRLKLYRLDGRFTIYVL ADRVGVHYSSGWSRNFKPRIDK LMILEEIFDENYREL</td>
</tr>
<tr>
<td>KJ451625</td>
<td>Bacillus phage Bcp1, complete genome</td>
<td>50.83</td>
<td>1.9608 1E-05</td>
<td>23</td>
<td>32</td>
<td>RMDAGHTIYSLADKLGVDTTVSNW ENGNKYPRNPNKTELEDLGVSREL</td>
</tr>
<tr>
<td>KT995480</td>
<td>Bacillus phage BM10, complete genome</td>
<td>50.83</td>
<td>1.9801 7E-05</td>
<td>23</td>
<td>32</td>
<td>RMDAGHTIYSLADKLGVDTTVSNW ENGNKYPRNPNKTELEDLGVSREL</td>
</tr>
<tr>
<td>JX094431</td>
<td>Bacillus cereus bacteriophage vB_BceM_Bc431v3, complete genome</td>
<td>50.44</td>
<td>2.1836 9E-05</td>
<td>23</td>
<td>35</td>
<td>RMDAGHTIYSLADKLGVDTTVSNW ENGNKYPRNPNKTELEDLGVSRYRL FEDLPED</td>
</tr>
<tr>
<td>KJ676859</td>
<td>Bacillus phage JBP901, complete genome</td>
<td>50.44</td>
<td>2.3385 3E-05</td>
<td>23</td>
<td>32</td>
<td>RMDAGHTIYSLADKLGVDTTVSNW ENGNKYPRNPNKTELEDLGVSREL</td>
</tr>
<tr>
<td>CP014859</td>
<td>Actinoalloteichus hymeniacidonis strain HPA177(T) (=DSM 45092(T)), complete genome</td>
<td>46.98</td>
<td>0.0003 45436</td>
<td>20</td>
<td>34</td>
<td>LRRALRNATGRSVRELANELIGAYGTI SRWGERVPQVESVARLLGALGAS HDDFER</td>
</tr>
<tr>
<td>accession</td>
<td>species/description</td>
<td>id</td>
<td>length</td>
<td>score</td>
<td>I</td>
<td>V</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------</td>
<td>----</td>
<td>--------</td>
<td>-------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>FN563149</td>
<td>Rhodococcus equi 103S chromosome</td>
<td>46.98</td>
<td>0.0003</td>
<td>55735</td>
<td>23</td>
<td>38</td>
</tr>
<tr>
<td>GQ876259</td>
<td>Uncultured organism clone VC1AA68TR genomic sequence</td>
<td>46.59</td>
<td>0.0004</td>
<td>48833</td>
<td>23</td>
<td>29</td>
</tr>
<tr>
<td>CP016077</td>
<td>Actinoalloteichus sp. GBA129-24, complete genome</td>
<td>46.20</td>
<td>0.0006</td>
<td>72402</td>
<td>21</td>
<td>31</td>
</tr>
<tr>
<td>CP002199</td>
<td>Cyanothece sp. PCC 7822 plasmid Cy782201, complete sequence</td>
<td>46.20</td>
<td>0.0007</td>
<td>86435</td>
<td>23</td>
<td>36</td>
</tr>
<tr>
<td>LT629777</td>
<td>Pseudomonas asplenii strain ATCC 23835 genome assembly, chromosome: I</td>
<td>46.20</td>
<td>0.0008</td>
<td>42325</td>
<td>22</td>
<td>38</td>
</tr>
<tr>
<td>LN679998</td>
<td>[Clostridium] sordellii genome assembly ATCC9714_. chromosome : 1</td>
<td>45.82</td>
<td>0.0009</td>
<td>75652</td>
<td>25</td>
<td>33</td>
</tr>
<tr>
<td>LN681234</td>
<td>[Clostridium] sordellii genome assembly JGS6382, chromosome : 1</td>
<td>45.82</td>
<td>0.0009</td>
<td>75652</td>
<td>25</td>
<td>33</td>
</tr>
<tr>
<td>CP016207</td>
<td>Streptococcus sp. oral taxon 064 strain W10853 sequence</td>
<td>45.82</td>
<td>0.0011</td>
<td>3008</td>
<td>25</td>
<td>34</td>
</tr>
<tr>
<td>CP015849</td>
<td>Streptomyces sp. SAT1, complete genome</td>
<td>45.82</td>
<td>0.0011</td>
<td>4124</td>
<td>21</td>
<td>35</td>
</tr>
<tr>
<td>CP004856</td>
<td>Enterococcus casseliflavus EC20, complete genome</td>
<td>45.82</td>
<td>0.0011</td>
<td>7527</td>
<td>25</td>
<td>32</td>
</tr>
<tr>
<td>FM201786</td>
<td>Streptococcus pneumoniae partial Tn5253-like transposon, strain SpnA213</td>
<td>45.43</td>
<td>0.0012</td>
<td>4494</td>
<td>25</td>
<td>34</td>
</tr>
<tr>
<td>LN774769</td>
<td>Lactococcus piscium MKFS47 genome assembly L_piscium, chromosome : I</td>
<td>45.43</td>
<td>0.0013</td>
<td>09</td>
<td>22</td>
<td>36</td>
</tr>
<tr>
<td>CP007145</td>
<td>Hymenobacter swuensis DY53, complete genome</td>
<td>45.43</td>
<td>0.0015</td>
<td>312</td>
<td>23</td>
<td>38</td>
</tr>
<tr>
<td>Accession</td>
<td>Organism</td>
<td>Strain/Sequence</td>
<td>tmin</td>
<td>tmax</td>
<td>Score</td>
<td>Query Cover</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------------------------------------------</td>
<td>------------------------------------------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>CP014067</td>
<td>Enterococcus gallinarum strain</td>
<td>FDAARGOS_163, complete genome</td>
<td>45.43</td>
<td>94</td>
<td>0.0016</td>
<td>399</td>
</tr>
<tr>
<td>LT622830</td>
<td>Streptococcus salivarius integrative and conjugative element ICE_Ssa F6-1, strain F6-1</td>
<td></td>
<td>45.05</td>
<td>42</td>
<td>0.0020</td>
<td>6942</td>
</tr>
<tr>
<td>CP006587</td>
<td>Hymenobacter sp. APR13, complete genome</td>
<td></td>
<td>45.05</td>
<td>42</td>
<td>0.0022</td>
<td>2202</td>
</tr>
<tr>
<td>CP014942</td>
<td>Rhodococcus sp. BH4 plasmid, complete sequence</td>
<td></td>
<td>44.66</td>
<td>9</td>
<td>0.0024</td>
<td>5079</td>
</tr>
<tr>
<td>CP012752</td>
<td>Kibdelosporangium phytohabitans strain</td>
<td>KLBMP1111, complete genome</td>
<td>44.66</td>
<td>9</td>
<td>0.0025</td>
<td>2391</td>
</tr>
<tr>
<td>CP020382</td>
<td>Rhodothermaceae bacterium RA, complete genome</td>
<td></td>
<td>44.66</td>
<td>9</td>
<td>0.0026</td>
<td>7676</td>
</tr>
<tr>
<td>AP012222</td>
<td>Sphingobium sp. SYK-6 DNA, complete genome</td>
<td></td>
<td>44.66</td>
<td>9</td>
<td>0.0027</td>
<td>0313</td>
</tr>
<tr>
<td>CP001726</td>
<td>Eggerthella lenta DSM 2243, complete genome</td>
<td></td>
<td>44.66</td>
<td>9</td>
<td>0.0027</td>
<td>0313</td>
</tr>
<tr>
<td>FR856862</td>
<td>Novosphingobium sp. PP1Y main chromosome, complete replicon</td>
<td></td>
<td>44.66</td>
<td>9</td>
<td>0.0027</td>
<td>0313</td>
</tr>
<tr>
<td>CP001131</td>
<td>Anaeromyxobacter sp. K, complete genome</td>
<td></td>
<td>44.66</td>
<td>9</td>
<td>0.0027</td>
<td>838</td>
</tr>
<tr>
<td>CP009922</td>
<td>Streptomyces xiamenensis strain 318, complete genome</td>
<td></td>
<td>44.66</td>
<td>9</td>
<td>0.0030</td>
<td>405</td>
</tr>
<tr>
<td>CP003876</td>
<td>Nocardia brasiliensis ATCC 700358, complete genome</td>
<td></td>
<td>44.28</td>
<td>38</td>
<td>0.0032</td>
<td>2468</td>
</tr>
<tr>
<td>CP010054</td>
<td>Hymenobacter sp. DG25B, complete genome</td>
<td></td>
<td>44.28</td>
<td>38</td>
<td>0.0035</td>
<td>22</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Assembly</td>
<td>Identity</td>
<td>Similarity</td>
<td>Coverage</td>
<td>Start</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>------------</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td>LN774769</td>
<td>Lactococcus piscium MKFS47 genome assembly L_piscium, chromosome : I</td>
<td>42.74</td>
<td>0.0114</td>
<td>23</td>
<td>3</td>
<td>203</td>
</tr>
<tr>
<td>CP014859</td>
<td>Actinoalloteichus hynemiacidonidis strain HPA177(T) (=DSM 45092(T)), genome</td>
<td>41.58</td>
<td>0.0284</td>
<td>19</td>
<td>7</td>
<td>301</td>
</tr>
<tr>
<td>LT622830</td>
<td>Streptococcus salivarius integrative and conjugative element ICE_Ssa F6-1_rpsI, strain F6-1</td>
<td>38.89</td>
<td>0.2682</td>
<td>21</td>
<td>1</td>
<td>57</td>
</tr>
<tr>
<td>CP014859</td>
<td>Actinoalloteichus hynemiacidonidis strain HPA177(T) (=DSM 45092(T)), genome</td>
<td>38.50</td>
<td>0.3867</td>
<td>17</td>
<td>58</td>
<td>56</td>
</tr>
<tr>
<td>LN679998</td>
<td>[Clostridium] sordellii genome assembly ATCC9714_., chromosome : I</td>
<td>37.73</td>
<td>0.8401</td>
<td>19</td>
<td>54</td>
<td>32</td>
</tr>
<tr>
<td>LN681234</td>
<td>[Clostridium] sordellii genome assembly JGS6382, chromosome : I</td>
<td>37.73</td>
<td>0.8737</td>
<td>19</td>
<td>54</td>
<td>95</td>
</tr>
<tr>
<td>CP016207</td>
<td>Streptococcus sp. oral taxon 064 strain W10853 sequence</td>
<td>37.35</td>
<td>1.0124</td>
<td>22</td>
<td>02</td>
<td>8</td>
</tr>
<tr>
<td>LN774769</td>
<td>Lactococcus piscium MKFS47 genome assembly L_piscium, chromosome : I</td>
<td>36.96</td>
<td>1.5446</td>
<td>20</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>CP012752</td>
<td>Kibdelosporangium phytohabitans strain KLBMP1111, complete genome</td>
<td>36.96</td>
<td>1.5753</td>
<td>19</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>CP016077</td>
<td>Actinoalloteichus sp. GBA129-24, complete genome</td>
<td>34.65</td>
<td>8.6251</td>
<td>20</td>
<td>38</td>
<td>1</td>
</tr>
<tr>
<td>CP003876</td>
<td>Nocardia brasiliensis ATCC 700358, complete genome</td>
<td>34.65</td>
<td>8.6251</td>
<td>19</td>
<td>38</td>
<td>3</td>
</tr>
</tbody>
</table>