ABSTRACT

CAÑEZ, CASSANDRA LYNN. Development and Applications of CRISPR-Cas Based Precision Antimicrobials (Under the direction of Dr. Rodolphe Barrangou).

The groundbreaking introduction of antibiotics as a therapeutic treatment hinged on the ability of these compounds to eliminate infectious disease-causing pathogens without harming the recipient. It is now established that the indiscriminate mode of action of broad-spectrum antibiotics has more detrimental effects to the host than previously thought, and results in disruption of commensal microbiomes, especially in the human gut. The growing epidemic of multi-drug resistant bacteria, in combination with research underscoring the importance of symbiotic microbial populations of various niches, emphasizes the necessity for an alternative solution to traditional antimicrobials.

CRISPR-Cas systems offer a novel platform for specifically targeting pathogenic bacteria while maintaining the existing beneficial microbial consortia. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and their associated genes (cas) are adaptive immune systems widespread in prokaryotes. These systems maintain a genetic record of detrimental phages, plasmids, and other harmful nucleic acids in a CRISPR array. These arrays provide the foundational sequence for guide RNAs that mature and combine with Cas proteins to target and destroy invasive nucleic acids. There is a wide diversity of these systems; primarily categorized into two classes defined by their effector module composition and classified further into types and subtypes based on Cas nucleases, accessory Cas proteins, crRNA biogenesis, target recognition and loci organization. These systems have been repurposed for several applications due to their programmable nature and ability to carry out an assortment of functions from genome editing in eukaryotes to strain vaccination in industry relevant bacteria.
Most recently, these systems have been investigated for their ability to target and cleave bacterial genomes as a basis for sequence-specific antimicrobial applications. These studies utilized native systems to observe outcomes of chromosomal targeting in bacteria using either Type I or Type II CRISPR-Cas systems. The characteristic cleavage methods of Cas effectors in Type I and Type II systems damage DNA by exonucleolytic degradation of one strand of DNA or by generating an endonucleolytic double stranded blunt cut, respectively. Given the different mechanisms which Cas effector complexes cleave DNA, it is unclear if chromosomal targeting by distinct system types elicits more deleterious effects when applied as an antibacterial agent.

To answer this question, the model organism *Streptococcus thermophilus* DGCC7710, a staple strain in the cultured dairy industry, was used to investigate the outcomes of endogenous self-targeting. By utilizing active endogenous CRISPR-Cas systems in DGCC7710 we were able to investigative the outcomes of chromosomal damage generated from targeted cleavage from either the Type I-E or Type II-A systems. Plasmids harboring engineered CRISPR arrays containing spacers targeting chromosomal sequences were transformed into *S. thermophilus* to redirect the cleavage mechanisms of either the Type I-E or Type II-A systems. These self-targeting experiments resulted in several survivors largely enabled by delivery of a defective plasmid lacking a target. The most notable genotype recovered was due to a large deletion brought on by homologous recombination (HR) of identical sequences in separate *galE* coding regions, resulting in a 2% loss of the genome. This consequence reveals how HR contributes to the plasticity and remodeling of bacterial genomes as well as the avoidance of CRISPR-based targeting. This research will aid in future efforts to predict these events *in silico* for development of reliable, efficacious CRISPR-based antimicrobial targets to selectively eliminate pathogenic bacteria from beneficial microbial consortia.
Development and Applications of CRISPR-Cas Based Precision Antimicrobials

by
Cassandra Lynn Cañez

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Functional Genomics

Raleigh, North Carolina

2018

APPROVED BY:

_______________________________
Dr. Rodolphe Barrangou
Committee Chair

_______________________________
Dr. Matthew Koci

_______________________________
Dr. Jean Ristaino
BIOGRAPHY

Cassandra Cañe was born in Woodland, CA not too long after her father completed his PhD at UC Davis - this is where her love of science began. Frequent trips to the greenhouses and fields at the company where her dad worked sparked her curiosity in plants and biology. When she was seven years old her father told her about DNA and how it is responsible for all living organisms. Her parents moved the family to Yardley, PA for several years and then relocated to Raleigh, NC. Cassandra finished high school in Raleigh where she first learned about the potential of GMOs when one of her teachers told her about Golden Rice. She decided that she wanted to learn more about genetic engineering and attended NC State University for her undergraduate degree in Biochemistry. Throughout her undergraduate career she had internships at BASF and Monsanto where she worked on developing AHAS gene mutants for herbicide tolerant crops and optimizing molecular tools for generating and high throughput screening of mutant traits in \textit{Arabidopsis thaliana}. After her undergraduate degree, she was employed as a Research Technical Associate for Product Safety at Syngenta, Research and Production analyst generating engineered meganucleases at Precision Biosciences, and as a Research Associate II at Monsanto developing high throughput screening assays for phenotypic characterization of plants.

Eventually she wanted to learn more about the different molecular techniques and tools for genetic engineering and agriculture applications. In 2015, she applied and was accepted to the Functional Genomics MS program at NCSU to work with Dr. Rodolphe Barrangou researching applications of CRISPR-Cas systems. After completion of her Master’s degree Cassandra plans to apply to PhD programs to further her education in molecular biology and host-microbe interactions of plants for a career in the agriculture industry.
ACKNOWLEDGMENTS

I would like to thank everyone that helped me grow and progress as a scientist while at NC State University during my Master’s program.

- Dr. Rodolphe Barrangou, thank you for the opportunity to join the Barrangou lab and creating a family out of strangers. I am sincerely grateful for this experience.
- Dr. Todd Klaenhammer for educating me on the history of the lab and inspiring my curiosity of sourdough bread.
- To the lab members before me, Dr. Brant Johnson, Emily Stout, Jeff Hymes, Mia Theilmann, and Dr. Joakim Andersen for making me feel welcomed during the lab transition in such a short amount of time.
- To Dr. Kurt Selle for his encouragement, selfless contribution to getting me started, and seemingly endless tolerance to my questions.
- To the current lab members, Allie Crawley, Courtney Klotz, Katelyn Brandt, Dr. Claudio Hidalgo, Matthew Nethery, and Echo Pan. I have been so lucky to have gotten to know you all and thank you for making this experience so much fun and meaningful.
- To the staff, Rosemary Dawes, Dr. Jun Goh, and Dr. Sarah O’Flaherty, thank you all for your guidance and help with everything. You are all truly amazing.
- To Valerie Lapham for the beautiful SEM images and Dana Ripperton for all that you do.
- To my climbing friends, though I suspect you’ll never read this, thank you for continually reminding me of the importance of getting outside and having fun.
- Finally, thank you to my parents for their unwavering support, guidance, feedback, constant supply of jokes, and all my lunches. I wouldn’t have made it without you both and I cannot say thank you enough.
# TABLE OF CONTENTS

LIST OF TABLES .............................................................................................................................................. v
LIST OF FIGURES ............................................................................................................................................... vi

CHAPTER 1: DEVELOPMENT AND APPLICATIONS OF CRISPR-BASED PRECISION ANTIMICROBIALS .............................................................................................................................................. 1

1.1 ABSTRACT ...................................................................................................................................................... 2
1.2 INTRODUCTION ............................................................................................................................................. 3
1.3 CRISPR-CAS SYSTEMS OVERVIEW ............................................................................................................. 8
1.4 TYPE I AND TYPE II CRISPR-CAS INTERFERENCE ...................................................................................... 10
1.5 CRISPR-BASED ANTIMICROBIALS ............................................................................................................ 12
1.6 ANTI-CRISPR .................................................................................................................................................. 19
1.7 CONCLUSIONS AND PERSPECTIVES ......................................................................................................... 21

CHAPTER 2: OUTCOMES OF CHROMOSOMAL TARGETING UTILIZING NATIVE CRISPR-CAS SYSTEMS IN STREPTOCOCCUS THERMOPHILUS .................................................................................. 23

2.1 ABSTRACT ...................................................................................................................................................... 24
2.2 INTRODUCTION ............................................................................................................................................. 26
2.3 MATERIALS & METHODS ............................................................................................................................. 30
2.3.1 Bacterial Strains and Growth Conditions .................................................................................................. 30
2.3.2 DNA Isolation, CRISPR Array Construction, and Cloning ....................................................................... 30
2.3.3 Transformation of Streptococcus thermophilus DGCC7710 ...................................................................... 32
2.3.4 Carbohydrate Utilization Assessment ....................................................................................................... 33
2.3.5 Pacific Biosciences (PacBio) Whole Genome Sequencing, Annotations, and Protein Predictions .................................................................................................................................................. 34
2.3.6 RNA Extraction and Transcriptional (RNAseq) Analysis .......................................................................... 35
2.3.7 Scanning Electron Microscopy Imaging .................................................................................................... 37
2.4 RESULTS ......................................................................................................................................................... 38
2.4.1 CRISPR self-targeting is lethal in S. thermophilus DGCC7710 ..................................................................... 38
2.4.2 Characterization of Surviving Populations ................................................................................................. 39
2.4.3 Identification of Large Genomic Deletion .................................................................................................. 41
2.4.4 Phenotypic Comparisons of Wild-type and lacZ-targeted Survivor Mutant Strains .................................. 42
2.4.5 RNA-seq Transcriptional Analysis ............................................................................................................ 43
2.5 DISCUSSION ................................................................................................................................................ 44

2.6 TABLES ......................................................................................................................................................... 48
2.7 FIGURES ....................................................................................................................................................... 56
2.8 REFERENCES ................................................................................................................................................. 67
LIST OF TABLES

Table 2.6.1  Bacterial strains and plasmids ................................................................. 48
Table 2.6.2  Primers for sequencing and plasmid construction ........................................ 50
Table 2.6.3  Primer sets used in genomic deletion screen ............................................... 51
Table 2.6.4  List of CDS contained within the genomic deletion region of the lacZ-
targeting survivor mutants ......................................................................................... 52
Table 2.6.5  Genes differentially expressed between DGCC7710 wild-type and Type II-A
lacZ targeting survivor mutant .................................................................................. 53
Table 2.6.6  Differentially expressed hypothetical proteins and Phyre2 results ............... 55
LIST OF FIGURES

Figure 2.7.1 Diagram of experimental layout ................................................................. 56
Figure 2.7.2 Dosage responses of transformed targeting plasmids ........................................ 57
Figure 2.7.3 Type I-E and Type II-A targets at 400 ng ...................................................... 58
Figure 2.7.4 Genotyping of plasmids isolated from Type I-E and Type II-A targeting ................................................................. 59
Figure 2.7.5 PCR identification of chromosomal lacZ deletion in targeting survivors .......... 60
Figure 2.7.6 PCR screening for genomic deletion boundaries encompassing lacZ region..... 61
Figure 2.7.7 Identification of genomic deletion junction ..................................................... 62
Figure 2.7.8 Growth in broth and light microscope imaging ............................................... 63
Figure 2.7.9 Scanning electron microscopy (SEM) imaging of DGCC7710 and survivor mutant .................................................................................................................. 64
Figure 2.7.10 Carbohydrate growth experiments ............................................................... 65
Figure 2.7.11 Overlay of RNA-seq transcriptome profiles of wild-type DGCC7710 and mutant strain ................................................................................................................. 66
CHAPTER 1: DEVELOPMENT AND APPLICATIONS OF CRISPR-BASED PRECISION ANTIMICROBIALS.
1.1 ABSTRACT

The introduction of antibiotics revolutionized medical treatment of infectious disease. The ability to effectively treat disease by specific removal of pathogenic bacteria drastically reduced mortality rates and increased overall life expectancy. Although the application of antibiotics reduced deaths cause by bacterial diseases, nearly 100 years after their introduction, the widespread overuse of antibiotics has led to a number of unintended consequences. The rise of antibiotic resistance and the spread of resistance genes to pathogenic bacteria has created a pandemic of multi-drug resistant strains. Mortality rates have increased due to infections from multi-drug resistant bacteria in addition to having immense negative economic costs [1]. The discovery of CRISPR-Cas systems in bacteria have significantly expanded molecular tools available for genome editing, but precision genetic applications, in particular. These deliverable systems are easily reprogrammed, which offers the potential to cleave a desired genetic sequence by taking advantage of their naturally evolved DNA targeting and degradation mechanisms. Recently, proof of concept has been provided for repurposing CRISPR-Cas systems as precision antimicrobials. These studies specifically target bacteria, and trigger efficient killing of select microbes enabled by programmability and specificity of Cas nucleases. The dispatch of these systems, via engineered bacteriophages, reduces the potential for off-target effects, providing specificity to bacterial pathogens responsible for infectious disease. Additionally, selection granted by phage host range and the ability to design targeted sequences, enables selective elimination of pathogenic bacteria while maintaining beneficial consortia. Current research is summarized here, showcasing the potential CRISPR-Cas systems have as novel sequence-specific antimicrobials.
1.2 INTRODUCTION

The Golden age of antibiotics ushered in a new era that drastically reduced the mortality rate cause by infectious diseases in the 19\textsuperscript{th} century. The “magic bullet” described by Paul Ehrlich in 1910 offered the first chemotherapeutic shown to be an effective antisyphilitic treatment, although the organoarsenic compound was not without its side-effects [2, 3]. Ehrlich and Hata’s identification of arspheme, later termed Salvarsan, as a treatment of syphilis was the beginning of antimicrobial drug therapy. A major turning point for antimicrobial discovery was in 1928 when Alexander Fleming described the antimicrobial properties of a blue mold, \textit{Penicillium notatum} [4]. Although Fleming was able to report the antimicrobial activity observed from this mold, he was unable to demonstrate its therapeutic potential. Later in 1930, Howard Florey and Ernst Chain, together with a group of scientists, were able to exhibit penicillin’s therapeutic value building off of Fleming’s work [5, 6]. Salvarsan was highly utilized up until the successful demonstration of penicillin as a therapeutic in 1943 by Florey. However, initial antimicrobial characterization experiments done by Fleming unknowingly created a path for surveying biological organisms for their antimicrobial properties [7]. This research paved the way for antibiotic discovery, which in turn lead to the breakthrough of sulfonamides, streptomycin, and cephalosporins and derivatives of these compounds. Eventually, widespread distribution of these antibiotics was enabled by large-scale commercial manufacturing in the late 1940s [7]. Broad implementation of antibiotic treatments, in conjunction with vaccinations, reformed disease treatment and prevention in the 20\textsuperscript{th} century. These advanced approaches had tremendous societal influence by increasing longevity and reducing morbidity.

The modes of action of these antimicrobials is what allowed for the selective targeting of bacteria, while leaving the patient unharmed. Most notably penicillin, a beta-lactam,
indiscriminately targets bacteria by inhibiting cell wall synthesis by preventing the peptidoglycan crosslinking [8, 9]. Irreversible attachment to transpeptidases, annotated as penicillin binding proteins (PBPs), activates autolysis of the cell wall, which leads to cell death. Penicillin was first used in 1942 to effectively treat a patient with streptococcal septicemia [10]. Early discovery of native resistance to penicillin occurred in 1940 when beta-lactamases were discovered in *Escherichia coli*. These enzymes inactivate penicillin by cleavage of the beta-lactam ring [11]. Though this observation was meaningful it did not outweigh the substantial benefits of utilizing antibiotics to combat a variety of infectious diseases. Beta-lactamase inhibitors were used in conjunction with broad-spectrum penicillins to circumvent the inactivation of these antimicrobials [12]. Thereafter, despite the successful use of penicillin, hospitals had become a source of resistant strains of *Staphylococcus aureus* by the 1950s [13].

The development of antibiotic resistance genes (Abr) was further exacerbated by horizontal gene transfer events among associated bacteria that led to the rapid dissemination of Abr genes [14]. The selection for these resistance genes also occurred by killing bacteria which had not yet acquired resistance genes, and physically being unable to penetrate the coat of sporulated bacteria [15]. In addition to harboring and horizontal transfer of multiple antibiotic resistance genes, bacteria are also capable of modulating the expression of efflux pumps and vertical transfer point mutations that prevent binding of these compounds [16, 17]. These resistance mechanisms lower the amount of antibiotics within the cell, below an effective inhibitory concentration, or inhibit their ability to bind to target proteins preventing their activity entirely.

The human gut is an ideal environment for the spreading and selection of resistance gene to commensal bacteria and opportunistic pathogens. This enclosed niche environment is
composed of bacteria, archaea, viruses, protozoa and even fungi which influences the nutritional and health status of the host [18]. Closely associated microbial communities, such as the gut microbiota, can exchange and transfer genetic material by transduction via viral vectors, conjugation between directly connecting bacteria, or natural uptake of exogenous DNA [15]. The gut microbiome contains multitudes of bacterial taxa and acts as a diverse reservoir which allows for horizontal gene transfer events to occur [19]. Repeated exposure and selection for resistant bacteria overtime has favored the emergence and proliferation of antibiotic resistance.

Additionally, use of antimicrobials in various fields for application, aside from medical therapeutics, food production in agriculture and animal husbandry have contributed to providing reservoirs of resistance [20-22]. Good stewardship practices in agriculture to food production can limit the spread of resistance genes, however, consistency is important. Reported outbreaks of shigellosis are connected to poor food production hygiene. Though monitoring of resistant bacterial pathogens has provided circumstantial evidence of the spread of resistant bacteria. Although the spread of *Shigella* is most often transmitted via person-to-person via a fecal-oral route the contamination of crops may be attributed to use of contaminated water used for irrigating fields [21]. A variety of factors impede the ability to monitor all anthropogenic activities that contribute to the spread of bacterial antibiotic resistance.

Although this showcases the potential accumulation resistance genes in different areas, antibiotic resistant genes are naturally found in many species and ecological niches [23, 24]. Microbial products from actinomycetes have shown to produce antibiotic activity such as streptomycin from produced from *Streptomyces griseus* and tetracycline from *Streptomyces rimosus* [25, 26]. The natural occurrence of these biosynthesis genes responsible for producing these small molecule raises the question of what their natural ecological role in microbial
communities might be. Antibiotics derived from naturally occurring compounds have been shown to modulate cell signaling among bacteria at sub inhibitory levels [27]. It was shown in exposing *P. aeruginosa* to tobramycin, tetracyclin, and norfloxacin at sub-inhibitory levels increased biofilm formation and in some cases increased motility and triggering of the type III secretion system [28].

Cell signaling genes which interact with these naturally occurring compounds suggesting that these genes are putative resistance genes that can mutate and give rise to antibiotic resistance genes [28]. This suggests that bacteria are capable of surviving and resisting other naturally-derived antibiotics in the future. In addition to the spread of resistance genes there is increased understanding of how antibiotics disrupt beneficial microbial communities. In the recent past, extensive focus on microbial communities associated with various human niches has deepened our understanding of the bacterial world beyond pathogens. Increased research efforts to characterize host-microbe interactions have shed light on the many beneficial functions of bacteria [29, 30]. Microbial imbalance of the human gut can increase susceptibility to disease, negatively impact immune system development, and alter host metabolism. These factors can contribute to obesity and an abundance of physiological conditions and metabolic diseases [31].

Given the indiscriminate targeting of antibiotics removing beneficial bacteria and proliferation of resistance genes reducing their effectiveness against pathogens the need for new technology to address these issues is present. *Clustered Regularly Interspaced Short Palindromic Repeats* (CRISPR) and their *CRISPR-associated genes* (*cas*) constitute unique DNA targeting molecular tool used for a number of applications based on their sequence-specific target recognition [32-34]. The innate function of these bacterial immune genes is to elicit DNA damage as a method for avoiding infection or uptake of harmful mobile genetic elements such as
phages and plasmids. These naturally occurring systems provide a foundation for exact and programmable targeting with precision not afforded by antibiotics. This review covers basic CRISPR-Cas biology, describes their recent adaptation as CRISPR-based antimicrobials, and discusses opportunities and challenges lying ahead for these next-generation antibacterials.
1.3 CRISPR-CAS SYSTEMS OVERVIEW

CRISPR-Cas immune systems are encoded in the chromosome of most archaea and some bacteria which contain an array, comprised of direct repeats (21-48 nts) and variable sequences termed ‘spacers’, along with their neighboring cas genes [35, 36]. A variety of CRISPR-Cas systems have evolved in archaea and many bacteria, that provide an adaptive immune response against invasive harmful DNA, and, in some circumstances, RNA [36-39]. These systems are classified into 6 types and 29 distinct subtypes [40] defined by their variable effector modules, sequence variation among cas genes, an cas genes necessary for accessory processes, and loci organization. The components of the CRISPR-Cas loci undertake three phases which are necessary for functional immunity: acquisition, expression, and interference [37]. These three phases together allow for the RNA-guided sequence–specific antiviral defense in prokaryotes.

The acquisition phase occurs when novel spacers are obtained from sequences sampled from foreign DNA [41] called “protospacers” which are then incorporated into the CRISPR array, in an iterative, chronological manner. This acquisition step is carried out by Cas1 and Cas2, which are universally conserved among most known CRISPR-Cas systems [40, 42, 43]. The Cas1-Cas2 complex is responsible for the polarized incorporation of spacers into the array, integrating the newest spacers at the beginning of the CRISPR array [44, 45]. These spacers sequences serve as a genetic vaccination record for the host allowing for protection against successive attacks by identical or even similar invasive species [44].

The expression phase encompasses transcription of the repeat-spacer array into pre-crRNAs (pre-CRISPR RNA) and subsequent processing to produce mature CRISPR RNAs (crRNAs). The maturation process of the full and complete crRNAs are processed by different Cas proteins depending on the system type. Fully processed crRNAs then combine with the
signature Cas protein to form crRNP (crRNA-Cas ribonucleoprotein) complexes [46-48]. These complexes are composed of either a multisubunit complex or a single-multi domain protein, which define Class 1 and Class 2 systems, respectively [36]. The crRNP complex surveys the cell for targets and then cleaves complementary viral or plasmid nucleotide sequences resulting in interference [49-52]. Successful interference of non-self DNA requires a two-factor authentication process between the Cas effector protein, the crRNA, and the target DNA. Accurate recognition of the target is mediated by an antigenic signature sequence known as the PAM (protospacer-adjacent motif) [53, 54] which varies for each system sub-type [55]. The specificity of the RNA guide combined with the safe guard of the PAM requirements of the Cas protein enables non-self-identification and precise cleavage of the target [52, 56].

The processes of acquisition, crRNA biogenesis, and interference are consistent among the CRISPR-Cas systems. There are distinct mechanistic differences among modes of action for interference in various classes, types and sub-types of CRISPR-Cas systems. In this review, we will narrow the scope to mechanisms of Type I and Type II interference, but comprehensive reviews of CRISPR-Cas mechanisms for other subtypes are available [57-59].
1.4 TYPE I AND TYPE II CRISPR-CAS INTERFERENCE

Though there are multiple CRISPR-Cas systems with distinct interference mechanisms, however, the initial excitement for the molecular application of these systems was generated by the genome editing potential of the Type II system [60]. Type II CRISPR-Cas systems are categorized as Class 2 systems, and distinctly characterized by their signature single multidomain effector nuclease, Cas9 [36]. These systems are presently cataloged into three subtypes based on the organization of the locus and their specific single-protein nuclease Cas9. The trait characteristic of the Type II locus is the presence of the cas9 gene proximal to cas1 and cas2 [36]. The crRNP (CRISPR RNA ribonucleoprotein) complex for Type II systems is comprised of a dual RNA guide [61], tracrRNA::crRNA (trans-activating CRISPR RNA) [47], which loads into Cas9 [47]. The seed sequence [47] and PAM recognition by Cas9 is essential for DNA cleavage to occur [47, 62]. Once identification of the correct PAM sequence has been established by the WED and PI domains of Cas9 [63], R-loop formation is permitted [58]. The conserved RuvC and HNH domains of Cas9 are responsible for the recognition and cleavage of target sequences [64]. These domains form the operative passage which accommodates the guideRNA:targetDNA heteroduplex. Hybridization of the crRNA to the target and displacement of the complimentary strand, drives a conformational rearrangement in Cas9 [56, 64]. This rearrangement places the RuvC and HNH nuclease domains in a catalytically active position which results in a blunt dsDNA break [65]. This double stranded dsDNA break occurs precisely 3 nucleotides away from the 3’ edge of the protospacer sequence [50].

Type I CRISPR-Cas systems involve a more cumbersome multi-subunit complex classifying them in Class 1 with their signature gene, cas3 [36]. Type I systems exhibit extensive sequence and functional variability, and have been separated out in seven different
subtypes to date [36]. The Type I systems also account for the majority of identified CRISPR-
Cas loci in bacterial and archaeal genomes. Although more complex than Type II systems the
recognition requirements of a PAM and seed sequence are conserved and necessary [66]. The
crRNA-effector module consists of a core complex (Cas5, 6, and 7) with two subunits (Cse1 and
Cse2) also referred to as CasABCDE which later transitioned into what is now known as
Cascade (CRISPR associated complex for antiviral defense) [38, 67]. The coupled validation of
PAM recognition and correct complement hybridization of seed sequence to the target is needed
to initiate and accomplish interference. The identification of the PAM initiates correct base-
pairing of the crRNA to the seed sequence which promotes unidirectional R-loop formation [68].
Correct identification of the PAM and target sequence stabilize the R-loop in Cascade and induce
a conformational change. The R-loop is then “locked” in place which then allows licensing for
Cas3-mediated DNA degradation [68, 69]. Cas3 is a multifunctional protein with ATP-
dependent helicase activity stimulated by ssDNA and nuclease activity located on its N-terminal
HD-like domain [70]. The correct placement of Cascade to the target positions Cas3 properly to
interact with the CasA (Cse1) subunit, and helicase-nuclease activity is stimulated by the
displacement of the non-target strand [67, 71]. Once these conditions are met, the Cas3 nuclease-
helicase preferentially degrades ssDNA from the 3’ to 5’ direction [52, 70], leading to extensive
target DNA damage.
1.5 CRISPR-BASED ANTIMICROBIALS

The unique specificity and mechanisms of CRISPR-Cas interference renders invading mobile genetic elements (MGEs) unable to infect their desired bacterial host. This sequence-specific disarming of once functional genetic elements exhibits potential to develop precise molecular tools for generating DNA breaks in bacterial chromosomes. Bacteria have relatively limited DNA repair pathways to survive such damage rendering most chromosomal damage to be fatal in such populations [72].

The first experimental evidence establishing the significance of spacer content in conferring CRISPR-Cas interference was done by Barrangou et al [44]. Altering of spacer content in Streptococcus thermophilus DGCC7710 CRISPR1 locus along with challenging the engineered strains with phages matching spacers, demonstrated the necessity of spacer content in identifying appropriate targets. This study successfully demonstrated the adaptive nature of CRISPR-Cas systems but critically showed the redirecting of targets by exchanging spacer content. The ability CRISPR-Cas systems to be heterologously expressed and maintain utility was confirmed by Sparanauskas et al. [73] highlighting the transferability of these systems. The CRISPR3-Cas system from S. thermophilus was transformed into E. coli ER2267 was successfully able to confer interference of plasmids and of phage lambda. This work further supports the importance of spacer content but also the necessity of the correct PAM sequence while opening the doors for the transfer of these systems to other bacteria.

Horizontal gene transfer events can occur in several different ways, which allows the spread of virulence factors and antibiotic resistance genes. Experiments performed by Bikard et al. [49] in S. pneumoniae showed that introduction of S. pyogenes CRISPR-Cas system can prevent horizontal transfer of virulence genes to avirulent pneumococci. This demonstrated the
potential of CRISPR-Cas to prevent transfer of undesired genes to other bacterial populations. This research also exhibited that active CRISPR-Cas loci harboring chromosomal spacers with 100% identity could not exist within the same cell without eliciting cell death. Any colonies recovered from their experiments ‘escaped’ CRISPR-Cas targeting via inactive CRISPR-Cas loci or mutations in essential sequences. This CRISPR-Cas induced death of the majority of recipient bacterial cells sparked the possibility of repurposing CRISPR-Cas systems for sequence-specific killing of bacteria.

The first experimental evidence of the ability to reprogram Cas9 to cleave with an engineered single transcript guide, creating a chimeric crRNA-tracrRNA structure, to cleave a desired dsDNA sequence *in vitro* was demonstrated by Jinek et al [47]. This ease of programmability was expanded further by utilizing Cas9-based genome editing in bacteria presented by Jiang et al [74]. This research was able to validate the ease of editing with Cas9 while also establishing the requirements for target specificity. Generation of a dsDNA break in the chromosome accompanied with a repair template harboring the desired edit enables cell survival via homologous recombination. This creates a new genotype containing the edited gene which no longer includes the original target. Delivery of chromosomal target utilizing the native Type II system in *S. pneumoniae* strain crR6 with a repair template demonstrated the potential for precise genome editing with Cas9. To test efficiencies in genome editing researchers designed an experiment to measure independent cleavage by RNA guided Cas9. An erythromycin sensitive strain containing a premature stop codon in the resistance gene was created. This strain was then used to measure the efficiency of Cas9 editing by rescuing the resistant phenotype. Comparison of non-edited cells recovered after delivery of a chromosomal target and a control (no target) with a kanamycin resistance marker revealed that chromosomal
cleavage induced death in non-edited cells. This showed that increased efficiency in recovering edited cells was caused by elimination of the population containing a target for cleavage but did not repair with the template, resulting in death due to DNA cleavage. The dual-RNA:Cas9 selection of correctly edited cells contributes to the efficiency of edited genomes recovered by removal of wild-type population. However, it also exhibits the lethality of precise Cas9-induced chromosomal breaks in bacteria.

Repurposing of native CRISPR-Cas systems was investigated to observe the inter-play of harboring CRISPR-Cas systems and the plasticity of the bacterial genome. Vercoe et al. transformed an arabinose inducible plasmids containing an engineered CRISPR array with a chromosomal target to utilize the native type I-F CRISPR1 system in *Pectobacterium atrosepticum* [75]. The targeting of a pathogenicity island, HAI2, resulted in surviving populations where the entire island was removed (97,875 bp deletion) due to recombination at *attR* and *attL* sites flanking the island. Other variable sized deletions were observed, the largest at 40,277 bp, where they proposed that micro-homologies that led to putative alternative end-joining (A-EJ) events producing these deletions. Although targeting is not without biological hurdles, point mutations in PAM sequences resulted in wild-type surviving populations. These data highlighted the underlying genotypes existing within a small subset of mixed bacterial populations as well as demonstrating the variable outcomes of self-targeting.

Research by Gomaa *et al.* [76] exhibited the capabilities of sequence-specific targeting has on manipulating mixed bacterial populations. A plasmid-based type I-E systems was utilized in *E. coli* K12 to show the selectable killing potential of type I systems. These systems were delivered with their chromosome specific target in pure culture to observe killing potential in *E. coli* K12, *E.coli* B, and *S. enterica*. Targets were designed to remove specific strains with over
99% genetic identity or to targeting homologous genes among strains, demonstrating the programmable selection of strains based on target design. Various targets were chosen and engineered to perturb different regions of the genome to determine if targeting efficiency was reliant on the target of interest. Gomaa *et al.* demonstrated that cell death was independent of essentiality of genes targeted, varied genomic context, and transcriptionally active and inactive regions. This research was the first to characterize the lethality of CRISPR-Cas based genome targeting and remodeling of bacterial consortia.

Later in 2014 a series of studies investigating CRISPR-Cas systems for their antimicrobial potential were published by Bikard *et al.* [77] and Citorik *et al.* [78]. Researchers engineered phagemids to carry exogenous SpyCas9 (*Streptococcus pyogenes* Cas9) with sequence specific targets to showcase the ease of delivery coupled with the deleterious effects of targeting. Antibiotic resistance genes were targeted in clinically relevant *E coli* EMG2 and *S. aureus* RN4220 strains in Bikard and Citorik’s studies, respectively. Bikard was able to demonstrate the manipulation of heterogeneous populations by specifically targeting antibiotic resistance genes harbored in the chromosome or plasmids. Further experiments where a mouse-skin model was employed provided evidence for the potential to decolonize antibiotic resistant bacteria from an animal. Citorik *et al.* infected *Galleria mellonella* with enterohemorrhagic *E. coli* 0157:H7 (EHEC) to investigate survival rates after administering the phagemid-delivered SpyCas9 target. The survival rates of the phagemid treatment vs. chloramphenicol showed that delivery of a genome-targeting crRNA via phagemid was significantly more effective. Citorik and colleagues also targeted plasmids harboring addictive toxin-antitoxin systems which illustrated that elimination of the antitoxin system displayed cytotoxic effects. In both cases there were some survivors which were a result of defective plasmids due to deletion of the targeting
spacer, tracrRNA, and mutations or deletions in *cas9*. The combination of these papers reinforces the potential of sequence specific targeting with the addition of delivery via phagemids especially with *in vivo* models.

CRISPR-based targeting design shifted to a new direction by repurposing native CRISPR-Cas type II systems in *S. thermophilus* LMD-9 to screen for *in silico* predicted genomic islands. Selle *et al.* [79] identified and mapped critical IS elements within the genome of LMD-9 to design targets for screening of these potential genomic island excision events. CRISPR-based targeted selection for these events revealed four large genomic island present within the LMD-9 genome, the largest (102 kb) of which accounted for ~5.5% of the genome. These mutations within the genome are spontaneous and CRISPR-based selection enables recovery of individual deletion genotypes. This CRISPR-based screening method allows for selective isolation of rare genomic island excision events occurring at a low frequency among a microbial population. This research highlights the malleability of bacterial genomes and the MGEs at play when investigating chromosomal targets.

Cui *et al.* [80] challenged the efficacy of Cas9 based targeting by experimenting with the interplay between DNA repair machinery and chromosomal cleavage. It was previously documented that recombination events in bacterial populations can be observed, and lead to selection of these sub-populations within a heterogeneous mix. Comparison of different targeted regions in the presence or absence of *recA* showed that homology directed repair can save part of the *E. coli* population from Cas9 cleavage. Some targets were able to achieve lethal results regardless of the presence of *recA*. However, less deleterious targets resulted in survival when *recA* was functional. This was suspected to be due to a sister chromosome being present after cleavage providing a repair template to enable survival. This was experimentally supported by
providing a repair template concurrently with targeting. Together, this showed that homology directed repair is a mechanism enabling survival in bacteria following Cas9 cleavage. To test whether this pathway was utilized due to the bacterial SOS response, a GFP promoter was used to monitor SOS levels. These experiments linked the elevated SOS response to all targeting spacers examined. Although DNA damage induced the SOS response, survival was not achievable without a repair template present.

Other research by Tong et al. had shown indels (insertions and deletions) can be incorporated into Cas9 targeted sequences in *Streptomyces coelicolor* via native non-homologous end joining (NHEJ) system [81]. This research investigated the effects of NHEJ pathway present in bacteria and showed that in the absence of a repair template or sister chromosome, the NHEJ pathway offers an alternative mechanism to recover from dsDNA breaks. Exogenous SpyCas9 with a target was delivered to an *E. coli* strain harboring a NHEJ machinery from *M. tuberculosis* in the chromosome. Targeting resulted in an extremely low survival rate with colonies recovered containing small deletions at the target loci. Overall, introduction of an NHEJ repair system did not significantly rescue *E. coli* from Cas9 mediated DNA damage, supporting that HDR is the favored DNA repair method in bacteria.

Most recently Bernheim et al. investigated the connection between DNA repair systems and CRISPR-Cas systems present among 5,563 fully sequence bacterial genomes [82]. They discovered that there was a negative correlation in Firmicutes between Type II-A CRISPR-Cas and NHEJ systems. It was proposed that NHEJ systems are not associated with CRISPR-Cas as they could compete with the same substrate(s) and therefore could inhibit CRISPR-Cas interference. The effects of NHEJ system from *B. subtilis* in the presence of Type II-A system from *S. pyogenes* was investigated using *S. aureus* as a model. The resulting *S. aureus* strain was
challenged with a phiNM4 phage which contained a target for the introduced CRISPR-Cas system. Overall, their results showed that there was no measurable outcome of preventing CRISPR-Cas interference in the presence of NHEJ system. It was also shown that the presence of an NHEJ system has no effect on spacer acquisition. Surprisingly, Csn2 prevents NHEJ as it competes for the same DNA substrate as Ku which is necessary to recruit LigD for repair. This implies that substrate competition between CRISPR-Cas and NHEJ systems would most likely occur during acquisition process rather than interference.

By and large, the main outcome of chromosomal targeting by Cas nucleases has been shown to drastically reduce bacterial populations, with limited means to recover from such damage. CRISPR-Cas targeting experiments have demonstrated the flexible remodeling capabilities of bacterial genomes and interplay between chromosomal targeting with DNA repair machineries. Additional chromosomal targeting research will expand on our knowledge of the applications and potential limitations of these systems.
1.6 ANTI-CRISPR

The evolution of CRISPR-Cas systems to protect bacteria against harmful elements has been described as an “arms-race” for survival. Avoidance of CRISPR-Cas interference has been shown to occur due to mutations in crucial sequences necessary for effective interference and drives evolution of phage genomes [83]. However, it was recently shown that phages have evolved a counter-measure for CRISPR-Cas systems: anti-CRISPR proteins (Acr). Several years after the discovery of CRISPR-Cas systems, it was shown that certain bacteriophage genes can inactivate the CRISPR-Cas bacterial immune system [84]. These phage genes were shown to inhibit the type I-F system present in Pseudomonas aeruginosa PA14. Further investigation into these proteins produced from phage genes demonstrated that they are capable of interacting with different Cas proteins providing multiple mechanistic ways of inhibiting type I-F CRISPR-Cas systems [85]. Bioinformatics analysis was used along with insights from previous research to discover additional families of anti-CRISPR proteins shown to inhibit type I-E systems in P. aeruginosa [86]. The majority of characterized anti-CRISPR proteins inhibit Type I systems, but are functionally and genetically diverse in addition to lacking shared sequence motifs making in silico identification of these proteins challenging. Fortunately, genomic context of Type I anti-CRISPRs that have been found to share a highly conserved putative transcriptional regulator gene downstream of anti-CRISPR genes. This is now known as the anti-CRISPR associated 1 or Aca1. The presence of the aca1 gene has led to a “guilt by association” method to identify new anti-CRISPRs [87]. Similar to the discovery of CRISPR-Cas the research of anti-CRISPR proteins is increasing. Currently, there are 21 distinct families of these proteins characterized that inhibit type I and type II systems. Anti-CRISPR genes have been found in myophages, siphophages, conjugative elements, pathogenicity islands, and other MGEs. CRISPR-Cas
systems can limit horizontal gene transfer events [49] but show no evolutionary adaptation in doing so, which still allows for transfer and incorporation of untargeted genetic material [88]. The potential exchange of MGEs harboring anti-CRISPR protein encoding genes could propagate through bacterial communities rendering native or deliverable CRISPR-Cas systems inactive.

Inhibition of CRISPR-based antimicrobials by anti-CRISPR proteins is among the most noteworthy obstacles that would prevent successful implementation of this technology, especially in genomes that harbor prophages. As research in the anti-CRISPR area advances potential new ways of utilizing anti-CRISPRs can be explored. It could be possible to couple anti-CRISPR genes to inactivate native systems of pathogenic bacteria while delivering exogenous systems to increase the efforts of targeting. Future research focusing in both CRISPR and anti-CRISPR will unlock new potential for CRISPR-Cas based technologies and applications.
1.7 CONCLUSIONS AND PERSPECTIVES

The accumulation of research that is currently available has demonstrated the powerful potential that CRISPR-based antimicrobials have as a novel tool in addressing the rise of multi-drug resistant bacteria. Investigation into utilization of both type I and type II based targeting provides a strong basis for employing either system. The ability to package and delivery these systems via phagemid has shown to achieve significant killing of bacteria \textit{in vitro} as well as \textit{in vivo} models. Investigation of potential DNA repair mechanisms via SOS response of introduction of NHEJ systems to bacteria provide a better understanding for active ways bacteria can survive chromosomal damage. Research has also shown the notable issues to be aware of when choosing and designing chromosomal targets. Excision of pathogenicity and genomic islands within a sub-population of a bacterial species could allow for the selection and survival of unwanted populations of bacteria. Multi-plexing of delivered targets could potentially be a solution for avoiding sub-populations that lack some targets of interest due to genomic remodeling. This also brings to light that even delivery of a CRISPR array containing one target occasionally leads to homologous recombination between repeats rendering the plasmid inactive. This is also an issue when it comes to point mutations in \textit{cas} genes or guide RNAs delivered with the plasmids. Increasing the amount of targeting plasmids delivered may maximize the possibility that bacterial cells received a fully functional CRISPR array. Mutations in native systems responsible for crRNA processing or interference could be a potential issue if utilizing endogenous systems to carry out targeting. This issue could possibly be avoided by utilizing multiple native systems, if available in the desired bacterium of interest. Design and stability of synthetically constructed CRISPR arrays and deliverable plasmids can be fine-tuned to help ensure successful targeting in cells receiving these constructs.
Finally, as the discovery of beta-lactamases served as a small but significant cautionary tale, the discovery of anti-CRISPR proteins should be acknowledged. Incorporation of known anti-CRISPR proteins with antimicrobial research can help better prepare researchers to design more efficacious sequence-based antimicrobials. Although this area of research is still growing further understanding of these proteins can only help add to the collective knowledge of CRISPR-Cas system biology and applications. Regardless of targeting escape or inhibition of Cas protein activity CRISPR-based antimicrobials offer promising platform for combating the spread and treatment of multidrug resistant bacterial pathogens.
CHAPTER 2: OUTCOMES OF CHROMOSOMAL TARGETING UTILIZING NATIVE CRISPR-CAS SYSTEMS IN STREPTOCOCCUS THERMOPHILUS.
2.1 ABSTRACT

CRISPR-Cas systems provide adaptive immunity against phages in bacteria via DNA-encoded, RNA-mediated, nuclease-dependent targeting. This machinery, most notably, has been repurposed in eukaryotes for precision genome editing. These systems hinge on Cas effector nucleases that drive sequence-specific DNA targeting and cleavage. This is accomplished by endonucleolytic double-stranded DNA breaks created by Cas9 or exonucleolytic degradation of ssDNA by Cas3. Due to inefficient and relatively limited DNA repair pathways in bacteria, CRISPR-Cas systems can be repurposed for lethal DNA targeting, as an innovative form of precision antimicrobials. Our objective was to compare the relative killing efficiencies of endogenous Type I and Type II CRISPR-Cas systems in the model organism *Streptococcus thermophilus* DGCC7710. Specifically, we aimed to determine whether endogenous targeting via Cas3-mediated DNA damage would elicit different effects than Cas9-directed cleavage. Targeting of lacZ was carried out by engineered plasmids containing CRISPR arrays with self-targeting spacers repurposing either the native Type I-E or Type II-A systems. We observed efficient killing by both systems, in a dose-dependent manner, when delivering 0.4-400ng of plasmid DNA. We next characterized survivors from both targeting strategies, via targeted PCR screening and genome sequencing, to determine the genetic basis enabling survival. Results showed that evasion of Type I-E self-targeting was primarily the result of low-frequency targeting-defective plasmids that lacked the targeting spacer, likely generated via deletion by homologous recombination between CRISPR repeats. In contrast, Type II-A survivors revealed mutations in the chromosome resulting in a loss of the target sequence. Surprisingly, we observed the excision of a large (37.4 kb) genomic segment. The genesis of the deletion presumably occurred via homologous recombination between internal conserved sequences (206
bp identical segments) in two UDP-glucose-4-epimerase coding regions. These results exhibit the potential for repurposing of endogenous CRISPR-Cas systems for screening of rare genomic remodeling events and provide insights into the development of efficacious sequence-specific antimicrobials.
2.2 INTRODUCTION

The ability of bacteria to alter their genomes by acquisition, loss, and high rates of mutation drive rapid evolution and enable adaptation to various niches and environmental conditions [89]. This genetic elasticity is likely a key contributor to their ubiquitous occurrence and global perseverance. An increasingly alarming example of their adaptive nature is the rapid surge of multi-drug resistant bacteria as a consequence to the widespread overuse of antibiotics [90]. Prevalence of antibiotic resistance has extended beyond hospitals and patients with connecting waterways combined with extensive use of antibiotics in agriculture [91] and animal husbandry [20]. This has significant human health and environmental implications [92] given that the broad-spectrum nature of antibiotics does not discern pathogenic bacteria from beneficial bacteria. Alterations to the naturally occurring human microbiome population due to stressors (i.e. antibiotics) cause taxonomic shifts that result in a variety of health-related issues [93]. These imbalances inhibit the symbiotic relationship of beneficial microbes with the host immune system to prevent colonization of pathogens in humans and livestock [94, 95]. For example, exposure to antibiotics cause rapid microbial community profile changes and can increase susceptibility to bacterial infections like C. difficile [96]. Consequently, the proliferation of antibiotic-resistant bacteria requires new and precise means of removing disease-causing pathogens from beneficial consortia.

Bacteria have evolved an adaptive immune system to endure persistent assaults from exogenous biological agents, which has proven to be a remarkable platform for developing various molecular tools. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) - Cas (CRISPR associated sequences) systems provide acquired immunity in bacteria and archaea [37]. These DNA-encoded systems keep a genetic vaccination record of sequences from previous
encountered harmful elements termed “spacers”, that are derived from harmful invading DNA or RNA from phages or plasmids [44]. These CRISPR arrays are expressed and processed into mature crRNAs (CRISPR RNAs) [97] that guide signature effector Cas nucleases to elicit nucleic acid damage [49]. The chromosomal presence of these spacer sequences avoid self-targeting through a unique two-factor authentication process. This requires both pairing of the CRISPR spacer with the complementary target protospacer, and the presence of an adjacent sequence defined as the protospacer-adjacent-motif (PAM) [53, 57, 66]. The presence and recognition of these two elements are required for the Cas nuclease to cut DNA [57].

The diversity of CRISPR-Cas systems has allowed for a wide breadth of applications encompassing genome editing in bacteria [98] and eukaryotes [34], along with strain typing and vaccination of bacteria against viruses and plasmids [99]. Although these systems provide protection by destroying foreign nucleic acid sequences, there are various types [36] that accomplish this strategy in mechanistically distinct ways [57]. Introduction of an engineered guide or CRISPR array for chromosomal targeting can result in edited genomes via imperfect DNA repair pathways, although, the primary outcome is cell death [76-78, 80]. CRISPR-Cas based killing of bacteria can be achieved by reprogramming endogenous systems or delivering exogenous cas genes with an engineered guide via transformation or engineered phagemid.

Given the various means by which different CRISPR-Cas systems confer interference, it is important to investigate the various outcomes of self-targeting to determine which systems could serve as more effective antimicrobials. The Type II CRISPR-Cas systems are characterized by the endonuclease, Cas9, which utilizes a dual guide RNA (tracrRNA::crRNA) for target identification [47, 100]. Interaction with the PAM by Cas9 initiates binding by forming a guideRNA:targetDNA complex which is established at the seed sequence. This results in strand
displacement and genesis of an R-loop, which positions the RuvC and HNH nickase domains to generate a blunt double stranded DNA break [56]. The Type I CRISPR-Cas systems method of interference is distinct from the Type II-A system, while the requirements for proper interference remain equivalent with the reliance of the guideRNA:target complex and the presence of a PAM. In contrast, the Type I system method of interference relies on a multi-subunit surveillance complex known as Cascade (CRISPR associated complex for antiviral defense) in which the crRNA resides [52, 67, 101]. The signature protein of the Type I systems is the Cas3 nuclease-helicase which causes exonucleolytic degradation of ssDNA from the 3’ to 5’ direction [52]. R-loop formation of the crRNA:target heteroduplex and strand displacement stabilizes Cascade which allows for recruitment of the Cas3 nuclease to access the ssDNA substrate [68]. The distinct mechanistic differences in interference between these two systems posit that a Type I system could potentially act as a more potent molecular tool than a Type II system. To compare these two systems, it is best to use a well-characterized organism in which reprogramming of endogenous, active systems can be achieved.

The lactic acid bacterium *Streptococcus thermophilus* is widely used in the dairy industry as a starter culture for yogurt and cheese production [102, 103], and several commercial and model strains have been characterized [104-107]. The *S. thermophilus* evolutionary adaptation to milk has occurred through significant gene loss, yet this species remains genetically similar to pathogenic streptococci [104]. Horizontal gene transfer events have significantly influenced the *S. thermophilus* genome, as well as the presence of mobile genetic elements (MGEs) creating genome variation among strains [105]. Though the malleability of genetic material has influenced the development of the *S. thermophilus* genome, the genetic stability of this species is maintained as a result of its homeostatic milk environment [104]. However, these genomes still
contain a high presence of MGEs suggesting that these genomes may be unstable and subjected to molecular perturbations [79]. Therefore, we were interested in using *S. thermophilus* DGCC7710 as a model strain, utilizing previously characterized endogenous CRISPR-Cas systems to determine the impact of chromosomal self-targeting events. This particular strain contains four active CRISPR-Cas systems across Types I, II, and III [108]. Previous studies utilizing native systems observed the outcomes of self-targeting with either Type I [75, 76] or Type II systems [77, 80]. Our specific interest lies in comparing and observing the consequences of self-targeting by the previously defined active Type I-E and Type II-A CRISPR-Cas systems [44, 52, 108]. Comparison of targeting and cell death observed should help elucidate the potential of killing enabled by these distinct and separate CRISPR-Cas systems.
2.3 MATERIALS & METHODS

2.3.1 Bacterial Strains and Growth Conditions

Bacterial strains used are listed in Table 2.6.1. Escherichia coli EC1000 was propagated aerobically at 37°C with aeration at 250 rpm (Thermo Scientific MaxQ 8000 incubator) in Luria-Bertani (Difco) broth or on solid LB medium with 1.5% (w/v) agar. Erythromycin (Fisher Scientific) and kanamycin (Fisher Scientific) at 150 µg/mL and 40 µg/mL, respectively, were used for antibiotic selection of E. coli transformants containing the recombinant plasmids.

Streptococcus thermophilus DGCC7710 (DuPont) was grown in aerobic static conditions at 42°C (Thermo Scientific Heratherm incubator) in modified Elliker(B-Elliker) broth which was supplemented with 1% beef extract (w/v) and 1.9% (w/v) β-glycerophosphate or on solid B-Elliker medium containing 1.5% (w/v) agar incubated anaerobically at 37°C. S. thermophilus transformed with pTRK669 and pORI28 or any targeting constructs were grown in the presence of 5 µg/mL erythromycin and 5 µg/mL chloramphenicol (Fisher Scientific) for antibiotic selection. All frozen stocks created or used in this study were kept at -80°C with a 15% glycerol v/v concentration.

2.3.2 DNA Isolation, CRISPR Array Construction, and Cloning

Genomic DNA, plasmid DNA, and PCR amplicons were purified using Qiagen Mini-Prep kit, Qiagen Gel Extraction kit, Zymo Research DNA Clean & Concentrator-5 kit, and Zymo Research Fungal/Bacterial DNA Miniprep kit, according to manufacturers’ specifications except for the DNA elution step; all elutions were done using pre-warmed (∼ 42°C) UltraPure distilled water (Invitrogen). Type II-A targeting plasmids was isolated from E. coli clones previously constructed by Selle et al ([79] and unpublished). All Type I-E targeting plasmid inserts were...
synthesized in gBlocks (Integrated DNA Technologies) in the following order: native leader sequence from the CRISPR4 locus, native repeat sequence from CRISPR4, spacer sequence to respective target, followed with another final native repeat sequence. The gBlocks were amplified via PCR which included BamHI (New England BioLabs) and SacI (New England BioLabs) restriction enzyme sites. The insert amplicons were then ligated into a BamHI/SacI digested pORI28 backbone and transformed via heat-shock into E. coli EC1000. Colonies were then screened via colony PCR to confirm the insert was present by generation of a 802 bp amplicon. A minimum of three colonies with the correct band size were chosen and grown overnight aerobically at 37°C in LB liquid medium, shaking at 250 rpm. Plasmids were isolated then performed and sent for DNA sequencing to confirm sequence integrity of engineered CRISPR arrays. The plasmids with the correct and intact constructed CRISPR arrays were then selected for self-targeting experiments in S. thermophilus DGCC7710. Previously constructed plasmid utilizing the native leader and repeat sequences from the Type II-A CRISPR3 locus of Streptococcus thermophilus LMD-9 (homologous to the Type II-A CRISPR3 locus in DGCC7710) containing protospacer targeting the 5’ end of the lacZ CDS was used for targeting in DGCC7710. The Type I-E lacZ protospacer was chosen by confining potential protospacers based on PAM sequences (AGG) within the lacZ CDS. Additional targeting plasmids ([79] and unpublished data, Table 2.6.1) for chromosomal targeting in LMD-9 were screened against sequences in DGCC7710 (Geneious) and targets with the compatible protospacer sequence and PAM sequences for CRISPR3 in DGCC7710 were used for further Type II-A targeting experiments. These plasmids, responsible for screening for GEI2 (prtS) and GEI3 (Cu ATPase; copper efflux) in LMD-9 were used along with plasmids targeting comS, xerS, dltA, RNA helicase gene, and an intergenic region, were validated and utilized for targeting. Targeting
plasmids used in LMD-9 resulted in screening of genomic islands or cell death, these spacers were chosen and used in DGCC7710 to determine if different outcomes would be observed. Type I-E targeting plasmids were created to target the cas3, CRISPR4 leader, CRISPR4 array, gyrB, and a chromosomal intergenic region. All plasmids previously constructed were sent for sequencing prior to transformations in *S. thermophilus* DGCC7710 to confirm sequence of the engineered arrays. All primers used for PCR amplification and sequencing are listed in Table 2.6.2 and Table 2.6.3.

### 2.3.3 Transformation of *Streptococcus thermophilus* DGCC7710

A frozen stock of *S. thermophilus* DCGG7710 previously transformed with the temperature sensitive helper plasmid pTRK669 (Table 2.6.1) was used to inoculate 5 mL of B-Elliker medium with 5µg/mL chloramphenicol to grow overnight at 42°C. The overnight culture (1%) was used to inoculate another 5 mL of B-Elliker the next day and left to grow overnight at 42°C. Competent cells were prepared the day of transformation by inoculating 2% of the subcultured DGCC7710 into 100 mL of B-Elliker and grown at 42°C. Penicillin G (Sigma), a cell wall weakening agent, was added to the cells once they reached an OD$_{600nm}$ (Thermo Scientific Spectronic 200) of 0.3 to a final concentration of 10 µg/mL, and the culture was grown for another two hours [109, 110]. Cells were then centrifuged (Thermo Scientific Legend X1R) at 3,220 x g for 10 minutes at 4°C and washed/resuspended with cold 10 mL of cold 3.5X SMEB electroporation buffer (1 M sucrose and 3.5 mM MgCl$_2$; vacuum filter sterilized with Thermo Scientific Nalgene Flow Filter 0.2µM) [111]. This wash step was performed two more times for a total of three washes. After the final wash step, cells were then centrifuged and resuspended in 1 mL of electroporation buffer for a concentrated working stock of competent cells. For
transformation of dosage-dependent targeting, 100 µL of competent cells were aliquoted into separate 1.5 mL microcentrifuge tubes (Eppendorf) containing 400 ng, 40 ng, 4 ng, and 0.4 ng of each targeting plasmid (Table 2.6.1) and the control vector pORI28. The cells with the plasmids were then gently mixed before transfer into 0.2 mm electroporation cuvettes. Electroporation was performed using the BioRad GenePulser Xcell at 2,500 V, 25 µF capacitance, and 200 Ω resistance and time constants were recorded. Electroporated cells were re-suspended with 900 µL of pre-warmed B-Elliker and then allowed to recover overnight at 42°C for ~12 hrs. The recovered cells were then plated on selective medium and incubated at 37 °C under anaerobic conditions. Colonies were enumerated after ~48 hrs of growth (Figure 2.7.1).

2.3.4 Carbohydrate Utilization Assessment

Elliker media without carbon sources (Elliker-CHO) was prepared using 2% w/v pancreatic digest of casein (Becton Dickinson), 0.5% w/v yeast extract (Sigma), 0.25% w/v gelatin (Sigma), 0.4% w/v sodium chloride (Fisher Scientific), 0.15% w/v sodium acetate (Fisher Scientific), and 0.05% w/v ascorbic acid (Sigma). Separate carbon sources were prepared in 25% w/v stock solutions of each glucose (Sigma), sucrose (Sigma), galactose (Sigma), and lactose (Fisher) dissolved in nuclease-free water (Invitrogen UltraPure Distilled water). Each sugar solution was then filtered through a 0.2µm syringe filter (Thermo Scientific Nalgene Syringe Filter). Frozen stocks of wild-type and mutant strains of DGCC7710 were inoculated into 5 mL of B-Elliker medium containing 5 µg/mL of each erythromycin and chloramphenicol (no antibiotic was used for growth of the wild-type strain) and incubated overnight at 42°C. A second transfer of the cultures was done the next day with Elliker-CHO supplemented with glucose (1% v/v final concentration) and cultures allowed to incubate overnight at 42°C. A clear
96-well Costar flat-bottom microtiter plate (Corning) containing Elliker-CHO supplemented with either glucose, galactose, sucrose, and lactose (1% v/v final concentration) along with B-Elliker and Elliker-CHO (200 µL media/well) was prepared for the growth experiments. The overnight cultures were then centrifuged at 3,220 x g, washed and resuspended with 1X phosphate buffered saline (PBS, pH 7.4; Gibco), and inoculated (4% inoculum) into each well of the prepared microtiter plate. The FLUOstar Omega (BMG LabTech) optical density plate reader was used to measure the OD_{600nm} to monitor growth on provided carbon sources for 24 hrs at 37°C. Samples were run in triplicate on the microtiter plate and two biological replicates were performed.

2.3.5 Pacific Biosciences (PacBio) Whole Genome Sequencing, Annotations, and Protein Predictions

Wild-type DGCC7710 was grown overnight at 42°C in 25 mL of B-Elliker. The culture was then divided into 5 mL aliquots and centrifuged at 4,694 x g for 10 min at 4°C. The culture supernatant was removed and cell pellets were kept frozen at -80°C. Frozen pellets were submitted to RTL Genomics (Lubbock, TX) for genomic DNA isolation, PacBio library preparation, and whole genome sequencing. Genomic DNA was extracted from samples using Qiagen’s MagAttract HMW DNA kit based on manufacturer’s specifications with the following exceptions. Samples were incubated at 37°C with shaking at 900 rpm for 1.5 hrs after adding the P1 buffer and lysozyme. The final DNA elution step was performed with 100 µL AE elution buffer. Samples were then quality checked using Quant-iT dsDNA Broad Range DNA kit (Thermo Fisher Scientific) on the Qubit Fluorometer 3.0 and Fragment Analyzer by Advanced Analytical Technologies using the High Sensitivity Large Fragment 50KB Analysis kit (Thermo Fisher Scientific). Next genomic DNA samples then were processed to create SMRTbell
Libraries using Pacific Biosciences protocol with the following modifications: samples were pooled in equimolar, total DNA used was 500 ng per sample, library adaptor ligations were extended overnight, and final elution was performed with 12 µL of EB buffer. Library quality check was performed using Qubit Fluorimeter 3.0 and Fragment Analyzer using High Sensitivity Large Fragment 50KB Analysis kit (Thermo Fisher Scientific). The library was then prepared for sequencing following Pacific Biosciences protocol for diffusion loading according to manufacturer’s specifications. A pre-extension time was added for 120 mins and the final loading was 6 pM.

Sequence assembly of PacBio reads was carried out using Canu v1.5 [112] with a reference genome size of 1.8 Mb. The assembly generated one full contig resulting in a complete genome closed sequence of 1,857,071 bp. This file was then annotated using the *S. thermophilus* DGCC7710 draft genome available from NCBI (NZ_AWVZ01000000) using Geneious version 11.0.2 [113]. Hypothetical proteins sequences were saved as fasta files and submitted to Phyre2 (Protein Homolog/analogY Recognition Engine V 2.0) for batch processing [114]. Several modeling predictions were obtained for each hypothetical protein, except for two in which premature termination codons were present within the amino acid sequence. Top model hits were selected from hit reports and listed in Table 2.6.6.

### 2.3.6 RNA Extraction and Transcriptional (RNAseq) Analysis

Frozen stocks of wild-type DGCC7710 and one mutant strain (NCK 2525) were grown in B-Elliker overnight at 42°C. The overnight cultures were inoculated into fresh B-Elliker (2% inoculum) and incubated at 42°C under aerobic condition. Once the cultures have reached log phase (OD$_{600nm}$ of 0.6-0.7), 10 mL aliquots of cells were centrifuged at 3,220 x g at 22°C for 5
minutes. Culture supernatant was then removed and cell pellets were then flash frozen in an ethanol-dry ice bath and stored at -80°C. The experiment was performed in two biological replicates. For RNA extraction, frozen cell pellets were then thawed and resuspended in 1 mL of TRI reagent (Thermo Fisher Scientific) and thereafter transferred into 1.5-ml bead beating conical tubes with 0.1-mm glass beads (BioSpec Products, Inc.), and cells were disrupted by six 1-min cycles (with 1 min on ice intermittently) with a Mini-Beadbeater 16 apparatus (BioSpec Products). RNA purification was executed using the Direct-zol RNA MiniPrep kit (Zymo Research) with on-column DNase I treatment followed by an additional Turbo DNase (Thermo Fisher) treatment of the eluted RNA. An additional purification step was performed using the RNA Clean and Concentrator-25 kit (Zymo Research). RNA quality and concentration were then analyzed using Agilent 2100 Bioanalyzer (Agilent Technologies) and Nanodrop 2000c Spectrophotometer (Thermo Scientific), respectively. The absence of genomic DNA was confirmed via PCR using S. thermophilus DCC7710 gene-specific primers. Library preparation and RNA sequencing were performed by the High-Throughput Sequencing and Genotyping Unit of the Roy J. Carver Biotechnology Centre, University of Illinois (Urbana-Champaign, IL). After rRNA removal (using a Ribo-Zero rRNA removal kit for bacteria; Illumina), library preparation was carried out using the TruSeq Stranded mRNA Sample Prep kit (Illumina). Single-read RNA sequencing was performed using a HiSeq 4000 ultrahigh-throughput sequencing system (Illumina) and the Illumina HiSeq 4000 sequencing kit version 1 (Illumina) with a read length of 150 nucleotides (nt). The raw reads were demultiplexed with the bcl2fastq conversion software (v 2.17.1.14; Illumina), trimmed for the adapter sequences, quality trimmed to remove sequence reads with an error probability threshold of 0.001 (Phred score, 30), and filtered to remove reads of 20 nt by using Geneious version 11.0.2 [113]. The quality of the reads was assessed by using
FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The resulting reads were then mapped to the *S. thermophilus* DGCC7710 complete genome (assembled form PacBio whole genome sequencing) by using the Geneious Mapper with default settings [113]. The RNA-sequencing coverage depths were calculated to be 616X to 700X, and expression level comparison was performed using the DESeq2 Package in Geneious with Manually assigned conditions setting the two wild-type biological replicates as the control expression levels. Differentially expressed genes were defined as having a log₂ ratio of ≤ -1 or ≥ 1 (p-value <0.05).

### 2.3.7 Scanning Electron Microscopy Imaging

Frozen stocks of wild-type DGCC7710 and one mutant strain (NCK2525) were used to inoculate 10 mL of B-Elliker broth and incubated at 42°C overnight. Next, 10 mL of fresh B-Elliker broth was inoculated with 2% of the overnight culture and grown at 37°C for 12 hours. Cells were harvested by centrifugation for 10 minutes at 2,500RPM and the culture supernatant was removed. Then, 10ml of 3% glutaraldehyde-tannic acid in 0.1M sodium cacodylate buffer (pH 5.5) was added to each tube to resuspend the cell pellet and the cell suspensions were stored at 4°C until processed. To prepare the bacteria for SEM processing, 3 mL syringes were fitted with 13 mm Swinney filter holders fitted with a 0.4µm Nucleopore filter. One mL of buffer was placed in each syringe, to which 10 drops of resuspended bacteria in fix were added and gently filtered through. The filters were removed from the holders and placed sample side down in a vial containing fresh buffer. Filters were washed with three 30-minute changes of 0.1M sodium cacodylate buffer (pH 5.5) and then dehydrated with 30-minute changes of cold 30%, 50%, 70% and 95% ethanol. The first 30-minute change of 100% ethanol was started cold and the samples warmed to room temperature. Dehydration was completed with two 30-minute changes of room
temperature 100% ethanol. Samples were critical point dried (Tousimis Samdri-795, Tousimis Research Corp) in liquid CO$_2$ and held for 10 minutes at critical point. Filters were mounted on stubs with double-stick tape and silver paint and then placed in a vacuum desiccator until coated. Samples were sputter coated (Hummer 6.2 sputtering system, Anatech USA) with 50Å Au/Pd. Bacterial cells were viewed using a JEOL JSM-5900LV SEM (JEOL USA). Images were acquired at a resolution of 1280 X 960 pixels at magnifications of 1,000x, 5,000x, 10,000x.

2.4 RESULTS

2.4.1 CRISPR self-targeting is lethal in S. thermophilus DGCC7710

To compare the efficiencies in CRISPR-Cas targeting using both native Type I-E and Type II-A systems, a varying range of plasmid concentrations was employed for the transformation experiment. The goal was to maximize the transformation efficiency by delivering enough of the targeting plasmid that at least one plasmid was transformed into each cell. Plasmid targeting *lacZ* utilizing either the native Type I-E (CRISPR4) or Type II-A (CRISPR3) CRISPR-Cas system and the control (pORI28) plasmid were transformed into separate aliquots of 100 µl of competent DGCC7710 cells starting with 400 ng of plasmid and decreasing by a factor of 10 down to 0.4 ng for three biological replicates. Plating of the transformed cells on erythromycin and chloramphenicol selective medium showed that the number of recovered cells exhibited a dose-response to the decreasing concentration of transformed plasmid (Figure 2.7.2). Survivors were observed from both the Type I and Type II targeting but only with the dosage of 400 ng of the targeting plasmid. No colonies were recovered on the antibiotic selective medium at the 0.4 ng dosage of plasmid transformations. Growth was observed on agar medium that did not contain antibiotic, indicating that the
competent cells were viable, and further reflecting the high efficiencies of the CRISPR-Cas targeting systems. The highest transformation efficiency was achieved when delivering 400 ng of the targeting plasmid, therefore additional targeting plasmids were delivered at the highest dosage.

Additional targeting experiments were performed using previously constructed plasmids ([79] and unpublished data) utilizing targeting spacers for *prtS*, copper efflux gene, *comS*, *xerS*, *dltA*, RNA helicase, and an intergenic region (Table 2.6.1). The spacers within each plasmid were analyzed against the DGCC7710 genome for the presence of a matching protospacer and correct PAM (NGGNG) to ensure targeting could be achieved with available constructs. For additional Type I-E targeting, plasmids constructed targeting the *gyrB*, *cas3*, CRISPR3 leader sequence, CRISPR3 array, and an intergenic region were used. All targeting using plasmids for Type II-A chromosomal targeting resulted in 100% cell death. Previously used targets for *prtS* and copper efflux resulted in surviving populations as a result of genomic island excision via recombination of IS elements in LMD-9 [79]; however, these genomic island excision events were not observed in strain DGCC7710. Targeting via the Type I-E CRISPR-Cas system largely resulted in cell death, although a total of 20 surviving colonies were recovered from targeting by *gyrB*, *cas3*, and the CRISPR3 leader (Figure 2.7.3).

### 2.4.2 Characterization of Surviving Populations

Surviving populations from the Type I-E targeting were observed at a much higher rate than the Type II-A targeting (Figure 2.7.2). In total, there were 30 surviving colonies recovered from *lacZ* targeting experiments; of which 27 surviving colonies were isolated from the Type I targeting whereas only 3 were isolated from Type II targeting. All colonies recovered were
grown in liquid B-Elliker containing erythromycin and chloramphenicol. Plasmid and genomic DNA were isolated for genotypic analysis of the surviving populations. All plasmids isolated from the Type I-E targeted survivors were used for PCR amplification and sequencing of the inserts containing the CRISPR targeting leader and repeat-spacer array. The results showed that all Type I-E plasmids were defective due to the absence of the targeting spacer and one repeat in the CRISPR array. This is suspected to have occurred due to homologous recombination caused by the palindromic nature of the Type I-E repeats which is necessary to form the hairpin structure in the mature guide RNA. PCR of plasmids recovered from the Type II-A targeted survivors contained intact targeting plasmids where no mutations were observed in the leader, repeats, or spacers on the plasmid (Figure 2.7.4). Further genotyping via PCR of the genomic DNA isolated from all survivors revealed that all type I-E targeted cells still contained intact lacZ. In contrast, all type II-A survivors no longer contained the lacZ gene (Figure 2.7.5).

Out of the 20 survivors observed from the other Type I-E targeting experiments, 16 were a result of a defective plasmid previously observed with the Type I-E targeting of lacZ. Although the other four surviving colonies (three from gyrB and one from CRISPR3 leader targeting) contained complete and correct CRISPR arrays within their respective plasmids (Figure 2.7.4). PCR amplicons of targeted regions in the genome were sequenced. Sequence analysis showed that there were no mutations in the protospacer sequences [66] nor any mutations in the PAM sequence [53] (data not shown) which could lead to targeting escape. Further experiments are necessary to determine the specific cause of the faulty targeting utilizing the CRISPR4 locus which may encompass mutations in genomic sequences needed for successful interference.
2.4.3 Identification of Large Genomic Deletion

Previous research has shown that self-targeting can screen for sub-populations of remodeled bacterial genomes with spontaneous genomic excisions [75, 79] that remove the target sequence. *In silico* analysis did not reveal any homologous IS elements within a 50 kb region flanking the *lacZ* gene that could be responsible for spontaneous removal of *lacZ* via genomic island excision. Multiple primer sets for PCR screening were then designed to scan for the boundaries of the potential excision of *lacZ* from the genome (Table 2.6.3). PCR products obtained from the Type II-A survivors in comparison to products produced from the wild-type genomic DNA showed that deletion boundaries existed between primer sets B - C and J - K within the DGCC7710 genome (Figure 2.7.6). Analysis of the regions encompassing the deletion boundaries indicated the presence of a 206 bp conserved sequence within the coding sequences of two separate UDP-glucose 4-epimerase genes (*galE*). Overall the homology between these two *galE* CDS regions is 86%, indicating that a genomic deletion junction may exist at this 206 bp region. Primers flanking the *galE* genes were designed to produce a 1,470 bp product, and the amplicons were purified and sequenced to confirm the recombination event occurring between the two genes (Figure 2.7.7). Sequence data confirmed that a recombination event occurred at the 206 bp conserved region between the two *galE* genes, resulting in an in-frame recombination of the 5’ segment of the first *galE* (DGCC7710_RS0100105) with the 3’ segment of the other *galE* (DGCC7710_RS10700), creating a 37.4 kb genomic deletion. This resulted in the loss of *lacZ* in addition to twenty-eight genes (Table 2.6.4) with twelve of them coding for hypothetical proteins (Table 2.6.6). Interestingly, PCR with wild-type genomic DNA as a template also showed that this deletion occurs at a low level within the wild-type population. Therefore, this does not appear to be an active DNA repair result in the targeted survivors’ response to CRISPR-Cas self-
targeting. Rather, the CRISPR-based targeting enabled screening for this rare event occurring within the wild-type population.

### 2.4.4 Phenotypic Comparisons of Wild-type and lacZ-targeted Survivor Mutant Strains

Visual comparison of stationary-phase broth cultures sedimentation profiles showed that the mutant strain had a more porous sedimentation profile, compared to the more compact cell sediment of the wild-type strain (Fig. 2.7A). This observation prompted further investigation into phenotypic differences between the two strains. Light microscope images revealed that the mutant strain produced significantly elongated chains and lends itself to more severe aggregation than the wild-type strain (Figure 2.7.8). Images obtained via scanning electron microscopy showed additional detailed morphological differences such as inconsistent cell size, asymmetrical division, noticeably longer chain formation, and significantly larger cell aggregates in the mutant strain than observed in the wild-type (Figure 2.7.9). Based on carbohydrate API fermentation profiling, wild-type DGCC7710 is capable of metabolizing sucrose, lactose, and glucose as sole carbon source, whereas the mutants were unable to ferment lactose. Growth experiments were performed on the wild-type and all three mutants with semi-defined Elliker medium supplemented with 1% of either sucrose, lactose, glucose, or galactose available. Both the wild-type and mutant showed that they were unable to metabolize galactose. All strains were capable of utilizing glucose and sucrose. As anticipated the wild-type strain was able to utilize lactose as a carbon source but the mutant strains were unable to grow with lactose as the sole sugar source, which is consistent with the loss of the lacZ gene. The various phenotypic differences observed most likely are related to the several genes lost in the removed genomic region, including various hypothetical proteins.
2.4.5 RNA-seq Transcriptional Analysis

RNAseq was performed on RNA extracted from log-phase wild-type cells and one of the Type II-A lacZ targeting surviving mutant strain, NCK2525. Only one mutant strain was used for RNAseq since DNA sequencing confirmed that all three mutant strains contained the same deletion junction, yielding identical genomic deletions. RNAseq data for the wild-type and mutant was filtered and then mapped to the reference wild-type genome. The mapped reads showed that the mutant and wild-type strains overall had similar transcriptome profiles but clearly indicates the genomic deletion at the 206 bp conserved regions within the galE genes upstream and downstream of lacZ. An overlay image of the mapped reads clearly reveals the expression levels of the wild-type (blue) and the mutant strain (yellow) harboring the 37.4 kb deletion (Figure 2.7.11). All genes with a differential log₂ expression level of ≤ -1 and ≥ 1 are listed in Table 2.6.5. Coding regions where transcription was significantly decreased in the mutant are shown ranging from log₂ of -1 to -11 compared to expression levels in the wild-type. There is a noticeable difference in expression of the hypothetical proteins harbored within the genomic deletion as well, but some more significantly than others, due to their various transcript levels (Table 2.6.6). To provide more insights into their potential function, Phyre2 [114] protein modeling and prediction analysis was used to investigate potential homologs for the hypothetical proteins annotated within the deleted region (Table 2.6.6). The top protein models revealed that most of the hypothetical proteins showed homology to transferase proteins and possible associations with cell cycle functions. In silico analysis provides some insights into how the loss of these hypothetical proteins may have been involved in generating the phenotype of the mutants. Not all hypothetical proteins within the deletion region were expressed at high levels in the wild-type strain. This suggests that the loss of some proteins in the mutant, which otherwise
were highly expressed in the parent strain, could be contributing more significantly to the mutant phenotype than other proteins. Individual gene knockouts would need to be performed to further explain the link between gene loss and the differences between the mutant and wild-type strain. However, not all genes where transcription levels decreased were encased within the genomic deletion. In addition, other genes downregulated in the mutant despite not being present in the deleted region indicate potential for these genes to be associated with the mutant’s growth phenotype growth and morphological differences. There are also five genes that showed a two-fold increased expression in the mutant vs. the wild-type strain. Differential expression of several genes beyond the deletion region could also potentially be attributed by the loss of hypothetical protein(s) or other genes within the deletion region in the mutant strain.

2.5 DISCUSSION

In this study, the native Type II-A and Type I-E CRISPR-Cas systems within *S. thermophilus* DGCC7710 were repurposed for intentional chromosomal targeting to observe outcomes of self-targeting by these separate systems. Results from targeting using native Type I-E and Type II-A systems showed that the primary outcome of self-targeting was cell death which is in agreement with previous antimicrobial studies performed [76-78, 80]. Survival was largely achieved by delivery of a faulty plasmid with a defective insert where the spacer had spontaneously been excised. The deletion of the targeting CRISPR spacer from delivered plasmid has been observed previously by Gomaa *et al.* [76] and Citorik *et al.* [78] from which the majority of escapes arose. Other survivors appeared due to point mutations in the tracrRNA [78], transposon insertions into *cas9* [78], or deletion of *cas9* gene entirely [77] from recovered plasmids. A few colonies were recovered that contained intact targeting plasmids and no
significant point mutations which would result in faulty targeting. This suggests the possibility of loss of function of the CRISPR system [77], which is suspected to have resulted in survivors from this study.

Other self-targeting research has highlighted the genome variation of bacterial sub-populations due to MGES and putative AE-J mechanisms in response to DNA damage. Work done by K. Selle et al. [79] showed how native Type II-A CRISPR-Cas systems can be repurposed for screening mutant sup-populations harboring large genomic deletions occurring due to IS elements in *S. thermophilus* LMD-9 [79]. Efficient screening for these large spontaneous excision events were enabled due to *in silico* analysis of potential targets which would most likely lead to recovery of sub-populations harboring predicted excisions. Vercoe et al. observed in *Pectobacterium atrosepticum* recovery of a sub-population where an entire pathogenicity island, HA12, was lost after the repurposing the native Type I-F system to target the island. Other variable sized deletions were seen as a result of suspected AE-J in response to the Cas-mediated DNA damage [75]. These studies show how MGEs play a significant role in genome remodeling and underscore potential circumvention of sequence-specific targeting using CRISPR-Cas systems.

*S. thermophilus* strains are highly similar, and 1,271 conserved core genes have been identified across 47 characterized strains [106]. In contrast, there is enough variation to posit the possibility that not all self-targets will yield equal results among strains. *In silico* analysis of the DGCC7710 genome revealed that IS elements responsible for predicted large genomic deletions in LMD-9 did not maintain the same loci in the DGCC7710. Therefore, it was speculated that targeting of the non-essential genes could lead to measureable differences in killing efficiency between the two CRISPR-Cas systems. The *prtS* proteinase gene is necessary for accelerated
acidification of milk and is capable of being horizontally acquired by other *S. thermophilus* strains [115] suggesting its potential for excision. CRISPR-Cas based self-targeting of this gene revealed a 11,932 bp genomic island excision event in LMD-9 [79]. However, the pair of IS elements responsible for the genomic island were not conserved in the DGCC7710 genome as a result self-targeting in DGCC7710 of *prtS* resulted in 100% cell death. The lack of IS elements was also the case observed when targeting the copper efflux gene (Cu ATPase) in DGCC7710 resulting in cell death but revealing another genomic island in LMD-9 [79]. Though the outcomes of other targets were dissimilar, self-targeting of *lacZ* in DGCC7710 revealed a low-frequency deletion event which resulted in a 2% loss of the genome. The resulting surviving mutants from Type II-A targeting shows that large genomic deletions can occur via homologous recombination between conserved sequences of orthologous genes. The conserved sequences within the two *galE* genes roughly 3 kb upstream and 30 kb downstream of *lacZ* shows that large genomic deletions can occur that are not associated with MGEs. Surprisingly, despite the 37.4kb genome loss the mutants maintained their ability to survive in the presence of sucrose and glucose as the sole carbon source. The surviving mutant strains recovered exhibited clear morphological and phenotypic differences in addition to notable metabolic deficiency compared to the wild-type strain. Although DGCC7710 and LMD-9 share 99.2% pairwise identity self-targeting of *lacZ* by homologous CRISPR-Cas systems lead to distinct large deletions generated by separate mechanisms. Overall, targeting essential and non-essential regions of the DGCC7710 chromosome with either CRISPR-Cas system mainly resulted in cell death.

Spontaneous homologous recombination events can remodel bacterial chromosomes contributing the unpredictable plasticity of bacterial genomes [89]. Bacterial genome diversity not only exists between strains but their plasticity also contributes to low frequency variation
within strains. This variation of the DGCC7710 genome revealed by CRISPR-based targeting shows that circumvention of targeting can occur from remodeling of the genome via homologous recombination. *In silico* analysis of MGEs provide a certain level of predictability in determining which chromosomal portions are capable of excision and therefore enabling the design more efficacious targets. Other homologous recombination events between conserved sequences of nearby genes could potentially be predicted through *in silico* analysis although they may not all result in a surviving population.

Sequence-specific CRISPR-Cas directed targeting has been shown to be a powerful tool for bacterial cell death using either Type I or Type II systems. The results presented in this research in addition to other chromosomal targeting work showcases the powerful cytotoxic effects CRISPR-Cas targeting can induce. This work also features the different genotypic outcomes that can result from same-gene-same-species targeting with CRISPR-Cas. The constant remodeling of bacterial genomes [75, 79] and SOS response DNA repair mechanisms [80] exhibit various ways through which bacteria are capable of circumventing targeting. Collectively it has been shown that targets are not only species-specific but also strain-specific [76], emphasizing the targeting precision provided by CRISPR-Cas systems. Successful implementation of CRISPR-based precision antimicrobials hinges on the delivery of a complete and stable CRISPR array and the reliable presence of the chromosomal target to elicit the desired cytotoxic effect. Further research of the outcomes of self-targeting escape that may result continues to add to our understanding of bacterial genome plasticity which better enable researchers to design more efficacious antimicrobials. Overall, the potential for CRISPR-Cas targeting for combating multi-drug resistant pathogens provides a novel and potent solution to address this global issue.
2.6 TABLES

**TABLE 2.6.1 Plasmids and bacterial strains**

<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>pTRK669 (NCK1391)</td>
<td>This study</td>
<td>Cm&lt;sup&gt;R&lt;/sup&gt; helper plasmid for replication of pORI28 in DGCC7710</td>
</tr>
<tr>
<td>pORI28 (NCK1609)</td>
<td>This study</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; control vector for targeting transformation experiments</td>
</tr>
<tr>
<td>pORI28:: CR3 5' lacZ (pTRK1104)</td>
<td>This study</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of lacZ in DGCC7710 via Type II-A (CRISPR3) system</td>
</tr>
<tr>
<td>pORI28:: CR4 5' lacZ (pTRK1125)</td>
<td>This study</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of lacZ in DGCC7710 via Type I-E (CRISPR4) system</td>
</tr>
<tr>
<td>pORI28:: gyrB (pTRK1174)</td>
<td>This study</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of gyrB in DGCC7710 via Type I-E (CRISPR4) system</td>
</tr>
<tr>
<td>pORI28:: cas3 (pTRK1171)</td>
<td>This study</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of cas3 in DGCC7710 via Type I-E (CRISPR4) system</td>
</tr>
<tr>
<td>pORI28:: CR3 leader (pTRK1173)</td>
<td>This study</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of leader sequence of the CR3 locus in DGCC7710 via Type I-E (CRISPR4) system</td>
</tr>
<tr>
<td>pORI28:: CR3 array (pTRK1172)</td>
<td>This study</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of two portions of the CR3 array in DGCC7710 via Type I-E (CRISPR4) system</td>
</tr>
<tr>
<td>pORI28::Intergenic (pTRK1175)</td>
<td>This study</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of an intergenic region in DGCC7710 via Type I-E (CRISPR4) system</td>
</tr>
<tr>
<td>pORI28::prtS (pTRK1107) K. Selle et al., 2015 PNAS</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of prtS in DGCC7710 via Type II-A (CRISPR3) system</td>
<td></td>
</tr>
<tr>
<td>pORI28::CuATP (pTRK1108) K. Selle et al., 2015 PNAS</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of Cu ATPase in DGCC7710 via Type II-A (CRISPR3) system</td>
<td></td>
</tr>
<tr>
<td>pORI28::comS (pTRK1176) K. Selle et al., 2015 Unpublished</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of comS in DGCC7710 via Type II-A (CRISPR3) system</td>
<td></td>
</tr>
<tr>
<td>pORI28::xerS (pTRK1177) K. Selle et al., 2015 Unpublished</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of xerS in DGCC7710 via Type II-A (CRISPR3) system</td>
<td></td>
</tr>
<tr>
<td>pORI28::RNAhelicase (pTRK1180) K. Selle et al., 2015 Unpublished</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of RNA helicase in DGCC7710 via Type II-A (CRISPR3) system</td>
<td></td>
</tr>
</tbody>
</table>
**TABLE 2.6.1 Plasmids and bacterial strains** (continued)

<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>pORI28::Intergenic</td>
<td>K. Selle <em>et al.</em>, 2015</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of an intergenic region in DGCC7710 via Type II-A (CRISPR3) system</td>
</tr>
<tr>
<td>(pTRK1178)</td>
<td>Unpublished</td>
<td></td>
</tr>
<tr>
<td>pORI28::dltA</td>
<td>K. Selle <em>et al.</em>, 2015</td>
<td>Em&lt;sup&gt;R&lt;/sup&gt; targeting of dltA in DGCC7710 via Type II-A (CRISPR3) system</td>
</tr>
<tr>
<td>(pTRK1179)</td>
<td>Unpublished</td>
<td></td>
</tr>
</tbody>
</table>

**Bacterial Strains**

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em></td>
<td>EC1000</td>
<td>Cloning host for CRISPR-Cas Type I-E and Type II-A targeting plasmids</td>
</tr>
<tr>
<td><em>Streptococcus</em></td>
<td>DGCC7710</td>
<td>Wild-type strain</td>
</tr>
<tr>
<td><em>thermophilus</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Streptococcus</em></td>
<td>DGCC7710::pTRK669</td>
<td>Used for Type I-E and Type II-A targeting experiments.</td>
</tr>
<tr>
<td><em>thermophilus</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Cm<sup>R</sup>, chloramphenicol resistance marker; Em<sup>R</sup>, erythromycin resistance marker
<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’ to 3’)</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC.M13F-40.BamHI</td>
<td>CTGCTGGGATCCGGTTTCCCCAGTCACG</td>
<td>gBlock amplification</td>
</tr>
<tr>
<td>CC.M13R-27.Sacl</td>
<td>CTGCTGGAGCTCGGAACAGCTATGACCATG</td>
<td></td>
</tr>
<tr>
<td>CC.pORI28.Inst.Seq.F_1</td>
<td>GCCCTTTTCCTGAGCCGAT</td>
<td>Insert amplification and sequencing</td>
</tr>
<tr>
<td>CC.pORI28.Inst.Seq.R_1</td>
<td>TCGCTAGTTCTAGAGCGGCG</td>
<td></td>
</tr>
<tr>
<td>CC.LacZDGCC7710.F</td>
<td>GGTGATCGTTGAAGTGGACA</td>
<td>Confirmation of lacZ in DGCC7710 genome</td>
</tr>
<tr>
<td>CC.LacZDGCC7710.R</td>
<td>GGAAGTAGCAACCCCTTGAATG</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 2.6.3: Primer sets used to genomic deletion screen

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’ to 3’)</th>
<th>Primer Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC.7710.Interval.R1</td>
<td>GTGCCGATTGTACGGAGTTTCA</td>
<td>A</td>
</tr>
<tr>
<td>CC.7710.Interval.F1</td>
<td>GAAATCGTAACCCCTTTCACATGG</td>
<td></td>
</tr>
<tr>
<td>CC.7710.Interval.R2</td>
<td>GGTGGATGTGCCATTCATTTGT</td>
<td>B</td>
</tr>
<tr>
<td>CC.7710.Interval.F2</td>
<td>GATGGCTTTGGGATCTTTTCAGG</td>
<td></td>
</tr>
<tr>
<td>CC.7710.Interval.R3</td>
<td>TCGGTTCATCACATGTTGATCGC</td>
<td>C</td>
</tr>
<tr>
<td>CC.7710.Interval.F3</td>
<td>CATAGTCACGAAACGTTATGGCPTC</td>
<td></td>
</tr>
<tr>
<td>CC.7710.Interval.R4</td>
<td>CTGACTTGGCTACCTCGAATTGAAGC</td>
<td>D</td>
</tr>
<tr>
<td>CC.7710.Interval.F4</td>
<td>GACAAGGTGACATAGAAGACGTCG</td>
<td></td>
</tr>
<tr>
<td>CC.7710.Interval.R5</td>
<td>GGGACACCGTGATGAATCATTAC</td>
<td>E</td>
</tr>
<tr>
<td>CC.7710.Interval.F5</td>
<td>CCGATGCCAACGTTTGATAAGAAC</td>
<td></td>
</tr>
<tr>
<td>CC.7710.IS.F_B</td>
<td>CCTGCTAATTTCATCGCTCCAAC</td>
<td>F</td>
</tr>
<tr>
<td>CC.7710.IS.R_G</td>
<td>GTGTAGAATGCTGCTGACTATGAG</td>
<td></td>
</tr>
<tr>
<td>CC.7710.IS.F_C</td>
<td>AGCGGTAAGGATGACATCCC</td>
<td>G</td>
</tr>
<tr>
<td>CC.7710.IS.R_F</td>
<td>CCGCTGGAACTAGTAATTACCC</td>
<td></td>
</tr>
<tr>
<td>CC.7710.IS.F_D</td>
<td>CCCAGCATATCTGAAAGTCCCT</td>
<td>H</td>
</tr>
<tr>
<td>CC.7710.IS.R_E</td>
<td>GCATTATTGAGGTGCTTTGGGA</td>
<td></td>
</tr>
<tr>
<td>CC.7710.Interval.R6</td>
<td>GGTATTTCATCTGCTAGGGGC</td>
<td>I</td>
</tr>
<tr>
<td>CC.7710.Interval.F6</td>
<td>TTCAAGGCATCGCATCACATCC</td>
<td></td>
</tr>
<tr>
<td>CC.7710.Interval.R7</td>
<td>CGCCAATTACTGCTATTGTTT</td>
<td>J</td>
</tr>
<tr>
<td>CC.7710.Interval.F7</td>
<td>CGAGACAACAGTTATCATGGGACG</td>
<td></td>
</tr>
<tr>
<td>CC.7710.Interval.R8</td>
<td>GATGAAGACCGACATGCTAT</td>
<td>K</td>
</tr>
<tr>
<td>CC.7710.Interval.F8</td>
<td>GCTGATTGATAGCAGCCATCTCC</td>
<td></td>
</tr>
<tr>
<td>CC.7710.Interval.R9</td>
<td>GCCAAGTGAGGAGGAACCTAAAAGT</td>
<td>L</td>
</tr>
<tr>
<td>CC.7710.Interval.F9</td>
<td>GAGGTGGTAGCAAGGACTCCA</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 2.6.4: List of CDS contained within the genomic deletion region of the \textit{lacZ}-targeting survivor mutants

<table>
<thead>
<tr>
<th>Name</th>
<th>locus_tag</th>
<th>CDS Length (bp)</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS0108775</td>
<td>1485</td>
<td>forward</td>
</tr>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS0108780</td>
<td>345</td>
<td>reverse</td>
</tr>
<tr>
<td>oxidoreductase CDS</td>
<td>DGCC7710_RS10705</td>
<td>$&gt;473$</td>
<td>forward</td>
</tr>
<tr>
<td>transcriptional regulator CDS</td>
<td>DGCC7710_RS10710</td>
<td>654</td>
<td>reverse</td>
</tr>
<tr>
<td>cell wall surface anchor protein CDS</td>
<td>DGCC7710_RS0108795</td>
<td>456</td>
<td>reverse</td>
</tr>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS0108780</td>
<td>345</td>
<td>reverse</td>
</tr>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS10715</td>
<td>212</td>
<td>forward</td>
</tr>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS0108805</td>
<td>360</td>
<td>reverse</td>
</tr>
<tr>
<td>DNA-directed DNA polymerase CDS</td>
<td>DGCC7710_RS0108810</td>
<td>1416</td>
<td>reverse</td>
</tr>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS0108815</td>
<td>270</td>
<td>reverse</td>
</tr>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS0108820</td>
<td>189</td>
<td>reverse</td>
</tr>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS0108825</td>
<td>288</td>
<td>forward</td>
</tr>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS0108830</td>
<td>249</td>
<td>forward</td>
</tr>
<tr>
<td>threonine transporter RhtB CDS</td>
<td>DGCC7710_RS0108835</td>
<td>618</td>
<td>forward</td>
</tr>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS0108840</td>
<td>360</td>
<td>reverse</td>
</tr>
<tr>
<td>glycosyl transferase CDS</td>
<td>DGCC7710_RS0108850</td>
<td>2094</td>
<td>reverse</td>
</tr>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS0100010</td>
<td>270</td>
<td>forward</td>
</tr>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS0100015</td>
<td>300</td>
<td>forward</td>
</tr>
<tr>
<td>sodium transporter CDS</td>
<td>DGCC7710_RS0100020</td>
<td>912</td>
<td>reverse</td>
</tr>
<tr>
<td>hypothetical protein CDS</td>
<td>DGCC7710_RS0100030</td>
<td>323</td>
<td>reverse</td>
</tr>
<tr>
<td>SPASM domain-containing protein CDS</td>
<td>DGCC7710_RS0100035</td>
<td>576</td>
<td>reverse</td>
</tr>
<tr>
<td>transporter CDS</td>
<td>DGCC7710_RS0100035</td>
<td>1092</td>
<td>reverse</td>
</tr>
<tr>
<td>radical SAM protein CDS</td>
<td>DGCC7710_RS09880</td>
<td>1321</td>
<td>reverse</td>
</tr>
<tr>
<td>integrase CDS</td>
<td>DGCC7710_RS0100055</td>
<td>813</td>
<td>reverse</td>
</tr>
<tr>
<td>transposase CDS</td>
<td>DGCC7710_RS09885</td>
<td>508</td>
<td>reverse</td>
</tr>
<tr>
<td>ATP-dependent dsDNA exonuclease CDS</td>
<td>DGCC7710_RS0100080</td>
<td>3180</td>
<td>reverse</td>
</tr>
<tr>
<td>exonuclease sbcCD subunit D CDS</td>
<td>DGCC7710_RS0100085</td>
<td>1209</td>
<td>reverse</td>
</tr>
<tr>
<td>beta-galactosidase CDS</td>
<td>DGCC7710_RS0100090</td>
<td>3081</td>
<td>reverse</td>
</tr>
<tr>
<td>PTS sugar transporter subunit IIA CDS</td>
<td>DGCC7710_RS0100095</td>
<td>1905</td>
<td>reverse</td>
</tr>
<tr>
<td>galactose mutarotase CDS</td>
<td>DGCC7710_RS0100100</td>
<td>1047</td>
<td>reverse</td>
</tr>
</tbody>
</table>
### TABLE 2.6.5: Genes differentially expressed between DGCC7710 wild-type and Type II-A lacZ targeting survivor mutant

<table>
<thead>
<tr>
<th>Name</th>
<th>Differential Expression Log₂ Ratio (mutant/WT)</th>
<th>Differential Expression p-value</th>
<th>locus_tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta-galactosidase CDS*</td>
<td>-11.29607652</td>
<td>0</td>
<td>DGCC7710_RS0100090</td>
</tr>
<tr>
<td>PTS sugar transporter subunit IIA CDS*</td>
<td>-10.22120998</td>
<td>0</td>
<td>DGCC7710_RS0100095</td>
</tr>
<tr>
<td>hypothetical protein CDS*</td>
<td>-7.622437039</td>
<td>1.30E-256</td>
<td>DGCC7710_RS0108775</td>
</tr>
<tr>
<td>hypothetical protein CDS*</td>
<td>-7.059813313</td>
<td>2.66E-207</td>
<td>DGCC7710_RS0108780</td>
</tr>
<tr>
<td>ATP-dependent dsDNA exonuclease CDS*</td>
<td>-6.86657188</td>
<td>1.97E-195</td>
<td>DGCC7710_RS0100080</td>
</tr>
<tr>
<td>glycosyl transferase CDS*</td>
<td>-6.496121984</td>
<td>2.35E-168</td>
<td>DGCC7710_RS0108850</td>
</tr>
<tr>
<td>galactose mutarotase CDS*</td>
<td>-5.988919909</td>
<td>7.46E-137</td>
<td>DGCC7710_RS0100100</td>
</tr>
<tr>
<td>oxidoreductase CDS*</td>
<td>-5.840900467</td>
<td>1.32E-128</td>
<td>DGCC7710_RS10705</td>
</tr>
<tr>
<td>transcriptional regulator CDS*</td>
<td>-5.332831054</td>
<td>1.33E-101</td>
<td>DGCC7710_RS10710</td>
</tr>
<tr>
<td>UDP-glucose 4-epimerase CDS</td>
<td>-3.661232443</td>
<td>1.09E-101</td>
<td>DGCC7710_RS0100105</td>
</tr>
<tr>
<td>exonuclease sbcCD subunit D CDS*</td>
<td>-5.041338366</td>
<td>2.60E-88</td>
<td>DGCC7710_RS0100085</td>
</tr>
<tr>
<td>galactose-1-phosphate uridylyltransferase CDS</td>
<td>-2.931925228</td>
<td>9.52E-88</td>
<td>DGCC7710_RS0100110</td>
</tr>
<tr>
<td>galactokinase CDS</td>
<td>-2.887217429</td>
<td>1.24E-77</td>
<td>DGCC7710_RS0100115</td>
</tr>
<tr>
<td>hypothetical protein CDS*</td>
<td>-4.375786176</td>
<td>5.00E-62</td>
<td>DGCC7710_RS0108840</td>
</tr>
<tr>
<td>DNA-directed DNA polymerase CDS*</td>
<td>-3.125298692</td>
<td>1.44E-28</td>
<td>DGCC7710_RS0108810</td>
</tr>
<tr>
<td>radical SAM protein CDS*</td>
<td>-2.984831064</td>
<td>6.44E-26</td>
<td>DGCC7710_RS09880</td>
</tr>
<tr>
<td>UDP-glucose 4-epimerase CDS</td>
<td>-2.268091157</td>
<td>6.41E-25</td>
<td>DGCC7710_RS10700</td>
</tr>
<tr>
<td>sodium transporter CDS*</td>
<td>-2.643021805</td>
<td>2.71E-20</td>
<td>DGCC7710_RS0100020</td>
</tr>
<tr>
<td>transcriptional regulator CDS</td>
<td>-1.292312421</td>
<td>7.20E-20</td>
<td>DGCC7710_RS0100120</td>
</tr>
<tr>
<td>hypothetical protein CDS*</td>
<td>-2.341803521</td>
<td>3.93E-16</td>
<td>DGCC7710_RS0100010</td>
</tr>
<tr>
<td>hypothetical protein CDS*</td>
<td>-2.195150949</td>
<td>2.32E-14</td>
<td>DGCC7710_RS0108805</td>
</tr>
<tr>
<td>threonine transporter RhtB CDS*</td>
<td>-2.150700399</td>
<td>7.47E-14</td>
<td>DGCC7710_RS0108835</td>
</tr>
<tr>
<td>transporter CDS*</td>
<td>-1.922385042</td>
<td>1.86E-11</td>
<td>DGCC7710_RS0100035</td>
</tr>
<tr>
<td>SPASM domain-containing protein CDS*</td>
<td>-1.821733833</td>
<td>1.66E-10</td>
<td>DGCC7710_RS0100030</td>
</tr>
<tr>
<td>hypothetical protein CDS*</td>
<td>-1.356626696</td>
<td>7.35E-07</td>
<td>DGCC7710_RS0108830</td>
</tr>
<tr>
<td>hypothetical protein CDS*</td>
<td>-1.3251008</td>
<td>1.23E-06</td>
<td>DGCC7710_RS0108825</td>
</tr>
<tr>
<td>hypothetical protein CDS*</td>
<td>-1.053253641</td>
<td>5.26E-05</td>
<td>DGCC7710_RS0100015</td>
</tr>
</tbody>
</table>
TABLE 2.6.5: Genes differentially expressed between DGCC7710 wild-type and Type II-A lacZ targeting survivor mutant (continued)

<table>
<thead>
<tr>
<th>Name</th>
<th>Differential Expression Log$_2$ Ratio (mutant/WT)</th>
<th>Differential Expression p-value</th>
<th>locus_tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>phosphoribosyl-AMP cyclohydrolase CDS</td>
<td>1.140207942</td>
<td>5.94E-06</td>
<td>DGCC7710_RS0107575</td>
</tr>
<tr>
<td>Queuine tRNA-ribosyltransferase CDS</td>
<td>1.016563707</td>
<td>9.14E-07</td>
<td>DGCC7710_RS0106685</td>
</tr>
<tr>
<td>MFS transporter CDS</td>
<td>1.259627757</td>
<td>1.66E-07</td>
<td>DGCC7710_RS0104980</td>
</tr>
<tr>
<td>1-(5-phosphoribosyl)-5-((5-phosphoribosylamino)methylideneamino)imidazole-4-carboxamide isomerase CDS</td>
<td>1.112870743</td>
<td>2.35E-08</td>
<td>DGCC7710_RS0107585</td>
</tr>
<tr>
<td>phosphorylase CDS</td>
<td>1.004296444</td>
<td>8.20E-09</td>
<td>DGCC7710_RS0102135</td>
</tr>
</tbody>
</table>

*Indicates genes within the chromosomal deletion region.
### TABLE 2.6.6: Differentially expressed hypothetical proteins and Phyre2 results

<table>
<thead>
<tr>
<th>locus_tag</th>
<th>Differential Expression Log₂ Ratio (mutant/WT)</th>
<th>Top Model</th>
<th>Confidence (%)</th>
<th>Coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGCC7710_RS0108775</td>
<td>-7.622437039</td>
<td>Unknown function (PDB: upf0371 protein dip2346)</td>
<td>100%</td>
<td>52%</td>
</tr>
<tr>
<td>DGCC7710_RS0108780</td>
<td>-7.059813313</td>
<td>6-phosphogluconate dehydrogenase C-terminal domain-like</td>
<td>29%</td>
<td>17%</td>
</tr>
<tr>
<td>DGCC7710_RS10715*</td>
<td>-0.184979273</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>DGCC7710_RS0108805</td>
<td>-2.195150949</td>
<td>cell cycle/IMC sub-compartment protein</td>
<td>65.80%</td>
<td>52%</td>
</tr>
<tr>
<td>DGCC7710_RS0108815</td>
<td>-0.380093503</td>
<td>HPr-like protein</td>
<td>34.90%</td>
<td>11%</td>
</tr>
<tr>
<td>DGCC7710_RS0108820</td>
<td>-0.239509379</td>
<td>Structural genomics, de novo protein (PDB: 0r446)</td>
<td>13.40%</td>
<td>26%</td>
</tr>
<tr>
<td>DGCC7710_RS0108825</td>
<td>-1.3251008</td>
<td>Uncharacterized protein</td>
<td>19.70%</td>
<td>30%</td>
</tr>
<tr>
<td>DGCC7710_RS0108830</td>
<td>-1.356626696</td>
<td>S-adenosyl-L-methionine-dependent methyltransferases</td>
<td>89.70%</td>
<td>17%</td>
</tr>
<tr>
<td>DGCC7710_RS0108840</td>
<td>-4.375786176</td>
<td>Fibrinogen beta chain</td>
<td>17.50%</td>
<td>10%</td>
</tr>
<tr>
<td>DGCC7710_RS0100010</td>
<td>-2.341803521</td>
<td>Transferase/chitin synthase 1</td>
<td>40.60%</td>
<td>26%</td>
</tr>
<tr>
<td>DGCC7710_RS0100015</td>
<td>-1.053253641</td>
<td>delta-Endotoxin, N-terminal domain</td>
<td>69.70%</td>
<td>15%</td>
</tr>
<tr>
<td>DGCC7710_RS09875*</td>
<td>-0.825218817</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

*Phyre2 unable to populate results; Pre-termination codon in AA sequence.*
2.7 FIGURES

Figure 2.7.1: Diagram of experimental layout

This diagram illustrates the plasmids containing the engineered CRISPR arrays for self-targeting, native systems used, PAMs required for each system, antibiotic selectable markers, and experimental steps. Separate transformations of plasmids constructed utilizing the native CRISPR3 (Type II-A) and CRISPR4 (Type I-E) in *Streptococcus thermophilus* DGCC7710 were performed. Additionally, a transformation of the vector control, pORI28, was also performed as a non-lethal control to help determine reduction of recovered transformants.
Figure 2.7.2: Dosage responses of transformed targeting plasmids

Plasmid targeting *lacZ* utilizing either the native Type I-E or Type II-A CRISPR-Cas system and the control (pORI28) plasmid were transformed into competent DGCC7710 cells. Which plasmid was used is indicated by a + or -, in the case where neither the Type I or Type II plasmid was used the control vector was transformed. Each plasmid was transformed at 400, 40, 4 or 0.4 ng per 100 µl of cells. Transformed cells plated on B-Elliker (1.5% agarose) plates containing 5 µg/mL of each erythromycin and chlorophemicol showed that recovered cells exhibited a dosage response to the decreasing amount of plasmid delivered. Survivors were observed in both the Type I and Type II targeting but only with the dosage of 400 ng of the targeting plasmid. Colonies were not recovered at the 0.4 ng dosage of plasmid transformations; however, growth was observed on agar medium that did not contain antibiotic indicating that the competent cells were viable.
Figure 2.7.3: Type I-E and Type II-A targets at 400 ng

Additional targeting was performed and all targeting plasmids were delivered at a 400 ng per 100 µl of competent cells. Targeted regions are listed below with the control plasmid pORI28 indicated by (+). Several more survivors were observed in the Type I-E targeting; however, 100% killing was seen with the intergenic region and CRISPR3 (abbreviated as CR3) array targets. Complete killing was observed with this particular set of Type II-A targets there was 100% killing.
Figure 2.7.4: Genotyping of plasmids isolated from Type I-E and Type II-A targeting

PCR screening of plasmids from survivors recovered from targeting with different Type I-E targets revealed a mix of fully intact inserts and defective plasmids (A). All plasmids isolated from the Type I-E targeting experiments for lacZ have a defective insert missing the repeat and spacer sequence whereas the Type II-A targeting plasmids recovered from survivors contained the full insert (B). Sequencing of all plasmids isolated from surviving colonies confirmed that the majority of the Type I-E targeting survivors were a result of defective CRISPR-Cas targeting plasmids in which a repeat and the targeting spacer were missing from the insert (C).
Figure 2.7.5: PCR identification of chromosomal \textit{lacZ} deletion in targeting survivors

Genomic DNA (gDNA) was isolated from all of the Type II-A survivors along with one Type I-E survivor, wild-type DCGG770, and DGCC7710::pTRK669. Primers were designed to flank both Type I-E and Type II-A targeted regions of \textit{lacZ} to produce a 642 bp product from intact \textit{lacZ} gene. PCR with genomic DNA revealed that all Type II-A survivors did not produce a PCR product whereas all other genomic DNA samples produced the expected 642 bp product. Absence of a \textit{lacZ} PCR product suggests that a potential genomic deletion has occurred in the Type II-A surviving population.
Figure 2.7.6: PCR screening for genomic deletion boundaries encompassing lacZ region

Primers sets (labeled A through L) were designed to generate PCR products from genomic DNA ranging from 370 bp to 699 bp (blue bars) to scan for boundaries of the putative genomic deletion encompassing lacZ (pink). PCRs were performed on all genomic DNA from Type II-A survivors and genomic DNA from wild-type DGCC7710. All PCR products expected were observed from the wild-type DGCC7710 genomic DNA with expected product sizes. Primer sets C through J did not generate products in any of the PCRs performed on genomic DNA isolated from the survivors. It is suspected that the boundaries of the putative deletion exist between the regions defined by primer sets B to C and J to K of the DGCC7710 genome, regions shown by the black brackets.
**Figure 2.7.7: Identification of genomic deletion junction**

*In silico* analysis of the deletion boundaries within the wild-type (WT) genome showed that regions flanking *lacZ* each contain one UDP-glucose 4-epimerase gene (*galE*). Overall sequence identity of the two *galE* CDS shows only 87% similarity (79.2% N-terminal domain CDS; and 86.4% C-terminal domain CDS). However, a 206 bp region with 100% sequence identity exists between both *galE* sequences (A). Primers (green) were designed to produced deletion junction amplicons from genomic DNA isolated from surviving populations suspected to be generated by homologous recombination between the identical 206 bp regions in the *galE* genes (B). A 1,470 bp amplicon was recovered from the Type II-A survivors and WT genomic DNA (C) and sequenced which confirmed the recombination event occurring in the two *galE* CDS. The presence of the deletion junction amplicon form the WT genomic DNA indicates that this recombination event is occurring at a low frequency within the WT population.
Figure 2.7.8: Growth in broth and light microscope imaging

Liquid culture of the wild-type (WT) vs. the survivor mutant strain exhibited visually different sedimentation profiles (A). The WT growth in B-Elliker broth formed compact sedimentation, compared to the loose sedimentation of the mutant strain. Light microscope images were taken at 40x magnification and show that the DGCC7710 WT cells produces relatively short chains (B) whereas the mutant strain appeared as significantly longer chains and tends to aggregate (C). These traits were observed across multiple samples and are consistent.
Figure 2.7.9: Scanning electron microscopy (SEM) imaging of DGCC7710 and survivor mutant

SEM images of DGCC7710 (A) and a mutant strain (B) were taken at 1,000x, 5,000x, and 10,000x magnification. At 1,000x magnification it is clear that the mutant strain tends to form extensive cell aggregates in much larger groups than observed in the wild-type strain. Closer examination showed the differences in cell surface morphologies, cell size, and shape. The mutant strain exhibited abnormal, asymmetrical cell division along with inconsistent cell surface morphologies.
Figure 2.7.10: Carbohydrate growth experiments

Growth of DGCC7710 wild-type and the three lacZ-targeting survived mutant strains on different sugars as measured by OD$_{600\text{nm}}$ for a 24-hr period at 37°C. All strains were capable of utilizing glucose and sucrose and unable to metabolize galactose. All mutants lost the ability to utilize lactose as a carbon source for growth.
Figure 2.7.11: Overlay of RNA-seq transcriptome profiles of wild-type DGCC7710 and mutant strain

RNA-seq expression data from DGCC7710 wild-type and one of the mutant strains confirms the large genomic deletion suspected from \textit{in silico} and sequencing analysis. The recombination boundaries are observed at the 206 bp of homologous regions (represented by orange triangles) of the \textit{galE} coding sequences present in the mutant strain. The overall transcriptome profiles from each strain were similar with the exception of the 37.4 kb deletion region in the mutant strain. A minor peak of mapped reads from the mutant strain within the deletion region is representative of homologous sequence present in other parts of the genome.
2.8 REFERENCES


60. Pennisi, E., *The CRISPR craze*. 2013, American Association for the Advancement of Science.


70. Sinkunas, T., et al., *Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system*. The EMBO journal, 2011. 30(7): p. 1335-1342.


