ABSTRACT

STOUT, EMILY AE-HUI. CRISPR-Cas Targeting and Escape in *Lactobacillus gasseri*. (Under the direction of Dr. Rodolphe Barrangou & Dr. Todd Klaenhammer).

Lactobacillus gasseri is a lactic acid-producing commensal bacterium found in the genital and gastrointestinal tracts of healthy humans. Due to its long history of safe consumption, it is Generally Regarded As Safe (GRAS) and is frequently used as a probiotic, supported by a number of substantiated health benefits. One of the primary threats that face lactic acid bacteria, commonly utilized in the food industry for their probiotic or starter culture capabilities, are lytic bacteriophages that destroy the cell upon replication. However, multiple strains of *L. gasseri* have evolved to possess a natural defense mechanism known as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and associated proteins (Cas). CRISPR-Cas systems are a DNA-encoded, RNA-mediated adaptive immune system capable of directing sequence-specific cleavage of invasive mobile genetic elements (MGE) such as bacteriophages or plasmids.

Although CRISPR-Cas targeting is generally very efficient, MGEs can occasionally escape targeting by the CRISPR-Cas system due to mutations that occur in sequences essential to the CRISPR interference process in either the MGEs or the host chromosome. To better understand CRISPR escape processes, a plasmid interference system was used to screen for mutant cells with a target plasmid that circumvented CRISPR-Cas targeting in *L. gasseri* strains JV-V03 and NCK1342. Plasmids containing a target sequence, a protospacer adjacent motif (PAM), and an erythromycin resistance gene were transformed into both strains for targeting by the native CRISPR-Cas system in the host over three independent biological replications. Mutants that possessed plasmids that had escaped CRISPR-Cas

targeting were recovered to investigate the genetic mechanisms of escape and their occurrence.

Deletions in the CRISPR array of the host, resulting in the loss of the spacer targeting the plasmid, were the dominant pattern of escape in both strains. Analysis of sequence mutations revealed internal deletions in the CRISPR array, characterized by polarized excisions from the leader end that systematically included the leader-proximal targeting spacer. This establishes that internal deletions of targeting spacers within CRISPR arrays constitute a key escape mechanism whereby cells adapt to evade CRISPR targeting, while maintaining both the target sequence and the functionality of the CRISPR-Cas system. This is in contrast to the widely observed mutation of the target sequence, commonly seen in phage escape of CRISPR immunity. We speculate that these internal deletions may occur via homologous recombination between identical repeats within CRISPR arrays. The nuclease role of Cas1 in the integration of new spacers may also play a role in the excision of spacers. CRISPR-Cas has been a revolutionary discovery in both starter cultures and probiotics that could vastly improve the industry's ability to manage bacteriophages detrimental to fermentation and fermented foods. Understanding CRISPR targeting processes will advance the use of CRISPR-based technologies for food science applications such as bacterial genotyping, phage resistance, vaccination against plasmid uptake, antimicrobial activity, and genome editing.

CRISPR-Cas Targeting and Escape in *Lactobacillus gasseri*

by
Emily Ae-Hui Stout

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

Food Science

Raleigh, North Carolina

2016

APPROVED BY:

_____	_____
Todd R. Klaenhammer, Ph.D. Committee Co-Chair Biotechnology Minor Chair	Rodolphe Barrangou, Ph.D. Committee Co-Chair

_____
Suzanne Johanningsmeier, Ph.D.

DEDICATION

To my husband, Mark Stout, for his constant love and support. Thank you for always reminding me why I love science and for making me laugh every day. Also, to my parents, Donald and Kyong-Sun DeKam, for giving me every opportunity to succeed and for encouraging me every step of the way.

BIOGRAPHY

Emily Stout spent most of her childhood on ten acres in Fennville, MI with her parents and three siblings where she played in the woods during the summer and shoveled much snow in the winter. After graduating from Hamilton High School in 2009 as co-valedictorian, she attended Brigham Young University-Provo. There she discovered her love for travel and language through studying abroad in Mexico and Spain and participating in the Pembroke-Kings Programme at the University of Cambridge. Emily learned she had a passion for service and had the opportunity to serve as the vice-Executive and Executive Chair of Choose to Give, teach free English classes in Spain, and visit Zambia, Tanzania, and Zanzibar during work with an education-focused non-profit. She also threw herself into the food science program at BYU, where she enjoyed teaching food science labs and taking an active role in the food science club, serving as Secretary, Vice-President, and President.

After research and development internships at Land O'Lakes and General Mills, as well as a summer of directing a food science-based research program in Malawi, Emily knew that she wanted to pursue a research-focused path in the food industry. After graduating from BYU in May 2014 with a B.S. in Food Science and minors in chemistry, Spanish, and international development, she began graduate school at North Carolina State University in Raleigh, NC. Under the direction of Dr. Todd Klaenhammer and Dr. Rodolphe Barrangou, she is currently pursuing a Master's degree in Food Science with a minor in biotechnology. She lives in Raleigh with her husband, Mark Stout, who is also currently pursuing a master's degree in Food Science at NC State. Upon graduation, she hopes to continue to work in the food industry with a focus on research and development.

ACKNOWLEDGMENTS

There are so many people to whom I owe this experience at North Carolina State University and to whom I would like to extend my heartfelt thanks. My growth here would not have been possible without you.

- First, Dr. Todd Klaenhammer for believing in me enough to offer me a position in this truly amazing lab as one of your last few graduate students. Thank you for your thoughtful mentorship as well as your passion for science and teaching. I will never be able to sufficiently convey the amount of respect and gratitude I have for you.
- Dr. Rodolphe Barrangou for being willing to co-advise me. Thank you for adopting me as a member of the CRISPR lab on Centennial campus, guiding my research, and providing advice on my future career path as well as the science. I look forward to seeing the lab continue to flourish under your care and guidance. It has been an honor to be one of your first few graduate students.
- Dr. Suzanne Johanningsmeier for serving on my committee, providing constructive feedback, and reminding me to think about the food science applications of my research.
- Rosemary Sanozky-Dawes, Dr. Yong Jun Goh, Dr. Sarah O'Flaherty, Dr. Emily Henriksen, and Evelyn Durmaz for answering all my many questions and teaching me to be a better scientist. You make this lab a superb place to work. A special shoutout to Rosemary, who has rescued me more times than I can count—you are wonderful.
- Brant and Kiersten (honorary member) Johnson, Katheryne Daughtry, Kurt Selle, Allie Briner, Jeff Hymes, Courtney Klotz, Casie Cañez, Katelyn Brandt, Mia Theilmann, and Dr. Joakim Andersen for all being a part of the Klaenhammer/Barrangou lab family

during my time here. Thanks for providing stimulating scientific discussions, answers to my questions, and a listening ear for the occasional vent session. You have all played a role in making North Carolina truly feel like home. Allie, thanks for mentoring me during my time at P2 as well as making our Dairy Bar shifts much more bearable with your contagious enthusiasm. Brant and Kiersten, thanks for introducing me to CookOut and Harris Teeter, hosting many a TV show marathon, and letting me play with Link and Eve to my heart's content. Kat, thank you for being one of my first good friends in North Carolina. I'll fondly remember those nights of dying after TRX, cramming for microbiology tests, and giving each other excellent grad school pep talks.

- Rosalind and Chris Burns for making your house a home away from home. Ros, I can't express how grateful I am for your sage advice, sense of humor, and constant thoughtfulness. Thank you for the many amazing meals you have fed me over the past year and a half. You have a way of making everything around you beautiful and you are a huge source of inspiration to me. You and Chris are incredible friends to us.
- To my family for *always* being there when I need you. Jonathan, Jeremiah, and Evelyn, I couldn't be prouder to be your sister and I love you all so much. Dad, thank you for your love, patience, and dedication to making sure that all of us kids had an amazing childhood. Your constant support has meant everything to me. Mom, you are one of the strongest women I know and I owe you so much of who I am today. Thanks for being a wonderful teacher, mother, and friend. And finally, Mark, you inspire me every day to be a better person. Thank you for filling my life with so much happiness. I love you more with each passing moment, and with you, the future looks incredibly bright.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Clustered regularly interspaced short palindromic repeats ..... CRISPR
CRISPR-associated sequences ..... Cas
CRISPR associated complex for antiviral defense ..... Cascade
CRISPR RNA ..... crRNA
Precursor CRISPR RNA ..... pre-crRNA
Trans-activating CRISPR RNA ..... tracrRNA
Protospacer adjacent motif ..... PAM
Single guide RNA ..... sgRNA
Ribonucleoprotein ..... RNP

CHAPTER 1. CRISPR-Cas technologies and applications in food bacteria

Emily A. DeKam, Todd R. Klaenhammer, & Rodolphe Barrangou

1.1 Abstract

Clustered regularly-interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins form adaptive immune systems that occur in many bacteria and most archaea. In addition to protecting bacteria from phages and other invasive mobile genetic elements, CRISPR-Cas molecular machines can be repurposed as a toolkit for applications relevant to the food industry. A primary concern of the food industry has long been the proper management of food–related bacteria, with a focus on both enhancing the outcomes of beneficial microorganisms such as starter cultures and probiotics as well as limiting the presence of detrimental organisms such as pathogens and spoilage microorganisms. This review introduces CRISPR-Cas as a novel set of technologies to manage food bacteria and offers insights into CRISPR-Cas biology. This review primarily focuses on the applications of CRISPR-Cas systems and tools in starter cultures and probiotics, encompassing strain-typing, phage resistance, plasmid vaccination, genome editing, and antimicrobial activity.

1.2 Introduction

The widespread occurrence of bacteria in food products has resulted in a host of documented outcomes, from negative consequences like food-borne illness and food spoilage to positive benefits like gut health and food preservation. Unsurprisingly, many food matrices constitute ideal habitats for bacterial growth due to the presence of valuable substrates including simple carbohydrates, lipids, proteins, minerals, and vitamins needed for human nutrition (Papadimitriou, Pot, & Tsakalidou, 2015). Techniques in food processing and preservation such as salting, pickling, drying, and fermenting were recorded as early as 6,000 B.C. to extend food longevity and safety (Caplice & Fitzgerald, 1999; Fox, 1993). Concurrently, beneficial bacteria were cultivated as starter cultures through back-slopping techniques to create specific outcomes in a variety of fermented products such as wines, cheeses, yogurts, pickled vegetables, sourdough breads, and sausages. Over time, various microbes evolved to thrive in their respective food environments. More recently, food fermentations via lactic acid bacteria have become an important method of food preservation, highly valued for their desirable sensory attributes (Caplice & Fitzgerald, 1999; Douglas & Klaenhammer, 2010). These traditional methods of unwittingly managing both detrimental and beneficial bacteria were maintained and democratized over thousands of years.

In a turn of events, the 19th century witnessed advances in the field of microbiology that unraveled the role of bacteria in food and health and was a tipping point for our appreciation and understanding of the critical roles bacteria play in healthy and palatable foods. In the 1850s-60s, Louis Pasteur established the role of microorganisms in fermentations and pioneered a method to kill contaminating bacteria using heat via

pasteurization. Multiple microbes involved in foods were studied, isolated, and characterized. Specific strains could now be utilized in defined starter cultures to create more reliable fermentation outcomes in particular products (Caplice & Fitzgerald, 1999; Johnson & Klaenhammer, 2014). Shortly thereafter, Eli Metchnikoff (1907) proposed that some microorganisms, now termed "probiotics", could play a critical role in promoting gastrointestinal health, and increase longevity in humans. Furthermore, new preventative measures for processing and packaging were developed that minimized the presence or growth of pathogens and spoilage microorganisms. The need to consistently deliver a safe, quality product was essential to the food industry, and that could only take place through being able to control the microbiota of that product throughout the food supply chain. Thus began the battle for the modern food industry to optimize, cultivate, and protect the beneficial bacteria in various fermented products while also creating conditions that are bacteriostatic or bactericidal to detrimental pathogens and spoilage microorganisms. A variety of technologies have been implemented over the years to better understand the role these organisms play in food as well as to discover how to reduce or exploit their respective metabolic outcomes. Recently, advances in molecular biology and genomics have vastly improved the industry's ability to monitor and manage the detrimental and beneficial microbiota in food products.

One such recent scientific advancement is the discovery of clustered, regularly interspaced, short, palindromic repeats (CRISPR) and CRISPR-associated sequences (Cas). CRISPR biology found its roots in food microbiology, demonstrating its role in adaptive immunity against phages of *Streptococcus thermophilus*, the primary starter culture for yogurt manufacturing (Barrangou et al., 2007). In the past ten years, research on CRISPR has

grown exponentially, primarily due to the system's ability to perform specific DNA cleavage and its potential for use in genome editing. With descriptors such as the "CRISPR craze" and the "CRISPR revolution", the speed at which the scientific community has not only understood and characterized CRISPR-Cas systems natively, but also developed genetic tools for gene deletion, insertion, and regulation in non-native systems is unprecedented (Barrangou, 2014; Barrangou, 2015a; Pennisi, 2013). The potential for CRISPR-related applications in prokaryotes in the food industry is also enormous. This review presents key concepts of CRISPR-Cas biology, introduces CRISPR as a novel tool to manage food-related bacteria, and discusses a number of potential CRISPR applications in starter cultures and probiotics in particular and also in pathogens and spoilage microorganisms.

1.3 CRISPR-Cas overview and biology

CRISPR repeat-spacer arrays together with Cas proteins form DNA-encoded, RNA-mediated adaptive immune systems in bacteria and archaea that protect against phages and other invasive mobile genetic elements (MGEs) via DNA or RNA cleavage (Barrangou et al., 2007; Brouns et al., 2008; Hale, et al., 2008; Marraffini & Sontheimer, 2008). The repeat-spacer array consists of a leader sequence that acts as a transcriptional promoter followed by a series of conserved palindromic repeats ranging from 21-48 nucleotides that flank unique spacers (Kunin, Sorek, & Hugenholtz, 2007; Kupczok & Bollback, 2013; Makarova et al., 2015). Repeat-spacer array sequences were first observed and reported in *Escherichia coli* K12 in 1987, although the term CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) was not coined to describe them until 2002 (Ishino, Shinagawa, Makino, Amemura,

& Nakata, 1987; Jansen, van Embden, Gaastra & Schouls, 2002). Although similar sequences were seen in other bacteria and archaea many times over the following years, the purpose and origin of repeat-spacer arrays were unknown until 2005, when three groups reported similarities between spacer sequences and bacteriophage and MGE sequences (Bolotin, Quinquis, Sorokin, & Ehrlich, 2005; Mojica, Diez-Villasenor, Garcia-Martinez, & Soria, 2005; Pourcel, Salvignol, & Vergnaud, 2005). The spacer sequences found between the repeats were indeed derived from previous exposure of the cell to exogenous DNA elements, and within a year, it was proposed that CRISPR was involved in heritable cell immunity through inferences based on the amino acid composition and domain functions of the Cas proteins (Makarova, Grishin, Shabalina, Wolf, & Koonin, 2006). One year later, Barrangou et al. (2007) established that CRISPR confers adaptive phage resistance and that CRISPR-associated *cas* genes were an integral part of both vaccination and immunity. Marraffini and Sontheimer (2008) demonstrated soon afterwards that CRISPR provides DNA-targeting and could prevent plasmid uptake as well, solidifying the role of CRISPR in cell immunity to exogenous DNA and sparking intense interest in the RNA-mediated mechanisms driving CRISPR-Cas.

The evolution of CRISPR-Cas over time has led to the discovery of a diverse set of CRISPR-Cas systems, distinguishable from each other by their "signature" genes, method of precursor CRISPR RNA (pre-crRNA) processing, and the mechanism by which target DNA or RNA is cleaved (Makarova et al., 2011; Makarova et al., 2015). The latest classification system has defined six major types of CRISPR-Cas systems (I-VI) with 19 subtypes that fall under two main classes defined by the crRNA-effector complex responsible for DNA/RNA

cleavage during the interference step (Figure 1). Class 1 systems contain a multi-subunit complex that mediates immunity whereas Class 2 systems utilize a single protein to perform all interference functions of the complex (Makarova et al., 2015; Shmakov et al., 2015). Types I, III, and IV fall under the Class 1 category whereas Types II, V and VI fall under the Class 2 category (Makarova et al., 2015; Shmakov et al., 2015). Despite their differences, cell immunity to exogenous invasive materials is mediated across almost every type of CRISPR-Cas system via a fundamental three-stage process: *adaptation* through the incorporation of additional spacers into the repeat-spacer array, *expression* of the repeat-spacer array and the consequent processing of that array into CRISPR RNAs (crRNAs), and finally, *interference*, during which invasive target sequences are recognized and destroyed by the crRNA-effector complex (Figure 2).

1.3.1 Adaptation

Adaptation is a two-step process whereby the CRISPR-Cas system acquires novel spacer sequences through sampling foreign DNA and then implanting the new target sequence into its repeat-spacer array (Barrangou et al., 2007). When the target sequence is present within the foreign DNA, it is referred to as a protospacer, whereas once it has been incorporated into the repeat-spacer array, it is termed a spacer (Deveau et al., 2008). Adaptation is mediated by the almost universal Cas1-Cas2 complex, although in certain types of CRISPR systems other Cas proteins may also be involved. All six types of CRISPR-Cas systems contain the *cas1* and *cas2* genes with the exception of Type IV systems, which are often encoded in plasmids or genomes with no repeat-spacer array (Makarova et al., 2015).

The Cas1-Cas2 complex consists of a hexameric structure that is comprised of two Cas1 dimers flanking a Cas2 dimer (Nuñez et al., 2014). The spacer selection process is driven by the protospacer adjacent motif (PAM), a unique set of 2-4 nucleotides, that flanks the protospacer and marks it as a target sequence in Type I and Type II systems (Deveau et al., 2008; Horvath et al., 2008, Mojica, Diez-Villasenor, Garcia-Martinez, & Almendros, 2009). Not only is the PAM essential for the selection and binding of the protospacer, it also ensures that a distinction can be made between the host and foreign DNA due to the presence of the PAM in the target DNA but not in the CRISPR locus (Marraffini & Sontheimer, 2010). In Type I systems, the PAM is located at the 5' end of the protospacer; a binding pocket present in each Cas1 dimer has been theorized to play a role in PAM specificity as it interacts with sequence complementary to the PAM (Deveau et al., 2008; Wang et al, 2015). In Type II systems, the PAM is located at the 3' end of the protospacer (Barrangou, 2015b; Deveau et al., 2008). PAM specificity in Type II systems appears to be driven by Cas9 which associates with the Cas1-Cas2 complex to bind the PAM sequence to ensure proper selection of the protospacer (Heler et al., 2015; Wei, Terns, & Terns, 2015). Type V systems also utilize a PAM-dependent system for protospacer selection, although adaptation has yet to be characterized in these systems (Shmakov et al., 2015; Zetsche et al. 2015). Type III systems are believed to utilize a PAM-independent adaptation process, although the mechanism is unknown (van der Oost, Westra, Jackson, & Wiedenheft, 2014). Adaptation in Type VI systems has not yet been characterized.

Once a protospacer has been located via interactions with the PAM, the complex processes the foreign DNA substrate into spacer precursors of a particular size (Swarts,

Mosterd, van Passel, & Brouns, 2012; van der Oost et al., 2014). During this sampling process, crystal structures of the Cas1-Cas2 complex revealed that Cas2 is responsible for securing the dsDNA of the target sequence within the complex through forming a positively-charged arginine clamp with Cas1 that interacts with the negatively charged phosphodiester backbone of the DNA (Nuñez, Lee, Engleman, & Doudna, 2015). The spacer precursors and the Cas1/Cas2 complex localize to the repeat-spacer array, where Cas1 makes a staggered cut across the repeat closest to the leader end, creating two single-stranded repeats (Arslan, Hermanns, Wurm, Wagner, & Pul, 2014). The spacer is integrated and ligated between the two repeats, whereupon DNA repair machinery complements the repeats, finishing the adaptation process (Arslan et al., 2014; Nuñez et al., 2015). As adaptation is a polarized, chronological process, with new spacers always being incorporated at the leader-proximal end, the repeat-spacer array is a historical record of immunization events the cell has faced over time (Andersson & Banfield, 2008; Barrangou et al., 2013; Tyson & Banfield, 2008). Because the content of the repeat-spacer array determines the extent to which the cell is protected from invading MGEs, adaptation is a critical part of CRISPR cell immunity.

Additional biases in spacer selection can exist, such as "priming," in which pre-existing spacers influence the acquisition of additional spacers from the same target (Datsenko et al., 2012; Richter et al., 2014). Priming has been observed in Type I-B, Type I-E, and Type I-F CRISPR-Cas systems and requires the presence of Cascade (CRISPR-associated complex for antiviral defense) and Cas3 in addition to the Cas1-Cas2 complex usually required for adaptation (Amitai & Sorek, 2016; Datsenko et al., 2012; Sternberg, Richter, Charpentier, & Qimron, 2016). The concept was born when observations were made

that the presence of a spacer in the repeat-spacer array appeared to trigger the acquisition of additional spacers from the same strand of DNA of the same target, appearing as though the presence of the spacer sequence "primed" the target DNA strand for further acquisition events (Datsenko et al., 2012; Paez-Espino et al., 2012; Swarts et al., 2012). A study confirmed that a ten-fold bias existed towards acquisition of spacers from a primed strand versus non-primed strand (Savitskaya, Semenova, Dedkov, Metlitskaya, & Severinov, 2013). In short, priming is yet another evolutionary strategy that cells can adopt to add another layer of longer-term protection against phages or plasmids that can mutate to avoid CRISPR-Cas targeting (Amitai & Sorek, 2016; Fineran et al., 2014).

1.3.2 Expression

Expression is a multi-step process responsible for the biogenesis of crRNA: first, transcription of the repeat-spacer array occurs, followed by processing of the repeat-spacer array transcript into small mature crRNAs, and finally, the assembling of mature crRNAs into their crRNA-effector complex. The leader region of the CRISPR locus, located directly upstream of the repeat-spacer array, often contains an AT-rich promoter that is responsible for initiating transcription of the array. (Hale et al., 2012; Carte et al., 2014; Plagens, Tjaden, Hagemann, Randau, & Hensel, 2012). Typically, the array is transcribed into a long pre-crRNA, which is then processed into smaller mature crRNAs via cleavage of the repeat sequences by a Cas endonuclease or RNaseIII. The mature crRNAs each contain a spacer sequence that may be flanked by partial repeats on either side. Although the entire pre-crRNA is typically transcribed in one transcript, small RNA sequences have shown that the

abundance of individual crRNAs can vary across the locus (Zoephel & Randau, 2013). However, generally the abundance of crRNA gradually decreases from the leader-proximal end to the leader-distal end of the array (Zoephel & Randau, 2013; Carte et al., 2014). Spacers at the leader-distal end of the array are less likely to benefit the cell as they may protect against threats that no longer face the cell in its current environment; interestingly, in addition to being less transcribed, older spacers at the trailer-end of the array appear more likely to undergo deletions from the repeat-spacer array (Briner & Barrangou, 2014; Horvath et al., 2008; Horvath & Barrangou, 2011; Weinberger et al., 2012). Biologically, spacers closest to the promoter are more likely to be expressed due to the increasing likelihood of premature termination and RNA polymerase stalling as transcript size lengthens (Zoephel & Randau, 2013). From both a biological and evolutionary standpoint, the incorporation of new spacers at the leader-proximal end with a transcriptional promoter benefits the cell as newer spacers are thus more likely to be highly transcribed, allowing the cell to guard itself effectively against more recent threats. Once transcription of the repeat-spacer array has occurred, processing of the pre-crRNA into mature crRNAs varies, depending on the type of CRISPR system.

The Class 1 systems, Type I and Type III, are similar in their methods of processing pre-crRNA. Although Type IV also technically belongs to Class 1, expression has not been characterized in this system and will not be discussed here (Makarova et al., 2015). Type I and Type III systems both utilize a Cas6-like protein to process the pre-crRNA. Cas6 is a metal-independent endoribonuclease that generally contains two RAMP (repeat associated mysterious protein) domains that often play a role in RNA binding or cleavage (Rouillon et

al., 2013; Reeks, Naismith, & White, 2013; Niewoehner, Jinek, & Doudna, 2014). Due to the palindromic nature of the repeats, hairpin structures are often formed within the repeat sequences of the pre-crRNA. Cas6 hydrolyzes a phosphodiester bond at the 3' end of each repeat hairpin structure, resulting in crRNAs that have eight nucleotides on the 5' end, a complete spacer sequence, and then a repeat hairpin of variable size on the 3' end (Reeks et al., 2013; van der Oost et al., 2014).

Once the pre-crRNA has been processed into crRNA, the Cas6 variants of Type I-A, I-B, and I-D systems as well as Type III systems transport them to their respective effector complexes where trimming of the 3' hairpin end of the crRNA occurs (Carte, Wang, Li, Terns, & Terns, 2008; Hatoum-Aslan, Samai, Maniv, Jjiang, & Marraffini, 2013; Sashital, Jinek, & Doudna, 2011; Staals et al., 2013). However, the Cas6 variants of Type I-E and Type I-F share close affinity with the crRNA, and once they transport the crRNA to the effector complex, they remain tightly bound to the 3' end hairpin structure and become an integral part of the effector complex (Niewoehner et al. 2014, van der Oost et al. 2014). In Type I-C systems, a Cas5 variant is used to cleave pre-crRNA instead of Cas6; however, Cas5 is a RAMP protein similar to Cas6 and the end result is very comparable, with the exception of there being eleven nucleotides on the 5' end instead of eight. Similar to Type I-E and I-F systems, the Cas5 protein also closely associates with the crRNA and becomes a part of the effector complex (Nam et al. 2012). Although Cas5 is present in other Type I systems, its primary role consists of participating in the effector complex through perhaps associating with the 5' end of the crRNA (Wiedenheft et al. 2011a). As Type I and Type III

systems are Class 1 systems, additional proteins needed to form the multi-subunit crRNA-effector complex are recruited to the complex at this point.

Class 2 systems, Type II, Type V, and Type IV not only differ from the Class 1 systems in their methods of processing pre-crRNA. Type II systems process pre-crRNA with the help of a transactivating-crRNA (tracrRNA) that is encoded near the Cas operon (Deltcheva et al., 2011). The tracrRNA contains sequence at its 5' end that is complimentary to the repeats in the pre-crRNA. The tracrRNA anti-repeat sequences associate with the repeats on the pre-crRNA, forming a structural module with six distinct features: the spacer, the lower stem, the upper stem, the bulge, the nexus, and the hairpins (Briner et al., 2014). While these features may vary in size and sequence across strains with Type II systems, the nexus and hairpins were shown to drive Cas9 specificity, and thus orthogonality between Type II systems (Briner et al., 2014; Esvelt et al., 2013). Furthermore, the bulge and nexus were shown to be key in recruiting and binding Cas9, the next step in processing pre-crRNA. The recognition (REC) lobe of Cas9 is responsible for recognizing and binding the crRNA:tracrRNA complex (Hirano et al., 2016; Nishimasu et al., 2014). At this point, RNase III, a housekeeping ribonuclease, processes the complex into individual crRNA-tracrRNA units by cleaving the repeat:anti-repeat section of each crRNA-tracrRNA unit (Deltcheva et al., 2011; Karvelis et al., 2013; van der Oost et al., 2014). A second unknown nuclease then trims the spacer sequence on the 5' end of the crRNA down to around 20 nucleotides, finishing the process (Deltcheva et al., 2011). Typically, the processed crRNA contains around 20-24 nucleotides of spacer sequence followed by approximately 22 nucleotides of direct repeat (Deltcheva et al., 2011; Chylinski et al., 2013). The Cas9 protein remains tightly

associated with the fully processed crRNA:tracrRNA complex, composing the single-protein Type II crRNA-effector complex (Deltcheva et al., 2011; Jinek et al., 2012).

Type V and VI systems were discovered relatively recently and expression has not been characterized in either type. However, it is known that the Type V-A system can process pre-crRNA without the aid of tracrRNA or RNase III using solely its signature protein Cpf1 (Fonfara, Richter, Bratovič, Le Rhun, & Charpentier, 2016, Zetsche et al., 2015). The mature crRNAs are 42-44 nucleotides long and contain first the repeat sequence (19 nt) followed by the spacer sequence (23-25 nt). The structure of the crRNA is relatively simple and contains only a single stem loop within the palindromic repeat sequence. It appears that the structural features, as well as certain nucleotides within the sequence of the stem loop, are key for Cpf1 recognition of the crRNA (Zetsche et al., 2015). Similar to the Type V-A systems, no putative tracrRNA sequences were discovered in the Type V-C systems; however, putative tracrRNA sequences have been discovered in several Type V-B systems and cleavage of target DNA was crRNA-specific and tracrRNA-dependent (Shmakov et al., 2015). Although putative tracrRNA sequences were also discovered in Type VI systems, no expression of the putative sequence was observed and crRNA maturation as well as target cleavage was found to be tracrRNA-independent processes (Abudayyeh et al., 2016; Shmakov et al., 2015).

1.3.3 Interference

The properly assembled crRNA-effector complex is responsible for interference, that is, the protecting the cell against invasive MGEs via sequence-specific recognition and

cleavage of target nucleic acid sequences. Interference occurs when the spacer sequence of the crRNA interacts with complementary sequences on target protospacers via Watson-Crick pairing; once the system recognizes the protospacer target, endonucleolytic Cas proteins then cleave the offending target sequence (Garneau et al., 2010; Gasiunas, Barrangou, Horvath, & Siksnys, 2012; Jinek et al., 2012; Sontheimer & Barrangou, 2015; Westra et al., 2012,). Due to similarities in their crRNA-effector complex, the Class 1 systems, Type I and III, will be discussed first, followed by the Class II systems, Types II, V, and VI. As interference has not been characterized in Type IV systems, it will not be discussed.

Interference in Type I-E systems has been well studied. The multi-subunit crRNA-effector complex, termed CRISPR-associated complex for antiviral defense (Cascade), involves Cas5, Cas6, Cas7, and the crRNA, as well as Cse1 and Cse2 to a lesser degree (van der Oost et al., 2014). The complex comes together to form a seahorse-shape in which Cas7 proteins form a helical backbone for the crRNA (Wiedenheft et al., 2011a). The complex surveys the cell for MGEs through searching for complementary DNA sequences (van der Oost et al. 2014).. Cse1 allows non-specific interactions with DNA during the surveillance process; however, studies have suggested that a flexible loop in Cse1 may be key in PAM recognition (Jore et al., 2011; Mulepati, Orr, & Bailey, 2012; Sashital, Wiedenheft, & Doudna, 2012; van der Oost et al. 2014). Once Cse1 recognizes a PAM sequence, signaling to the crRNA-effector complex that the DNA is non-self, the DNA duplex is destabilized, allowing the crRNA to access the target DNA (Sashital et al., 2012). The spacer sequence in the crRNA attempts to base pair with the target sequence. Exact complementarity between the seed sequence (positions 1-5, 7, and 8) of the crRNA and the target sequence are critical

(Zhao et al., 2014). Due to the manner in which Cas7 holds the crRNA, every sixth nucleotide is inaccessible for binding, meaning that mismatches can occur in that position without affecting interference (Fineran et al., 2014; Jackson et al., 2014; Mulepati et al., 2014). Mismatches outside the seed sequence may also be tolerated (Maier et al., 2013; Fineran et al., 2014). If the target sequence does not match the spacer sequence of the crRNA, the Cascade complex will detach and DNA cleavage will not occur. However, if sufficient base pairing does occur between the crRNA and the target sequence, an R-loop is generated which causes a major conformational change of the Cascade complex and the target DNA (Rutkauskas et al., 2015; Szczelkun et al., 2014; Westra et al., 2012). This appears to recruit or activate the signature Cas3 nuclease-helicase to the Cascade complex which associates with the loop in Cse1 tied to PAM recognition and subsequently drives progressive and extensive DNA degradation (Hochstrasser et al., 2014; Huo et al., 2014; van der Oost et al., 2014; Westra et al., 2012).

The subtypes of the Type III systems, III-A and III-B, utilize a complex of Csm and Cmr proteins respectively, although both complexes exploit the Type III signature gene *cas10*. The Csm and Cmr complexes share structural similarities to the Cascade complex, with Csm3 in Type III-A systems and Cmr4 in Type III-B systems acting as a helical protein backbone for the crRNA, much like Cas7 does in Type I systems (Spilman et al., 2013; Staals et al., 2013). Interference in Type III systems has also been shown to be uniquely dependent upon transcription across target sequences, thus granting the system some flexibility in terms of allowing conditional tolerance to non-self DNA elements such as temperate phages (Deng, Garrett, Shah, Peng, & She, 2013; Goldberg, et al., 2014). Unlike

Type I and Type II systems, Type III systems do not utilize PAMs to distinguish between self and non-self DNA during the interference process. Rather, Type III-A systems utilize a "tag" on the 5' end of the crRNA that consists of eight nucleotides of repeat sequence (Samai et al., 2015). If the tag binds to a template strand of DNA along with the crRNA, the cell recognizes the sequence as self and interference will not occur, thus preventing auto-cleavage of the CRISPR locus and the host chromosome. Cas10 may be involved in this process of DNA recognition and binding in Csm complexes (Hatoum-Aslan et al., 2014). Once the Csm complex recognizes a target sequence as foreign, studies have shown that it will then degrade the DNA (Marraffini & Sontheimer, 2008). It has been suggested that Csm6 (formerly named Csx1) may work in conjunction with Cas10 to perform DNA cleavage and that it may be recruited to the complex in a similar manner to Cas3 in Type I systems (Deng et al., 2013; van der Oost et al., 2014).

Although the Cmr complex in Type III-B systems may be less defined than the Csm complex, it was the first CRISPR-Cas crRNA-effector complex discovered that could perform RNA cleavage (Hale et al., 2009). The Cmr complex seeks out RNA that is complimentary to the crRNA sequence in the complex and cleaves the target RNA multiple times at six nucleotide intervals (Staals et al., 2013). The complex appears to utilize the Cmr4 protein to perform cleavage of the RNA (Benda et al., 2014; Ramia et al., 2014; Zebec, Manica, Zhang, White, & Schleper, 2014). Although it was initially presumed that the Type III-A Csm systems target DNA and Type III-B Cmr systems target RNA, recent studies have revealed a much more multifaceted story. The Type III-A Csm systems in both *S. thermophilus* and *Thermus thermophilus* have been shown to target and degrade RNA in a

manner very similar to that of the Cmr complex, using the Csm3 protein as its endoribonuclease (Staals et al., 2014; Tamulaitis et al., 2014). On the other hand, the Cmr complex in the Type III-B system of *Solfolobus islandicus* was shown to degrade DNA as well as RNA, making it the first CRISPR system to exhibit dual targeting (Deng et al., 2013; Peng, Feng, Feng, Liang, & She, 2015). Furthermore, it appears that in some systems that contain both types of Type III systems, such as *T. thermophilus*, that the Csm and Cmr complexes may be able to share crRNA (Staals et al., 2014). Needless to say, much remains to be explored in the interference processes of Type III systems.

Interference in Type II systems is distinct from Type I and Type III systems in the simplicity of its crRNA-effector complex. Other than the crRNA, the only other components are the tracrRNA and the Cas9 endonuclease (Barrangou, 2015b). The Cas9 protein is a bi-lobed protein; the recognition (REC) lobe is responsible for interacting with the crRNA:tracrRNA complex and surveying target DNA whereas the nuclease (NUC) domain is responsible for performing double-stranded cleavage of the target DNA via its HNH and RuvC nuclease domains (Jiang et al., 2015a; Nishimasu et al., 2014; Nishimasu et al., 2015). When Cas9 binds the tracrRNA:crRNA complex during the formation of the crRNA-effector complex step, it undergoes conformational changes that allows its REC lobe and NUC lobe to separate from one another and form a positively-charged channel for active target DNA surveillance (Anders, Niewoehner, Duerst, & Jinek, 2014; Jinek et al., 2014; Jiang et al., 2015a; Jinek et al., 2014; Nishimasu et al., 2014; Nishimasu 2015). The Type II interference complex surveys DNA through indiscriminately colliding with DNA sequences and scanning for PAM sequences (Sternberg et al., 2014). When found, the PAM sequence is

accommodated between the PAM-interacting (PI) domain and the Wedge (Wed) domain; amino acid residues in the PI domain then interact with the PAM sequence to verify its accuracy (Anders et al., 2014; Nishimasu et al., 2014; Nishimasu et al., 2015).

Once the PI domain recognizes and binds with the PAM on the non-complementary strand, destabilization of the target sequence occurs and R-loop formation is triggered (Jiang et al., 2016; Nishimasu et al., 2014; Sternberg et al., 2014). The strand of DNA containing the PAM sequence is displaced, whereas the strand containing the sequence complementary to the PAM is incorporated into the central channel of the Cas9 protein where it hybridizes with the crRNA (Jinek et al., 2014; Nishimasu et al., 2014). The target DNA sequence within the crRNA-effector complex is then interrogated from the 3' to 5' direction to ensure that it correctly complements the spacer sequence across the crRNA. Although up to six neighboring mismatches can occur with no consequence at the 5' end of the protospacer, if any mismatches are discovered in the first 8-12 base pairs (the seed region) of the protospacer, the crRNA-effector complex will detach from the sequence and no DNA cleavage will occur (Jinek et al., 2012, Semenova et al., 2011; Sternberg et al., 2014; Wiedenheft et al., 2011b). However, if the crRNA-effector complex can confirm sufficient complementarity, double-stranded cleavage of the target will occur via the NUC lobe of the Cas9 protein. Double-stranded cleavage of bacteriophage and plasmid DNA by Cas9 was first observed *in vivo* in 2010 (Garneau et al., 2010). Soon afterwards, it was determined that Cas9 was the only Cas protein necessary for Type II CRISPR interference and that cleavage activity of the protein relied on two domains, RuvC and HNH, that each cleave a site-specific location on opposite DNA strands (Gasiunas, Barrangou, Horvath, & Siksnys, 2012;

Sapranauskas et al., 2011). The HNH domain is responsible for nicking the strand within the crRNA-effector complex exactly 3 base pairs upstream of sequence complementary to the PAM (Jinek et al., 2012). Similarly, RuvC nicks the displaced strand 3 base pairs upstream of the PAM, generating a blunt-end cut (Jinek et al., 2012). It was soon shown that the Cas9 protein can be guided by an engineered guide-RNA that replaces the crRNA:tracrRNA complex thus simplifying the interference process to an easily programmable two-component system (Jinek et al., 2012). Unlike Type I and Type III systems which utilize multi-subunit complexes to degrade target DNA in multiple locations, the simplicity and specificity of interference in both native and engineered Type II systems make it an especially viable tool in the world of genome editing, which continues to drive much of the CRISPR research today.

Type V systems were identified in *Francisella* subspecies as early as 2013; however, only very recently has interference in Type V-A and V-B systems been characterized (Schunder, Rydzewski, Grunow, & Heuner, 2013; Shmakov et al., 2015; Yamano et al., 2016; Zetsche et al., 2015). Type V systems are similar to Type II systems in their exploitation of a PAM sequence, the use of a single protein in their crRNA-effector complexes, and the specific manner in which DNA cleavage occurs. However, several key differences remain. For instance, the PAM for Type V systems is located at the 5' end of the protospacer and is a T-rich motif whereas PAMs in Type II systems are generally G-rich and located at the 3' end (Zetsche et al., 2015). The Type V-B system appears to operate very similarly to Type II systems, cleaving DNA through using a tracrRNA:crRNA complex that interacts with its signature C2c1 endonuclease to seek out protospacer sequences that are

located near a PAM (Shmakov et al., 2015). Experiments have confirmed that sgRNAs could be designed for Type V-B systems through fusing the 3' end of the tracrRNA with the 5' end of the crRNA.

Interference in Type V-A systems, characterized by their signature Cpf1 protein, has been more thoroughly studied. The seed region of the Type V-A system's crRNA is generally found within the first five nucleotides of the 5' end of the spacer sequence, and no mismatches may occur in this area. The crRNA-effector complex interacts with the strand complementary to the spacer sequence in the crRNA while the non-target strand with the PAM sequence is displaced. Crystal structures have determined that the Cpf protein can adopt a bilobed or triangular architecture with a large, positively-charged central channel in which the crRNA and protospacer are bound (Dong et al., 2016; Yamano et al., 2016). Once the crRNA-effector complex has determined there is sufficient complementarity between the protospacer and the crRNA spacer sequence, it appears as though a novel nuclease domain may cleave the target strand whereas a RuvC domain cleaves the strand not containing the protospacer; evidence suggests that the cleavage of the non-target strand by RuvC may be a prerequisite for cleavage of the target strand by the novel nuclease domain (Yamano et al., 2016). The targeted strand complementary to the crRNA spacer sequence is cleaved after the 18^{th} base whereas the non-targeted strand containing the PAM sequence is cleaved after the 23^{rd} base (Zetsche et al., 2015). In essence, a staggered cut is created in the protospacer with a five nucleotide 5' overhang on the opposite end of the protospacer from the PAM and seed region. Type V-A systems offer an intriguing alternative to Type II systems in terms of genome editing. The absence of tracrRNA and the smaller size of Cpf1 in comparison to

Cas9 allow for smaller effector complexes that may be cheaper to manufacture and utilize in genome editing applications. Staggered DNA cleavage as opposed to blunt-ended cleavage may make forced-directional cloning a possibility. Finally, the fact that the cleavage site is located far from the PAM and seed sequence may improve homology-directed repair (HDR) through giving Cpf1 a second chance to possibly re-cleave and re-initiate HDR if it did not occur the first time (Fagerlund et al., 2015; Zetsche et al., 2015).

Type VI systems are characterized by their C2c2 signature protein, which targets ssRNA phages. The C2c2 protein is guided by a crRNA guide and cleavage occurs via two HEPN domains when the effector complex enounters ssRNA targets containing protospacers matching the crRNA sequence (Abudayyeh et al., 2016). The effector complex binds ssRNA at the target site, but cleavage location is dependent on secondary structure of the ssRNA with the cleavage site most likely occuring in exposed ssRNA loop regions (Abudayyeh et al., 2016). The seed region of the crRNA appears to be in the center of the crRNA rather than the 3'- or 5'- end, with double mismatches in the center region highly reducing the system's efficiency (Abudayyeh et al., 2016). Interestingly, it appears as though C2c2 and crRNA effector complex is activated when it successfully binds to a targeted ssRNA; once activated, the effector complex can thereafter cleave other collateral ssRNAs even when not specifically targeted. This promiscuous targeting of ssRNA suggests that the Type VI system could potentially play a role in dormancy induction upon infection by invasive genomes that could provide the cell with additional time to fight the infection (Abudayyeh et al., 2016).

1.3.4 Perspective

In short, CRISPR-Cas systems are highly diverse in both their CRISPR locus content and architecture, and the molecular mechanisms through which they provide immunity (Table 1). Present in around 84% of analyzed archaea genomes and approximately 45% of bacterial genomes, CRISPR-Cas presents a constantly evolving picture of strain evolution and divergence in the microbial world (CRISPRdb updated 8/05/2014; Grissa, Vergnaud, & Pourcel, 2007). A clear grasp of CRISR-Cas biology is invaluable to gaining a broader perspective of where the fields of micro- and molecular biology are headed; however, most importantly, an understanding of the biology is critical in the development of relevant tools and implementation of corresponding applications. The power of CRISPR-Cas relies not only on its ability to protect the cell and lend background to strains of interest, but also the manner in which it can be harnessed for practical genomic applications in industries across the board. The food industry, in particular, has much to gain from understanding the potential that CRISPR-Cas possesses in managing and influencing microbial populations in food production.

1.4 CRISPR-Cas applications in food bacteria: Starter cultures, probiotics, pathogens and spoilage organisms

Fermented foods are a staple in the modern diet, with milk, meat, cucumber, grains, and cabbage being the most common substrates. Starter cultures play an essential role in transforming these substrates into their respective products, including cheeses, yogurt, sausage, salami, pickles, and sauerkraut (Caplice & Fitzgerald, 1999). Other microbes play a

role as probiotics, defined as “live microorganisms, which when administered in adequate amounts, confer a health benefit upon the host” (FAO/WHO, 2002; Hill et al., 2014). Probiotics may be added as adjuncts in fermented products or may be available as dietary supplements. Starter cultures and probiotics are both beneficial microorganisms that play multiple roles in food processing, including 1) preservation of foods via the production of organic acids, hydrogen peroxide, and bacteriocins, 2) improving food safety through pathogen inhibition, 3) enhancing the nutritional value of the food product, 4) developing the organoleptic qualities of the food product, and/or 5) administering health benefits to the host via delivery of live organisms to the gut (Bourdichon et al., 2012; Hill et al., 2014). Although honed to an art for centuries (perhaps millennia), the fermentation processes utilizing these beneficial organisms are not perfect and issues such as product reliability/quality, culture purity and composition, and fermentation failures due to bacteriophage attack, commonly occur. Microbes that produce lactic acid, prevalent in both starter cultures and probiotics, have an especially high occurrence of CRISPR-Cas systems, with loci appearing in 62.9% of analyzed lactobacilli genomes and 77% of bifidobacteria genomes (Briner et al., 2015; Sun et al., 2015a). The distribution of CRISPR-Cas systems in these microbes and the diversity of these systems provide a historical view into phage-microbial ecosystems of large-scale fermentations. In addition, CRISPR-Cas can be a powerful tool in the management of fermentation processes with applications in strain-typing, phage resistance, plasmid vaccination, genome editing, and antimicrobial activity.

Several of these CRISPR-Cas applications, specifically strain-typing and antimicrobial activity, can be used not only to manage beneficial microorganisms, but also to track and

minimize the presence of detrimental microorganisms. Since the origins of food processing and agriculture, a continual effort has been made to reduce the effect of detrimental bacteria, due to the heavy toll they enact on the food industry and human health. Food pathogens are responsible for approximately 9.4 million foodborne illnesses and are associated with an annual economic cost of $77.7 billion in the United States alone (Scallen et al., 2011; Scharff, 2012). Food spoilage microorganisms are also responsible for a heavy financial burden as over a quarter of the world's food supply has been estimated to be lost through microbial activity alone (Hui in't Veld, 1996). In an industry where consumer awareness is high and perception is key, the food industry must not only struggle with the tremendous waste of resources and financial loss involved with recalling a food product should microbial contamination occur, but also with the damage done to their brand reputation due to the loss of consumer trust. All of these factors lead to an imperative need for the food industry as well as various federal agencies such as the US Department of Agriculture (USDA) and the Food and Drug Administration (FDA) to 1) develop effective monitoring systems to detect the presence of such organisms and 2) tools to combat them should they be present. CRISPR-Cas is a novel technology that could be exploited to manage these two challenges in particular with pathogens and spoilage microorganisms. In short, CRISPR-Cas offers the food industry a wide span of powerful applications that can be utilized to manage both beneficial and detrimental food bacteria (Table 2).

1.4.1 Strain-Typing

The adaptive nature of CRISPR-Cas allows for a dynamic picture of the evolutionary trajectory of a strain over time. As outside threats are encountered, novel spacers are added to the repeat-spacer array in a chronological manner at the leader-proximal end (Arslan et al. 2014, Barrangou et al. 2007, Barrangou et al. 2013). As spacer acquisition is primarily driven by the threats present at a particular time, location, or setting, spacer sequences within the repeat-spacer array of a host genome can lend important historical, geographical, and environmental insights into a particular strain (Andersson & Banfield, 2008; Tyson & Banfield, 2008). This dynamic, external-driven aspect of CRISPR-Cas makes it a powerful tool for not only identifying strains, but also for studying strain divergence and relatedness, exploring microbial ecology and evolution, lending epidemiological insights, and understanding population-level genotypes in complex environmental samples (Barrangou & Horvath, 2012; Barrangou & Dudley, 2016).

The starter culture industry has a strong need for a method that allows for accurate identification of strains. The ability to differentiate between strains is essential to ensuring that the correct strains are present within a particular starter culture and to optimizing the process of isolating and identifying potential new starter strains for use in the industry. A number of methods including pulsed field gel electrophoresis, repetitive element PCR, and 16S rDNA sequencing have been investigated with the end goal of being able to find a quick and cost-effective method of determining species and strain variation (Selle & Barrangou, 2015a). Strain-typing through the amplification and sequencing of CRISPR repeat-spacer arrays is a novel and effective tool that can be added to the toolbox. Obviously, the presence

of CRISPR-Cas in strains of interest is the biggest limitation to its utilization in this application. However, due to the unusually high occurrence of CRISPR-Cas in many fermentation microorganisms, the starter culture industry is situated in a particularly favorable position to utilize CRISPR-Cas in strain-typing applications (Barrangou & Horvath, 2012).

CRISPR-Cas has already successfully been used to type and characterize multiple industry-relevant microorganisms. In *S. thermophilus*, commonly used in cheese and yogurt production, repeat-spacer arrays were used to type strains and provide insights into the relatedness of various strains as well as their ecology (Horvath et al., 2008). Repeat-spacer arrays in conjunction with *cas* genes were used to characterize strains of *Lactobacillus casei* and *Lactobacillus paracasei*, microorganisms commonly utilized in dairy fermentations or as probiotics, showing that CRISPR contributed to genomic variation within the strains (Broadbent et al., 2012; Smokvina et al., 2013). CRISPR-based genotyping in *Lactobacillus rhamnosus*, another microorganism commonly found in artisanal and industrial dairy products, could differentiate between closely-related strains; furthermore, various sub-lineages were confirmed by distinct groupings in repeat-spacer array sequences that were shared by particular strains (Douillard et al., 2013). CRISPR-Cas has also been used in the typing of non-dairy starter organisms, such as *Enterococcus faecalis*, which is often used in fermented meats (Hullahalli et al., 2015). Analysis of spacer sequences in the *Lactobacillus gasseri* and the *Bifidobacterium* genus, well known for their commensal and probiotic roles, provided insights into the evolutionary development and genetic divergence of particular strains (Briner et al., 2015, Sanozky-Dawes, Selle, O'Flaherty, Klaenhammer, & Barrangou,

2015). Although CRISPR-Cas genotyping cannot be applied on a species level, it can be utilized to differentiate between two otherwise seemingly identical clonal strains. In short, CRISPR-Cas genotyping has been shown to be an effective method of identifying beneficial microbes in the food industry while simultaneously yielding valuable insights into their phylogeny, evolution, and ecology.

Due to the historical, geographical, and environmental insights that CRISPR-based strain-typing provides, it can also be a particularly useful tool in tracking detrimental microorganisms. When present within an organism, the speed and ease with which CRISPR-Cas subtyping can be performed make it an ideal candidate for strain identification in the event of a contamination issue or outbreak. The presence of spoilage microorganisms, common contaminants in many fermentation processes, could be monitored via CRISPR-Cas strain-typing (Barrangou & Dudley, 2016). For example, CRISPR-Cas was utilized to successfully type *Lactobacillus buchneri*, a spoilage microorganism that universally contains a Type II system and is commonly found in cucumber fermentations (Briner & Barrangou, 2014). Theoretically, CRISPR strain-typing could be used to monitor and track the presence and load of other spoilage microorganisms that contain CRISPR-Cas over the food production process, thus increasing the quality and shelf-life of the final product (Barrangou & Dudley, 2016).

CRISPR-Cas has also been shown to be useful in tracking and identifying food-borne pathogens. Salmonellosis is the most common cause of food-borne diarrheal disease, and non-typhoidal *Salmonella* infections are responsible for ~15,000 hospitalizations and 400 deaths annually in the United States (Voetsch et al., 2004). Extensive CRISPR-Cas

genotyping of *Salmonella* from both clinical isolates and food illness outbreaks has provided an increased understanding of the microevolution of the strains as well as valuable epidemiological insights (DiMarzio, Shariat, Kariyawasam, Barrangou, & Dudley, 2013; Fabre et al., 2012; Fricke et al., 2011; Liu et al., 2011a; Liu et al., 2011b; Pettengill et al., 2014; Shariat et al., 2013a; Shariat et al., 2013b; Shariat, Sandt, DiMarzio, Barrangou, & Dudley, 2013c; Shariat, Timme, Pettengill, Barrangou, & Dudley, 2014; Timme et al., 2013; Wehnes, Rehberger, Barrangou, & Smith, 2014). In addition to *Salmonella,* many other pathogens have also undergone CRISPR-based genotyping including *Campylobacter jejeni* (Kovanen, Kivistö, Rossi, and Hänninen, 2014), *Clostridium difficile* (Hargreaves, Flores, Lawley, & Clokie, 2014), *Corynebacterium diptheriae* (Mokrousov et al., 2009), *Enterococcus faecalis* (Lindenstrauss et al., 2011), *Enterococcus faecium* (Tremblay, Charlebois, Masson, & Archambault, 2013), *Erwinia amylovora* (Rezzonico, Smits, & Duffy, 2011), *Escherichia coli* (Jiang, Yin, Dudley, & Cutter, 2015b; Yin et al., 2013), *Legionella pneumophila* (D'Auria, Jimenez-Hernandez, Peris-Bondia, Moya, & Latorre, 2010), *Microcystis aeruginosa* (Kuno, Yoshida, Kaneko & Sako, 2012), *Mycobacterium tuberculosis* (Groenen, Bunschoten, van Soolingen, & van Embden, 1993), *Propionibacterium acnes* (Brüggemann, Lomholt, Tettelin, & Kilian, 2012), *Streptococcus agalactiae* (Lier et al., 2015), *Staphylococcus aureus* (Kinnevey et al., 2013), *Vibrio parahemolyticus* (Sun et al., 2015b), *Xanthomonas aeruginosa* (Semenova, Nagornykh, Pyatnitskiy, Artamonova,& Severinov, 2009), and *Yersinia pestis* (Riehm et al., 2012). In short, CRISPR-Cas subtyping, used on its own or in conjunction with existing typing tools, can provide additional discriminatory power in the identification of specific pathogenic

outbreak isolates. Whether tracking spoilage microorganisms or identifying the source of food pathogens, CRISPR-Cas strain-typing can be utilized to help further ensure the safety and quality of food products.

1.4.2 Phage Resistance

Phages are the most abundant biological entities on earth, outnumbering bacteria 10 to 1 (Edmond & Moineau, 2007). As lytic phages replicate via the destruction of their respective bacterial hosts, it is no surprise that they present a challenge to fermentation processes. In an effort to minimize the economic losses that each phage infection entails, the food industry has come up with multiple ways to limit phage exposure including the modification of factory design, sanitation measures, ventilation systems, process protocols, starter culture mediums, and culture rotation procedures (Marcó, Moineau, & Quiberoni, 2012). Yet, phage infection of starter cultures remains the number one reason for slow or failed fermentations in the industry (Marcó et al., 2012). Phage infection has furthermore been shown to negatively impact the quality of the end product in terms of taste and texture (Samson & Moineau, 2013). Phages are often present in raw materials and cannot be eliminated without modifying the functional properties of that material (Samson & Moineau, 2013). Furthermore, many phages are resistant to various treatments including thermal, high pressure, ionizing radiation, and pasteurization, making it nearly impossible to eliminate them entirely (Samson & Moineau, 2013). As working with sterile raw materials and a sterile working environment is not practical in most fermentation processes, the most practical long-term solution appears to be utilizing phage-resistant microorganisms and hoping that the

arms race will play out favorably. Cells have evolved to rise to this challenge, with many containing natural defense mechanisms against bacteriophages, including various phage adsorption defense strategies, restriction-modification systems, and abortive infection systems (Djordjevic, O'Sullivan, Walker, Conkling, & Klaenhammer, 1997; Durmaz & Klaenhammer, 2007; Higgins, Sanozky-Dawes, & Klaenhammer, 1988; Labrie, Samson, & Moineau, 2010; Sanders & Klaenhammer, 1981; Sanders & Klaenhammer, 1983). CRISPR-Cas is yet another resistance mechanism that cells have developed to protect the cell from invasive DNA elements. Given the predominance of phages in fermentation processes, the high occurrence of CRISPR-Cas systems in fermentation microorganisms is not surprising.

The adaptive nature of CRISPR-Cas denotes its ability to face new phages, acquire a spacer, and be immune to that phage in any future encounters. So far, there is no observed limit to the number of new immunization events that can occur, although internal deletions of ancestral spacers may occasionally occur (Barrangou et al., 2013; Levin, Moineau, Bushman, & Barrangou, 2013; Weinberger et al., 2012). Strains of interest that contain native CRISPR-Cas systems may be iteratively exposed to phages that pose a risk. When a new spacer is incorporated that matches the phage with which it was challenged, the strain is effectively vaccinated against any future encounters with that phage or any other related phages carrying the same protospacer sequence. This process has been successfully demonstrated in *S. thermophilus*, with researchers exposing a strain to four iterative rounds of phage exposure. Four different lytic phages were utilized, and after each round, screening for phage-insensitive mutants was performed to ensure that spacer acquisition occurred. By the end of the fourth round, the strain was immune to all four phages, with four new spacers

incorporated at the leader ends of their repeat spacer arrays, which definitively demonstrated for the first time the adaptive immune function of CRISPR (Barrangou et al., 2013). The beauty, and perhaps power, of CRISPR-Cas adaptive immunity is that it is a natural process that already takes places in many fermentation organisms. It is a relatively simple process to prompt the system to adapt new spacers and no engineering must take place to exploit the immunity-bestowing effects of natural CRISPR-Cas systems. With that being said, however, not all strains contain CRISPR-Cas. Although more complex, it is also possible to engineer CRISPR-Cas adaptive immunity in valuable starter strains that lack it.

Unsurprisingly, CRISPR-Cas systems are rare in lactococcal species due to their activity against plasmids (Millen et al., 2012). *Lactococcus lactis*, an organism frequently utilized in dairy fermentations, possesses a number of plasmid-encoding milk adapting genes that have allowed it to thrive in milk environments (Millen et al., 2012). Furthermore, they already possess a number of mechanisms to defend themselves against phage infection, the majority of which are also plasmid-encoded (Coffey & Ross, 2002). Nonetheless, a plasmid-encoded, self-transmissible Type III CRISPR-Cas system was discovered in a strain of *L. lactis* that confers phage resistance (Millen et al., 2012). Spacers on the plasmid were observed to match phages common in the environment; however, it was also discovered that phage resistance could be engineered through direct introduction of a synthetically designed spacer. Interestingly, the system lacks a *cas2* gene and it is possible that the system is non-adapting, which would allow the lactococci to maintain plasmids beneficial to their survival. Furthermore, researchers were able to exploit this particular CRISPR-Cas system through transferring the plasmid to various other lactococcal strains, thus bestowing phage immunity

to other previously at-risk strains (Millen et al., 2012). The ability to transfer CRISPR-Cas immunity to lacking strains is valuable as it allows the industry to preemptively protect valuable starter strains that may lack CRISPR-Cas and are thus more vulnerable to a range of phage threats.

1.4.3 Plasmid Vaccination

In addition to targeting bacteriophage, CRISPR-Cas systems have also been shown to prevent the uptake of plasmids through cleavage of the DNA (Garneau et al., 2010; Marraffini & Sontheimer, 2008). In fact, the wide proliferation of CRISPR-Cas in *S. thermophilus* has been suggested to be responsible for the relative lack of plasmids within the strain (Barrangou & Horvath, 2012). In theory, plasmids all have a "Malthusian fitness burden" attributed to them which is passed on to the host cell due to the extra energy the cell must exert to replicate and propagate the plasmid; the greater this fitness burden, the greater the negative effect on the host cell (Levin, 2010). However, when a plasmid encounters a cell that contains CRISPR-Cas immunity, the cell samples a protospacer sequence in the plasmid and is subsequently vaccinated against that plasmid's uptake, hence the general evolutionary advantage gained by strains that are able to enact CRISPR-mediated plasmid vaccination. A second advantage gained from this sort of vaccination is that it limits the uptake of unwanted DNA elements such as pathogenicity islands or antibiotic resistance genes that are often transferred via plasmids (Edgar & Qimron, 2010; Nozawa et al., 2011; Palmer & Gilmore, 2010; Shimomura et al., 2011). While vaccination against such elements may happen

naturally, the CRISPR-Cas system can also be artificially engineered to target such genetic elements.

Of particular concern to the food industry is the spread of antibiotic resistance. It is possible that strains which contain antibiotic resistance genes may act as a reservoir for transferable antibiotic resistance elements; as such, surveillance of gene transfer of these antibiotic elements in the food chain is considered a priority in the safety demonstration of microorganisms (Bourdichon et al., 2012). The presence of CRISPR-Cas has been linked to a marked absence of antibiotic resistance markers in enterococci (Palmer & Gilmore, 2010). In the same way it is possible to engineer phage resistance in strains which contain native CRISPR-Cas systems, it is possible to engineer CRISPR-Cas targeting of antibiotic resistance genes contained in plasmids. This has been performed successfully in *S. thermophilus* through presenting the strain with a plasmid containing an antibiotic resistance gene and selecting for strains that lost the plasmid. Strains that had acquired a spacer that matched the antibiotic resistance gene were resistant to the uptake or dissemination of antibiotic resistance genes, as any sequences matching the spacer underwent CRISPR-Cas targeting (Garneau et al., 2010). The *S. thermophilus* CRISPR-Cas immune system has been shown to be effective in transferring heterologous immunity when transferred into *Escherichia coli* (Sapranauskas et al., 2011). Theoretically, the transfer of CRISPR-Cas systems into CRISPR-free strains could also effectively vaccinate other fermentation microorganisms against not only phages, but also specific undesirable elements such as antibiotic resistance genes that are often carried on exogenous DNA such as plasmids.

1.4.4 Genome Editing

Much of the publicity surrounding CRISPR-Cas comes from the ease with which Type II systems can be repurposed to act as a genome editing tool. Type II systems are preferred for genome editing over the other systems for a number of reasons including the simplicity of its interference complex as well as the sequence-specific manner in which blunt-ended, double-stranded cleavage of target DNA is performed (Doudna & Charpentier, 2014; Jinek et al., 2012). In order for genome editing to occur via CRISPR-Cas, four elements must be in place. The first element is Cas9, the effector protein that enacts interference. The second and third elements are the crRNA and the tracrRNA which form the crRNA:tracrRNA complex responsible for interrogating DNA for the target sequence and interacting with Cas9 (Briner et al., 2014). Finally, the fourth element is a DNA repair mechanism within the host organism to repair the damage enacted via CRISPR-Cas and incorporate the desired genomic changes (Selle & Barrangou, 2015b). In most cases, the first three elements will need to be delivered to the organism. Even if the organism contains a native Type II CRISPR-Cas with Cas9, it most likely does not contain a spacer that targets the exact location in the genome where editing is desired. A ground-breaking discovery in 2012 established that Cas9 could be guided by a single chimeric RNA molecule that essentially fused the 3' end of the crRNA to the 5' end of the tracrRNA, forming a single guide (Jinek et al., 2012). This molecule, later termed a single-guide RNA (sgRNA), revolutionized the field of genome editing as the sgRNA could easily be engineered to efficiently target any sequence in a genome of interest when delivered to an organism with the Cas9 protein. The discovery of the sgRNA greatly streamlined CRISPR-facilitated

genome editing through providing researchers with a simple two-component system—Cas9 and sgRNA—that was easy to program, package, and transfer into cells of interest.

The vast majority of CRISPR-based genome editing has taken place in eukaryotes (Barrangou & van Pijkeren, 2016; Sander & Joung, 2014). Typically, a sgRNA is designed to attack a specific sequence within the genome of interest. It is then delivered to the organism along with Cas9, where it enacts double-stranded cleavage at the desired location. Genome editing then takes place when the cell repairs the damage through utilizing endogenous DNA repair pathways such as non-homologous end-joining or homology-directed repair, which often enact mutations at the CRISPR-Cas cleavage site. However, when dealing with prokaryotes, the approach for using CRISPR-Cas as a tool for genome editing must account for bacterial physiology and DNA homeostasis mechanisms; for example, double-stranded cleavage of bacterial genomes most frequently leads to cell death due to the lack of robust endogenous DNA repair pathways (Selle & Barrangou, 2015b). This would make the delivery of sgRNA and Cas9 to the organism of choice an inefficient method of bacterial genome editing (Selle & Barrangou, 2015b). Rather, the way forward appears to be grounded in what is already known: first, that genome editing in bacteria should be driven primarily by recombination events, whether they occur naturally over time or intentionally via recombineering technology; and second, that CRISPR-Cas is an exceptionally suited tool for the screening and selection of low-frequency desired genotypes (Barrangou & van Pijkeren, 2016; Selle et al., 2015). The exploitation of CRISPR-Cas to select for specific genotypes has already been utilized in a fermentation organism, *S. thermophilus*, and a probiotic, *Lactobacillus reuteri* (Oh & van Pijkeren, 2014; Selle et al., 2015).

In *S. thermophilus*, CRISPR-Cas was used to select for naturally occurring genotypes in heterogeneous populations (Selle et al., 2015). It was found that four genomic islands exist in *S. thermophilus* that possessed no essential genes and were flanked by insertion sequences (Goh, Goin, O'Flaherty, Altermann, & Hutkins, 2011; Selle et al., 2015). Homologous recombination occurred at the insertion sites flanking these genomic islands, leading to natural excision of the islands at low levels. Through using CRISPR-Cas self-targeting, it was possible to select for specific genotypes that had excised these particular genomic islands. As *S. thermophilus* already encodes the *cas9* gene and tracrRNA, only the crRNA needed to be provided to the system. A spacer was designed that matched a specific sequence on the genomic island of choice and transformed into *S. thermophilus* on a plasmid along with the necessary repeat and leader sequences. CRISPR-Cas then enacted sequence-based removal of target-sequence genotypes through the targeting and killing of cells that contained the genomic island, and thus the protospacer, while sparing cells that had previously excised the genomic island. Once the desired genotype was obtained, in this case one without a particular genomic island, further testing could be done to characterize the genotype. For example, the *S. thermophilus* genotype lacking the fourth and largest genomic island at ~100kb which contained the *lacZ* gene for β-galactosidase, was found to no longer possess the ability to acidify milk, a key component of fermentation (Selle et al., 2015). Theoretically, CRISPR-Cas could be used to select for any number of naturally occurring genotypes to further explore their functionality in industry-relevant endeavors.

In *L. reuteri*, CRISPR-Cas was used to select for mutants that had undergone single-strand DNA (ssDNA) recombineering (Oh & van Pijkeren, 2014). Oh & van Pijkeren (2014)

describe a dual-step approach utilizing ssDNA recombineering and CRISPR-Cas targeting to maximize the number of recoverable recombinants in *L. reuteri*. First, the oligonucleotide is transformed into a strain of *L. reuteri* that contains a plasmid encoding *cas9* and tracrRNA as well as a plasmid encoding *recT* and is allowed to recover overnight. This step allows the oligonucleotide to be incorporated into the host genome and begin to propagate the mutation. Secondly, the transformant cells are made competent once again and this time plasmids containing crRNA designed to attack non-mutated cells are transformed into the cells. The crRNA associates with the tracrRNA and Cas9 targets cells that have not incorporated the mutation into their genome. Those that have incorporated the mutation no longer have a protospacer that matches the spacer sequence on the crRNA and thus escape CRISPR-based targeting. The combination of ssDNA recombineering and CRISPR-Cas was also shown to enable targeted codon mutagenesis as well as the recovery of very low-efficiency events (Oh & van Pijkeren, 2014). However, although ssDNA recombineering is a powerful tool, it has limited applicability as recombineering technology must be optimized for each individual strain. A focus on utilizing CRISPR-Cas to select for homologous recombination events may be more effective in furthering genome editing applications in bacteria. Nonetheless, the presence of CRISPR-Cas in genome editing applications greatly expands the ability of the industry to enact small- and large- scale mutations on bacterial genomes and recover them for further characterization. Not only will this capability allow the industry to better understand gene functionality within starter cultures and probiotics, it may also play a timely and effective role in the improvement of various strain properties to better suit industry needs.

1.4.5 Antimicrobial Activity

Due to prokaryotes often lacking robust endogenous DNA repair pathways, double-stranded cleavage of bacterial genomes via CRISPR-Cas is often deadly to the cell (Selle & Barrangou, 2015b). Although this lethality presents several interesting challenges in the arena of bacterial genome editing, it also allows for the use of CRISPR-Cas as an "antimicrobial." Essentially, CRISPR-Cas systems can be co-opted to enact lethal damage on particular bacterial populations through sequence-specific self-targeting of the genome, leading to the reduction of undesirable environmental populations. If a strain that contains a CRISPR-Cas system is provided with a self-targeting spacer sequence, CRISPR-Cas will destroy its own host's genome. The lethal effect of CRISPR targeting on bacterial genomes was first observed in 2013, showing that self-targeting of the host chromosome led to a non-reversible ~10^5 reduction in viable counts (Vercoe et al. 2013). There are two strategies to using CRISPR-Cas to specifically target the genomes of particular strains, and they depend on whether or not the strain intended for elimination contains a native CRISPR-Cas system. If the strain of interest contains a native CRISPR-Cas system, it can be eliminated through simply providing the cell with a plasmid that contains guide RNA with spacer sequence that matches a unique protospacer located on the bacteria's genome. If the strain of interest does not contain a native CRISPR-Cas system, a plasmid must be provided that not only includes the synthetic guide but also the Cas proteins necessary for interference.

The lethality and specificity of CRISPR-Cas antimicrobials could be harnessed to the fermentation industry's benefit in particular, especially in the management of mixed cultures (Beisel, Gomaa, & Barrangou, 2014). Mixed cultures are frequently used in the fermentation

industry for a number of reasons, including reducing the risk of phage infection, performing multi-step transformations of the substrate, or increasing product yield (Hesseltine, 1992). However, although mixed-culture fermentations do present a number of benefits, there are a number of disadvantages to consider. If the mixed-culture is yet uncharacterized (e.g. an artisanal culture), it can be difficult to isolate strains as many strains share similar physiological properties. Furthermore, contamination of a mixed culture by outside organisms can be difficult to control and eliminate. Finally, it is often problematic to maintain the ideal balance of strains in the mixed-culture as population dynamics are often unique to specific blends of organisms (Hesseltine, 1992; Sieuwerts et al., 2008). Traditional methods of isolating strains or controlling the composition of a mixed-culture population such as defined growth conditions, antibiotics, antimicrobial peptides, and bacteriophages, have thus far offered only partial solutions.

While CRISPR antimicrobials have not yet been applied to fermentation organisms, proof of concept in other organisms demonstrate the potential that CRISPR antimicrobials holds for use in the fermentation industry. The use of CRISPR-Cas as an antimicrobial exploited the lethality of self-targeting events and was first pioneered in strains that contained native CRISPR-Cas systems through providing the cells with a self-targeting spacer that was transmitted on a plasmid (Gomaa et al., 2014). It was shown that CRISPR-Cas antimicrobials could effectively separate two substrains of *E coli*, K-12 and B, that shared 99% sequence homology and nearly identical cellular processes. When *E. coli* K-12 was desired, the mixed-culture was presented with a plasmid containing a spacer designed to target a sequence that was only present in *E. coli* B, thus eliminating *E. coli* B from the

culture due to lethal CRISPR-Cas self-targeting (Gomaa et al., 2014). This same principal could be utilized to eliminate a known contaminant in a mixed culture or isolate closely related strains. Because the spacer sequence provided to the strain of interest can be engineered to focus on areas where genomes differ, CRISPR antimicrobials provide an unmatched level of specificity. The study also demonstrated that CRISPR-Cas antimicrobials can be utilized to modify the composition of a mixed culture when complete elimination is not desirable. This can be accomplished through controlling the "dose" of antimicrobial via creating a pre-determined blend of targeting plasmid and non-targeting plasmid to be transformed into a mixed population culture. When this was done in a mixed culture of *E. coli*, the proportion of targeted plasmid had a nearly linear association with the selective removal of that particular strain, showing that antimicrobials could be utilized to modulate culture populations (Gomaa et al., 2014). In short, it was shown that native CRISPR-Cas systems can function as a powerful and programmable antimicrobial with the ability to selectively eliminate specifically targeted strains or control mixed culture populations in a dose-effect manner.

The concept of CRISPR antimicrobials was also tested on strains that did not possess native CRISPR-Cas systems; these studies utilized phagemids to transmit Cas9, tracrRNA, and a crRNA designed to target the host chromosome into the cells (Bikard et al., 2014; Citorik, Mimee & Lu, 2014). The specificity of CRISPR antimicrobials was confirmed as they were able to distinguish between a single base pair difference between two strains (Citorik et al., 2014). The antimicrobials could also be manipulated to specifically target undesirable genes such as antibiotic resistance or virulence genes, leading to the elimination

of strains that adopt such traits (Bikard et al., 2014; Citorik et al., 2014). Both studies also showed that multiple spacers could be programmed in the phagemid to create multiple crRNAs that allows for simultaneous targeting of a variety of undesirable targets. Overall, CRISPR-Cas antimicrobials offer the food industry a comprehensive, timely method of singling out particular strains without affecting other closely-related strains in the future. Futhermore, they could be utilized for a variety of purposes including the isolation of a particular strain for further study or use, elimination of a known contaminant from a mixed population, screen against the presence of undesirable genes, or the regulation of culture composition for optimal product quality.

The lethality of self-targeting events in CRISPR-Cas makes it a powerful antimicrobial tool in managing the presence of detrimental microorganisms such as pathogenic or spoilage microorganisms as well. A plasmid or phagemid containing the *cas9* gene and a sgRNA can be programmed to target a desired pathogen/spoilage organism and then provided to the organisms within a particular environment. The CRISPR antimicrobial then cleaves the genomes of the targeted organism, thus greatly minimizing the presence of the pathogen or spoilage organism. CRISPR antimicrobials have already been tested in *E. coli*, Enterobacteriaceae, and *Staphylococcus aureus* (Bikard et al., 2014; Citorik et al., 2014; Gomaa et al., 2014). In short, antimicrobials could be designed to target specific spoilage or pathogenic microorganisms to eliminate their presence from food cultures or products. In addition, if a more general approach were needed, antimicrobials could be designed to target any microorganism that contains a specific undesirable gene, such as virulence factors or antibiotic resistance.

Due to their specificity and programmability, CRISPR antimicrobials present an appealing alternative to antibiotics, which tend to indiscriminately target detrimental and beneficial bacteria alike. They could be utilized in beneficial bacteria to minimize the spread of antibiotic resistance or virulence factors across strains. Furthermore, they could allow for unprecedented control in managing populations and eliminating contaminants within mixed food cultures with no negative effects on non-targeted strains. They could also be utilized to target detrimental bacteria such as pathogens or spoilage microorganisms to limit their presence in food products. However, a number of hurdles remain in optimizing this application for use in the food industry. First, an appropriate delivery vehicle is needed to make CRISPR-Cas antimicrobials a viable option for managing strain populations, especially those that lack native CRISPR-Cas systems. Although phages have been utilized, delivery of the antimicrobial remains a challenge as phages often possess a very narrow host range. Ideally, the delivery vehicle would be easily transmittable to a broad range of strains so that the antimicrobial can reach all possible cells in a culture and the host strain cells would hopefully contain necessary surface receptors to facilitate infection by the delivery vehicle (Beisel et al., 2014). Even if the delivery vehicle successfully delivers the antimicrobial, it must still get past any defense systems that the host may possess (Beisel et al., 2014). In addition, CRISPR targeting is not 100% effective, and sequences often mutate both in the host and target to escape CRISPR-Cas cleavage (Selle & Barrangou, 2015b; Stout et al., 2016- Unpublished work). Further work may need to be done to explore causes behind escape of CRISPR targeting in order to improve the efficacy of CRISPR antimicrobials. A better understanding of DNA repair pathways and their activity and distribution would also

be helpful in understanding antimicrobial activity in bacterial cultures (Barrangou & van Pijkeren 2016).

1.5 Conclusions

CRISPR-Cas has fascinated the scientific community since the discovery of unusual repeat-spacer sequences in archaeal and bacterial genomes. The establishment of its role in adaptive immunity in 2007 opened up new avenues for the field, leading to the eventual characterization of six distinct types of CRISPR-Cas systems, each with their own signature Cas proteins and peculiar mechanisms of interference. The tipping point occurred over the past 4 years, when the CRISPR-Cas machinery was repurposed for genome editing. The food industry has much to gain from adopting these tools and technologies for the genesis of next-generation food cultures and for controlling and modulating mixed microbial populations. The prevalence of CRISPR-Cas in beneficial food microorganisms such as starter cultures and probiotics make fermentation processes an unusually equipped industry for many CRISPR-Cas applications including the identification and distinction of closely related strains, protection of important starter cultures against phage infection, the vaccination of strains against unnecessary plasmid uptake, the isolation or modulation of specific strains within a mixed-culture, and the editing of bacterial genomes to create industry-relevant bacterial workhorses. Furthermore, CRISPR-Cas has also proven to be useful in working with detrimental organisms such as pathogens and spoilage microorganisms, making CRISPR-Cas an effective tool in the management of microbial populations in food production processes across the board. However, further work must be done to optimize

these applications and apply them to a broad spectrum of other industry-relevant organisms. What remains to be seen is when the food industry will readily embrace the potential that CRISPR-Cas holds as well as its role in the discovery of CRISPR-Cas and choose to drive the development of these applications to the benefit of food science as a whole. Although this review has primarily focused on applications in food bacteria, CRISPR-Cas has begun to impact food manufacturing processes from farm to fork, with CRISPR-based genome editing already being utilized to increase crop hardiness and nutrition as well as improve herd genetics. The scale and pace at which CRISPR-Cas has driven the development of relevant applications in not only food-related bacteria but also the fields of agriculture, husbandry, and environmental monitoring strongly suggest that it could soon be an established technology that could be utilized to drive further research and development in addition to maximizing the safety and quality of food products during the manufacturing process.

1.6 References

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., ... Zhang, F. 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*. 10.1126/science.aaf5573

Amitai, G. & Sorek, R. (2016). CRISPR-Cas adaptation: Insights into the mechanism of action. *Nature Reviews Microbiology*, 14, 67-76.

Anders, C., Niewoehner, O., Duerst, A., Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 513, 569-573.

Andersson, A.F. & Banfield, J.F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*, 320(5879), 1047-1050.

Arslan, Z. Hermanns, V., Wurm, R., Wagner, R., & Pul, U. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Research*, 42(12), 7884-7893.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., ... Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819), 1709-1712.

Barrangou, R. & Horvath, P. (2012). CRISPR: New horizons in phage resistance and strain identification. *Annual Review of Food Science and Technology*, 3, 143-162.

Barrangou, R., Coûté-Monvoisin, A.-C., Stahl, B., Chavichvily, I., Damange, F., Romero, D.A., ... Horvath, P. (2013). Genomic impact of CRISPR immunization against bacteriophages. *Biochemical Society Transactions*, 41(6), 1383-1391.

Barrangou R. 2014. Cas9 targeting and the CRISPR revolution. *Science* 344(6185): 707-708.

Barrangou R. 2015a. The roles of CRISPR-Cas systems in adaptive immunity and beyond. *Current Opinion in Immunology*, 32, 36-41.

Barrangou, R. 2015b. Diversity of CRISPR-Cas immune systems and molecular machines. *Genome Biology*, 16, 247.

Barrangou, R. & Dudley, E.G. (2016). CRISPR-based typing and next-generation tracking. *Annual Review of Food Science and Technology*, 7, 395-411.

Barrangou, R. & van Pijkeren, J.-P. (2016). Exploiting CRISPR-Cas immune systems for genome editing in bacteria. *Current Opinion in Biotechnology*, 37, 61-68.

Benda, C., Ebert, J., Scheltema, R.A., Schiller, H.B., Baumgartner, M., Bonneau, F., ... Conti, E. (2014). Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4. *Molecular Cell*, 56(1), 43-54.

Beisel, C.L., Gomaa, A.A., & Barrangou, R. (2014). A CRISPR design for next-generation antimicrobials. *Genome Biology*, 15, 516.

Bickard, D., Euler, C.W., Jiang, W., Nussenzweig, P.M., Goldberg, G.W., Duportet, X., ... Marraffini, L.A. (2014). Exploiting the CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nature Biotechnology*, 32(11), 1146-1150.

Bolotin, A., Ouinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology-Sgm*, 151, 2551-2561.

Bourdichon, F., Casaregola, S., Farrokh, C.,Frisvad, J.C., Gerds, M.L., Hammes, W.P., ... Hansen, E.B. (2012). Food fermentations: Microorganisms with technological beneficial use. *International Journal of Food Microbiology*, 154(3), 87-97.

Briner, A.E. & Barrangou, R. (2014). *Lactobacillus buchneri* genotyping on the basis of clustered regularly interspaced short palindromic repeat (CRISPR) locus diversity. *Applied and Environmental Microbiology*, 80(3), 994-1001.

Briner, A.E., Donohoue, P.D., Gomaa, A.A., Selle, K., Slorach, E.M., Nye, C.H., ... Barrangou, R. (2014). Guide RNA functional modules direct Cas9 activity and orthogonality. *Molecular Cell*, 53, 333-339.

Briner, A.E., Lugli, G.A., Milani, C., Duranti, S., Turroni, F., Gueimonde, M., ... Barrangou, R. (2015). Occurrence and diversity of CRISPR-Cas systems in the genus *Bifidobacterium*. *PLoS ONE*, 10(7), e0133661.

Broadbent JR,, J.R., Neeno-Eckwall, E.C., Stahl, B., Tandee, K., Cai, H., Morovic, W., . Steele, J.L. (2012). Analysis of the *Lactobacillus casei* supragenome and its influence in species evolution and lifestyle adaptation. *BMC Genomics*, 13, 533.

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A., ... van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 321(5891), 960-964.

Brüggemann, H., Lomholt, H.B., Tettelin, H., & Kilian, M. (2012). CRISPR/Cas loci of type II *Propionibacterium acnes* confer immunity against acquisition of mobile elements present in type I *P. acnes*. *PLoS ONE*, 7, e34171

Caplice, E., & Fitzgerald, G. F. (1999). Food fermentations: role of microorganisms in food production and preservation. *International Journal of Food Microbiology*, 50(1–2), 131-149.

Carte, J., Wang, R., Li, H., Terns, R.M., & Terns, M.P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & Development*, 22, 3489-3496.

Carte, J., Christopher, R.T., Smith, J.T., Olson, S., Barrangou, R., Moineau, S., Glover III, C.V.C., Gravely, B.R., Terns, R., & Terns, M.P. (2014). The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in *Streptococcus thermophilus*. *Molecular Microbiology*, 93(1), 98-112.

Chylinski, A., Le Rhun, E., & Charpentier, E., (2013). The tracrRNA and Cas9 families of Type II CRISR-Cas immunity systems. *RNA Biology*, 10(5), 726-737.

Citorik, R.J., Mimee, M., & Lu, T.K. (2014). Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nature Biotechnology*, 32, 1141-1145.

Coffey, A. & Ross, R.P. (2002). Bacteriophage-resistance systems in dairy starter strains: molecular analysis to application. *Antonie Van Leeuwenhoek*, 82(1-4), 303-321.

Datsenko, K.A., Pougach, A., Tikhonov, A., Wanner, B.L., Severinov, K., & Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Communications*, 3, 945.

D'Auria, G., Jimenez-Hernandez, N., Peris-Bondia, F., Moya, A., & Latorre, A. (2010). *Legionella pneumophila* pangenome reveals strain-specific virulence factors. *BMC Genomics*, 11, 181.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y.J., Pirzada, Z.A., ... Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), 602-607.

Deng, L., Garrett, R.A., Shah, S.A., Peng, X., & She, Q. (2013). A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Molecular Microbiology*, 87(5), 1088-1099.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., ... Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of Bacteriology*, 190, 1390-400.

DiMarzio, M., Shariat, N., Kariyawasam, S., Barrangou, R., & Dudley, E.G. (2013). Antibiotic resistance in *Salmonella* Typhimurium associates with CRISPR sequence type. *Antimicrobial Agents and Chemotherapy*, 57(9), 4282-4289.

Djordjevic, G.M., O'Sullivan, D.J., Walker, S.A., Conkling, M.A., & Klaenhammer, T.R. (1997). A triggered-suicide system designed as a defense against bacteriophages. *Journal of Bacteriology*, 179(21), 6741-6748.

Dong, D., Ren, K., Qiu, Z., Zheng, J., Guo, M., Guan, X., ... Huang, Z. (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature*. 532, 522-526.

Douglas, G.L. & Klaenhammer, T.R. (2010). Genomic evolution of domesticated microorganisms. *Annual Review of Food Science and Technology*, 1, 397-414.

Doudna, J.A. & Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213), 1258096.

Douillard, F.P., Ribbera, A., Kant, R., Pietilä, T.A., Järvinen, H.M., Messing, M., ...de Vos, W.M. (2013). Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. *PLoS ONE*, 9(8), e1003683.

Durmaz, E. & Klaenhammer, T.R. (2007). Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected *Lactococcus lactis*. *Journal of Bacteriology*, 189(4), 1417-1425.

Edgar, R., & Qimron, U. (2010). The *Escherichia coli* CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. *Journal of Bacteriology*, 192, 6291-6294.

Edmond, E. & Moineau, S. (2007). Bacteriophages and food fermentations. In McGrath, S. & van Sinderen, D., eds. *Bacteriophage: Genetics and Molecular Biology* (pp. 93-124). Horizon Scientific Press/Caister Academic Press.

Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J., Church, G.M. (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods*. 10(11), 1116-1121.

Fabre, L., Zhang, J., Guigon, G., Hello, S.L., Guibert, V, Accou-Demartin, M., ... Weill, F.-X. (2012). CRISPR typing and subtyping for improved laboratory surveillance of Salmonella infections. *PLoS ONE*, 7(5), e36995.

Fagerlund, R.D., Staals, R.H.J., & Fineran, P.C. (2015). The Cpf1 CRISPR-Cas protein expands genome-editing tools. *Genome Biology*, 16, 251.

FAO/WHO. (2002). Guidelines for the evaluation of probiotics in food. FAO/WHO, London, ON, Canada.

Fineran, P.C., Gerritzen, M.J.H., Suarez-Diaz, M., Kunne, T., Boekhorst, J., Van Hijum, S.A.F.T., ... Brouns, S.J.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. *PNAS*, 111(16), E1629-1638.

Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., & Charpentier E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, 532, 517-521.

Fox, P. F. (1993). Cheese: An Overview. In P. F. Fox (Ed.), *Cheese: Chemistry, Physics and Microbiology* (pp. 1-36): Springer US.

Fricke, W.F., Mammel, M.K., McDermott, P.F., Tartera, C., White, D.G., LeClerc, J.E., ... Cebula, T.A. (2011). Comparative genomics of 28 *Salmonella enterica* isolates: Evidence for CRISPR-mediated adaptive sublineage evolution. *Journal of Bacteriology*, 193(14), 3556-3568.

Garneau, J.E., Dupuis, M.-E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., ... Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468, 67-71.

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS USA*, 109(39), E2579-E2586.

Goh, Y.J., Goin, C., O'Flaherty, Altermann, E., & Hutkins, R. (2011). Specialized adaptation of a lactic acid bacterium to the milk environment: the comparative genomics of *Streptococcus thermophilus* LMD-9. *Microbial Cell Factories*, 10(Supple 1), S22.

Goldberg, G.W., Jiang, W., Bikard, D., & Marraffini, L.A. (2014). Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature*, 514, 633-637.

Gomaa, A.A., Klumpe, H.E., Luo, M.L., Selle, K., Barrangou, R., Beisel, C.L. (2014). Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *mBio*, 5(1), e00928-13.

Grissa, I., Vergnaud, G., & Pourcel, C. (2007). CRISPRFinder: A webtool to identify clustered regularly interspersed short palindromic repeats. *Nucleic Acids Research*, 35, W52-57.

Groenen, P.M., Bunschoten, A.E., van Soolingen, D., & van Embden, J.D. (1993). Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Molecular Microbiology,* 10, 1057–1065.

Hale, C., Kleppe, K., Terns, R.M., & Terns, M.P. (2008). Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA*, 14(12), 2572-2579.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Gravely, B.R., Wells, L., Terns, R.M. & Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, 139(5), 945-956.

Hale, C.R., Majumdar, S., Elmore, J., Plister, N., Compton, M., Olson, S., Resch, A.M., Glover III, C.V.C., Gravely, V.R., Terns, R.M., & Terns, M.P. (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Molecular Cell*, 45(3), 292-302.

Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W., & Marraffini, L.A. (2013). A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *Journal of Biological Chemistry*, 288, 27888-27897.

Hatoum-Aslan, A., Maniv, I., Samai, P., & Marraffini, L.A. (2014). Genetic characterization of antiplasmid immunity through a Type III-A CRISPR-Cas system. *Journal of Bacteriology*, 196(2), 310-317.

Hargreaves, K.R., Flores, C.O., Lawley, T.D., & Clokie, M.R.J. (2014). Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in *Clostridium difficile* strains and prophages target multiple phage types within this pathogen. *mBio*, 5(5), e01045-13.

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., & Marrafini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature*, 519, 199-202.

Hesseltine, C.S. (1992). Mixed culture fermentations. *Applications of biotechnology to fermented foods: Report of an ad hoc panel of the board on science and technology for international development*. Washington (DC): National Academies Press (US). Available from: http://www.ncbi.nlm.nih.gov/books/NBK234678/.

Higgins, D.L., Sanozky-Dawes, R.B., & Klaenhammer, T.R. (1988). Restriction and modification activities from *Streptococcus lactis* ME are encoded by a self-

transmissible plasmid, pTN20, that forms cointegrates during mobilization of lactose-fermenting ability. *Journal of Bacteriology*, 170(8), 3435-342.

Hill, C., Guarner, F., Reid, G., Gibson, G.R., Merenstein, D.J., Pot, B., ...Sanders, M.E. (2014). Expert consensus document: The International Scientific Association for Probiotics and Prebiotics consensus statement on the score and appropriate use of the term probiotic. *Nature Reviews Gastroenterology & Hepatology*, 11, 506-514.

Hirano, H., Gootenberg, J.S., Horii, T., Abudayyeh, O.O., Kimura, M., Hsu, P.D., ... Nureki, O. (2016). Structure and engineering of *Francisella novicida* Cas9. *Cell*, 164(5), 950-961.

Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., & Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *PNAS*, 111(18), 6618-23.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., ... Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *Journal of Bacteriology*, 190(4), 1401-1412.

Horvath, P. & Barrangou, R. (2011). Protection against Foreign DNA. In G. Storz & R. Hengge (Eds.), *Bacterial Stress Responses*, 2nd edition (333-348). Washington, DC: ASM Press.

Huis in't Veld, J.H. (1996). Microbial and biochemical spoilage of foods: An overview. *International Journal of Food Microbiology*, 33(1), 1-18.

Hullahalli, K., Rodrigues, M., Schmidt, B.D., Li, X., Bhardwaj, P. & Palmer, K.L. (2015). Comparative analysis of the orphan CRISPR2 locus in 242 *Enterococcus faecalis* strains. *PLoS ONE*, 10(9), e0138890.

Huo, Y., Nam, K.H., Ding, F., Lee, H., Wu, L., Xiao, Y., ... Ke, A. (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nature Structural & Molecular Biology*, 21, 771-777.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, 169(12), 5429-5433.

Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., ... Wiedenheft. B. (2014). Structural Biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science*, 345(6203), 1473-1479.

Jansen, R., van Embden, J.D.A., Gaastra, W., & Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, 43(6), 1565-1575.

Jiang, F., Zhou, K., Ma, L., Gressel, S., & Doudna, J.A. (2015a). A Cas9-guide RNA complex preorganized for target DNA recognition. *Science*, 348(6242), 1477-1481.

Jiang, Y., Yin, S., Dudley, E.G., & Cutter, C.N. (2015b). Diversity of CRISPR loci and virulence genes in pathogenic *Escherichia coli* isolates from various sources. *International Journal of Food Microbiology*, 204, 41-46.

Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, ... Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, 351(6275), 867-871.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-821.

Jinek, M., Jiang, D.W., Taylor, S.H., Sternberg, E., Kaya, E., Ma, C., ... Doudna, J.E. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, 343, 1247997.

Johnson, B.R. & Klaenhammer, T.R. (2014). Impact of genomics on the field of probiotic research: historical perspectives to modern paradigms. *Antonie Van Leeuwenhoek*, 106(1), 141-156.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., ... Brouns S.J.J. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Structural & Molecular Biology*, 18, 529-536.

Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P., & Siksnys, V. (2013). crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus. RNA Biology*, 10(5), 841-851.

Kinnevey, P.M., Shore, A.C., Brennan, G.I., Sullivan, D.J., Ehricht, R., Monecke, S., ... Coleman, D.C. (2013). Emergence of sequence type 779 methicillin-resistant *Staphylococcus aureus* harboring a novel pseudo staphylococcal cassette chromosome *mec* (SCC*mec*)-SCC-SCC_{CRISPR} composite element in Irish hospitals. *Antimicrobial Agents and Chemotherapy*, 57, 524–531.

Kovanen, S.M., Kivistö, R.I., Rossi, M., & Hänninen, M.-L. (2014). A combination of MLST and CRISPR typing reveals dominant *Campylobacter jejuni* types in organically farmed laying hens. *Journal of Applied Microbiology*, 117(1), 249-257.

Kunin, V. Sorek, R., & Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biology*, 8(4), R61.

Kuno, S., Yoshida, T., Kaneko, T., & Sako, Y. (2012). Intricate interactions between the bloom-forming cyanobacterium *Microcystis aeruginosa* and foreign genetic elements, revealed by diversified clustered regularly interspaced short palindromic repeat (CRISPR) signatures. *Applied and Environmental Microbiology*, 78, 353–60

Kupczok, A. & Bollback, J.P. (2013). Probabilistic models for CRISPR spacer content evolution. *BMC Evolutionary Biology*, 13, 54.

Labrie, S.J., Samson, J.E., & Moineau, S. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*, 8, 317-327.

Levin, B.R. (2010). Nasty viruses, costly plasmids, population dynamics, and the conditions for establishing and maintaining CRISPR-mediated adaptive immunity in bacteria. *PLoS Genetics*, 6, e1001171.

Levin, B.R., Moineau, S., Bushman, M., & Barrangou, R. (2013). The population and evolutionary dynamics of phage and bacteria with CRISPR-mediated immunity. *PLoS Genetics*, 9(3), e1003312.

Lier, C., Baticle, E., Horvath, P., Haguenoer, E., Valentin, A.S., Glaser, P., ... Lanotte, P. (2015). Analysis of the type II-A CRISPR-Cas system of *Streptococcus agalactiae* reveals distinctive features according to genetic lineages. *Frontiers in Genetics*, 6, 214

Lindenstrauss, A.G., Pavlovic, M., Bringmann, A., Behr, J., Ehrmann, M.A., & Vogel, R.F. (2011). Comparison of genotypic and phenotypic cluster analyses of virulence determinants and possible role of CRISPR elements towards their incidence in *Enterococcus faecalis* and *Enterococcus faecium*. *Systematic and Applied Microbiology*, 34, 553-560.

Liu, F., Barrangou, R., Gerner-Smidt, P. Ribot, E.M., Knabel, S.J., & Dudley, E.G. (2011a). Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Applied and Environmental Microbiology*, 77(6), 1946-1956.

Liu, F., Kariyawasam, S., Jayarao, B., Barrangou, R., Gerner-Smidt, P., Ribot, E.M., ... Dudley, E.G. (2011b). Subtyping of *Salmonella enterica* serovar Enteritidis isolates from different sources by using sequence typing based on virulence genes and clustered regularly interspaced short palindromic repeats (CRISPRs). *Applied and Environmental Microbiology*, 77(13), 4520-4526.

Maier, L.-K., Lange, S.J., Haas, K.A., Fischer, S., Fischer, E., Duchardt-Ferner, E., ... Marchfelder, A. (2013). Essential requirements for the detection and degradation of invaders by the *Haloferax volcanii* CRISPR/Cas system I-B. *RNA Biology*, 10(5), 865-874.

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*, 1(7).

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., ... Koonin, E.V. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology*, 9(6), 467-77.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.A., ... Koonin, E.V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*, 13, 722-736.

Marcó, M.B., Moineau, S., & Quiberoni, A. (2012). Bacteriophages and dairy fermentations. *Bacteriophage*, 2(3), 1-10.

Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in Staphylococci by targeting DNA. *Science*, 322(5909), 1843-1845.

Marrafini, L.A. & Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, 463, 568-571.

Metchnikoff, E. (1907). The prolongation of life: optomistic studies. William Heinemann, London.

Millen, A.M., Horvath, P., Boyaval, P., Romero, D.A., & Dennis, A. (2012). Mobile CRISPR/Cas-mediated bacteriophage resistance in *Lactococcus lactis*. *PLoS ONE*, 7(12), e51663.

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution*, 60(2), 174-182.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, 155, 733-40.

Mokrousov, I., Vyazovaya, A., Kolodkina, V., Limeschenko, E., Titov, L., & Narvaskaya, O. (2009). Novel macroarray-based method of *Corynebacterium diptheriae* genotyping: evaluation in a field study in Belarus. *European Journal of Clinical Microbiology & Infectious Diseases*, 28, 701-703.

Mulepati, S., Orr, A. & Bailey, S. (2012). Crystal structure of the largest subunit of a bacterial RNA-guided immune complex and its role in DNA target binding. *Journal of Biological Chemistry*, 287, 22445-22449.

Mulepati, S., Heroux, A., & Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science*, 345(6203), 1479-1484.

Nam, K.H., Haitjema, C., Liu, X., Ding, F.,Wang, H., DeLisa, M.P., & Ke, A. (2012). Cas5d protein processes pre-crRNA and assembles into a Cascade-like interference complex in Subtype I-C/Dvulg CRISPR-Cas system. *Structure*, 20(9), 1450-1452.

Niewoehner, O., Jinek, M., Doudna, J.A. (2014). Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Research*, 42, 1341-1353.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., ... Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 156(5), 935-949.

Nishimasu, H., Cong, L., Yan, W.X., Ran, F.A., Zetsche, B., Li, Y., ... Nureki, O. (2015). Crystal structure of *Staphylococcus aureus* Cas9. *Cell*, 162(5), 1113-1126.

Nozawa, T., Furukawa, N., Aikawa, C., Watanabe, T., Haobam, B., Kurukawa, K., ... Nakagawa, I. (2011). CRISPR inhibition of prophage acquisition in *Streptococcus pyogenes*. *PLoS ONE*, 6(5), 1-9.

Nuñez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., & Doudna, J.A. (2014).
Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature Structural & Molecular Biology*, 21, 528-534.

Nuñez, J.K., Lee, A.S.Y., Engleman, A. & Doudna, J.D. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature*, 519, 193-198.

Oh, J.-H. & van Pijkeren, J.-P. (2014). CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*. *Nucleic Acids Research*, 42(17): e131.

Paez-Espino, D., Morovic, W., Sun, C.L., Thomas, B.C., Ueda, K-I., Stahl, B., ... Banfield, J.F. (2012). Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nature Communications*, 4, 1430.

Palmer, K.L. & Gilmore, M.S. (2010). Multidrug-resistant enterococci lack CRISPR-*cas*. *MBio*, 1, e00227-10.

Papadimitriou, K., Pot, B., & Tsakalidou, E. (2015). How microbes adapt to a diversity of food niches. *Current Opinion in Food Science*, 2, 29-35.

Peng, W., Feng, M., Feng, X., Liang, Y.X., & She, Q. (2015). An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference. *Nucleic Acids Research*, 43(1), 406-417.

Pennisi, E. (2013). The CRISPR Craze. *Science*, 341(6148), 833-836.

Pettengill, J.B., Timme, R.E., Barrangou, R., Toro, M., Allard, M.W., Strain, E., ... Brown, E.W. (2014). The evolutionary history and diagnostic utility of the CRISPR-Cas system within *Salmonella enterica* ssp. e*nterica*. *PeerJ*, 2, e340.

Plagens, A., Tjaden, B., Hagemann, A., Randau, L., & Hensel, R. (2012). Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax. Journal of Bacteriology*, 194(10), 2491-2500.

Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology-Sgm*, 151, 653-663.

Ramia N.F., Spilman, M., Tang, L., Shao, Y., Elmore, J., Hale, C., ... Stagg, S.M. (2014). Essential structural and functional roles of the Cmr4 subunit in RNA cleavage by the Cmr CRISPR-Cas complex. *Cell Reports*, 9(5), 1610-1617.

Reeks, J., Naismith, J.H., & White, M.F. (2013). CRISPR interference: A structural perspective. *Biochemical Journal*, 453, 155-166.

Rezzonico, F., Smits, T.H.M., & Duffy, B. (2011). Diversity, evolution and functionality of clustered regularly interspaced short palindromic repeat (CRISPR) regions in the fire blight pathogen *Erwinia amylovora*. *Applied and Environmental Microbiology*, 77, 3819-3829.

Richter, C., Dy, R.L., Mckenzie, R.E., Watson, B.N.J., Taylor, C., Chang, J.T., ... Fineran, P.C. (2014). Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Research*, 42, 8516-8526.

Riehm, J.M., Vergnaud, G., Kiefer, D., Damdindorj, T., Dashdavaa, O., Khurelsukh, T., ... Scholz, H.C. (2012). *Yersinia pestis* lineages in Mongolia. *PLoS ONE*, 7, e30624

Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., ... White, M.F. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with Cascade. *Molecular Cell*, 52(1), 124-134.

Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnys, V.,& Seidel, R. (2015). Directional R-loop formation by the CRISPR-Cas surveillance complex provides efficient off-target site rejection. *Cell Reports*, 10(9), 1534-1543.

Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A., & Marraffini, L.A. (2015). Co-transcriptional DNA and RNA cleavage during Type III CRISPR-Cas immunity. *Cell*, 161(5), 1164-1174.

Samson, J.E. & Moineau, S. (2013). Bacteriophages in food fermentations: New frontiers in a continuous arms race. *Food Science and Technology*, 4, 347-368.

Sander, J.D. & Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating, and targeting genomes. *Nature Biotechnology*, 32, 347-355.

Sanders, M.E. & Klaenhammer, T.R. (1981). Evidence for plasmid linkage of restriction and modification in *Streptococcus cremoris* KH. *Applied and Environmental Microbiology*, 42(6). 944-950.

Sanders, M.E. & Klaenhammer, T.R. (1983). Characterization of phage-sensitive mutants from a phage-insensitive strain of *Streptococcus lactis*: evidence for a plasmid determinant that prevents phage adsorption. *Applied and Environmental Microbiology*, 46(5), 1125-1133.

Sanozky-Dawes, R., Selle, K., O'Flaherty, S., Klaenhammer, T., & Barrangou, R. (2015). Occurrence and activity of a type II CRISPR-Cas system *in Lactobacillus gasseri. Microbiology*, 161, 1752-1761.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas systems provides immunity in *Escherichia coli. Nucleic Acids Research*, 39(21), 9275-9282.

Sashital, D.G., Jinek, M., & Doudna, J.A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nature Structural & Molecular Biology*, 18, 680-687.

Sashital, D.G., Wiedenheft, B., & Doudna, J.A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Molecular Cell*, 46, 606-615.

Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., & Severinov, K. (2013). High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Biology,* 10(5), 716-725.

Scallen, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.-A., Roy, S.L., ... Griffin, P.M. (2011). Foodborne illness acquired in the United States—Major pathogens. *Emerging Infectious Diseases*, 17(1), 1-15.

Scharff, R.L. (2012). Economic burden from health losses due to foodborne illness in the United States. *Journal of Food Protection*, 75, 123-131.

Schunder, E., Rydzewski, K., Grunow, R., & Heuner, K. (2013). First indication for a functional CRISPR/Cas system in *Francisella tularensis. International Journal of Medical Microbiology*, 303(2), 51-60.

Selle, K., Klaenhammer, T.R., & Barrangou, R. (2015). CRISPR-based screening of genomic island excision events in bacteria. *PNAS*, 112(26), 8076-8081.

Selle, K. & Barrangou, R. (2015a). CRISPR-based technologies and the future of food science. *Journal of Food Science*, 80(11), R2367-R2372.

Selle, K. & Barrangou, R. (2015b). Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends in Microbiology*, 23(4), 225-232.

Semenova, E., Nagornykh, M., Pyatnitskiy, M., Artamonova, II, & Severinov, K. (2009). Analysis of CRISPR system function in plant pathogen *Xanthomonas oryzae*. *FEMS Microbiology Letters,* 296, 110–116.

Seminova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., ... Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is guided by a seed sequence. *PNAS*, 108(25), 10098-10103.

Shariat N., DiMarzio, M.J., Shuang, Y., Dettinger, L., Sandt, C.H., Lute, J.R., ... Dudley, E.G. (2013a). The combination of CRISPR-MVLST and PFGE provides increased

discriminatory power for differentiating human clinical isolates of *Salmonella enterica* subsp. *enterica* serovar Enteritidis. Food Microbiology, 34(1), 164-173.

Shariat, N., Kirchner, M.K., Sandt, C.H., Trees, E., Barrangou, R., & Dudley, E.G. (2013b). Subtyping of *Salmonella enterica* serovar Newport outbreak isolates by CRISPR-MVLST and determination of the relationship between CRISPR-MVLST and PFGE results. *Journal of Clinical Microbiology*, 51(7), 2328-2336.

Shariat, N., Sandt, C.H., DiMarzio, M.J., Barrangou, R., & Dudley, E.G. (2013c). CRISPR-MVLST subtyping of *Salmonella enterica* subsp. *enterica* serovars Typhimurium and Heidelberg and application in identifying outbreak isolates. *BMC Microbiology*, 13, 254.

Shariat, N., Timme, R.E., Pettengill, J.B., Barrangou, R., & Dudley, E.G. (2014). Characterization and evolution of *Salmonella* CRISPR-Cas systems. *Microbiology*, 161, 374-386.

Shimomura, Y., Okumura, K.,Yamagata, M.S., Yagi, J., Ubukata, K., Kirikae, T., & Miyoshi-Akiyama, T. (2011). Complete genome sequencing and analysis of a Lancefield group G Streptoccus dysgalactiae subsp. Equisimilis strain causing streptoccal toxic shock syndrome (STSS). *BMC Genomics*, 12(17).

Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., ... Koonin, E.V. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Molecular Cell*, 60(3), 385-397.

Sieuwerts, S., de Bok, F.A.M., Hugenholtz, J., & van Hylckama Vlieg, J.E.T. (2008). Unraveling microbial interactions in food fermentations: from classical to genomics approaches. *Applied and Environmental Microbiology*, 74(16), 4997-5007.

Smokvina, T., Wels, M., Polka, J., Chervaux, C., Brisse, S., Boekhorst, J., ... Siezen, R.J. (2013). *Lactobacillus paracasei* comparative genomics: Towards species pan-genome definition and exploitation of diversity. *PLoS ONE*, 8(7), e68731.

Sontheimer, E., & Barrangou, R. (2015). The bacterial origins of the CRISPR genome-editing revolution. *Human Gene Therapy*, 26(7), 413-424.

Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., ... Stagg, S. (2013). Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Molecular Cell*, 52(1), 146-152.

Staals, R.H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W., van Duijn, E., ... Shinkai, A. (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas Complex of *Thermus thermophilus. Molecular Cell*, 52(1), 135-145.

Staals, R.H.J., Zhu, Y., Taylor, D.W., Kornfield, J.E., Sharma, K., Barendregt, A., ... van der Oost, J. (2014). RNA targeting by the Type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Molecular Cell*, 56(4), 518-530.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., & Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490), 62-67.

Sternberg, S.H., Richter, H., Charpentier, E., & Qimron, U. (2016). Adaptation in CRISPR-Cas systems. *Molecular Cell*, 61(6), 797-808.

Stout, E.A., Sanozky-Dawes, R.B., Klaenhammer, T.R., & Barrangou, R. (2016). Deletion-based escape of CRISPR-Cas targeting in *Lactobacillus gasseri*. Manuscript in preparation.

Sun, Z., Harris, H.M.B., McCann, A., Guo, C., Argimón, S., Zhang, W., ... O'Toole, P.W. (2015a). Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nature Communications*, 6, 8322.

Sun, H., Li, Y., Shi, X., Lin, Y., Qiu, Y., Zhang, J., Liu, Y., Jiang, M., Zhang, Z., Chen, Q., Sun, Q., & Hu, Q. (2015b). Association of CRISPR/Cas evolution with *Vibrio parahaemolyticus* virulence factors and genotypes. *Foodborne Pathogens and Disease*, 12, 68–73.

Swarts, D.C., Mosterd, C., van Passel, M.W., & Brouns, S.J. (2012). CRISPR interference directs strand specific spacer acquisition. *PLoS ONE*, 7(4), e35888.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., ... Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *PNAS*, 111(27), 9798-9803.

Tamulaitis, G., Kazlauskienne, M., Manakova, E., Venclovas, C., Nwokeoji, A.O., Dickman, M.J., ... Siksnys, V. (2014). Programmable RNA shredding by the Type III-A CRISPR-Cas system of *Streptoccocus thermophilus*. *Molecular Cell*, 56(4), 506-517.

Timme, R.E., Pettengill, J.B., Allard, M.W., Strain, E., Barrangou, R., Wehnes, C., ... Brown, E.W. (2013). Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. *enterica* inferred from the genome-wide reference-free SNP characters. *Genome Biology and Evolution*, 5(11), 2109-2123.

Tremblay, C.L., Charlebois, A., Masson, L., & Archambault, M. (2013). Characterization of hospital-associated lineages of ampicillin-resistant *Enterococcus faecium* from clinical cases in dogs and humans. *Frontiers in Microbiology*, 4, 245.

Tyson, G.W. & Banfield, J.W. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environmental Microbiology*, 10(1), 200-207.

van der Oost, J., Westra, E.R., Jackson, R.N., & Wiedenheft. B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature Reviews Microbiology*, 12(7), 479-92.

Vercoe, R.B., Chang, J.T., Dy, R.L., Taylor, C., Gristwood, T., Clulow, J.S., ... Fineran, P.C. (2013). Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genetics*, 9(4), e1003454.

Voetsch, A.C., Van Gilder, T.J., Angulo, F.J., Farley, M.M., Shallow, S., Marcus, R., ... Tauxe, R.V., Emerging Infections Program FoodNet Working Group. (2004). FoodNet estimate of the burden of illness caused by nontyphoidal Salmonella infections in the United States. *Clinical Infectious Diseases*, 38(Supplement3), S127-134.

Wang, J., Li, J, Zhao, H., Sheng, G., Wang, M.,Yin, M., & Wang, Y. (2015). Structural and mechanistic basics of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell*, 163(4), 840-853.

Wehnes, C.A., Rehberger, T.G., Barrangou, R., Smith, A.H. (2014). Determination of *Salmonella* clustered regularly interspaced short palindromic repeats (CRISPR) diversity on dairy farms in Wisconsin and Minnesota. *Journal of Dairy Science*, 97(10), 6370-6377.

Wei, Y., Terns, R.M., & Terns, M.P. (2015). Cas9 function and host genome sampling in type II-A CRISPR-Cas adaptation. *Genes & Development*, 29, 356-361.

Weinberger, A.D., Sun, C.L., Pluciński, M.M., Denef, V.J., Thomas, B.C., Horvath, P., ... Banfield, J.F. (2012). Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Computational Biology*, 8(4), e1002475.

Westra, E.R., Van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H., Seegers, C.L., ... van der Oost, J. (2012). CRISPR immunity relies on the consecutive binding and degradation

of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular Cell*, 46(5), 595-605.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J.J., van der Oost, J., ... Nogales, E. (2011a). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*, 477, 486-489.

Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., ... Doudna, J.A. (2011b). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *PNAS*, 108(25), 10092-10097.

Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.M., Li, Y., ... Nureki, O. (2016). Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell*, 165(4), 949-62.

Yin, S., Jensen, M.A., Bai, J., Debroy, C., Barrangou, R., & Dudley, E.G. (2013). The evolutionary divergence of shiga toxin–producing *Escherichia coli* is reflected in clustered regularly interspaced short palindromic repeat (CRISPR) spacer composition. *Applied and Environmental Microbiology*, 79, 5710–5720.

Zebec, Z., Manica, A., Zhang, J., White, M.F., & Schleper, C. (2014). CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus. Nucleic Acids Research*, 42(8), 5280-5288.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., ... Zheng, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 163(3), 759-771.

Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., ... Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli. Nature*, 515(7525), 147-150.

Zoephel, J., & Randau, L. (2013). RNA-Seq analyses reveal CRISPR RNA processing and regulation patterns. *Biochemical Society Transactions*, 41, 1459-1463.

Table 1.1 | Cas proteins involved in CRISPR immunity mechanisms[1,2]

	Adaptation	Expression	Interference		
	Spacer acquisition	**Pre-crRNA processing**	**crRNA-RNP complex assembly and surveillance**	**Target degradation nuclease**	**Degradation characteristics**
Type I-E	Cas1/Cas2	Cas6	Cascade (Cas5, Cas6, Cas7, Cse1, Cse2) + crRNA	Cas3* (HD nuclease and type A superfamily 2 helicase)	dsDNA unwinding and ssDNA nicking of the non-target strand
Type II-C	Cas1/Cas2	RNaseIII, tracRNA	Cas9* + crRNA:tracrRNA complex	Cas9* (RuvC and HNH domains)	Blunt-ended double-stranded DNA cleavage
Type III-A	Cas1/Cas2	Cas6	Csm complex (Cas10*, Csm4/Cas5, Csm3/Cas7, Csm5/Cas7, Csm2)	Csm6 (DNase), Csm3 (RNase subunit)	DNA cleavage (*in vivo*) and RNA shredding
Type IV	N/A	N/A	Csf1 complex (Csf1*, Cas7, Cas5)	Csf1*	Uncharac-terized
Type V-A	Cas1/Cas2	Cpf1*	Cpf1* + crRNA	Cpf1* (RuvC domain and putative novel nuclease domain)	Staggered DNA double-stranded DNA cleavage
Type VI	Cas1/Cas2	Unknown	C2c1* + crRNA	C2c2* (2 HEPN domains)	Sequence and secondary structure-driven ssRNA cleavage

[1]Genes in red are present in the CRISPR loci and are hypothesized to play a role in the mechanism although the mechanism has not been characterized yet.
[2]Asterisks denote the signature gene of that particular CRISPR-Cas system type.

Table 1.2 | Food Science Applications of CRISPR-Cas

	Application	Starter Cultures	Probiotics	Pathogens	Spoilage Organisms
Strain-typing	• Identify and differentiate strains	✓	✓	✓	✓
Phage resistance	• Reduce risk of failed fermentations due to phage attack	✓			
Plasmid vaccination	• Reduce fitness burden on cell • Reduce uptake of undesirable genes (e.g. antibiotic resistance/virulence genes)	✓	✓		
Genome editing	• Engineer strains for increased functionality • GRAS organisms for use in genome editing	✓	✓		
Anti-microbial	• Isolate nearly identical strains • Modify composition of mixed culture • Target contaminants • Target removal of unwanted organisms • Target removal of undesirable genes	✓	✓	✓	✓

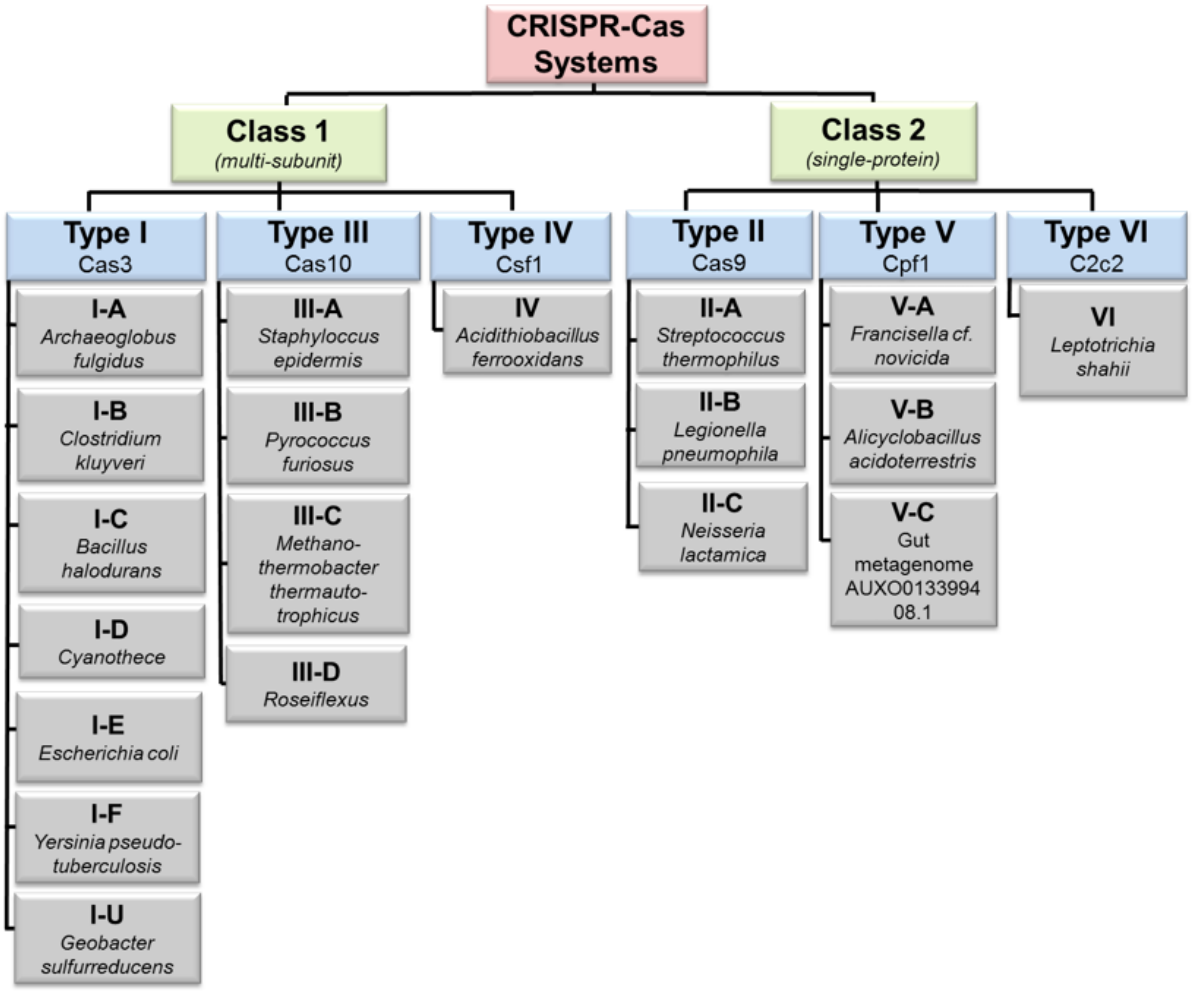

Figure 1.1 | Classification of CRISPR-Cas systems

CRISPR-Cas systems are categorized by class, type, and subtype. The two classes are defined by their crRNA-effector protein complex whereas the six major types are differentiated through the presence of a signature protein. The subtypes vary from each other via CRISPR loci organization, the presence of additional *cas* genes, and/or variation/inactivation of various cas genes. Shown are the six major types of CRISPR-Cas systems and and their corresponding subtypes with their model organisms.

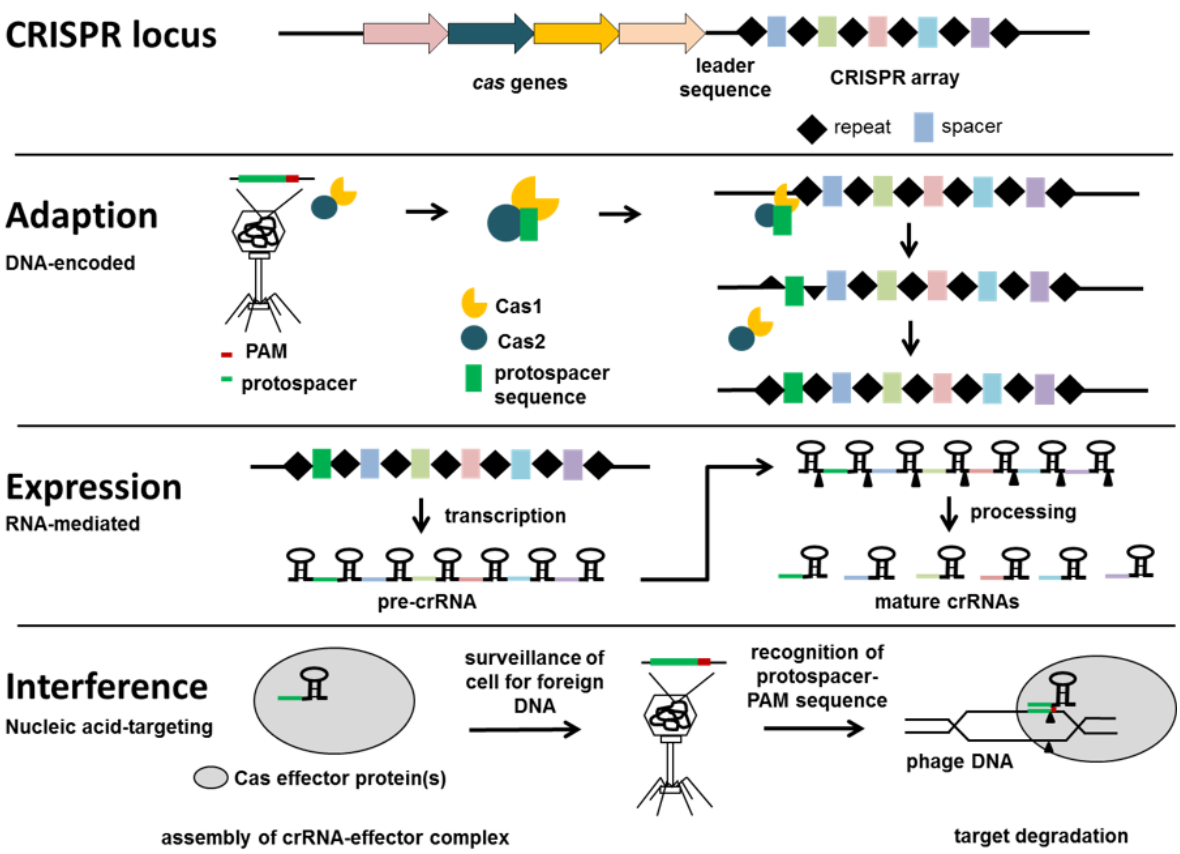

Figure 1.2 | CRISPR-based adaptive immunity

Top panel: CRISPR-Cas systems are composed of *cas* genes, a promoter leader sequence, and the CRISPR array which contains alternating repeat and spacer sequences. Second panel: adaptation occurs when the CRISPR-Cas systems samples the DNA of an invading element such as a phage and incorporates a small portion of that DNA into its repeat-spacer array as a spacer sequence. Third panel: expression takes place when the repeat-spacer array is transcribed as a long pre-crRNA sequence and processing occurs to form mature crRNAs that each posess a spacer sequence that will target foreign DNA containing that sequence. Bottom panel: interference is driven by Cas nucleases that interact with the crRNA to recognize and cleave invading nucleic acid sequences.

CHAPTER 2. CRISPR-Cas Targeting and Escape in *Lactobacillus gasseri*

Emily A. DeKam, Rosemary Sanozky-Dawes, Todd R. Klaenhammer, & Rodolphe Barrangou

2.1 Abstract

Lactobacillus gasseri is a lactic acid-producing commensal bacterium that often contains CRISPR-Cas, an adaptive immune system composed of CRISPR repeat-spacer arrays and CRISPR-associated proteins (Cas) that protect the cell from invasive mobile genetic elements (MGE). However, MGEs occasionally escape targeting by CRISPR-Cas due to DNA mutations that occur in sequences essential to the CRISPR interference process. To better understand CRISPR escape processes, a plasmid interference system was used to screen for mutants with plasmids that had escaped CRISPR-Cas targeting in *L. gasseri* strains JV-V03 and NCK 1342. Plasmids containing a target sequence, a protospacer adjacent motif (PAM), and an erythromycin resistance gene were transformed into both strains for targeting by the native CRISPR-Cas system. *L. gasseri* mutants containing plasmids that had escaped CRISPR targeting were recovered on selective media to investigate the genetic mechanisms of escape and their relative frequencies. Deletions in the CRISPR array were the dominant pattern of escape in both strains, accounting for 70% and 52% of the mutants in JV-V03 and NCK1342, respectively. Analysis of sequence mutations revealed internal deletions in the host CRISPR array, characterized by polarized excisions from the leader end that ranged from 1-15 spacers and that systematically included the leader-proximal targeting spacer. This established that deletions of targeting spacers within CRISPR arrays constitute a key escape mechanism whereby cells adapt to evade CRISPR targeting, while maintaining both the target sequence and the functionality of the CRISPR-Cas system. We speculate that internal deletions may occur via homologous recombination between identical repeats within CRISPR arrays.

2.2 Introduction

Lactobacillus gasseri is an autochthonous organism that frequently colonizes the oral cavity, the human gastrointestinal tract, and the vagina (Dal Bello & Hertel, 2006; Delgado, Suárez, & Mayo, 2007; Hernández-Rodríguez et al., 2011). Due to their many shared morphological and taxonomic characteristics, *L. gasseri* was regularly classified as *L. acidophilus* until 1980 when it could be differentiated via its unique DNA/DNA hybridization patterns (Lauer & Kandler, 1980). Further studies of *L. gasseri* have revealed high intraspecies diversity between strains and a number of niche-related characteristics that enable it to thrive in the gastrointestinal tract, including its ability to survive in highly acidic gastric juice, tolerate the bile salts commonly found in the GIT, and produce mucus-binding proteins that allow it to adhere to intestinal epithelial cells (Azcarate-Peril et al., 2008). Due to its long history of safe consumption, *L. gasseri* is considered a Generally Regarded As Safe (GRAS) commensal organism that is frequently utilized as a probiotic, that "when administered in adequate amounts, confer a health benefit upon the host" (FAO/WHO, 2002; Selle & Klaenhammer, 2013). *L. gasseri* is distinctively equipped to impact human health due to its capability to produce antimicrobial compounds that can limit the presence of detrimental bacteria, decrease the absorption of intestinal oxalate which has been linked to various pathological disorders, and stimulate an immune response that can impact host health and has led to the development of *L. gasseri* as a potential oral vaccine vector (Azcarate-Peril et al., 2008; Selle & Klaenhammer, 2013; Stoeker et al., 2011). As bacteriophages naturally populate the gastrointestinal tract and are a constant threat to many lactobacilli, it is no surprise that a number of temperate bacteriophages specific to *L. gasseri* have been

discovered and studied (Ackermann & Krisch, 1997; Baugher, Durmaz, & Klaenhammer, 2014; Dabrowska, Switala-Jelen, Opolski, Weber-Dabrowska, & Gorski, 2005; Ismail, Neve, Geis, & Heller, 2009; Raya, Kleeman, Luchansky, & Klaenhammer, 1989). One of the more common methods of combatting bacteriophage that is also utilized in *L. gasseri* is the presence of a restriction and modification (R/M) system that can target phage DNA (Azcarate-Peril et al., 2008). A second, more recently-discovered method of limiting phage infection is that of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and its CRISPR-associated (Cas) proteins.

CRISPR-Cas systems provide adaptive, RNA-directed, sequence-specific immunity to bacteria and archaea that protect against invading mobile genetic elements such as phages or plasmids (Barrangou et al., 2007; Brouns et al., 2008; Marraffini & Sontheimer, 2008). All CRISPR-Cas systems drive immunity via a general three-step process: adaption, expression, and interference (Barrangou & Marraffini, 2014). Adaption expands cell immunity through mediating the polarized incorporation of novel spacers into the repeat-spacer array when the cell is faced with new invasive mobile genetic elements (MGE) (Barrangou et al., 2007). Expression takes place as the repeat-spacer array is transcribed and processed into mature crRNAs that will associate with certain Cas proteins to form a crRNA-effector complex (Brouns et al., 2008; Hale et al., 2008). Finally, interference occurs when the crRNA-effector complex recognizes and cleaves matching target sequences (Hale et al., 2009; Garneau et al., 2010).

Lactic acid bacteria possess a rich reservoir of CRISPR-Cas systems, with 62.9% of analyzed genomes possessing CRISPR-Cas as opposed to the general occurrence rate of 46%

(Sun et al., 2015). An unusually high occurrence of Type II systems has also been observed in LABs, with occurence at a rate of 36% as opposed to the general occurrence rate of 5% (Sun et al., 2015). *L. gasseri* closely mirrors the observed general trends in lactic acid bacteria with six out of seventeen (35%) analyzed strains containing Type II CRISPR-Cas systems (Sanozky-Dawes, Selle, O'Flaherty, Klaenhammer, & Barrangou, 2015). Although all CRISPR-Cas systems play a role in providing adaptive immunity to the cell, Type II systems are especially valued due to their overall rarity and potential for exploitation in genome editing and transcriptional control applications. The abundance of Type II systems in *L. gasseri* as well as their confirmed *in vivo* functional activity make *L. gasseri* a prime candidate for further investigation (Sanozky-Dawes et al., 2015).

Type II systems take a minimalistic approach to cell immunity, utilizing only four (frequently three) Cas proteins: Cas1, Cas2, Csn2/Cas4, and Cas9, the "signature gene" of Type II systems. Adaption is driven by Cas1 and Cas2 (and Csn2 or Cas4 when present) which form a complex that can screen and sample invading MGE sequences. Selection of a protospacer for incorporation into the repeat-spacer array is based on the presence of a protospacer adjacent motif (PAM), a unique set of 2-4 nucleotides that flanks appropriate protospacers and marks them as target sequences (Deveau et al., 2008; Mojica et al., 2009; Shah, Erdmann, Mojica, & Garrett, 2013). The PAM also allows CRISPR-Cas to distinguish between host and foreign DNA as the PAM is located on the target but not in the CRISPR locus (Marraffini & Sontheimer, 2010). Cas9 associates with the Cas1-Cas2 complex to bind the PAM sequence to ensure that the correct protospacer is sampled and subsequently incorporated into the repeat-spacer array (Heler et al., 2015). Transcription of the repeat-

spacer array into pre-crRNA is driven by a promoter located in the leader sequence directly upstream of the repeat-spacer array (Hale et al., 2012; Plagens et al., 2012; Carte et al., 2014). The pre-crRNA is processed with the help of two bacterial factors unique to Type II systems, tracrRNA, and RNase III, that assist in producing mature tracrRNA:crRNA complexes; Cas9 then associates with the tracrRNA:crRNA complex to form the effector complex responsible for interference (Briner et al., 2014; Deltcheva et al., 2011; van der Oost et al., 2014). The effector complex searches for invasive MGEs through arbitrarily colliding with random DNA sequences and scanning for PAM sequences (Sternberg, Reddings, Jinek, Greene, & Doudna, 2014). If Cas9 recognizes a PAM sequence, R-loop formation is triggered and the effector complex interrogates the target sequence from 3' to 5' to confirm complementarity with the spacer sequence in the crRNA of the complex (Sternberg et al., 2014). If sufficient complementarity exists between the PAM and the protospacer sequence, the RuvC and HNH endonuclease regions of Cas9 will perform blunt-end, double-stranded cleavage of the target three bases upstream of the PAM (Gasiunas, Barrangou, Horvath, & Siksnys, 2012; Jinek et al., 2012).

Although CRISPR-Cas targeting is generally very efficient, occasionally MGEs can escape targeting by the CRISPR-Cas system. This was demonstrated in *L. gasseri* when the Type II CRISPR-Cas systems of two strains of *L. gasseri* were shown to reduce uptake of a targeted plasmid by three logs when compared to a control (Sanozky-Dawes et al., 2015). However, a number of colonies were still observed that had retained the plasmid despite an active CRISPR-Cas system, demonstrating that failed CRISPR-Cas targeting is very possible (Sanozky-Dawes et al., 2015). MGEs primarily escape targeting via CRISPR-Cas through

mutations or deletions in sequences that are essential to the CRISPR interference process in both the target and/or the host. For example, mutations in the PAM of the target will allow the target to evade CRISPR-Cas targeting through becoming unrecognizable to the CRISPR effector complex (Deveau et al., 2008; Jiang, Bikard, Cox, Zhang, & Marraffini, 2013a; Stern, Keren, Wurtzel, Amitai, & Sorek, 2010; Sun et al., 2012; Vercoe et al., 2013). Previous studies have also demonstrated that deletion of the protospacer or even minute mutations in the seed sequence of the protospacer will prevent Cas cleavage of the target (Deveau et al., 2008; Jiang et al., 2013a; Semenova et al., 2011; Vercoe et al., 2013; Wiedenheft et al., 2011; Zhao et al., 2014). Phages can also interact with each other to undergo recombination-based genome rearrangements that allow them to eliminate CRISPR-targeted sequences (Paez-Espino et al., 2015). Organisms have also been shown to escape CRISPR self-targeting events through eliminating the protospacer via large-scale deletions of their own chromosome through homologous recombination (Selle, Klaenhammer, & Barrangou, 2015). Alternatively, CRISPR-Cas cleavage of a target can be prevented through mutations in the host. In *L. gasseri*, it was shown that transformants could escape CRISPR-Cas targeting when it possessed a repeat-spacer array that had deleted the targeting spacer (Sanozky-Dawes et al., 2015). CRISPR-Cas targeting can also be prevented via loss of all *cas* genes, such as in the case of *Lactobacillus acidophilus* which contains a self-targeting spacer (Stern et al., 2010). A systematic analysis of transformants that had escaped the Type III CRISPR-Cas system in *Staphylococcus epidermidis* revealed deletions of the targeting spacer, partial or complete deletion of the CRISPR locus, mutations and deletions of *cas* genes, and transposon insertions in the *cas* genes (Jiang et al., 2013b).

To date, no work has systematically characterized circumvention of CRISPR-Cas targeting in Type II CRISPR-Cas systems. In this report, mechanisms of CRISPR escape are examined in two strains of *L. gasseri*, NCK 1342 and JV-V03, both of which contain active Type II CRISPR-Cas systems (Sanozky-Dawes et al., 2015). Plasmids that contained a validated protospacer-PAM sequence as well as an antibiotic resistance gene were transformed into both *L. gasseri* strains and transformants were grown up on selective media. Transformants that retained the plasmid due to evasion of CRISPR-Cas targeting were recovered and analyzed for mutations in the protospacer-PAM region of the plasmid, the endonuclease regions of Cas9, and the repeat-spacer array. Through using the endogenous CRISPR-Cas system in the wild-type *L. gasseri* host, we are able to study targeting and escape of CRISPR-Cas as it occurs natively. Collectively, our results indicate that spacer deletions within the CRISPR array are the major mechanism by which Type II systems escape CRISPR-Cas targeting and suggest that CRISPR arrays are in a constant state of flux with spontaneous deletions of spacers occurring at low frequencies that facilitate survival of the cell in a variety of circumstances.

2.3 Materials and Methods

2.3.1 Bacterial strains and culture conditions

All strains and plasmids used in this study are listed in Table 2.1. Bacterial cultures were cryopreserved in a suitable growth medium with 13% glycerol (vol/vol) and stored at -80ºC. *E. coli* strains were propagated with aeration in Luria-Bertani (Difco) broth supplemented with 150 µg ml^{-1} erythromycin (Em) at 37 ºC. *Lactobacillus gasseri* strains

were statically propagated in deMan, Rogosa and Sharpe (MRS) (Difco Laboratories, Inc.) broth in tightly capped tubes or on MRS agar (1.5% [w/v], Difco) under anaerobic conditions at 37 ºC for 72 hours. Selection for transformants took place through supplementing the MRS agar with 5µg ml^{-1} Em (Fisher Scientific, Pittsburg PA).

2.3.2 DNA isolation and manipulations

A plasmid interference model was used to select for mutants that had escaped CRISPR targeting (Figure 2.1). Plasmid DNA containing PAM-protospacer sequences specific to each strain of *L. gasseri* was isolated from the *E. coli* transformant strains using a QIAprep Spin miniprep kit (Qiagen Inc., Valencia, CA). The resulting plasmids, pTRK1090 and pTRK1092, were transformed as previously described into competent *L. gasseri* NCK 1342 and JV-V03, respectively (Walker et al., 1996). Three separate transformations were performed with each strain of *L. gasseri*. After each transformation, cells were recovered overnight in MRS before being plated on selective media. Colonies that had escaped CRISPR targeting and retained the plasmid were labeled and cryopreserved.

2.3.3 DNA Sequencing and Screening of *L. gasseri* Mutants

A selection of colony mutants from each transformation that had escaped CRISPR targeting was designated to undergo genotyping. Ten colonies were selected from the first transformation, fifteen colonies from the second transformation, and 100 colonies were selected from the third transformation. Colony PCR was used to screen these mutant colonies for deletions in the repeat-spacer array, the HNH endonuclease region of *cas9*, the RuvC

endonuclease region of *cas9*, and the protospacer-PAM region of pGK12. PCR amplification of the Type II-A repeat-spacer array in *L. gasseri* JV-V03 and NCK 1342 was performed using primers RSA-F (5'- GCTCAGGTAGGGATGTTAAATG-3') and RSA-R (5'-GATGCACTTAAACCACATAC-3'). PCR amplification of the HNH endonuclease region in Cas9 was performed using primers HNH-F (5'-CGTTGGTGCAGTGTATAATG-3') and HNH-R (5'-TGACGCCATTCGATCTAC-3'). PCR amplification of the RuvC endonuclease region in Cas9 was performed using primers RUVC-F (5'-TCTAATGAATTAGCATCGGAG-3') and RUVC-R (5'-ATAGGGAATCACATTGGTG-3'). The protospacer-PAM region of the plasmid was amplified through using primers pGK12-F (5'-CAAGGGTAAAATGGCCTTTTCC-3') and pGK12-R (5'-ATTCCCTCATACTCCCTTGAG-3'). All of the PCR primers were synthesized by Integrated DNA Technologies (Coralville, IA). Standard PCR protocols were followed using Choice-Taq Blue DNA Polymerase (Denville Scientific Inc., Metuchen, NJ) for amplicon generation. PCR products were analyzed using 1% agarose gel electrophoresis with EZ Vision Three DNA Dye (Fisher Scientific, Pittsburg, PA) as a molecular weight ladder. PCR amplicons that demonstrated a difference in size from the wild-type control were sequenced. DNA sequencing of PCR amplicons was performed by Eton Bioscience (Durham, NC).

2.3.4 Genotyping of CRISPR Repeat-Spacer Arrays

Sequenced repeat-spacer array sequences that had been screened for deletions were depicted and analyzed as previously described using a macro excel program that assigns spacer sequences unique color combinations (Horvath et al., 2008).

2.4 Results

2.4.1 Type II-A CRISPR-Cas systems in *L. gasseri* target plasmids.

A plasmid interference model was used to screen for mutant cells that had escaped CRISPR-Cas targeting in *L. gasseri* JV-V03 and NCK 1342 as both strains have been shown to possess active Type II-A CRISPR-Cas systems that limit plasmid uptake (Sanozky-Dawes et al., 2015). Plasmids that contained an erythromycin resistance gene, a validated PAM sequence, and a protospacer sequence that matched the leader-proximal spacer on either JV-V03 or NCK 1342 were transformed into competent JV-V03 or NCK 1342, respectively (Figure 2.1). The interference complexes of each strain, composed of natively-expressed crRNA, tracrRNA, and Cas9, sought out the protospacer-PAM sequence on the plasmids to perform double-stranded cleavage (Figure 2.1). As transformants were grown on erythromycin-enriched media, cells that had undergone CRISPR-Cas targeting were not able to survive due to loss of the plasmid containing the antibiotic resistance gene. The transformation procedure was independently replicated three times with both strains, and the reduction in transformation efficiencies was compared to that of a control plasmid lacking the protospacer-PAM sequence. Although there was some variation from replication to replication, transformation efficiencies of the test plasmid containing the appropriate protospacer-PAM sequence decreased an average ~3 logs for each strain when compared to the control (Figure 2.2). While CRISPR-Cas targeting in *L. gasseri* JV-V03 and NCK 1342 was confirmed to be effective in reducing plasmid uptake, a number of transformants in the test transformations were observed that had managed to maintain their erythromycin resistance, likely through escaping CRISPR-Cas targeting of the plasmid.

2.4.2 CRISPR-Cas targeting can be circumvented via spacer deletions

A selection of CRISPR transformants that had escaped targeting were screened for mutations in sequences vital to the CRISPR interference process. Fifteen transformant colonies from each strain were selected from each of the first two transformations with the exception of the first transformation in JV-V03, where only seven transformants were present for further analysis. Colony PCR was used to amplify the protospacer-PAM region on the plasmid, the RuvC endonuclease of *cas9*, the HNH endonuclease region of *cas9*, and the host repeat-spacer array. PCR amplicons were then run on an agarose gel with a wild-type control to screen for deletions in these essential areas. Sequencing was used to verify any deletions observed on the gel. No deletions were observed in the protospacer-PAM region on the plasmid nor in the RuvC or HNH endonuclease regions of *cas9*. Although no deletions were indicated, amplicons detailing the PAM-protospacer region were sequenced to verify that no mutations such as single nucleotide polymorphisms (SNPs) were present in the PAM or seed sequence of the protospacer that could affect targeting. Nonetheless, no SNPs were observed in the sequences of either region. On the other hand, multiple deletions were observed in the repeat-spacer arrays of both *L. gasseri* strains in both sets of transformations (Figure 2.4). Sequencing of the repeat-spacer arrays across both strains revealed polarized excisions from the leader end of the repeat-spacer array that always included the targeted leader-proximal spacer. In JV-V03, instances of internal spacer deletions and recombination leading to the copying and pasting of older, ancestral spacers at the forefront of the array were detected as well. Certain deletions patterns were observed more than others, suggesting that some bias may exist in the manner that repeat-spacer array deletions occur.

2.4.3 Spacer deletions are the major mechanism of escape in Type II systems.

Once host repeat-spacer array deletions had been established as a method whereby plasmidscan escape CRISPR-Cas targeting, a third transformation was performed and 110 and 117 transformants were recovered from JV-V03 and NCK 1342, respectively. The transformants from both strains were screened for repeat-spacer array deletions. The percentage of transformants that escaped due to host repeat-spacer array deletions was calculated and compiled from each of the three transformations (Figure 2.3). An overview of all three transformation replications and the analyzed transformants is detailed in Table 2.2. Overall, repeat-spacer array deletions in the host chromosome were shown to be a substantial mechanism whereby *L. gasseri* cells can mutate to avoid CRISPR-Cas targeting of plasmids, accounting for 70% of the transformants in JV-V03 and 52% of the transformants in NCK 1342 (Figure 2.3). However, the difference in percentages between the two strains is not significant, suggesting that strain-type or specific spacers within the array do not play a role in dictating the frequency of repeat-spacer array deletions as a response to CRISPR-Cas targeting.

2.4.4 Certain deletion patterns are consistently observed in mutants that had escaped CRISPR-Cas targeting.

The 110 transformants from JV-V03 and the 117 transformants from NCK1342 from the third transformation were utilized in a quantitative analysis of deletion patterns in the repeat-spacer arrays of both strains. Fifty out of the 110 transformants screened from JV-V03 and 103 out of 117 transformants screened from NCK 1342 demonstrated deletions in their

repeat-spacer array. Repeat-spacer array amplicons from the 50 JV-V03 transformants and the 103 NCK 1342 transformants that had demonstrated deletions in their array were then sequenced and analyzed to provide an overview of common deletion patterns (Figure 2.5). Thirteen deletion patterns were observed in the 50 JV-V03 transformants, whereas only six deletion patterns were observed in the 117 NCK 1342 transformants. This could be due in part to the fact that NCK 1342 has a shorter repeat-spacer array than JV-V03, and thus fewer options for potential deletion patterns (Figure 2.1). Certain deletion patterns were found more commonly than others in both strains. JV-V03 appeared to favor deletions of one spacer and seven spacers, observed 8 times and 10 times, respectively. Both of these deletion patterns had also been observed in previously-screened transformants, with the loss of one spacer being observed in the first set of tested transformants and the loss of seven spacers being observed in the first and second set of transformants. These two deletion patterns account for 36% of the 50 transformants in JV-V03, with the remaining ten deletion patterns being observed six or less times each.

In NCK 1342, the loss of 5 spacers was observed 29 times and the loss of 10 spacers was observed 61 times. These two deletion patterns account for escape from CRISPR-Cas targeting of plasmids in 77% of the 117 transformants analyzed. The remaining four deletion patterns were observed no more than six times each. Although the loss of 5 spacers was also observed in transformants from the second transformation, the loss of 10 spacers was unique to this particular transformation. Interestingly, when the NCK 1342 transformants were grown up in MRS overnight, it was observed that 66% (19/29) of the transformants that had lost five spacers had a broth phenotype of clear media with a translucent pellet on the bottom

of the tube. Similarly, 95% (58/61) of the transformants that had lost ten spacers had a broth phenotype that was turbid throughout. These results indicate that deletions in the repeat-spacer array may impact phenotypic characteristics of the cell.

The 60 transformants from JV-V03 and the 14 transformants from NCK 1342 that did not exhibit deletions in their repeat-spacer array were also screened for deletions in the RuvC and HNH endonuclease regions of *cas9* and the protospacer-PAM region of the plasmid. However, no deletions were observed, confirming prior results that suggested that deletions in *cas9* or the protospacer-PAM region are not the primary method of escape from CRISPR-Cas targeting in *L. gasseri*. Interestingly, in four of the 110 analyzed transformants of JV-V03, no repeat-spacer array could be amplified; furthermore, in three of those four transformants, the endonuclease regions of Cas9 could not be amplified, suggesting that larger-scale deletions of the CRISPR-Cas loci could be a rare mechanism of escape in Type II systems.

2.5 Discussion

Type II CRISPR-Cas systems protect the cell from invasive mobile DNA elements and can be manipulated for use in a variety of applications, including genome editing. Recently it was shown that *L. gasseri*, a microbe that positively impacts human health as a probiotic, possessed Type II CRISPR-Cas systems in six out of seventeen analyzed strains and that the native CRISPR-Cas systems in strains JV-V03 and NCK 1342 were capable of reducing plasmid uptake by 3-4 logs (Sanozky-Dawes et al., 2015). Our experiments confirmed the efficiency of the Type II CRISPR-Cas system in both *L. gasseri* strains, with

plasmid uptake being reduced by an average of 3 logs. However, our primary interest lay in the few cells that had somehow circumvented CRISPR-Cas targeting of the introduced plasmids. We surveyed mutant cells whose plasmids had escaped CRISPR-Cas targeting for deletions in the protospacer-PAM region of the plasmid, the endonuclease regions of Cas9, and the repeat-spacer array. Here we submit the first report of a systematic review of deletion-based CRISPR-Cas circumvention in native Type II systems.

Interestingly, no deletions or insertions were observed in the HNH or RuvC endonuclease regions of Cas9, although such mutations in the interference proteins were commonly observed in the Type III system of *Staphylococcus epidermis* (Jiang et al., 2013b). Complete or partial loss of the CRISPR-Cas loci has also been observed in *S. epidermis* and *L. acidophilus,* accounting for 50% of transformants that had escaped CRISPR-Cas targeting in *S. epidermis* (Jiang et al., 2013b; Stern et al., 2010). However, this appears to occur only very rarely in the Type II systems with only four transformants being observed in JV-V03 that may have escaped via larger CRISPR-Cas deletions and none being observed in NCK 1342. Similarly, no mutations were observed in the protospacer-PAM region of the targeted plasmid. We theorize that this is most likely due to plasmid mutations occurring at a much lower rate than spontaneous mutations of the CRISPR-Cas loci in locations that can inhibit targeting (Jiang et al., 2013b).

Rather, deletions in the repeat-spacer array accounted for over 50% of how transformants from both strains circumvented CRISPR-Cas targeting. This is a much higher occurrence than was observed in Type III systems, where spacer deletion accounted for only 13% of analyzed transformants (Jiang et al., 2013b). Polarized excisions that always included

the leader-proximal spacer that targeted the protospacer-PAM sequence on the plasmid were primarily observed, although cases of internal deletions (separate from the polarized excision) and homologous recombination leading to copying and pasting of ancestral spacers at the leader-proximal end were also witnessed. As Cas9 and the protospacer-PAM region do not appear to be a common area for deletion or insertion-based mutations to occur in Type II systems, it is natural to wonder what accounts for escape of CRISPR-Cas targeting in the remaining transformants that did not contain repeat-spacer array deletions. It is possible that mutations could have also occurred in the tracrRNA or leader regions as we did not screen those sequences. Furthermore, as we primarily screened for deletion or insertion-based mutation events, it is possible that single nucleotide polymorphisms could have occurred in the Cas9 protein sequence that resulted in a nonsense mutation or frameshift that would result in a non-functional effector protein.

Prior studies have suggested that mutations which allow circumvention of CRISPR-Cas targeting occur spontaneously and are selected for via CRISPR-Cas targeting rather than being induced due to transfer of the exogenous DNA element (Gudbergsdottir et al., 2011; Jiang et al., 2013b; Selle et al., 2015). In this case, it would appear that to a certain degree, spontaneous mutations that excise leader-end spacers occur at a low frequency within populations that contain a native Type II system and account for the primary manner in which Type II systems are able to circumvent targeting of a protospacer-PAM sequence. Deletion of spacers is not an uncommon phenomenon in bacteria with CRISPR-Cas systems and has been theorized to play a role in limiting the size of the CRISPR locus (Horvath & Barrangou, 2011; Nozawa et al., 2011; Pourcel, Salvignol, & Vergnaud, 2005; Tyson &

Banfield, 2007). However, an evolutionary bias has been observed that appears to favor deletions of the older ancestral spacers at the opposite end of the repeat-spacer array from the leader sequence; such spacers would be less likely to be benefit the cell as they may provide resistance to exogenous DNA elements that no longer threaten the cell in its current environment (Briner & Barrangou, 2014; Horvath et al., 2008; Horvath & Barrangou, 2011; Weinberger et al., 2012).

The results presented here suggest that deletion of newer leader-proximal spacers also spontaneously occur at low frequencies, despite the supposed evolutionary disadvantage of losing spacers that protect against recently encountered threats. Due to the presence of certain deletion patterns in multiple replications, it appears as though a bias may exist towards certain deletion patterns. Although certain patterns were observed many times within a single replication, it is difficult to determine whether the mutation occurred spontaneously multiple times or whether it occurred once and the cell had time to replicate.

In this study, we examined escape of CRISPR-Cas targeting in a natural setting as opposed to utilizing an engineered system in an artificial host. We were able to examine the natural interplay between the native Type II CRISPR immune system and its target and found that the CRISPR repeat-spacer array spontaneously excises spacer sequences as their primary mechanism of escape from CRISPR targeting. Spontaneous mutations of the CRISPR-Cas system that allow circumvention of CRISPR-Cas targeting in a select subset of the population have been hypothesized to aid in the survival of the overall population in a wide variety of circumstances (Jiang et al., 2013b). This could be due to a majority of the population maintaining its resistance to negative mobile genetic elements such as phages

while allowing another subset to uptake plasmids or other genetic elements that increase its fitness (Jiang et al., 2013b). However, many times these spontaneous mutation inactivate the CRISPR-Cas system entirely (e.g. mutations in *cas* genes, deletion of the CRISPR-Cas loci), which leaves those cells vulnerable to any invasive mobile genetic elements. Here we showed that Type II systems utilize spacer deletions as their primary method of escape which is particularly favorable to the cell, as it allows the cell to uptake beneficial exogenous DNA that contains a target sequence *while* maintaining a functional CRISPR-Cas system. That these deletions can occur anywhere within the repeat-spacer array, even at the leader end with the most recently acquired spacers is especially powerful. Homologous recombination between identical repeats has been hypothesized drive excision of spacers from the repeat-spacer array (Deveau et al., 2008; Gudbergsdottir et al., 2011; Horvath et al., 2008; Horvath & Barrangou, 2011). We also speculate that Cas1, the nuclease responsible for cleaving the repeat sequence during integration of new spacers, could potentially play a role in cleaving repeat sequences during the excision of spacers. Current work is investigating the roles of both mechanisms in spacer deletion. This study advances our understanding of CRISPR-Cas targeting and escape in Type II systems, and has shown that the primary method of escape in the *L. gasseri* Type II system not only allows for conservation of the target sequence but also maintainance of a completely active CRISPR-Cas system.

2.6 Acknowledgements

This work was funded by the North Carolina Agricultural Foundation. The authors thank Dr. Yong Jun Goh, Dr. Sarah O'Flaherty, Dr. Kurt Selle, Allie Briner, and Dr. Brant Johnson for their technical assistance and constructive insights. We are also grateful to Dr. Philippe Horvath at DuPont for providing the CRISPR spacer Excel Macro for visualization of CRISPR spacer patterns.

2.7 References

Ackermann, H.-W., & Krisch, H. M. (1997). A catalogue of T4-type bacteriophages. *Archives of Virology*, 142, 2329-2345.

Azcarate-Peril, M.A., Altermann, E., Goh, Y.J., Tallon, R., Sanozky-Dawes, R.B., Pfeiler, E.A., ... Klaenhammer, T.R. (2008). Analysis of the genome sequence of *Lactobacillus gasseri* ATCC 33323 reveals the molecular basis of an autochthonous intestinal organism. *Applied and Environmental Microbiology*, 74(15), 4610-4625.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., ... Horvarth, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819), 1709-1712.

Barrangou, R. & Marraffini, L.A. (2014). CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Molecular Cell*, 54(2), 234-244.

Baugher, J.L., Durmaz, E., & Klaenhammer, T.R. (2014). Spontaneously induced prophages in *Lactobacillus gasseri* contribute to horizontal gene transfer. *Applied and Environmental Microbiology*, 80(11), 3508-3517.

Briner, A.E. & Barrangou, R. (2014). *Lactobacillus buchneri* genotyping on the basis of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) locus diversity. *Applied and Environmental Microbiology*, 80(3), 994-1001.

Briner, A.E., Donohoue, P.D., Gomaa, A.A., Selle, K., Slorach, E.M., Nye, C.H., ... Barrangou, R. (2014). Guide RNA functional modules direct Cas9 activity and orthogonality. *Molecular Cell*, 53, 333-339.

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., ... van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 321(5891), 960-964.

Carte, J., Christopher, R.T., Smith, J.T., Olson, S., Barrangou, R., Moineau, S., ... Terns, M.P. (2014). The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in *Streptococcus thermophilus*. *Molecular Microbiology*, 93(1), 98-112.

Dabrowska, K., Switala-Jelen, K., Opolski, A., Weber-Dabrowska, B., & Gorski, A. (2005). Bacteriophage penetration in vertebrates. *Journal of Applied Microbiology*, 98, 7-13.

Dal Bello, F. & Hertel C. (2006). Oral cavity as natural reservoir for intestinal lactobacilli. Systematic and Applied Microbiology, 29, 69-76.

Delgado, S., Suárez, A., & Mayo, B. (2007). Dominant cultivable *Lactobacillus* species from the feces of healthy adults in northern Spain. *International Microbiology*, 10, 141-145.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y.J., Pirzada, Z.A., ... Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), 602-607.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., ... Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of Bacteriology*, 190(4), 1390-400.

FAO/WHO. (2002). Guidelines for the evaluation of probiotics in food. FAO/WHO, London, ON, Canada.

Garneau, J.E., Dupuis, M.-E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., ... Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468, 67-71.

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS USA*, 109(39), E2579-E2586.

Gudbergsdottir, S., Deng, L., Chen, Z., Jensen, J.V.K., Jensen, L.R., She, Q., & Garrett, R.A. (2011). Dynamic properties of the *Solfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Molecular Microbiology*, 79(1), 35-49.

Hale, C., Kleppe, K., Terns, R.M., & Terns, M.P. (2008). Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA*, 14, 2572-2579.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., ... Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, 139(5), 945-956.

Hale, C.R., Majumdar, S., Elmore, J., Plister, N., Compton, M., Olson, S., ... Terns, M.P. (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Molecular Cell*, 45(3), 292-302.

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., & Marrafini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaption. *Nature*, 519, 199-202.

Hernández-Rodríguez, C., Romero-González, R., Albani-Companario, M., Figueroa-Damián, R., Meraz-Cruz, N., & Hernández-Guerrero, C. (2011). Vaginal microbiota of healthy pregnant Mexican women is constituted by four Lactobacillus species and several vaginosis-associated bacteria. *Infectious Diseases in Obstetrics and Gynecology*, 2011, 851485.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., ... Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *Journal of Bacteriology*, 190(4), 1401-1412.

Horvath, P. & Barrangou, R. (2011). Protection against Foreign DNA. In G. Storz & R. Hengge (Eds.), *Bacterial Stress Responses*, 2nd edition (333-348). Washington, DC: ASM Press.

Ismail, E.A., Neve, H., Geis, A., & Heller, K.J. (2009). Characterization of temperate *Lactobacillus gasseri* phage LgaI and its impact as prophage on autolysis of its lysogenic strains. *Current Microbiology*, 58, 648-653.

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L.A. (2013a). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, 31, 233-239.

Jiang, W., Maniv, I., Arain, F., Wang, Y., Levin, B.R., & Marraffini, L.A. (2013b). Dealing with the evolutionary downside of CRISPR immunity: Bacteria and beneficial plasmids. *PLoS Genetics*, 9(9), e1003844.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-821.

Kok, J., van der Vossen, J.M., Venema, G. (1984). Construction of plasmid cloning vectors for lactic streptococci which also replicate in *Bacillus subtilis* and *Escherichia coli*. *Applied and Environmental Microbiology*, 48, 726-731.

Lauer, E. & Kandler, O. (1980). *Lactobacillus gasseri* sp. nov., a new species of the subgenus Thermobacterium. *Zentralblatt für Bakteriologie Mikrobiologie und Hygiene: I. Abt. Originale C*, 1, 75-78.

Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in Staphylococci by targeting DNA. *Science*, 322(5909), 1843-1845.

Marrafini, L.A. & Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, 463, 568-571.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, 155, 733-40.

Nozawa, T., Furukawa, N., Aikawa, C., Watanabe, T., Haobam, B., Kurokawa, K., ... Nakagawa, I. (2011). CRISPR inhibition of prophage acquisition of Streptococcus pyogenes. *PLoS ONE*, 6(5), e19543.

Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B.C., Barrangou, R., & Banfield, J.F. (2015). CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. *mBio*, 6(2), e00262-15.

Plagens, A., Tjaden, B., Hagemann, A., Randau, L., & Hensel, R. (2012). Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *Journal of Bacteriology*, 194(10), 2491-2500.

Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151, 653-663.

Raya, R.R., Kleeman, E.G., Luchansky, J.B., & Klaenhammer, T.R. (1989). Characterization of the temperate bacteriophage phi adh and plasmid transduction in *Lactobacillus acidophilus* ADH. *Applied and Environmental Microbiology*, 55(9), 2206-2213.

Sanozky-Dawes, R., Selle, K., O'Flaherty, S., Klaenhammer, K., & Barrangou, R. (2015). Occurrence and activity of a type II CRISPR-Cas system in *Lactobacillus gasseri*. *Microbiology*, 161, 1752-1761.

Selle, K. & Klaenhammer, T.R. (2013). Genomic and phenotypic evidence for probiotic influences of *Lactobacillus gasseri* on human health. *FEMS Microbiology Reviews*, 37(6), 915-935.

Selle, K., Klaenhammer, T.R., & Barrangou, R. (2015). CRISPR-based screening of genomic island excision events in bacteria. *PNAS*, 112(26), 8076-8081.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., ... Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *PNAS*, 108(25), 10098-10103.

Shah, S.A., Erdmann, S., Mojica, F.J., & Garrett, R.A. (2013). Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biology*, 10, 891-9.

Stern, A., Keren, L., Wurtzel, O., Amitai, G., & Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity. *Trends in Genetics*, 26(8), 335-340.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., & Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490), 62-67.

Stoeker, L., Nordone, S., Gunderson, S., Zhang, L., Kajikawa, A., LaVoy, A., ... Dean, G.A. (2011). Assessment of *Lactobacillus gasseri* as a candidate oral vaccine vector. *Clinical and Vaccine Immunology*, 18(11), 1834-1844.

Sun, C.L., Barrangou, R., Thomas, B.C., Horvath, P., Fremaux, C., & Banfield, J.F. (2012). Phage mutations in response to CRISPR diversification in a bacterial population. *Environmental Microbiology*, 15(2), 463-470.

Sun, Z., Harris, H.M.B., McCann, A., Guo, C., Argimón, S., Zhang, W., ... O'Toole, P.W. (2015). Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nature Communications*, 6, 8322.

Tyson, G.W. & Banfield, J.F. (2007). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environmental Microbiology*, 10(1), 200-207.

van der Oost, J., Westra, E.R., Jackson, R.N., & Wiedenheft. B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature Reviews Microbiology*, 12(7), 479-92.

Vercoe, R.B., Chang, J.T., Dy, R.L., Taylor, C., Gristwood, T., Clulow, J.S., ... Fineran, P.C. (2013). Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genetics*, 9, e1003454.

Walker, D.C., Aoyama, K., & Klaenhammer, T.R. (1996). Electrotransformation of *Lactobacillus acidophilus* group A1. *FEMS Microbiology Letters*, 138, 233-237.

Weinberger, A.D., Sun, C.L., Pluciński, M.M., Denef, V.J., Thomas, B.C., Horvath, P., ... Banfield, J.F. (2012). Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Computational Biology*, 8(4), e1002475.

Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., ... Doudna, J.A. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *PNAS*, 108(25), 10092-10097.

Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., & Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature*, 515(7525), 147-150.

Table 2.1 | Bacterial strains and plasmids used in this study

STRAIN or PLASMID	SOURCE	REFERENCE
Strains		
L. gasseri		
NCK 2323 (JV-V03)	Female Genitourinary Tract	BEI Resources
NCK 1342 (JG141)	Patient Endoscopy	Kullen et al., 2000
E. coli		
NCK 2346	MC1061 Transformant with pTRK1090	Sanozky-Dawes et al., 2015
NCK 2348	MC1061 Transformant with pTRK1092	Sanozky-Dawes et al., 2015
Plasmids		
pGK12	*ori* (pWV01), Em^r Cm^r $RepA^+$ Gram-positive shuttle vector	Kok et al., 1984
pTRK 1090	pGK12 with protospacer-PAM sequence that corresponds to the CRISPR locus in *L. gasseri* NCK 1342	Sanozky-Dawes et al., 2015
pTRK 1092	pGK12 with protospacer-PAM sequence that corresponds to the CRISPR locus *in L. gasseri* JV-V03	Sanozky-Dawes et al., 2015

Table 2.2 | Overview of Transformation Replications and Analyzed Transformants

***L. gasseri* strain**	**Rep**	**Analyzed Transfor-mants**	**# that contained CRISPR deletions**	**% of transformants that contained CRISPR deletions**	**# of deletion patterns observed**
JV-V03	1	7	5	71%	1
	2	15	13	87%	6
	3	110	50	45%	12
NCK 1342	1	15	7	47%	2
	2	15	3	12%	2
	3	117	103	88%	6

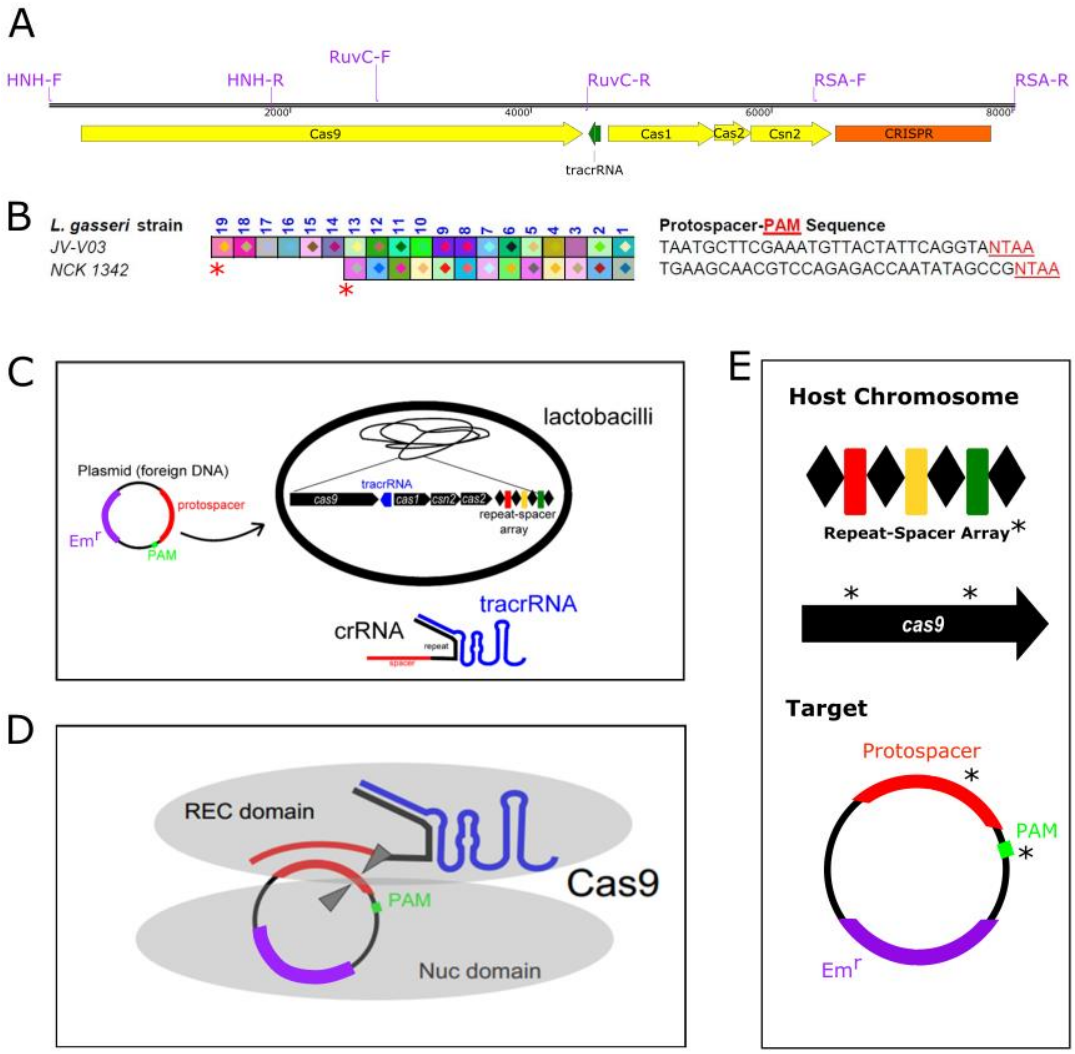

Figure 2.1 | Overview of Methodology

A) CRISPR locus of JV-V03 showing locations of *cas* genes, tracrRNA, array and screening primers B) Visualization of the repeat-spacer array of JV-V03 and NCK 1342. Each square and unique color combination represents the presence of a particular spacer; spacers are numbered in order of acquisition. Red asterisks denote the spacer sequence that was utilized in the target. The protospacer-PAM sequence for both strains is also shown C) Plasmids containing a protospacer-PAM sequence and an Em resistance gene are transformed into a strain with a Type II-A CRISPR-Cas system; meanwhile, crRNA:tracrRNA complexes are generated by the cell that match the protospacer sequence found on the plasmid D) The crRNA:tracrRNA complex associates with Cas9 to form an effector complex; when the effector complex associates with the protospacer-PAM sequence on the plasmid, double-stranded cleavage of the plasmid occurs E) Black asterisks denote regions in the host and target that were screened for mutations.

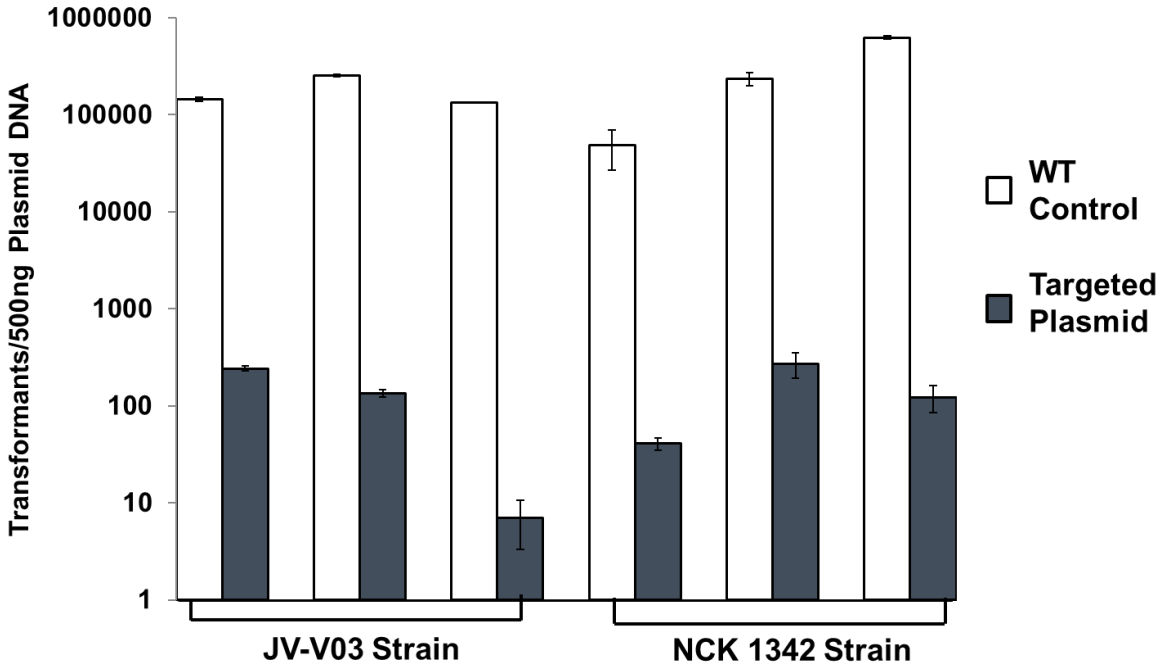

Figure 2.2 | CRISPR Interference and Transformation Efficiencies in *L. gasseri*

L. gasseri strains JV-V03 and NCK 1342 were transformed with 500ng plasmid DNA. The "WT control" was transformed with vector plasmid pGK112 whereas the "Targeted Plasmid" was transformed with an interference plasmid (pTRK1090 or pTRK1092) that contained a protospacer-PAM sequence that corresponded to the particular CRISPR locus of each strain. Three independent biological replications of the transformation were performed and are represented above. Transformation efficiencies of the "Targeted Plasmid" were reduced by an average of ~3 logs in both strains. Error bars represent standard error over technical replicates within each biological replication. All "WT control" means from both strains are significantly different from their corresponding "Targeted Plasmid" means ($p<0.0001$).

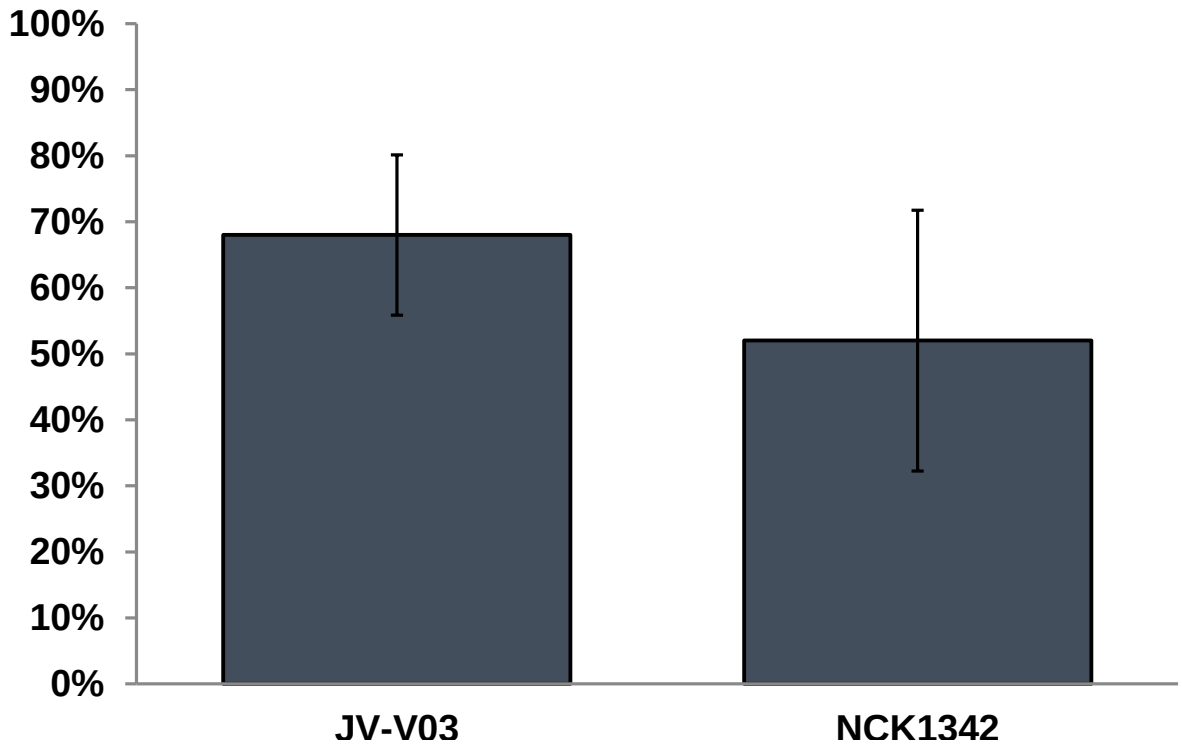

Figure 2.3 | Percentage of Mutations that Occur in the CRISPR Repeat-Spacer Array

A number of transformants from each independent biological transformation (see Table 2.2) was screened for deletions in the repeat-spacer array. The number of transformants that contained repeat-spacer array deletions over the total number of transformants analyzed was calculated for each transformation. The average percentage of transformants that contained spacer deletions is depicted above. Error bars represent standard error over three independent biological experiments. No statistical difference was detected between JV-V03 and NCK 1342 ($p=0.4975$).

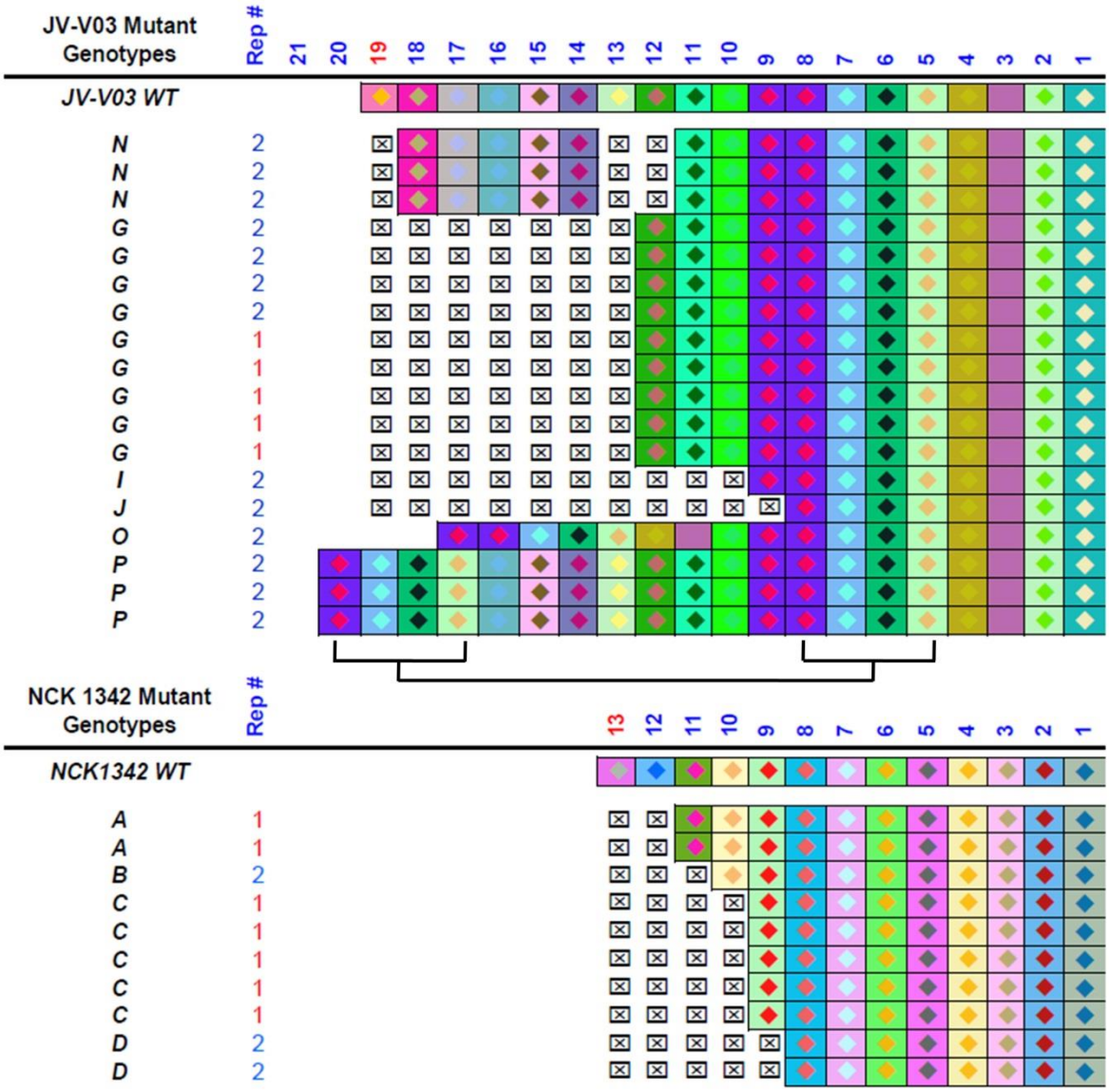

Figure 2.4 | Preliminary Screening and Visualization of CRISPR Spacer Content

CRISPR arrays were visualized from the WT parent strain and mutants that had escaped CRISPR-Cas targeting in *L. gasseri* strains JV-V03 and NCK 1342. Each square, along with its unique color code combination, represents the presence of a particular spacer sequence; spacers are numbered in order of acquisition to the locus. The targeting spacer number is indicated in red. Deletions are presented by "x." Conserved repeats are not represented. Copying and pasting of older spacers to the forefront of the CRISPR array in strain JV-V03 is indicated by the black lines. Mutant spacer sequences were assigned a genotype letter based on their deletion pattern as well as a replication number to designate from which replication they originated.

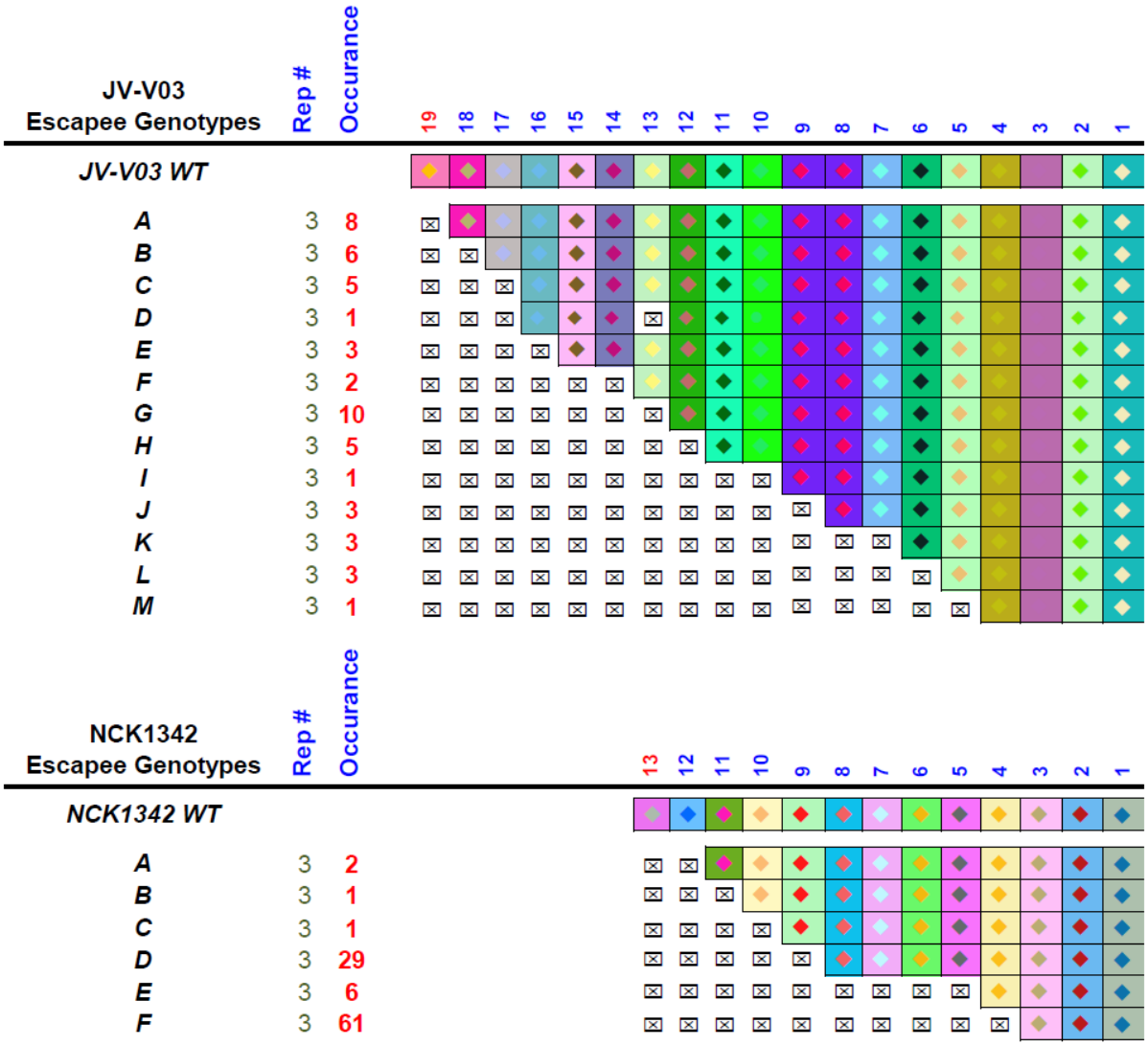

Figure 2.5 | In-Depth Visualization of Mutant Spacer Deletion Genotypes

Over 100 mutants that had escaped CRISPR-Cas targeting were screened for mutations in the repeat-spacer array in *L. gasseri* strains JV-V03 and NCK 1342. Fifty out of 110 mutants analyzed in the JV-V03 strain were found to have deletions in the repeat-spacer array while 103 out of 117 analyzed mutants in the NCK 1342 strain were found to have deletions in the repeat-spacer array. The targeting spacer number is indicated in red. Each deletion pattern was assigned a genotype letter that corresponds with Figure 2.4. Each deletion pattern was also assigned a replication number to designate they were all from replication 3 as well as an occurrence number that indicates how many times the genotype was observed.

CHAPTER 3. Conclusions

3.1 Perspective and Future Directions

Advances in genetics and genomics have revolutionized the way the food industry approaches the field of food microbiology. Additional tools are continually being developed to more effectively maximize the effects of beneficial microbes while minimizing the presence of detrimental microbes in food products. CRISPR-Cas, a DNA-encoded, RNA-mediated, and nucleic acid-cleaving immune system that protects against foreign invasive elements such as phages, is one such tool that could have tremendous impacts on the food industry. CRISPR-Cas systems, found in many bacteria and archaea, mediate immunity via a general three-step process: *adaption* of the repeat-spacer array to incorporate additional spacers, *expression* of the repeat-spacer array and its consequent processing into crRNAs, and *interference* or cleavage of foreign DNA targeted by the crRNA-effector protein complex. An in-depth understanding of CRISPR-Cas biology is critical to the development of additional tools that can be utilized in the industry. Although many of the research applications of CRISPR-Cas have focused on genome editing in eukaryotic systems, the food industry is more than capable of exploiting these tools and technologies for the genesis of next-generation food cultures.

Lactic acid bacteria, commonly utilized as starter cultures or probiotics, possess an especially high occurrence of native CRISPR-Cas systems, placing the starter culture industry in a particularly favorable position to exploit CRISPR-Cas for a number of food science-related applications. Specifically, relevant applications include phage resistance, plasmid vaccination, antimicrobial activity, and genome editing. At the heart of all these applications is the ability of the CRISPR-Cas system to consistently cleave targeted

sequences. However, it has occasionally been observed that targets such as phages or plasmids are able to escape CRISPR-Cas cleavage. Although a number of escape mechanisms have been noted in the literature, the frequency with which these various escape mechanisms occur is poorly understood. Unfortunately, not only does a lack of CRISPR-Cas targeting lead to increased vulnerability of the cell to attacking phages or plasmids in host systems, it also diminishes the ability of the CRISPR-Cas system to be used as an effective antimicrobial or gene editing tool.

Interestingly, data characterizing plasmid CRISPR escape mechanisms in *Lactobacillus gasseri* JV-V03 and NCK 1342 revealed that plasmid targets most commonly escaped Type II CRISPR-Cas targeting due to the cell mutating its host repeat-spacer array. Internal deletions of the repeat-spacer array account for over 50% of failed instances of CRISPR-Cas cleavage. It was also observed that the targeted region on the plasmids as well as the endonuclease regions of Cas9 were unlikely to undergo insertion/deletion mutations. Although this study was key to understanding escape of CRISPR targeting in Type II systems, further studies should done to confirm that these results are consistent across other strains that contain Type II systems. It should also be determined whether these internal deletions occur due to spontaneous mutation events or whether they are induced by internal machinery for survival purposes upon uptake of the plasmid. Finally, the mechanisms mediating these internal deletions in the repeat-spacer array should be further investigated. Internal deletions could be facilitated by RecA during homologous recombination between identical repeats. Alternatively, Cas1, a nuclease responsible for cleaving repeat sequences during incorporation of new spacers, may also play a role in excising spacers.

Although Type II systems are of particular interest due to their potential in genome editing applications, future research examining escape of CRISPR targeting across all six types of CRISPR-Cas systems would greatly clarify the stability of host CRISPR-Cas elements versus the target sequence. In short, CRISPR targeting plays an essential role in almost all CRISPR-Cas applications utilized in the food industry. A better understanding of targeting efficacy of native CRISPR-Cas systems as well as the causes behind circumvention of CRISPR targeting will be essential in designing more effective tools to manage food microbiota.

APPENDICES

Appendix A

Assembly of the CRISPR array and *in vivo* validation of Type II CRISPR-Cas activity in *Lactobacillus jensenii*

A.1 Introduction

Lactobacillus jensenii is a lactic acid bacterium that is a major part of the vaginal microbiota (Martin et al., 2015). *L. jensenii* produces hydrogen peroxide and lactic acid, both of which work to maintain vaginal health by preventing detrimental organisms from populating the vagina (Martin et al, 2015, Lagenaur et al., 2011). The strain of *L. jensenii* used for this project, DSM 20557, is a strain that was isolated from human vaginal discharge by Gasser, Mandel, and Rogosa (1970), and a draft genome of the strain is available (Genome ID: 1423762.3).

The presence of *cas* genes towards the end of scaffold 27 (15,932 nt) in the draft genome of *L. jensenii* DSM 20557 indicates the presence of a CRISPR-Cas system, which provides adaptive immunity to the cell against invading mobile genetic elements (Horvath & Barrangou, 2010; Terns & Terns, 2011). Due to the presence of a *cas9* gene (9585-13790 nt), a *cas1* gene (14013-14921 nt), a *cas2* gene (14878-15204 nt), and a *csn2* gene (15201-15869 nt), *L. jensenii* DSM 20557 presumably contains a Type II-A CRISPR-Cas system. However, scaffold 27 ends soon after the leader promoter sequence and immediately before the customary starting point of a CRISPR repeat-spacer array. The lack of a repeat-spacer array in the draft genome is most likely due to the reads composing the repeat-spacer array region being thrown out during the genome assembly step. Due to the repetitive nature of the palindromic repeats, repeat-spacer arrays are often difficult for computer programs to assemble.

Although the draft genome lacked a CRISPR repeat-spacer array, the tentative presence of a functional Type II CRISPR-Cas system in *L. jensenii* was further investigated

due to its potential for genome editing. The simplicity and utility of the effector complex in Type II systems as well as its ability to perform RNA-guided blunt-end cleavage at very specific sequence-driven locations make them an ideal candidate for genome editing applications (Chylinski et al., 2014). Only six Type II CRISPR-Cas systems have been functionally characterized in terms of *in vivo* activity to date: *Lactobacillus gasseri, Streptococcus thermophilus, Streptococcus agalactiae, Neisseria meningitides, Francisella tularensis* subsp. *novicida,* and *Streptoccocus pyogenes* (Bondy-Demony & Davidson, 2014; Sanozky-Dawes et al., 2015). Three out of these six systems are recognized as human pathogens which may limit future genome editing applications in human cells, agriculture, or food-related bacteria. As *L. jensenii* is a commensal organism that does not exhibit pathogenic activity, it holds great potential as another Type II "scalpel" that can be added to the genome editing toolbox. Characterization of multiple Type II systems delivers added versatility to genome editing applications through providing additional protospacer adjacent motifs (PAM) that can be located next to target sequences. Furthermore, the presence of a wide array of Type II CRISPR-Cas systems in the toolbox provides the opportunity for multiple orthogonal systems to be utilized simultaneously to mediate various activities, such as transcriptional activity or gene editing (Esvelt et al., 2013).

Over the course of this project, the repeat-spacer array for *L. jensenii* DSM 20557 was assembled using small RNA reads which included precursor CRISPR RNA (pre-crRNA) and CRISPR RNA (crRNA) that contain partial repeat and spacer sequences. Once the repeat-spacer array was correctly assembled, we validated a PAM sequence for *L. jensenii* DSM 20557 and determined the functionality of its Type II-A CRISPR-Cas system *in vivo*

through utilizing a plasmid interference assay. Plasmids that contained an erythromycin resistance gene, a tentative protospacer adjacent motif (PAM), and a protospacer sequence matching one of the spacers in the repeat-spacer array were transformed into the DSM 20557 strain. Transformants were recovered on selective media; transformation efficiencies were then calculated and compared to a number of controls to calculate the efficiency of CRISPR-Cas cleavage. Collectively, our results indicate that the Type II-A CRISPR-Cas system in *L. jensenii* DSM 20557 is functional and capable of reducing plasmid uptake by at least 3 logs when the correct PAM sequence (nGG) is present next to the target sequence.

A.2 Materials and Methods

A.2.1 Bacterial strains and culture conditions

All strains and plasmids used in this study are listed in Table A.1. Bacterial cultures were cryopreserved in a suitable growth medium with 13% glycerol (vol/vol) and stored at -80 ºC. E. *coli* strains were propagated with aeration in Luria-Bertani (Difco) broth supplemented with 150 µg ml-1 Em at 37 ºC. *E. coli* transformants were selected on brain heart infusion agar (1.5%, w/v; Difco) with 150 µg erythromycin (Em) ml^{-1} and IPTG/Xgal (Thermo-Fischer). *L. jensenii* was statically propagated in deMan, Rogosa and Sharpe (MRS) (Difco Laboratories, Inc.) broth in tightly capped tubes or on MRS agar (1.5% [w/v], Difco) under anaerobic conditions at 37 ºC for 72 hours. Selection for *L. jensenii* transformants took place through supplementing the MRS agar with 5µg ml-1 Em (Fisher Scientific, Pittsburg PA).

A.2.2 *In silico* analyses

Prior to this project, the small RNAs of *L. jensenii* DSM 20557 were extracted and sequenced. Analysis of data quality was done in Fast QC. The reads underwent trimming and filtering in Geneious to clean up the reads and ensure that all reads had a phred score of over 30. In total, over 10,000,000 small RNA reads were analyzed utilizing Geneious. Reads were extracted that contained the leader sequence or a full repeat/ partial repeat sequence. Spacer fragments were attached to either side of the repeats, and due to repeats that contained partial spacers on either side, groupings of spacers in the correct order could be created. Assembly was attempted from both the leader end as well as the ancestral end. Twenty complete spacers were discovered, and eight groupings of spacers were assembled. A group of three spacers was at the leader end, and another group of three was located at the ancestral end; six groupings of spacers lay between those two ends: one group of 4, one group of 3, three groups of 2, and one spacer that stood alone.

A.2.3 DNA sequencing of *L. jensenii* DSM 20557 repeat-spacer array

Primers were designed to amplify the repeat-spacer array region of *L. jensenii* DSM 20557. The forward primer was located in the *csn2* gene and the reverse primer was located in the first ancestral spacer. Genomic DNA was extracted from *L. jensenii* DSM 20557 with a DNeasy Blood & Tissue Kit (Qiagen) and was utilized with PCR to generate an amplicon (~1,200bp) containing the entire repeat-spacer array. Standard PCR protocols were followed using Q5 High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA). Two additional sequencing primers were designed from spacers that were known to be located

somewhere in the middle of the repeat-spacer array. All PCR primers were synthesized by Integrated DNA Technologies (Coralville, IA) and are listed in Table A.2. DNA sequencing of amplicons was performed at the NC State Genomic Science Lab (Raleigh, NC). Sequence reads were trimmed and assembled on Geneious. The repeat-spacer array was assembled and all repeats and spacers were annotated. The 20 spacers discovered from the RNA-seq reads were verified and placed in their proper order. The newly assembled repeat-spacer array from *L. jensenii* DSM 20557 was visualized and compared to the repeat-spacer arrays of other *L. jensenii* strains as previously described using a macro excel program that assigns spacer sequences unique color combinations (Horvath et al., 2008).

A.2.4 Construction of target plasmids

Prior to this study, a tentative PAM for *L. jensenii* had been predicted through BLASTing known *L. jensenii* spacers against publically available databases. Nucleotide regions that flanked the protospacer hits were extracted and submitted to WebLogo (Crooks, Hon, Chandonia, & Brenner, 2004) for sequence motif identification (Figure A.1). Six tentative PAMs were extrapolated from the WebLogo. Six extended oligos and their reverse complements were ordered from Integrated DNA Technologies that each contained one of the tentative PAM sequences, a protospacer sequence that matched the leader-proximal spacer on the repeat-spacer array of *L. jensenii* DSM 20557, and HindIII/SpeI restriction sites. Two additional oligos and their reverse complements were ordered as controls; one contained a protospacer sequence but no PAM and other contained a PAM but a non-targeting protospacer sequence that was not found in the repeat-spacer array. All oligo

sequences are listed in Table A.3. The oligo sequences and their reverse complement were annealed to form a double-stranded insert that was than digested with HindIII and SpeI. Vector pTRK563 plasmids were obtained from *E. coli* NCK 1123 using a QIAprep Spin miniprep kit (Qiagen) and were also digested with HindIII and SpeI enzymes. The target plasmids were constructed through ligating the six digested protospacer-PAM inserts as well as the two digested control inserts into the digested pTRK563 vector plasmids using Quick T4 DNA Ligase (New England Biolabs). The constructs were transformed into rubidium chloride competent *E. coli* D10 and plated on selective media. The target plasmids were then isolated from the *E. coli* transformants and screening was performed via colony PCR to check for the presence of the insert. Sanger sequencing over the multiple cloning site of the target plasmids was used to verify the PAM and protospacer sequence. Sequencing was performed by the NC State Genomic Science Lab (Raleigh, NC).

A.2.5 Transformation of Target Plasmids into *L. jensenii*

An optimized transformation protocol was developed for *Lactobacillus jensenii* through combining two protocols utilized for *Lactobacillus* species (Heravi, Nasiraii, Sankian, Kermanshahi, & Varasteh, 2012; Speer 2012). Overnight cultures were inoculated into 100 mL of MRS containing 0.5M sucrose and 2% (w/v) glycine to an OD_{600}nm of 0.75-0.10. Cultures were grown to an OD_{600} of 0.45-0.5 when they were pelleted by centrifugation at 4000xg for 5 minutes at 4°C (all centrifugations took place under these conditions). The cell pellet was resuspended in 10mL 50mM EDTA and incubated on ice for 5 minutes before the addition of 40mL cold ddH_2O. The cells were then centrifuged, resuspended in 50mL

cold ddH_2O, re-centrifuged, resuspended in 25mL cold ddH_2O, and re-centrifuged. The cell pellet was then suspended in a final wash of 25mL 0.3M sucrose solution. The cells were centrifuged again and the final pellet was suspended in 1mL of 0.3M sucrose and placed on ice. 100µL of competent cells were added to 600ng of plasmid DNA. The cell and plasmid mixture was then pipetted into a chilled 2mm gap electroporation cuvette. Upon electroporation at a constant voltage of 1.5kV, the cells were transferred to 900µL of pre-warmed MRS to recover for 2-3 hours. Cells were plated on MRS agar supplemented with erythromycin and colonies were counted after they had grown for 2 days under anaerobic conditions. The transformation efficiencies of all target and control plasmids were calculated. Three independent biological replications were performed for all six interference plasmids as well as the three control plasmids (pTRK563 with no insert, pTRK563 with protospacer but no PAM, and pTRK563 with PAM but non-targeting protospacer). Standard error was calculated based on the three replications.

A.3 Results

A.3.1 Identification of repeat-spacer array in *L. jensenii* DSM 20557

Access to small RNA reads from *L. jensenii* DSM 20557 that contained pre-crRNA and crRNA allowed for identification of 20 spacer sequences as well as a tentative order of spacers. Primers were then identified to amplify and sequence the repeat-spacer array to confirm the presence and order of all spacers (Table A.4). The repeat-spacer array of *L. jensenii* DSM 20557 contains 20 spacers and was compared to the repeat-spacer arrays of six other *L. jensenii* strains. Although six genotypes were observed, there were two distinct sets

of strains. DSM 20557 shares multiple spacers with strains SJ-7A-US, 1153, and 269-3 (Figure A.2). DSM 20557 appears to have retained the majority of its ancestral spacers whereas SJ-7A-US, 1153 and 269-3 possessed internal deletions which were largely observed at the trailer end of the repeat-spacer array. The location of the deletions is consistent with previously observed results (Briner & Barrangou, 2014; Horvath et al., 2008; Weinberger et al., 2012). Overall, it appears as though DSM 20557 followed a similar evolutionary path to these three strains although the presence of internal deletions and novel spacers allows for hypervariability between strains. The second set of strains, JV-V16, 272-CHN, and 115-3-CHN do not share any spacer homology with the previous set. Strains 272-CHN and 115-3-CHN contain identical repeat-spacer arrays and share ancestral spacers with JV-V16, suggesting that this second set of strains followed an entirely distinct evolutionary path from the first set of strains.

A.3.2 Functionality of Type II-A CRISPR-Cas system confirmed *in vivo*

A number of components need to be present in a host Type II CRISPR-Cas system for double-stranded cleavage of foreign DNA elements to occur via the effector complex: the Cas9 endonuclease, tracrRNA, and crRNA from the repeat-spacer array. The draft genome of *L. jensenii* DSM 20557 insinuated the existence of a Type II-A CRISPR-Cas system due to the presence of all necessary *cas* genes and a potential tracrRNA sequence. Although the draft genome lacked a repeat-spacer array, the discovery and assembly of a previously unknown 20-spacer array strongly suggested that the CRISPR-Cas system located in DSM 20557 was functional and capable of protecting the strain from mobile genetic elements such

as phages and plasmids. CRISPR-Cas activity was tested via a plasmid interference assay that provided the native system with a target containing a protospacer and a tentative PAM sequence. A WebLogo had previously been created that suggested possible PAM sequences for *L. jensenii* strains and was utilized to select six tentative PAMs for testing. The PAMs along with a known protospacer sequence taken from the leader-proximal spacer in the repeat-spacer array were cloned into a vector plasmid with an Em resistance gene. If the CRISPR-Cas system is functional and the PAM correct, the effector complex would seek out the protospacer, perform double-stranded cleavage of the plasmid, and the cell would lose its ability to survive on media that contains Em. The transformation efficiencies of a targeted plasmid compared to a non-targeted plasmid would be severely reduced when plated on selective media (Gomaa et al., 2014; Marraffini & Sontheimer, 2008; Sanozky-Dawes et al., 2015).

All controls experienced a transformation efficiency of ~3.5 logs upon being plated on selective medium, indicating the necessity of a PAM sequence for cleavage as well as the need for the protospacer sequence to match a sequence found in the repeat-spacer array. Three of the tentative PAMs experienced transformation efficiencies that were reduced by ~3.5 logs (Figure A.3). The PAM sequences were nGGnA, nGG, and nGGC from the Lje1, Lje2, and Lje7 inserts, respectively. These results indicate that we have validated 5'-nGG-3' as the correct PAM sequence for CRISPR-Cas cleavage activity in *L. jensenii* DSM 20557. Double-stranded cleavage of the plasmid was 100% efficient when the nGG PAM was present, with zero colonies being observed when the culture was plated on selective medium. The remaining three PAM sequences experienced transformation efficiencies similar to the

controls, showing that mutations in the PAM sequence will severely affect CRISPR-Cas cleavage. Specifically, these mutated PAMs showed that no cleavage will occur if the first and/or the second G is missing in nGG. The sensitivity of this PAM is consistent with other studies which have found that target sequences can avoid CRISPR-Cas targeting through mutations in the PAM sequence (Jiang et al., 2013; Stern, Keren, Wurtzel, Amitai, & Sorek, 2010; Sun et al., 2012; Vercoe et al., 2013). In short, we have validated a PAM sequenced for a previously unexamined Type II CRISPR-Cas system and verified a high rate of *in vivo* interference activity of that system with transformation efficiencies being reduced by at least 3.5 logs.

A.4 Discussion

Identification and activity of CRISPR-Cas systems are reliant on the presence of relevant *cas* genes, any needed bacterial factors (e.g. tracrRNA, RNaseIII), and a repeat-spacer array. The advances that have been made in genome sequencing and bioinformatics have made it easier to identify the presence of tentative CRISPR-Cas systems in bacterial genomes. However, computer programs often face difficulty in assembling the repeat-spacer array region of the CRISPR locus, which can lead to sequence reads composing the repeat-spacer array being thrown out. Here we show an approach to assembling a repeat-spacer array that utilizes a combination of small RNA reads and DNA sequencing. This methodology could facilitate the detection and assembly of missing or unsequenced repeat-spacer arrays in other bacterial genomes.

Once the presence of all necessary *cas* genes, tracrRNA, and the complete repeat-spacer array had been verified, we validated a PAM sequence for *L. jensenii* and verified *in vivo* plasmid interference activity via the Type II-A CRISPR-Cas system. Plasmid uptake was reduced by ~3.5 logs when a protospacer and valid PAM sequence were present, suggesting that the Type II CRISPR-Cas locus in *L. jensenii* DSM 20557 is both transcribed and biochemically active at functional levels. It is possible that the CRISPR-Cas system could reduce plasmid efficiency by more than 3.5 logs; our study was constrained by the transformation efficiencies of our controls. The fact that zero cells with the correct protospacer-PAM sequence escaped CRISPR targeting suggests that the system may be more active than could be observed in this particular study. It is also possible that the repeat-spacer array of DSM 20557 is resistant to mutation via spacer deletions as it was previously observed that spacer deletions in the host are the primary method by which plasmids escape Type II targeting (Chapter 2). Interestingly, although the repeat-spacer array of DSM 20557 shares ancestral spacers with a set of three other *L. jensenii* strains, it is the only strain in the set to not possess internal deletions (Fig A.2).

The validated PAM for *L. jensenii*, 5'-nGG-3', is not an uncommon PAM being shared by *Streptococcus pyogenes*, *Listeria innocua*, *Streptococcus mutans*, and *Francisella novicida* (Chylinski, Le Rhun, & Charpentier, 2013; Fonfara et al., 2014; Hirano et al., 2016; Jinek et al., 2012; Serbanescu et al., 2015). Nonetheless, it appears to be very stringent with cleavage activity being completely eliminated upon replacement of either G with a different nucleotide. Although popular, the simplicity of this PAM sequence makes it valuable due to the flexibility it offers with selecting target sequences for genome editing. Furthermore, the

shared PAM may provide a greater number of options due to access to a number of Cas9 proteins from various strains that are all capable of cleaving the same target. Different Cas9 proteins could be selected from various strains for specific projects depending on the needs of that project. For example, a number of considerations could include pathogenicity of the strain, cleavage efficiency, frequency of off-target mutagenesis, or the availability of an engineered version for transcriptional control or single-strand nicking. In short, the commensal, non-pathogenic properties of *L. jensenii* DSM 20557 combined with the simplicity of its PAM sequence and the targeting efficacy of its native effector complex suggest that its Type II-A CRISPR-Cas system will be a valuable addition to the toolbox of genome editing.

A.5 References

Bondy-Denomy J. & Davidson A. R. (2014). To acquire or resist: the complex biological effects of CRISPR-Cas systems. Trends in Microbiology, 22, 218-225.

Briner, A.E. & Barrangou, R. (2014). *Lactobacillus buchneri* genotyping on the basis of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) locus diversity. *Applied and Environmental Microbiology*, 80(3), 994-1001.

Chylinski, K., Le Rhun, A., & Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biology*, 10(5), 726-737.

Chylinski, K., Makarova, K.S., Charpentier, E., & Koonin, E.V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research*, 42(10), 6091-6105.

Crooks, G.E., Hon, G., Chandonia, J.M., & Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Research,* 14(6), 1188-1190.

Esvelt, K.M., Mali, P., Braff, J.L.,Moosburner, M., Yaung, S.J. & Church, G.M. (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods*, 10(11), 1116-1121.

Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lécrivain, A.L., Bzdrenga, J., ... Charpentier, E. (2014). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Research,* 42(4), 2577-2590.

Gasser, F., Mandel, M., & Rogosa, M. (1970). *Lactobacillus jensenii* sp. nov., a new representative of the subgenus Thermobacterium. *Journal of General Microbiology*, 62, 219-222. http://www.ncbi.nlm.nih.gov/pubmed/5493596

Gomaa, A. A., Klumpe, H. E., Luo, M. L., Selle, K., Barrangou, R. & Beisel, C. L. (2014). Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *MBio*, 5, e00928-13.

Heravi, R.M., Nasiraii, L.R., Sankian, M., Kermanshahi, H., & Varasteh, A.R. (2012). Optimization and comparison of two electrotransformation methods for lactobacilli. *Biotechnology*, 11, 50-54.

Hirano, H., Gootenberg, J.S., Horii, T., Abudayyeh, O.O., Kimura, M., Hsu, P.D., ... Nureki, O. (2016). Structure and engineering of *Francisella novicida* Cas9. *Cell*, 164(5), 950-961.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., ... Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *Journal of Bacteriology*, 190(4), 1401-1412.

Horvath P. & Barrangou R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science*, 327, 167-170.

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, 31, 233-239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., & Charpentier E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096), 816-821.

Lagenaur, L.A., Lee, P.P., Hamer, D.H., & Sanders-Beer, B.E. (2011). Demonstration of vaginal colonization with GusA-expressing *Lactobacillus jensenii* following oral delivery in rhesus macaques. *Research in Microbiology*, 162(10), 1006-1010.

Marraffini, L. A. & Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science, 322, 1843–1845.

Martin, R., Sanchez, B., Urdaci, M.C., Langella, P., Suarez, J.E., & Bermudez-Humaran, L.G. (2015). Effect of iron on the probiotic properties of the vaginal isolate *Lactobacillus jensenii* CECT 4306. *Microbiology*, 161, 708-718.

Russell, W.M. & Klaenhammer, T.R. (2001). Identification and cloning of gusA, encoding a new β-glucuronidase from *Lactobacillus gasseri* ADH. *Applied and Environmental Microbiology*, 67, 1253-1261.

Serbanescu, M.A., Cordova, M., Krastel, K., Flick, R., Beloglazova, N., Latos, A., ... Cvitkovitch, D.G. (2015). Role of the *Streptococcus mutans* CRISPR-Cas systems in immunity and cell physiology. *Journal of Bacteriology*, 197(4), 749-761.

Sanozky-Dawes , R., Selle, K., O'Flaherty, S., Klaenhammer, T., & Barrangou, R. (2015). Occurrence and activity of a type II CRISPR-Cas system in *Lactobacillus gasseri*. *Microbiology*, 161, 1752-1761,

Speer, M. (2012). Lactobacillus transformation (Speer 2012). *OpenWetWare*. Retrieved from http://openwetware.org/index.php?title=Lactobacillus_transformation_%28Speer_2012%29&oldid=567428.

Stern, A., Keren, L., Wurtzel, O., Amitai, G., & Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity. *Trends in Genetics*, 26(8), 335-340.

Sun, C.L., Barrangou, R., Thomas, B.C., Horvath, P., Fremaux, C., & Banfield, J.F. (2012). Phage mutations in response to CRISPR diversification in a bacterial population. *Environmental Microbiology*, 15(2), 463-470.

Terns M.P. & Terns R.M. (2011). CRISPR-based adaptive immune systems. *Current Opinion in Microbiology*, 14, 321-327.

Vercoe, R.B., Chang, J.T., Dy, R.L., Taylor, C., Gristwood, T., Clulow, J.S., ... Fineran, P.C. (2013). Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genetics*, 9, e1003454.

Weinberger, A.D., Sun, C.L., Pluciński, M.M., Denef, V.J., Thomas, B.C., Horvath, P., ... Banfield, J.F. (2012). Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Computational Biology*, 8(4), e1002475.

Table A.1 | Bacterial strains and plasmids used in this study

STRAIN or PLASMID	SOURCE/CHARACTERISTICS	REFERENCE
Strains		
L. jensenii DSM 20557	Human vaginal discharge	Gasser, Mandel, & Rogosa, 1970
E. coli NCK 1123	DH5α strain with pTRK563	Russell & Klaenhammer, 2001.
Plasmids		
pTRK563	Em^r; Δcat derivative of pGK12 with lacZ	Russell & Klaenhammer, 2001.
pTRK563-Lje1	pTRK563 with Lje1 insert	This study
pTRK563-Lje2	pTRK563 with Lje2 insert	This study
pTRK563-Lje3	pTRK563 with Lje3 insert	This study
pTRK563-Lje4	pTRK563 with Lje4 insert	This study
pTRK563-Lje5	pTRK563 with Lje5 insert	This study
pTRK563-Lje6	pTRK563 with Lje6 insert	This study
pTRK563-Lje7	pTRK563 with Lje7 insert	This study
pTRK563-Lje8	pTRK563 with Lje8 insert	This study

Table A.2 | List of Primers

Primers	**Sequence (5’ to 3’)**	**Amplification/Analysis Region**
Lje.RSA-F	GAGGATTTTGTCGACTGGTA	Repeat-Spacer Array
Lje.RSA-R	ACTAACCTTAAATATTCATCGACTAC	Repeat-Spacer Array
Lje.SpacerX-R	GCTTGACAGAGAAATTGAAAGTAG	Repeat-Spacer Array
Lje.SpacerY.R	CATCTAATGCTTCAGGCTTG	Repeat-Spacer Array
pTRK563-F	CCAGGGTTTTCCCAGTCACG	Multiple cloning site
pTRK563-R	TCACTCATTAGGCACCCCAG	Multiple cloning site

Table A.3 | Lje Insert Oligonucleotide Sequences

Insert Name	Tested PAM	5' Flank and HindIII Restriction Site	Protospacer	PAM	SpeI Restriction Site and 3' Flank
Lje1	No PAM	ATAAAGCTTC TCGAG	ATTAGGATACTGTGTAGGATATACCTTCCA	ATTTTT	CTGCAGACT AGTATA
Lje2	nGGnA	ATAAAGCTTC TCGAG	ATTAGGATACTGTGTAGGATATACCTTCCA	TGGCAA	CTGCAGACT AGTATA
Lje3	nGG	ATAAAGCTTC TCGAG	ATTAGGATACTGTGTAGGATATACCTTCCA	TGGTTT	CTGCAGACT AGTATA
Lje4	nnGnA	ATAAAGCTTC TCGAG	ATTAGGATACTGTGTAGGATATACCTTCCA	TTGTAA	CTGCAGACT AGTATA
Lje5	nnnnA	ATAAAGCTTC TCGAG	ATTAGGATACTGTGTAGGATATACCTTCCA	TTTTAA	CTGCAGACT AGTATA
Lje6	nGnnA	ATAAAGCTTC TCGAG	ATTAGGATACTGTGTAGGATATACCTTCCA	TGTTAA	CTGCAGACT AGTATA
Lje7	nGGC	ATAAAGCTTC TCGAG	ATTAGGATACTGTGTAGGATATACCTTCCA	TGGCTT	CTGCAGACT AGTATA
Lje8	nGGnA	ATAAAGCTTC TCGAG	ACGACATAGAGGTTATTCATTCCATTTCA C	TGGCAA	CTGCAGACT AGTATA

Table A.4 | *L. jensenii* DSM 20557 CRISPR-Cas leader, repeat, and spacer sequences

Name	Sequence
CRISPR Leader Sequence	TTTTAAATCTGAGAAGCTGATGATAAAATGTCG
Repeat Units	GTTTTAGAAGGTTGTTAAATCAGTAAGTTGAAAAAC
Spacer 1 (newest)	ATTAGGATACTGTGTAGGATATACCTTCCA
Spacer 2	TTGTAAAATTTGTGATAATGATTTTCGTGT
Spacer 3	TTCCAATGGTGAAACAAGAGTTATCAAATA
Spacer 4	TGTTGCAAGCGCCGACCCTTATGGTGGTAT
Spacer 5	AGTAGCTCAACAACTTCAAAGCGAAGTTGG
Spacer 6	TTTTTGAGCTGTTTTATCAAGCCAAGAAGG
Spacer 7	GGATACTGGCAGGCCTTCCATTTATCAGTT
Spacer 8	GGATACTGGCAGGCCTTCCATTTATCAGTT
Spacer 9	TGTTTAAAGTAGTAAACAAGACTATTGAGC
Spacer 10	CTCTATTTTAATCTGCATACATACATGTAA
Spacer 11	AAGACAATACGTAATAAGGATATTGCACAA
Spacer 12	AATATATCAAGCCTGAAGCATTAGATGGCT
Spacer 13	ATATCTTTGTGGTTCAACATGGTTGTGTAT
Spacer 14	GATGCAGACCATGCATCCCTGGTACACCAC
Spacer 15	TAAGCTTGAATTAGCATTAAGAGAAGGTTA
Spacer 16	ACTACTTTCAATTTCTCTGTCAAGCTTCTT
Spacer 17	TACGCCTAGTGGCTATATTGACTTAACTAG
Spacer 18	AAATTTTTGCAGGAAATTGCCGATGAGCTA
Spacer 19	GAAAACTTGAAGTCGCAACTCAATAGCGAG
Spacer 20 (ancestral)	ATTAGGATACTGTGTAGGATATACCTTCCA

Figure A.1 | WebLogo sequence motif identification for tentative PAMs in *L. jensenii*

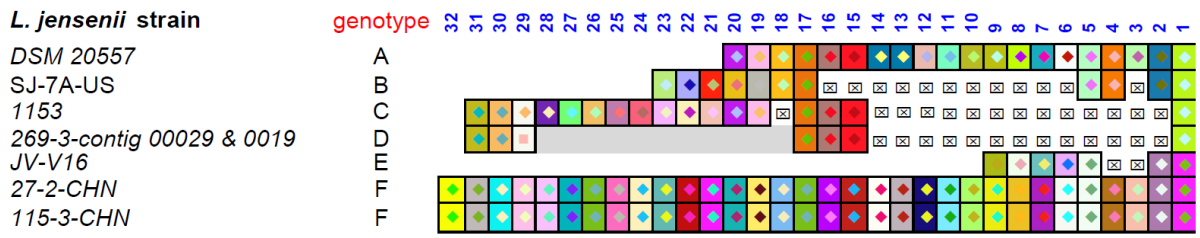

Figure A.2 | Comparison of CRISPR-Cas Spacer Content in *L. jensenii* strains

The assembled CRISPR array of *L. jensenii* DSM 20557 was visualized along with the CRISPR arrays of six other *L. jensenii* strains. Each square with its unique color code combination represents the presence of a particular spacer sequence; spacers are numbered in order of acquisition to the locus. Deletions are presented by "x." Conserved repeats are not represented. Each unique CRISPR array was assigned a genotype.

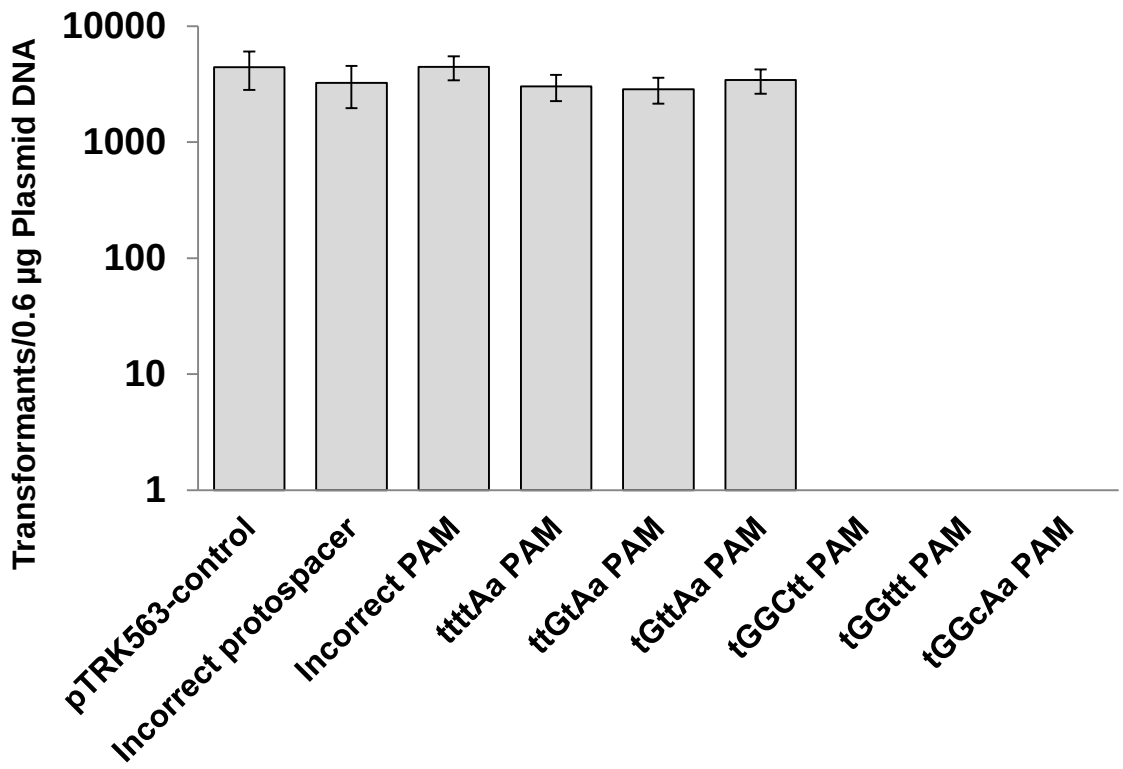

Figure A.3 | CRISPR interference and Transformation Efficiencies

Plasmids containing an insert with a protospacer sequence and one of six PAM variants were transformed into *L. jensenii* DSM 20557. A plasmid with no insert, a plasmid with a PAM but an incorrect protospacer, and a plasmid with a protospacer but an incorrect PAM were also transformed into *L. jensenii* as controls. The transformation efficiencies of the various plasmids to validate a PAM sequence and determine interference activity of the Type II-A system in *L. jensenii*. Standard error was calculated from three independent biological replications.

Appendix B

Isolation of Rough and Smooth Morphotypes from *Lactobacillus buchneri* CD034

B.1 Introduction

Lactobacillus buchneri is an obligate heterfermentative lactic acid bacterium that is often utilized in food and animal feed fermentations (Heinl et al., 2012). It has been found in food products such as pickled juice (Zeng , Pan, & Guo, 2010), acidcoagulating cheese (Radovanovic & Katic, 2009) and spoiled cucumber fermentations (Franco, Perez-Diaz, Johanningsmeier, & McFeeters, 2012; Johanningsmeier, Franco, Perez-Diaz, & McFeeters, 2012). It has also been isolated from an industrial ethanol fermentation plant (Liu et al., 2011) and the human gastrointestinal tract (Koll et al., 2010). As *L. buchneri* commonly inhabits decomposing plants and is frequently utilized as a silage starter culture, it is no surprise that *L. buchneri* CD034 was isolated from stable grass silage (Heinl et al., 2012). Additional insights into the biology and fermentation capabilities of *L. buchneri* CD034 can be extrapolated from its completely sequenced genome, making it an ideal organism for further study.

Upon plating of the CD034 strain, two colony morphotypes, rough and smooth, were observed. Contamination was assumed to be responsible. Various freezer stocks of the strain (the lab seed stock and personal working stocks) were plated to determine the origin of the contamination. However, all of the freezer stocks consistently displayed the two morphotypes. Over the course of a month of continuous re-streaking, the two morphotypes were separated and stocked as "Lbu Rough" and Lbu Smooth." A number of genomic and molecular analyses were performed to characterize the two morphotypes, including description of their phenotypic characteristics, microscopic observations, 16s rDNA genotyping, Diversilab typing, and CRISPR array analysis. Results indicate that

contamination is not responsible for the two morphotypes; rather, *L. buchneri* CD034 is a pleomorphic strain with distinct rough and smooth phenotypes.

B.2 Phenotypic Description and Microscopy Observations

Rough and smooth *L. buchneri* CD034 morphotypes were statically propagated on deMan, Rogosa, and Sharpe (MRS) agar (1.5% [w/v], Difco Laboratories, Inc.) or in MRS broth (Difco). When plated on MRS agar under anaerobic conditions, the smooth morphotype exhibits a circular, raised white colony, with a ropy, sticky consistency. The rough morphotype colony is more of an off-white color with fuzzy, irregular edges. It is more flat than the smooth and has a very dry, almost waxy/granular consistency. When colonies were examined under a microscope, the smooth morphotype consisted of consistently distributed long rods whereas the rough morphotype possessed many shorter rods. The two morphotypes were also plated under aerobic conditions. The smooth morphotype colonies plated under aerobic conditions looked identical to those that had been plated under anaerobic conditions. Interestingly, when the rough morphotype was plated under aerobic conditions, the colonies were a unique blend of both the rough and smooth morphotypes. The colonies were very white and had a stickier, ropier consistency; however, they also had slightly rough/fuzzy edges. This suggests that the presence of oxygen may play a role in the conversion of the rough morphotypes to smooth (Figure B.1).

A colony from the A) smooth morphotype grown under anaerobic conditions, B) the rough morphotype grown under anaerobic conditions, and C) the rough morphotype grown under aerobic conditions was selected to inoculate three separate tubes of 9-10 mL MRS. The

tubes were all capped tightly and allowed to grow overnight at 37 degrees Celsius. The smooth morphotype culture (A) was very turbid throughout and when the culture was examined under the microscope, both long and short rods were evenly distributed. The rough morphotype culture from a colony grown under anaerobic conditions (B) contained clear supernatant with large white clumps of aggregated growth that were often suspended in the culture. The rough morphotype culture from a colony grown under aerobic conditions (C) looked identical to the other rough morphotype culture. It is possible that the cells reverted back to their rough phenotype due to the lack of oxygen. When both rough morphotype cultures were examined under the microscope, they both displayed rods that were intertwined with each other to form clumps, and there were very few free-floating rods (Figure B.2).

Both the rough and smooth colony and culture morphology types from CD034 have been observed previously in *L. buchneri* strains; in a study characterizing 36 *L. buchneri* isolates, eight colony morphologies and four culture cellular morphologies were documented (Figure B.3, Figure B.4) (Daughtry, 2015). The smooth colony morphotype from CD034 closely resembles the colony morphology from isolate LA0030, with its white color, circular form, convex elevation, entire margins, mucoid surface, and butyrous consistency. On the other hand, the rough colony morphotype from CD034 closely resembles the colony morphology from isolate LA1167, with an off-white color, irregular form, undulate margin, granular surface, and powdery texture. When grown in a culture, the cellular morphology of the smooth and rough morphologies from CD034 also mirrors those of LA0030 and LA1167, respectively (Figure B.4). The smooth morphology from CD034 and LA0030 possess a cellular morphology that consists of rods of a variety of sizes that are evenly spread

throughout. The rough morphology from CD034 and LA1167 have a cellular morphology of rods that have aggregated together to form clumps with very few free-floating rods.

The fact that the rough morphotype changed its phenotype slightly when exposed to oxygen suggests that the CD034 strain is capable of pleomorphic behavior. The presence of both phenotypes in other *L. buchneri* isolates suggests that both the rough and smooth phenotypes are not unusual for *L. buchneri* and it is possible that CD034 could exhibit both phenotypes. However, at this point of the study, it was also possible that the presence of the two phenotypes in the CD034 strain could be due to contaminants such as other *L. buchneri* strains as both phenotypes could be accounted for by various *L. buchneri* strains.

B.3 Molecular and Genomic Analyses

Both the rough and smooth morphotypes were grown up in MRS for 16 hours in a tightly sealed tube at 37°C. A genomic DNA extraction was performed using an UltraClean microbial DNA isolation kit (Mo Bio) and quantified using a Nanodrop 1000 spectrophotometer (Thermo-Scientific). The genomic DNA was utilized to amplify and sequence the 16S rDNA gene and the CRISPR repeat-spacer array. Standard PCR protocols were followed using High-Fidelity 2x Master Mix; primers are listed in Table B.1. All sequencing was performed by Eton BioSciences. The genomic DNA was also utilized in Diversilab typing to compare the rough and smooth morphotypes to each other.

16S rRNA gene sequences have long been utilized as a marker gene to identify microbial taxa due to their high level of sequence conservation and ubiquitous presence in prokaryotic cells (Clarridge, 2004; Garrity & Holt, 2001). The conserved and variable

regions of the gene, which evolve at various rates, possess enough interspecific polymorphisms to infer phylogenetic relationships in bacteria and assign taxonomy at the genus and oftentimes the species level (Clarridge, 2004; Smit, Widmann, & Knight, 2007; Srinivasan et al., 2015). Approximately 500 base pairs of the 16S rRNA gene containing the variable v1 and v2 regions were amplified and sequenced. The 16S rRNA gene sequences for the rough and smooth morphotypes were identical. They shared 100% homology to the published *L. buchneri* CD034 strain (Accession No.: CP003043) 16S rDNA sequence. Although this high sequence homology to the CD034 strain does not necessarily confirm that the rough and smooth morphotypes are definitively the CD034 strain, it does indicate that both morphotypes belong to the *L. buchneri* species.

Rep-PCR *Lactobacillus*-based DiversiLab typing (bioMérieux, Durham, NC) was utilized to type the rough and smooth morphotype for comparison purposes against each other to confirm whether they belong to the same strain. First, rep-PCR for amplification of DNA was performed using the Lactobacillus DiversiLab kit (bioMérieux), AmpliTaq DNA polymerase (Carlsbad, CA), and a Bio-Rad MyCycler thermal cycler (Bio-Rad, Hercules, CA). The thermocycler was programmed for an initial denaturation at 94°C for 120 seconds, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, and extension at 70°C for 90 seconds. A final extension step was performed at 70°C for 180 seconds. The reaction mixture was then added along with the appropriate DiversiLab DNA reagents and supplies to the DiversiLab system chip according to manufacturer's instructions. The DiversiLab examined the chip samples using software version 3.4 and a unique barcode was generated for each sample. The barcodes for both the rough and smooth

morphotypes were identical to each other, further confirming that both morphotypes are likely two distinct phenotypes of the same strain (Figure B.5).

The adaptive nature of CRISPR-Cas immunity leads to a CRISPR array that naturally evolves over time. As the cell encounters outside threats, novel spacers are added in a chronological manner to the leader-proximal end of the repeat-spacer array (Andersson & Banfield, 2008; Tyson & Banfield, 2008). As these threats are often driven by a particular time or environment, the spacers can present a dynamic picture of strain relatedness and divergence. CRISPR repeat-spacer arrays are a powerful tool in the identification of strains and have been utilized to genotype multiple bacterial strains relevant to the food industry (Barrangou & Horvath, 2012; Barrangou & Dudley, 2016; Horvath et al., 2008; Sun et al., 2015). Moreover, CRISPR-based genotyping can be instrumental in distinguishing different clones derived from a single strain, making it a useful tool for analysis and identification of the rough and smooth morphotypes. CRISPR-based genotyping has already been utilized to strain-type 26 industry isolates of *L. buchneri* and compare them to sequenced *L. buchneri* strains CD034 and NRRL-30929 (Briner & Barrangou, 2014) (Figure B.6). This study confirmed hypervariability between even closely related *L. buchneri* strains and showed that strain CD034 had a unique repeat-spacer array pattern of 16 spacers that was not duplicated in any of the other 27 *L. buchneri* strains (Briner & Barrangou, 2014). When the repeat-spacer array of both the smooth and rough morphotypes were amplified and sequenced, sequence analysis revealed that both morphotypes contained identical repeat-spacer arrays that possessed 16 spacers. The repeat-spacer arrays of both the rough and smooth morphotype were identical to the published repeat-spacer array of *L. buchneri* CD034 and

distinct from the repeat-spacer arrays of all the other *L. buchneri* strains, which strongly indicates that the rough and smooth morphotype are both phenotypes of the CD034 strain (Briner & Barrangou, 2014) (Figure B.6).

B.4 Conclusion

Pleomorphism within single-strain populations has been noted multiple times in lactic acid bacteria populations (Kandler & Weiss, 1986; Klaenhammer & Kleeman, 1981; Rogosa & Mitchell, 1950; Reinheimer, Morelli, Bottazzi, & Suarez, 1995; Tareb, Bernardeau, Horvath, & Vernoux, 2015). Pleomorphic variations have been theorized to play an important role in strain adaption and survival in fluctuating environments that may experience changes in oxygen presence, pH, temperature, or growth medium (Davidson & Surette, 2008; Reinheimer et al., 1995; Wolf, Vazirani, & Arkin, 2005). In this study, *L. buchneri* CD034 was identified as a lactic acid-producing pleomorphic strain with a rough and smooth phenotype. Each phenotype possesses distinct physical characteristics (both macro- and microscopically) that can be observed when grown up on a plate or in a culture. 16S rRNA sequencing revealed both phenotypes to belong to the *L. buchneri* species and DiversiLab typing did not detect any differences between the genomes of the two morphotypes. CRISPR-Cas genotyping established that the CRISPR arrays of both morphotypes were identical to the array of the CD034 strain, further confirming that both morphotypes are variants of the CD034 strain. Although the cause for this pleomorphic behavior was not identified, the presence of oxygen appeared to affect the rough phenotype with colonies adopting some of the smooth phenotype qualities when grown on MRS plates.

B.5 References

Andersson, A.F. & Banfield, J.F. (2008). Virus Population Dynamics and Acquired Virus Resistance in Natural Microbial Communities. *Science*, 320(5879), 1047-1050.

Barrangou, R. & Horvath, P. (2012). CRISPR: New horizons in phage resistance and strain identification. *Annual Review of Food Science and Technology*, 3, 143-162.

Barrangou, R. & Dudley, E.G. (2016). CRISPR-based typing and next-generation tracking. *Annual Review of Food Science and Technology*, 7, 395-411.

Briner, A.E. & Barrangou, R. (2014). *Lactobacillus buchneri* genotyping on the basis of Clustered regularly interspaced short palindromic repeat (CRISPR) locus diversity. *Applied and Environmental Microbiology*, 80(3), 994-1001.

Clarridge, III J.E. (2004). Impact of 16s rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clinical Microbiology Reviews*, 17(4), 840-862.

Davidson, C.J. & Surette, M.G. (2008). Individuality in bacteria. *Annual Review of Genetics*, 42, 253-268.

Daughtry, K.V. (2015). *Phenotypic and genotypic characterization of* Lactobacillus buchneri *strains isolated from spoiled fermented cucumber*. (Master's thesis). Retrieved from NCSU Institutional Repository (Theses and Dissertations).

Franco, W., Perez-Diaz, I.M., Johanningsmeier, S.D., & McFeeters, R.F. (2012). Characteristics of spoilage-associated secondary cucumber fermentation. *Applied and Environmental Microbiology*, 78(4), 1273-1284.

Garrity, G.M. & Holt, J.G. (2001). The road map to the manual. In D.R. Boone & R.W. Castenholz (Eds.), *Bergey's Manual of Systematic Bacteriology* (119-141). Springer: New York.

Heinl, S., Wibberg, D., Eikmeyer, F., Szczepanowski, R., Blom, J., Linke, B., ... Schluter, A. (2012). Insights into the completely annotated genome of *Lactobacillus buchneri* CD034, a strain isolated from stable grass silage. *Journal of Biotechnology*, 161(2), 153-166.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., ... Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus. Journal of Bacteriology*, 190(4), 1401-1412.

Johanningsmeier, S.D., Franco, W., Perez-Diaz, I., & McFeeters, R.F. (2012). Influence of sodium chloride, pH and lactic acid bacteria on anaerobic lactic acid utilization during fermented cucumber spoilage. *Journal of Food Science*, 77(7), M397-404.

Kandler, O. & Weiss, N. (1986). Genus Lactobacillus Beijerinck 1901, 212. In P.H.A. Sneath, N.S. Mair, M.E. Sharpe, J.G. Holt (Eds.), Bergey's Manual of Systematic Bacteriology (pp. 1209-1234). Williams & Wilkins: Baltimore.

Klaenhammer, T.R. & Kleeman, E.G. (1981). Growth characteristics, bile sensitivity, and freeze damage in colonial variants of *Lactobacillus acidophilus. Applied and Environmental Biology*, 41, 1461-1467.

Koll, P., Mandar, R., Smidt, I., Hutt, P., Truusalu, K., Mikelsaar, R.H.., ... Mikelsaar, M. (2010). Screening and evaluation of human intestinal lactobacilli for the development of novel gastrointestinal probiotics. *Current Microbiology*, 61(6), 560-566.

Kullen, M.J., Sanozky-Dawes, R.B., Crowell, D.C., & Klaenhammer, T.R. (2000). Use of the DNA sequence of variable regions of the 16S rRNA gene for rapid and accurate identification of bacteria in the Lactobacillus acidophilus complex. Journal of Applied Microbiology, 89(3), 511-516.

Liu, S., Leathers, T.D., Copeland, A., Chertkov, O., Goodwin, L., & Mills, D.A. (2011). Complete genome sequence of *Lactobacillus buchneri* NRRL B-30929, a novel strain from a commercial ethanol plant. *Journal of Bacteriology*, 193(15), 4019-4020.

Radovanovic, R.S. & Katic, V. (2009). Influence of lactic acid bacteria isolates on *Staphylococcus aureus* growth in skimmed milk. *Bulgarian Journal of Agricultural Science*, 15(3), 196-203.

Reinheimer, J.A., Morelli, L., Bottazzi, V., & Suarez, V. (1995). Phenotypic variability among cells of *Lactobacillus helveticus* ATCC 15807. *International Dairy Journal*, 5, 97-103.

Rogosa, M. & Mitchell, J.A. (1950). Induced colonial variation of a total population among certain lactobacilli. *Journal of Bacteriology*, 59, 303-308.

Smit, S., Widmann, J., & Knight, R. (2007). Evolutionary rates vary among rRNA structural elements. *Nucleic Acids Research*, 35(10), 3339-3354.

Srinivasan, R., Karaoz, U., Volegova, M., MacKichan, J., Kato-Maeda, M., Miller, S., ... Lynch, S.V. (2015). Use of 16s rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. *PLoS ONE*, 10(2), e0117617.

Sun, Z., Harris, H.M.B., McCann, A., Guo, C., Argimón, S., Zhang, W., ... O'Toole, P.W. (2015). Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nature Communications*, 6, 8322.

Tareb, R., Bernardeau, M., Horvath, P., & Vernoux, J.-P. (2015). Rough and smooth morphotypes isolated from *Lactobacillus farciminis* CNCM I-3699 are two closely-related variants. *International Journal of Food Microbiology*, 193, 82-90.

Tyson, G.W. & Banfield, J.W. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environmental Microbiology*, 10(1), 200-207.

Wolf, D.M., Vazirani, W., & Arkin, A.P. (2005). Diversity in times of adversity: Probalistic strategies in microbial survival games. *Journal of Theoretical Biology*, 234(2), 227-253.

Zeng, X.Q., Pan, D.D., & Guo, Y.X. (2010). The probiotic properties of *Lactobacillus buchneri* P2. *Journal of Applied Microbiology*, 108(6), 2059-2066.

Table B.1 | List of PCR Primers

Primer Name[1]	Primer Sequence (5' to 3')	Source	Amplification Region
Lbu.16S-F	AGAGTTTGATCCTGGCTCAG	Kullen et al., 2000	16s rRNA gene
Lbu.16S-R	GGCTGCTGGCACGTAGTTAG	Kullen et al., 2000	16s rRNA gene
Lbu.RSA-F	CCAGAATGAATGATCTGTTG	Briner et al., 2014	CRISPR array
Lbu.RSA-R	CATCGACGAGAACTTTG	Briner et al., 2014	CRISPR array

[1] F, forward; R, reverse

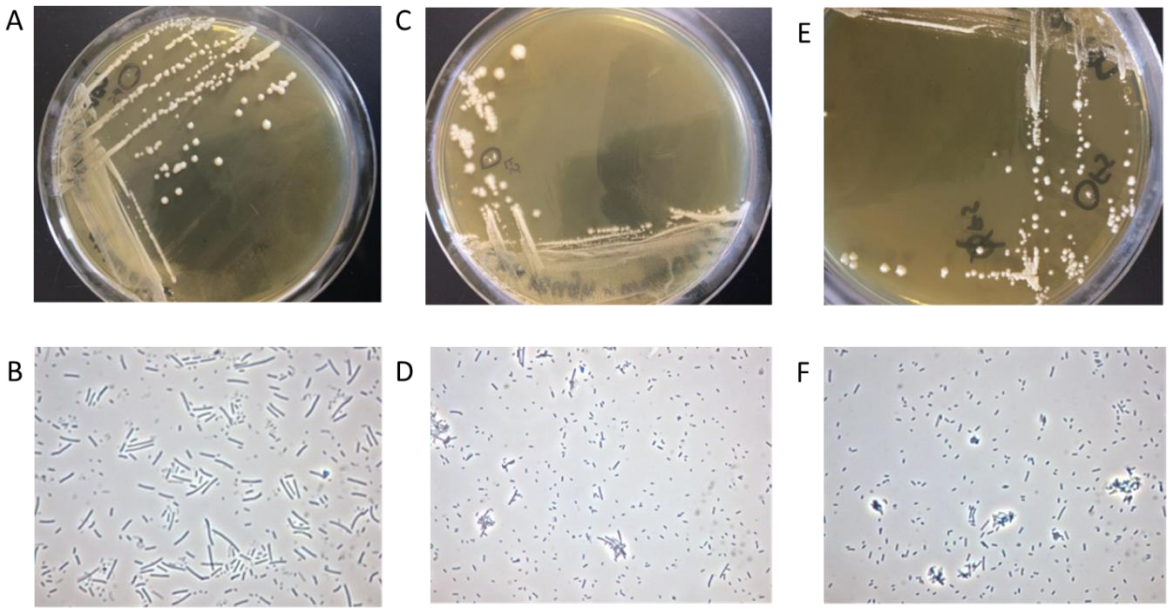

Figure B.1 | Rough and Smooth Colony Phenotype

A) Smooth morphotype colonies, grown under anaerobic conditions B) Microscope image of smooth morphotype colony grown under anaerobic conditions C) Rough morphotype colonies, grown under anaerobic conditions D) Microscope image of rough morphotype colony grown under anaerobic conditions E) Rough morphotype colonies, grown under aerobic conditions F) Microscope image of rough morphotype colony grown under aerobic conditions

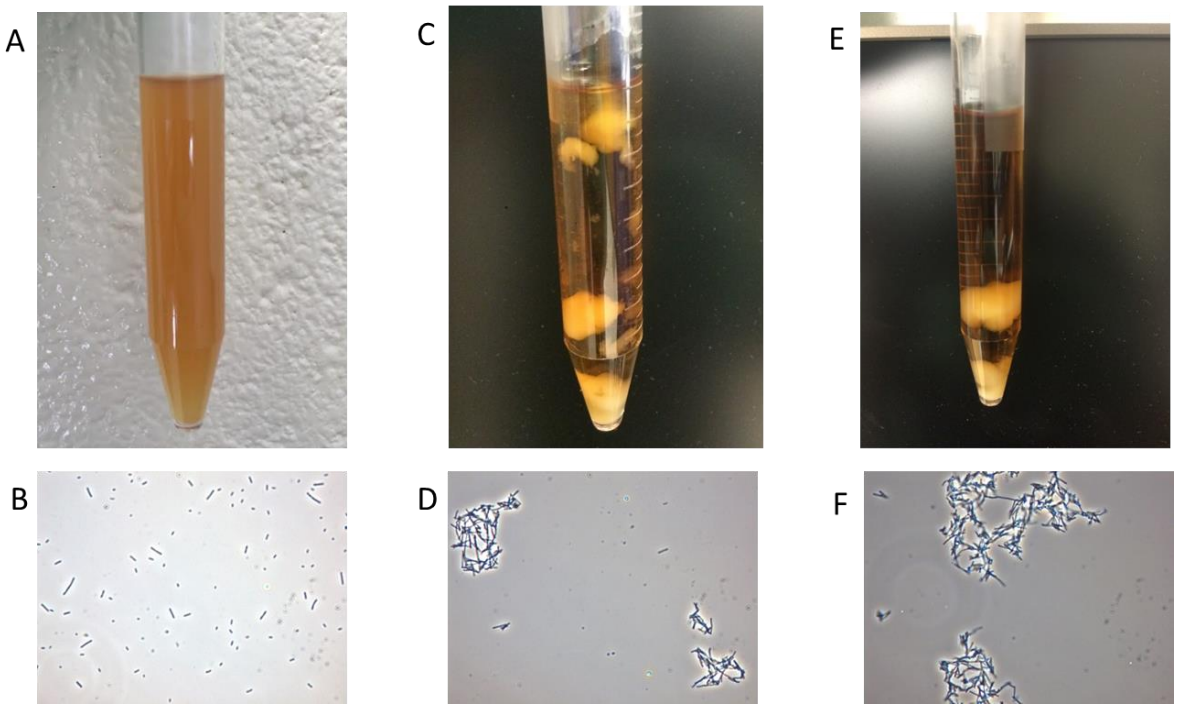

Figure B.2 | Rough and Smooth Culture Phenotype

A) Smooth morphotype culture, inoculated from colony from smooth morphotype plate grown under anaerobic conditions (Fig B.1.A) B) Microscope image of smooth morphotype culture, inoculated from colony from smooth morphotype plate grown under anaerobic conditions C) Rough morphotype culture, inoculated from colony from rough morphotype plate grown under anaerobic conditions (Fig B.1.C) D) Microscope image of rough morphotype culture, inoculated from colony from rough morphotype plate grown under anaerobic conditions E) Rough morphotype culture, inoculated from colony from rough morphotype plate grown under aerobic conditions (Fig B.1.E) F) Microscope image of rough morphotype culture, inoculated from colony from rough morphotype plate grown under aerobic conditions

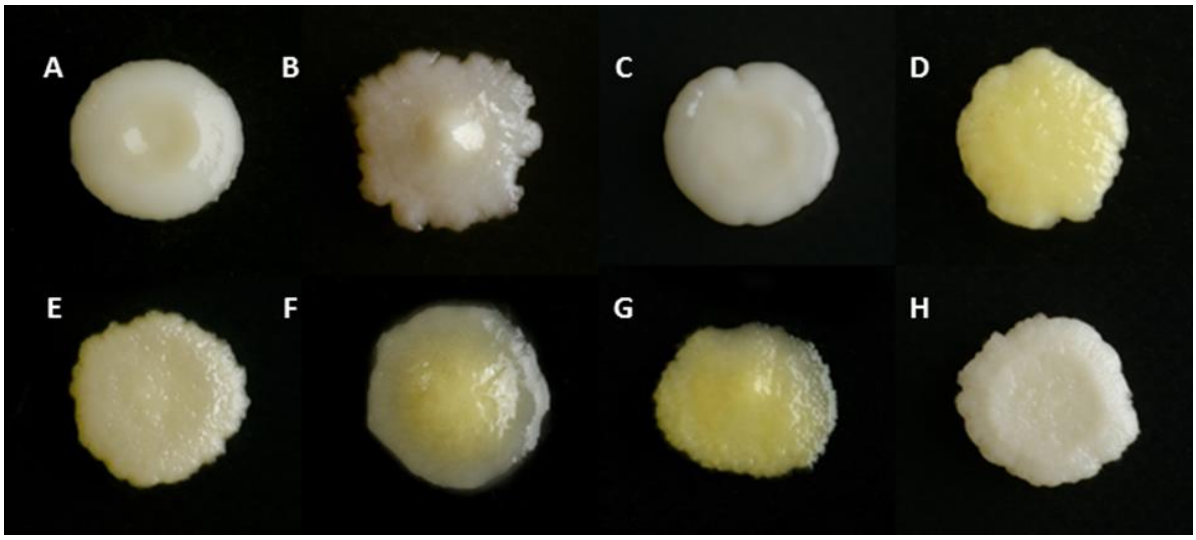

Figure B.3 | *L. buchneri* colony morphologies from eight isolates

Eight colony morphologies from selected *L. buchneri* isolates: (A) LA0030, (B) LA1147 (C) LA1161B (D) LA1184 (E) LA1167 (F) LA1175D (G) LA1181 (H) LA1161C reveal distinct phenotypic differences among *L. buchneri* strains.

Reprinted from *Phenotypic and genotypic characterization of* Lactobacillus buchneri *strains isolated from spoiled fermented cucumber* (Master's thesis) by K.V. Daughtry, 2015. Copyright 2015 by Katheryne Virginia Daughtry. Reprinted with permission.

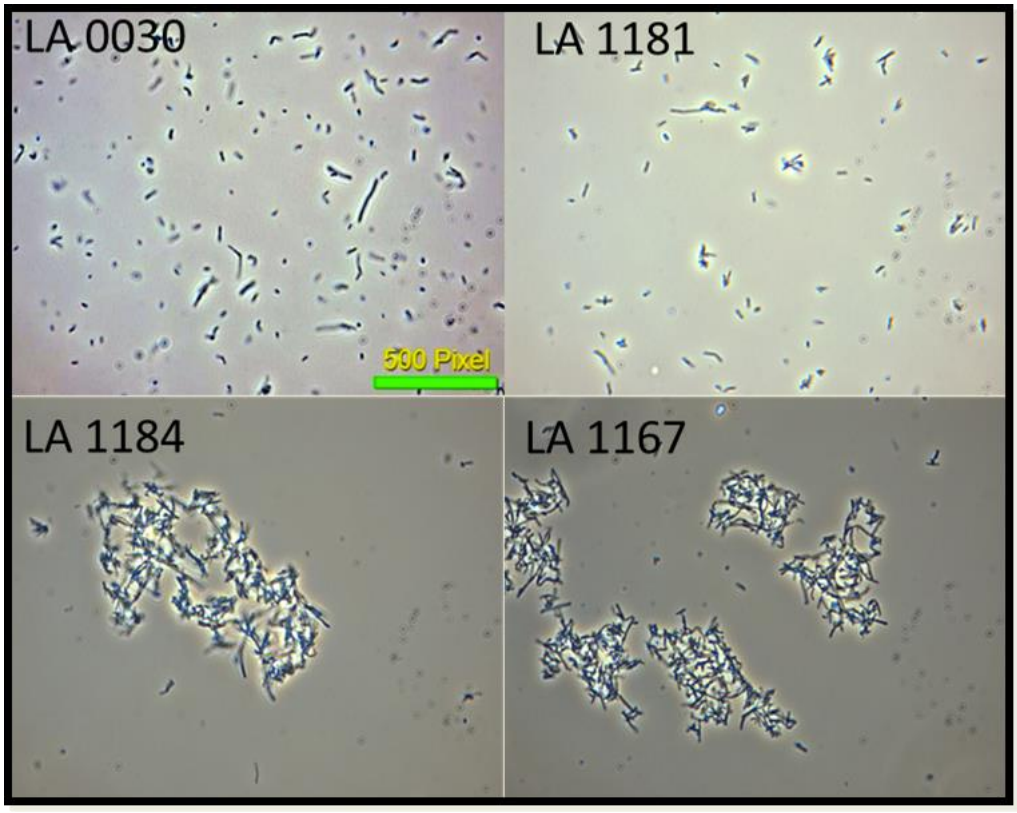

Figure B.4 | *L. buchneri* cellular morphologies from four isolates

Four cellular morphologies from selected *L. buchneri* isolates cultured in MRS: LA0030, LA1181, LA1184, and LA1167.

Reprinted from *Phenotypic and genotypic characterization of* Lactobacillus buchneri *strains isolated from spoiled fermented cucumber* (Master's thesis) by K.V. Daughtry, 2015. Copyright 2015 by Katheryne Virginia Daughtry. Reprinted with permission.

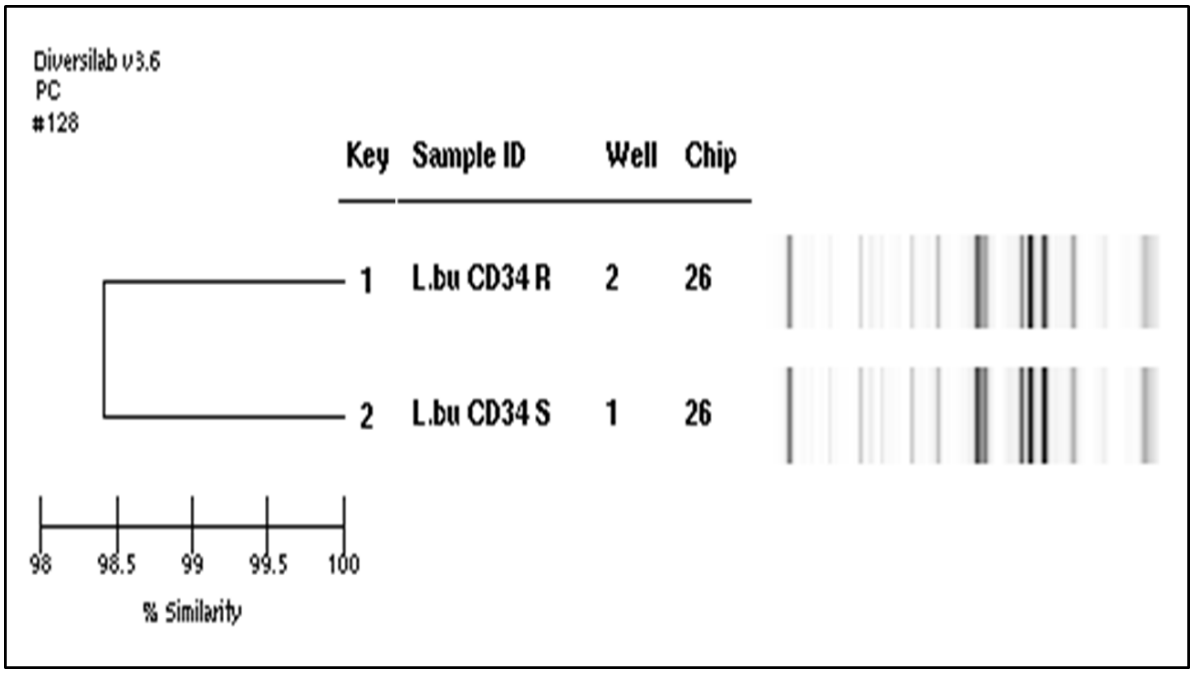

Figure B.5 | Diversilab barcodes of *L. buchneri* CD034R(ough) and CD034S(mooth).

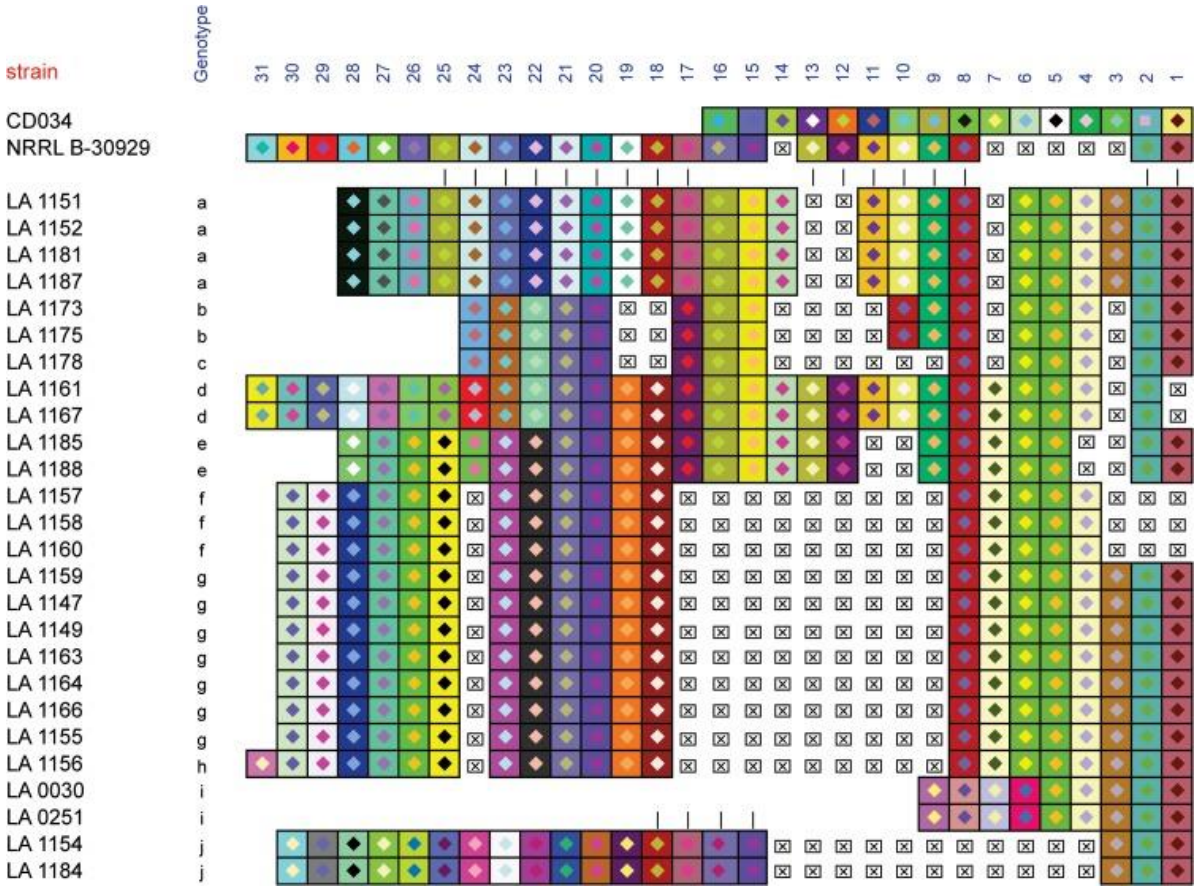

Figure B.6 | CRISPR spacer overview of *L. buchneri* strains

Each square, along with its unique color code combination, represents the presence of a particular spacer sequence; spacers are numbered by order of acquisition to the locus. Deletions are represented by "x." Conserved repeats are not represented. Each unique spacer combination was assigned a genotype (letters).

Reprinted from "*Lactobacillus buchneri* genotyping on the basis of clustered regularly interspaced short palindromic repeat (CRISPR) locus diversity," by A.E. Briner and R. Barrangou, 2014, *Applied and Environmental Microbiology*, 80(3), 994-1001. Copyright 2014 by the American Society for Microbiology. Reprinted with permission.