ABSTRACT

SELLE, KURT MICHAEL. Functional Genomic Studies of Lactic Acid Bacteria. (Under the direction of Dr. Todd Klaenhammer and Dr. Rodolphe Barrangou).

Lactic acid bacteria (LAB) are a diverse clade of Gram-positive eubacteria that share several microbiological characteristics but are defined by their fermentative conversion of carbohydrates into lactic acid. LAB have served anthropological roles in food preservation since the advent of agrarian societies and domestication of both macro- and micro-organisms. Certain LAB also have the capacity to inhabit or colonize human mucosal niches and positively impact the host through a variety of mechanisms. The ability to preserve foods and/or colonize the host ultimately stems from rapid genome reduction, with some acquisition of niche-specific traits conferring fitness in food and human environments.

Modern day bacteriological techniques combined with understanding the genetic basis for these traits have led to widespread use of LAB in commercial applications in food manufacture and in medical interventions through administration of probiotics. Methods in genome sequencing, comparative transcriptomics and proteomics are fundamental to defining the molecular processes governing beneficial LAB activity. In this dissertation, these approaches are applied to Streptococcus thermophilus, a starter culture, and Lactobacillus acidophilus, a probiotic microorganism.

Streptococcus thermophilus is a non-pathogenic thermophilic Gram-positive bacterium that catabolizes lactose to lactic acid in the syntrophic production of yogurt and various cheeses. S. thermophilus genomes are enriched in clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated sequence (Cas) systems, the prokaryotic adaptive immune system against invasive genetic elements. CRISPR arrays
encode small interfering RNAs that guide Cas effector proteins to cognate sequences in target DNA, ultimately resulting in destruction of the invasive DNA. The protective role of CRISPR-Cas systems in *S. thermophilus* against plasmids and bacteriophages is well-established, but there is a paucity of data concerning the genetic outcomes of targeting the bacterial host genome. We identified four genomic islands in *S. thermophilus* LMD-9 and designed synthetic plasmid-based repeat-spacer arrays to elicit independent CRISPR-Cas targeting of each island. Genome sequencing revealed deletion of the 102 kbp Lac operon encoding genomic island and screening confirmed deletions at each of the other three genomic islands. Mutants lacking the 102 kbp island exhibited longer generation times in rich growth medium and failed to acidify skim milk. Sequencing of the deletion junctions revealed chimeric insertion sequence (IS) element footprints generated by homologous recombination between IS elements flanking each island.

*L. acidophilus* NCFM is a widely consumed probiotic strain that has been at the forefront of genomic characterization of probiotic functionality. Genetic modification or removal of lipoteichoic acid (LTA) from *L. acidophilus* confers an increased capacity to alleviate mucosal inflammation in disease models of colitis and cancer. However, functions of LTA in lactobacilli are poorly understood. To assign compensatory cell responses to removal of LTA, we assessed two strains of LTA-deficient *L. acidophilus* for phenotypes associated with LTA function and surveyed their global transcriptional and exoproteome profiles. The LTA mutants exhibited elongated cellular morphology and showed an LTA-dependent sensitivity to elevated Manganese concentrations. Differentially expressed genes observed in the LTA-deficient strains numbered 24 and included a predicted heavy metal resistance operon and several putative peptidoglycan hydrolases. Cell surface protein surveys
revealed distinct changes in the composition and relative abundances of several extracellular proteins and showed a bias of intracellular proteins in LTA-deficient strains of *L. acidophilus*. 
Functional Genomic Studies of Lactic Acid Bacteria

by
Kurt Michael Selle

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Functional Genomics

Raleigh, North Carolina
2016

APPROVED BY:

_______________________________  _______________________________
Dr. Todd R. Klaenhammer  Dr. Rodolphe Barrangou
Committee Co-Chair  Committee Co-Chair

_______________________________  _______________________________
Dr. David Bird  Dr. Amy Grunden
DEDICATION

To my parents, Mark and Linda Selle, for their unwavering support and encouragement of my scientific interests from an early age well into my professional career.
BIOGRAPHY

Kurt Selle grew up in West Bend, Wisconsin and attended the University of Wisconsin-Madison, receiving a B.S. in Microbiology in 2010. At Wisconsin, Kurt was an undergraduate research assistant in the employment of Dr. Jim Steele. Kurt then enrolled in graduate school at North Carolina State University, receiving a M.S. in Food Science under the direction of Dr. Todd Klaenhammer in 2013. Kurt then joined the Functional Genomics program to pursue a doctorate degree under the direction of Dr. Klaenhammer and Dr. Rodolphe Barrangou.
ACKNOWLEDGMENTS

Dr. Todd Klaenhammer’s incredible legacy has had a tremendous impact in the fields of applied microbiology, fermentation science, and molecular genetics of lactic acid bacteria. His forty year career has consistently produced leaders in the food science and microbiology disciplines across academia and industry alike. Todd’s ability to provide constructive guidance and support while fostering independence has been invaluable to me over the past five years. I cannot think of any more suitable way to thank him than to contribute to his personal and professional legacy as I establish my own path.

Dr. Rodolphe Barrangou is a pre-eminent scientist in CRISPR biology, and I have had the privilege of learning from him in this amazing field of research. I have truly gained valuable scientific approaches and professional opportunities from studying under Rodolphe.

The CRISPR and TRK labs have been like a family, and I’d like to express my gratitude and elation for having shared my experiences, triumphs, and failures with them. I’d specifically like to thank Rosemary Sanozky-Dawes, Dr. Emily DeCrescenzo Henriksen, Evelyn Durmaz, Dr. Yong Jun Goh, Dr. Sarah O'Flaherty, Dr. Joakim Andersen, Brant Johnson, Jeff Hymes, Katelyn Brandt, Courtney Klotz, Casie Canez, Mia Theilmann, Kat Daughtry, Allie Briner, and Emily Stout, who were both friends and colleagues. I along with Brant Johnson consider myself lucky to be one of Todd’s last Ph.D. students and am grateful to have him as my colleague. There are several close friends who have always encouraged me in ways that a few sentences cannot adequately describe; Chad Jordan, Caroline Campbell, Thomas Clawson, and Jon Baugher. Each of you has contributed significantly to
the quality of life I have had as a graduate student. Madelyn, your companionship has provided me with balance, endearment and heartening. I cherish it.

I also thank the Dannon Company Inc., Du Pont Nutrition and Health, and the NC Agriculture Foundation for financial support.
# TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................. ix
LIST OF FIGURES ................................................................................................................ x

CHAPTER I ............................................................................................................................ 1
  1.1 Abstract ................................................................................................................... 2
  1.2 Introduction ............................................................................................................. 2
  1.3 CRISPR basics and background ............................................................................. 4
  1.4 Applications across the food bacterial spectrum .................................................... 8
  1.5 CRISPR-based genotyping ..................................................................................... 8
  1.6 Vaccination of industrial microbes ....................................................................... 10
  1.7 Antimicrobials ....................................................................................................... 12
  1.8 Genome editing and remodeling in bacteria ........................................................ 14
  1.9 The CRISPR revolution as it applies to the food chain ........................................ 15
  1.10 Conclusions and perspective ............................................................................. 16
  1.11 References ......................................................................................................... 17

CHAPTER II ......................................................................................................................... 24
  2.1 Abstract ................................................................................................................. 25
  2.2 CRISPR-Cas systems and bacterial genome editing ............................................ 26
  2.3 Lethality of targeting genomes ............................................................................. 28
  2.4 CRISPR-Cas targeting escape strategies .............................................................. 30
  2.5 Lessons from bacterial genome editing studies .................................................. 31
    2.5.1 Homology directed repair in *S. pneumoniae* ................................................. 32
    2.5.2 Efficient and targeted mutagenesis with recombineering in *Lactobacillus reuteri* ...................................................................................................................... 33
    2.5.3 Alternative end joining in *Pectobacterium atrosepticum* ......................... 34
  2.6 Design considerations ........................................................................................... 35
    2.6.1 Design of spacer sequences ......................................................................... 35
    2.6.2 Increasing transformation efficiency for optimized system delivery .......... 36
    2.6.3 DNA repair mechanisms in bacteria ............................................................ 37
  2.7 Future Applications ............................................................................................... 38
### Chapter II

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.7.1</td>
<td>Cas9 and editing template delivery through transformation of linear nucleic acids</td>
<td>38</td>
</tr>
<tr>
<td>2.7.2</td>
<td>Exploitation of endogenous and orthogonal systems</td>
<td>39</td>
</tr>
<tr>
<td>2.7.3</td>
<td>Understanding bacterial genome biology</td>
<td>40</td>
</tr>
<tr>
<td>2.8</td>
<td>Concluding remarks</td>
<td>41</td>
</tr>
<tr>
<td>2.9</td>
<td>References</td>
<td>42</td>
</tr>
</tbody>
</table>

### Chapter III

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Abstract</td>
<td>54</td>
</tr>
<tr>
<td>3.2</td>
<td>Introduction</td>
<td>55</td>
</tr>
<tr>
<td>3.3</td>
<td>Results</td>
<td>57</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Identification of expendable genomic regions</td>
<td>57</td>
</tr>
<tr>
<td>3.3.2</td>
<td>CRISPR-Cas targeting of lacZ selects for large deletion events</td>
<td>60</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Genomic deletions occur through recombination between homologous IS elements</td>
<td>62</td>
</tr>
<tr>
<td>3.4</td>
<td>Discussion</td>
<td>64</td>
</tr>
<tr>
<td>3.5</td>
<td>Materials and Methods</td>
<td>67</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Bacterial Strains</td>
<td>67</td>
</tr>
<tr>
<td>3.5.2</td>
<td>DNA Isolation and Cloning</td>
<td>68</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Selection and design of CRISPR spacers</td>
<td>69</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Transformation</td>
<td>69</td>
</tr>
<tr>
<td>3.5.5</td>
<td>Growth and activity assessment</td>
<td>70</td>
</tr>
<tr>
<td>3.6</td>
<td>References</td>
<td>71</td>
</tr>
</tbody>
</table>

### Chapter IV

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Abstract</td>
<td>87</td>
</tr>
<tr>
<td>4.2</td>
<td>Introduction</td>
<td>87</td>
</tr>
<tr>
<td>4.3</td>
<td>Materials and Methods</td>
<td>90</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Bacterial Strains</td>
<td>90</td>
</tr>
<tr>
<td>4.3.2</td>
<td>DNA Isolation, Manipulation, and Transformation</td>
<td>90</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Construction of an <em>L. acidophilus</em> NCFM ΔslpB ΔslpX ΔltaS deletion mutant</td>
<td>92</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Microscopy and growth curves</td>
<td>94</td>
</tr>
<tr>
<td>4.3.5</td>
<td>Transcriptional analysis</td>
<td>95</td>
</tr>
</tbody>
</table>
4.3.6 Extraction and identification of extracellular, non-covalently bound cell surface proteins ................................................................. 97
4.4 Results ......................................................................................................................... 97
4.4.1 LTA maintains cell shape and provides resistance to Manganese ............... 97
4.4.2 LTA-deficiency elicits compensatory transcriptional changes in L. acidophilus ......................................................................................... 99
4.4.3 LTA mutants exhibit altered cell surface protein profiles ................. 101
4.5 Discussion ............................................................................................................... 103
4.6 References ............................................................................................................. 108
Chapter V ......................................................................................................................... 125
  5.1 Conclusions and Perspectives ............................................................................. 125
APPENDICES ...................................................................................................................... 128
  Appendix I .............................................................................................................. 129
  Appendix II ......................................................................................................... 136
  Appendix III ......................................................................................................... 145
  Appendix IV .......................................................................................................... 152
  Appendix V .......................................................................................................... 162
  Appendix VI .......................................................................................................... 170
  Appendix VII ....................................................................................................... 181
LIST OF TABLES

Table 3.1. Primers and oligomers used in this study............................................................. 76
Table 3.2. Genomic island characteristics and CRISPR targets. ........................................ 77
Table 3.3. Bacterial Strains and Plasmids used in this study. .............................................. 78
Table 4.1. Primers used in this study. .................................................................................. 113
Table 4.2. Bacterial strains and Plasmids used in this study.................................................. 114
Table 4.3. List of differentially expressed genes in LTA deficient strains. ......................... 115
Table 4.4. Operons containing genes found to be differentially expressed in LTA-deficient strains. ............................................................................................................. 116
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>CRISPR-based adaptive immunity</td>
<td>22</td>
</tr>
<tr>
<td>1.2</td>
<td>Applications of CRISPR-Cas systems across the food chain</td>
<td>23</td>
</tr>
<tr>
<td>2.1</td>
<td>CRISPR-Cas9 targeting of DNA</td>
<td>49</td>
</tr>
<tr>
<td>2.2</td>
<td>Manipulation of microbial composition in defined consortia</td>
<td>50</td>
</tr>
<tr>
<td>2.3</td>
<td>Bacterial mechanisms of escape from CRISPR-Cas targeting</td>
<td>51</td>
</tr>
<tr>
<td>2.4</td>
<td>DNA repair outcomes of genome targeting CRISPR-Cas systems</td>
<td>52</td>
</tr>
<tr>
<td>3.1</td>
<td>Map of essential genes, insertion sequences, and genomic islands in Streptococcus thermophilus</td>
<td>79</td>
</tr>
<tr>
<td>3.2</td>
<td>Dendrogram of transposon coding sequences distributed throughout the genome of S. thermophilus LMD-9</td>
<td>80</td>
</tr>
<tr>
<td>3.3</td>
<td>Geneious alignments of transposon coding sequences</td>
<td>81</td>
</tr>
<tr>
<td>3.4</td>
<td>Organization of genomic islands and corresponding CRISPR targets</td>
<td>82</td>
</tr>
<tr>
<td>3.5</td>
<td>Transformants recovered after electroporation of control and self-targeting plasmids</td>
<td>83</td>
</tr>
<tr>
<td>3.6</td>
<td>Genome sequencing and phenotypic analysis of Lac− clones</td>
<td>84</td>
</tr>
<tr>
<td>3.7</td>
<td>Depiction of recombination events between ISs</td>
<td>85</td>
</tr>
<tr>
<td>4.1</td>
<td>Microscopy images of wild-type Lactobacillus acidophilus</td>
<td>117</td>
</tr>
<tr>
<td>4.2</td>
<td>Growth profiles of Lactobacillus acidophilus NCFM derivatives</td>
<td>118</td>
</tr>
<tr>
<td>4.3</td>
<td>Comparison of TPM values</td>
<td>119</td>
</tr>
<tr>
<td>4.4</td>
<td>RNA-seq transcriptional profiles of Lactobacillus acidophilus derivatives</td>
<td>120</td>
</tr>
<tr>
<td>4.5</td>
<td>RNA-seq transcriptional profiles of predicted peptidoglycan turnover proteins in Lactobacillus acidophilus</td>
<td>121</td>
</tr>
<tr>
<td>4.6</td>
<td>Extracellular protein profiles of Lactobacillus acidophilus LTA derivatives</td>
<td>122</td>
</tr>
<tr>
<td>4.7</td>
<td>Distribution of normalized spectral counts for proteins with and without secretion signals</td>
<td>123</td>
</tr>
<tr>
<td>4.8</td>
<td>Comparison of normalized spectral counts of select extracellular proteins across Lactobacillus acidophilus derivatives</td>
<td>124</td>
</tr>
</tbody>
</table>
CHAPTER I

CRISPR-based technologies and the future of food science.

Selle K, Barrangou R.
1.1 Abstract

The on-going CRISPR craze is focused on the use of Cas9-based technologies for genome editing applications in eukaryotes, with high potential for translational medicine and next-generation gene therapy. Nevertheless, CRISPR-Cas systems actually provide adaptive immunity in bacteria, and have much promise for various applications in food bacteria that include high-resolution typing of pathogens, vaccination of starter cultures against phages, and the genesis of programmable and specific antibiotics that can selectively modulate bacterial population composition. Indeed, the molecular machinery from these DNA-encoded, RNA-mediated, DNA-targeting systems can be harnessed in native hosts, or repurposed in engineered systems for a plethora of applications that can be implemented in all organisms relevant to the food chain, including agricultural crops trait-enhancement, livestock breeding and fermentation-based manufacturing, and for the genesis of next-generation food products with enhanced quality and health-promoting functionalities. CRISPR-based applications are now poised to revolutionize many fields within food science, from farm to fork. In this review, we describe CRISPR-Cas systems and highlight their potential for the development of superior foods.

1.2 Introduction

Food science may be generally defined as the application of many scientific fields and disciplines to food products and processing. While many distinct sciences have historically been applied to the genesis, formulation, processing, storage, enhancement and enjoyment of food products over time, food science is arguably in a renaissance stage, which
is fueled by the availability of key technologies that allow food scientists and engineers to
develop health-promoting products. Notwithstanding the many advances we have witnessed
over the last century, the field of genetics has contributed critical advances in the recent past.
In particular, the use of recombinant genetic technologies has profoundly impacted food
science, agriculture, ecology, animal husbandry, and medicine. Tangible and impactful
improvements in the human condition, due to the industrial biosynthesis of vitamins,
enzymes, pharmaceuticals, antibiotics, and bioactive peptides, have been enabled by
advances in genetic methodologies. Forthcoming DNA technologies have accelerated the rate
of molecular biology research in diverse backgrounds, where few technologies previously
existed or were suboptimal. Specifically, clustered regularly interspaced short palindromic
repeats (CRISPR) and CRISPR-associated sequences (Cas) are an adaptive immune system
against invasive genetic elements in bacteria (1), which has been co-opted for genome editing
in a diverse set of organisms ranging from scientific models to industrial workhorses (2). The
ability to cleave and edit DNA with CRISPR-Cas systems has reinvigorated reductionist
biology during a scientific era defined by sequencing, and the rate of functional genomics has
accelerated to new limits as a result. However, early investigation of CRISPR-Cas systems
developed applications outside of genome editing, including manipulation of microbial
consortia (3), designed vaccination of microorganisms against invasive genetic elements (4),
and typing of bacterial strains (5). Many of the early applications of CRISPR-Cas systems
arose from food science-driven research during characterization of industrial starter culture
bacteria for improving milk fermentation processes (6). Food science is a growing field
investigating all biological, chemical and physical processes to improve production of safe
and sustainable food for a growing world population. Overall, given the pace at which CRISPR-Cas technology is being developed, one can already envision how applications of CRISPR-Cas systems may further be harnessed or engineered to address challenges related to the food and agriculture industries at every level of food manufacturing, from farm to fork.

1.3 CRISPR basics and background

Two genetic elements constitute the adaptive immune system in bacteria: CRISPR arrays which confer immunological memory and surveillance, and _cas_ genes, which encode effector proteins in all stages of immunity (Figure 1.1). CRISPR-Cas mediated immunity is categorized into three temporally overlapping but mechanistically distinct molecular processes: acquisition, expression, and interference (1, 7). Acquisition occurs via sampling of foreign genetic elements by the universal Cas1/Cas2 surveillance complex, from which short sequences, termed spacers, are integrated in a polarized fashion into the CRISPR array (6). New spacer sequences are added at the leader end of the array, resulting in an ordinal record of foreign DNA exposures that spans from most recent at the 5’ end to most ancient at the 3’ end. Expression of CRISPR arrays is constitutive under standard conditions, but is also inducible during phage infection (8). The array is transcribed as a long pre-CRISPR RNA (crRNA), and is typically further processed via Cas proteins and host ribonucleases into mature, small interfering crRNAs (9). Mature crRNAs guide Cas proteins to target DNA via sequence-specific complementarity for recognition and cleavage of target nucleic acids, causing direct interference of exogenous DNA elements (10, 11).
CRISPR arrays consist of highly-conserved, partially palindromic repeats that alternate with variable short spacer sequences. A compelling and defining feature of CRISPR-Cas systems, repeat sequences function in the formation of hairpins necessary for structure-dependent RNA processing during biogenesis of crRNAs. Repeat sequences also exhibit partial complementarity to trans-activating CRISPR RNAs (tracrRNA) and leader sequences, suggesting roles in acquisition and in Cas protein recognition. Specific to certain systems, tracrRNAs are enigmatic non-coding RNAs that elicit processing of crRNAs and for target DNA recognition and cleavage by Cas proteins (12, 13). By contrast, the origin and function of most spacer sequences is unknown, as only approximately 1% exhibit reliable identity to foreign invasive elements or chromosomal sequences (14). This may be because of extinction of genetic elements or the lack of environmental surveillance (sequencing) of the elements to which spacers correspond. Most of the known spacer targets correspond to plasmids and bacteriophages, but many also appear to target either self or foreign chromosomal sequences (14). Although the intuitive function of an adaptive immune system is targeting invasive genetic elements, many spacer sequences exhibit self-complementarity, and many do not appear to be transcribed, suggesting that CRISPR-Cas systems may play additional roles in microbial physiology beyond targeting of genetic elements (15). An essential feature of Type I and Type II systems is the protospacer adjacent motif (PAM), a short conserved sequence proximate to the spacer sequence in the target DNA (16-18). The PAM governs both the acquisition and interference processes, as it determines viable protospacers in the target sequence, and functions in differentiation of the target from the
CRISPR array. PAM sequences vary between CRISPR types and orthogonal systems within CRISPR types, but must be characterized in order to fully exploit functional systems.

Cas proteins are fundamental to each stage of CRISPR-based immunity, as they are responsible for acquisition of new spacers, processing of crRNAs, and recognition and degradation of sequences complementary to crRNAs. However, convergent evolution has resulted in a myriad of DNA recognition and cleavage mechanisms in keeping with microbial diversity, necessitating categorization of CRISPR-Cas based on gene content, operon organization, and distinct clusters of sequence homology (19). The most recent definitive description of CRISPR-Cas systems highlights two classes based on the composition of the immune effector complex (single vs. multi-subunit). The systems are further delineated into five types, specifically depending on the presence of signature cas genes (19). The hallmark features of Type II systems are the large multifunctional endonucleolytic Cas9, the tracrRNA, and ribonuclease III processing of crRNAs (12). Type I and Type III systems encode Cascade, Csy, and Csm proteins that constitute the multi-subunit effector complexes responsible for target nucleic acid recognition. The signature gene of Type I systems is Cas3, a single stranded nickase with 3’-5’ exonuclease activity, which is recruited to the target via the CRISPR-associated complex for antiviral defense (9, 20). In contrast to other CRISPR system Types, Type III systems target either or both DNA and RNA, and the signature gene is Cas10 (19) Estimates indicate that 46% of bacterial and 84% of archaeal genomes contain at least one CRISPR-Cas system (19). Despite the high distribution of CRISPR-Cas systems in bacteria, the relative youth of the field means that very few of the systems have been characterized for activity in each of the three stages of immunity: acquisition, expression and
interference. To date, only 14 systems have demonstrated activity in interference, highlighting the need for investigating orthogonal systems for holistic understanding of CRISPR-Cas mechanisms and applications (21).

The study of CRISPR-Cas systems had relatively humble origins in food science-related research, with its biological function being discovered during investigation of phage resistance mechanisms of dairy starter cultures (6). In particular, *Streptococcus thermophilus* notoriously undergoes attack by predatory bacteriophage during yogurt and cheese fermentation, and is highly enriched for active CRISPR-Cas systems (4). *S. thermophilus* encodes up to four CRISPR-Cas systems, two of which are innately active in both acquisition and interference, likely due to the high selective pressure of bacteriophage in dairy processing environments. Accordingly, functional genomic analysis of a food-grade starter culture bacterium and its bacteriophages established the role of CRISPR-Cas systems in phage/DNA protection (6). Several seminal contributions to the field were conducted in *S. thermophilus*, because it harbored highly active CRISPR-Cas systems, the genome sequence was available, and there were characterized lytic phages, all of which was related to its widespread industrial use as a starter culture in food fermentations. Basic food science research therefore led to determination of spacer origin (6), inference of PAM sequences (16, 18), unraveling of phage-host dynamics (22, 23), demonstration of Cas9 endonuclease activity (11, 24, 25), characterization of tracrRNA structural motifs governing function and orthogonality (12, 26), and recently, designed removal of large genomic islands (27).
1.4 Applications across the food bacterial spectrum

*Lactobacillus* spp. are scientifically, industrially, and medically relevant microorganisms that are propagated at high levels for fermentation processes or to elicit health benefits as probiotic microorganisms. Thus, as constituents of the human microbiome, or as fermentative bacteria, exposure to bacteriophage is highly likely, suggesting that bacterial phage resistance mechanisms would be abundant in these environments. Indeed, *in silico* surveys have revealed that Type II systems are disproportionately present in lactobacilli (14, 26), making them a reservoir for CRISPR-Cas systems. In this light, the human and food microbiomes are a relatively unexplored trove of new and diverse CRISPR-Cas systems that are suitable for use in food-grade systems. Given that bacteria are ubiquitous throughout the production and consumption of food, CRISPR technologies have the potential to impact all classes of bacteria across the food spectrum, including pathogenic, commensal, fermentative, probiotic, and spoilage organisms. CRISPR-based technologies with applications in food science include genotyping of bacteria, manipulation of microbial consortia, vaccination against phages and genome editing (Figure 1.2).

1.5 CRISPR-based genotyping

Identification and typing of bacterial strains is a considerable challenge due to the inherent diversity of microorganisms and their tendency to undergo horizontal gene transfer. Although genome sequencing can be considered the ‘gold standard’ for identification and typing of bacterial strains in terms of resolution, it is a costly, analytically-challenging and time-intensive process not suitable for high-throughput or rapid applications. The same
limitations apply to pulsed field gel electrophoresis, although fingerprinting of restriction digestion profiles is still performed for strain typing during foodborne outbreaks. Recently, repetitive-element PCR-based genotyping using high-resolution microfluidics has proven to be rapid and reliable in strain differentiation, but identification of strains requires a database of fingerprint data for comparison. 16S rDNA sequencing, although not a typing tool, is relatively fast and affordable for rough identification of bacteria, but can be unreliable even for applications requiring resolution down to the species level. By contrast, CRISPR array genotyping offers a rapid, affordable and high-resolution means of typing bacterial strains within a species (5). Due to the unique polarized nature of spacer acquisition in CRISPR-Cas systems, the highly ordinal composition of CRISPR arrays provides a means of typing with high resolution at the strain level (28), requiring a few PCR amplifications and sequencing of the array to provide a clear comparison of spacer content. Despite this, proof of concept for CRISPR-based typing has only been provided for a limited set of bacteria. In order to effectively be used as a tool for identification and typing, the same CRISPR-array locus must be enriched or ubiquitous within defined taxonomic groups (Genus or species) and the spacer content of the array must be diverse across all strains in a given subset. The presence or absence of a CRISPR array may also be used to differentiate strains, but is only reliable when it correlates with the phylogeny of the organism. In order for adequate spacer comparison to occur, within a given CRISPR locus, the most ancient spacers must share a common origin, which then diverges over the course of the array. In other words, there must be some shared and some disparate spacers in order to effectively type strains based on array content. Thus, the process is largely contingent on having had active spacer acquisition machinery at some
point in evolutionary history, although degeneracy of CRISPR arrays can also add to polymorphisms in spacer content. Of course, CRISPR-based typing also depends on the presence of CRISPR-Cas systems in the genomes of genera and species of interest, and these loci have been identified in most archaea and many bacteria, but only documented to occur in approximately 46% of bacterial genomes. To date, CRISPR-based typing schemes have been effectively employed in foodborne pathogens such as *Salmonella* (28) and *Escherichia coli* (29, 30), industrial fermentation starter cultures such as *S. thermophilus* (16), probiotics such as *Lactobacillus casei* (31), and spoilage organisms such as *Lactobacillus buchneri* (32), illustrating the broad potential of CRISPR-based genotyping across the bacterial spectrum.

1.6 Vaccination of industrial microbes

Mobile genetic elements (MGEs) are a class of DNA entities encompassing plasmids, bacteriophages, transposable elements, and integrative and conjugative elements. MGEs exhibit high rates of transfer and hijack bacterial DNA homeostasis pathways, causing continuous challenges to both population and genetic stability of bacteria. To cope with the permanent threat of predatory bacteriophages and selfish genetic elements, bacteria have evolved both innate and adaptive immune systems targeting exogenous genetic elements. Innate immune mechanisms include cell-wall modification, restriction/modification systems, and abortive phage infection (33). In the food industry, predatory bacteriophages constitute a significant threat to efficiency of preservation and continue to be a major source of inconsistent quality or loss in dairy fermentations. Many strategies have thus been developed to combat the ever-present and dynamic phage populations present in processing plant
environments. Starter cultures are especially susceptible to lytic phage infection due to a mono-culture population, wherein a single infective phage type can cause the crash of an entire population. Furthermore, the high rate of mutation in phages necessitates the use of multiple resistance mechanisms and control strategies to compensate for their high capacity for adaptation to the host. Specifically, both native biological mechanisms of resistance and environmental control are employed to prevent phage proliferation. Starter culture rotation, growth in the presence of chelators, multi-strain starter formulations, and steam sterilization of manufacturing equipment are all means of controlling phage in the dairy processing environment, whereas genetic transfer of plasmids containing native phage resistance mechanisms and/or CRISPR can be used to combat phage in the bacterial population (5). CRISPR provides unique advantages in vaccination of starters against predatory bacteriophage (4). Specifically, the process of adding spacers corresponding to phages is iterative, which means that additional spacers can always be acquired to target-emerging phages. Moreover, resistance is sequence-specific, which means that the resistance mechanism can be as broad, or specific as desired, especially if conserved functional sequences are targeted in phage genomes. One spacer may therefore be able to confer resistance to multiple phages if the respective phage genomes contain the same sequence targeted by the spacer. Lastly, unlike innate immune mechanisms of phage resistance, the target sequence of the phage must mutate in order to circumvent CRISPR-Cas as a mechanism, which can lead to detrimental mutations in phage machinery (22, 23).

Antibiotic resistance of pathogens is an alarming issue in the medicinal community, leading to extensive efforts in reducing the uptake and transmission of antibiotic resistance
genes in bacteria. To this end, the presence of transmissible antibiotic resistance genes is prohibitive during selection of starter culture and probiotic bacteria. Antibiotic resistance genes can be encoded genomically, or by plasmids, bacteriophages, and transposable elements, all of which can be targeted by CRISPR-Cas systems. Spontaneous mutations conferring antibiotic resistance can also be corrected using template-mediated genome editing. Similarly, limiting transfer of antibiotic resistance genes into food-grade bacteria during fermentation is highly desirable. In strains that contain an active CRISPR-Cas system, it is possible to introduce the antibiotic resistance gene on a plasmid and screen for CRISPR-based loss of the plasmid due to targeting of the antibiotic resistance gene (11). Thus, it is a natural means to vaccinate food-grade bacteria against transmissible antibiotic resistance genes, which can be achieved through incorporating a spacer sequence corresponding to that of the coding sequence for antibiotic resistance. Conversely, it is also possible to heterologously introduce an active CRISPR-Cas system into organisms lacking an endogenous system and vaccinate the recipient against the uptake of undesirable genetic content (25).

1.7 Antimicrobials

Self-targeting events of CRISPR-Cas systems are highly lethal, which has been determined experimentally and observed in vivo (34, 35). The lethality of self-targeting events relates to the nature of DNA destruction induced by CRISPR-Cas interference mechanisms. Type I systems elicit extensive DNA damage through the exonuclease activity of Cas3, introducing deletions that span approximately 40 kb in some experiments. DNA
damage of this nature is reparable at low frequency \((10^{-5})\), likely by an alternative end-joining mechanism (34). Type II systems elicit double-stranded DNA breaks via Cas9 activity, for which few DNA repair mechanisms exist in bacteria (11, 24, 36). Bacteria typically use the high fidelity pathway of homologous recombination to repair double-stranded DNA breaks, but restoration of the target locus to the wild-type does not circumvent targeting by CRISPR-Cas systems (36). Pathways of dsDNA break repair in bacteria also include the low-fidelity alternative end-joining, and non-homologous end-joining, although both occur at low frequency, and the latter is generally present in spore-forming bacteria and Mycobacterium spp. Due to a paucity of high-frequency DNA damage repair pathways that can cope with CRISPR-Cas targeting, population-wide depletion of cells exhibiting the target genotype occurs, generally on the order of 3-5 logs under experimental conditions. Self-targeting with Type I systems is highly efficacious, as lethality is not dependent on chromosomal location, expression level, or strand bias (3). Moreover, it was demonstrated that proper design of self-targeting spacers could lead to differentiation of highly related strains (3). Several experiments repurposing endogenous or delivered CRISPR-Cas systems have been performed in pathogenic organisms related to the food industry (3, 35, 3, 38).

Control of microbial consortia in processing facilities and in food products is fundamental to protecting the food supply from contamination or proliferation of foodborne pathogens. Self-targeting CRISPR Cas systems therefore present a novel and high-potential means to deplete microbial populations in a sequence-specific manner, sparing the innocuous native microbiota present in foods. As CRISPR-based microbiome engineering technologies further evolve, we anticipate several application avenues will be generated across the food supply.
chain, to optimally manage the composition of various microbial populations associated with soil, plants, livestock, manufacturing environments and the consumer.

1.8 Genome editing and remodeling in bacteria

Bacterial genomes exhibit site-specific plasticity that belies the linearity of their interpretation as straightforward sequences (39). The ability to reprogram CRISPR-Cas systems to target any sequence in the genome offers promising applications towards defining minimal bacterial genomes, determining essential genes, and characterizing genetically heterogeneous bacterial populations (36, 40). Recently, CRISPR-Cas targeting was used to show that mobile genetic elements contribute to genomic plasticity in *Streptococcus thermophilus* (27). Specifically, recombination between insertion sequences of high identity caused spontaneous deletion of large genomic islands, spanning from 8 to 102 kbp in length. Targeting the genomic islands with an endogenous CRISPR-Cas system enabled selection and recovery of naturally occurring mutants lacking genes necessary for acidification and preservation of milk. The approach also confirmed the non-essentiality of genes encoded on the genomic islands, ultimately resulting in excision of 7% of the genome of *S. thermophilus*. This approach could similarly be applied for removal of genomically encoded MGEs, increasing genome stability. Moreover, CRISPR-based removal of pathogenicity islands and/or virulence factors is an attractive method for neutralizing pathogenic bacteria. Thus, CRISPR-Cas systems facilitate characterization of mobile genetic elements and elucidation of bacterial genome plasticity. Similarly, this technology can also be harnessed in
combination with single-strand DNA recombineering to drive genome editing in probiotic strains such as *Lactobacillus reuteri* (41, 42).

1.9 The CRISPR revolution as it applies to the food chain

The use of CRISPR-based technologies has revolutionized the field of genetics in general, and genome editing of eukaryotes in particular. To date, this approach has been successfully employed for targeted mutagenesis of a plethora of genomes including *Homo sapiens, Mus musculus, Danio rerio, Drosophila melanogaster, Caenorhabditis elegans, Oryza sativa*, and *Saccharomyces cerevisiae* (2). The streamlined and multifunctional nature of Cas9 from Type II systems is practical for programmable genome editing through precise and directed targeting of chromosomal loci. The tipping point for genome editing was arguably the provision of a synthetic guide molecule that combines the functions of a native crRNA and tracrRNA (43) and the development of a corresponding two-component sgRNA:Cas9 genome editing system. The programmable specificity of exacted chromosomal cleavage is facilitated by selection and design of a spacer sequence unique to the target allele. Specificity is compounded by the PAM, a short conserved sequence that must be proximate to the protospacer in the target sequence (17, 18). Cas9-induced mutagenesis in eukaryotes occurs subsequent to cleavage and is typically mediated through the imperfect DNA repair mechanism of non-homologous end joining (NHEJ). Following the genesis of doublestranded breaks, NHEJ yields efficient recovery of insertion and deletion knockout clonal genotypes.
CRISPR-based genome editing has already been applied to organisms of interest across food science, including yeast, corn, rice and tomatoes. Genome editing of crops has applications for targeted engineering to improve growth under drought conditions, application of insecticide, low nutrition/fertilizer conditions, and also to improve the nutrition potential of food crops. Similarly, genome editing can improve yield in animal breeding through desirable alteration and selection of herd genetics. Moreover, there is the potential to increase the disease-resistance of both crops and cattle, but despite the promising outcomes of genome editing using CRISPR-Cas systems, the practical implications of doing so are yet to be unanimously defined.

1.10 Conclusions and perspective

While much of the on-going CRISPR craze (44, 45) has been focused on genome editing applications in human cells, and the potential of Cas9-based gene therapies for clinical applications (46), CRISPR-Cas systems and CRISPR-based technologies hold much promise for a broad range of applications across food science. For applications in food bacteria, native CRISPR-Cas systems present opportunities for genotyping of pathogens, for vaccination of cultures against phages, and as next-generation antimicrobials. Furthermore, engineered systems can be harnessed for genome-editing applications in crops and livestock for trait-enhancement in next-generation breeding approaches. Building off recent advances in their exploitation in agriculture, husbandry and industrial fermentations, we envision that CRISPR-Cas technologies will drive research and development in many food products, and open new avenues for the future of food science.
1.11 References


Figure 1.1. CRISPR-based adaptive immunity.
**Figure 1.2.** Applications of CRISPR-Cas systems across the food chain.
CHAPTER II

Harnessing CRISPR-Cas systems for bacterial genome editing.

Selle K, Barrangou R.
2.1 Abstract

Manipulation of genomic sequences facilitates identification and characterization of key genetic determinants in the investigation of biological processes. Genome editing via CRISPR-Cas constitutes a next-generation method for programmable and high throughput functional genomics. CRISPR-Cas systems are readily reprogrammed to induce sequence-specific DNA breaks at target loci, resulting in fixed mutations via host-dependent DNA repair mechanisms. Although bacterial genome editing is a relatively unexplored and underrepresented application of CRISPR-Cas systems, recent studies provide valuable insights for widespread future implementation of this technology. This review summarizes recent progress in bacterial genome editing and identifies fundamental genetic and phenotypic outcomes of CRISPR targeting in bacteria, in the context of tool development, genome homeostasis, and DNA repair.
2.2 CRISPR-Cas systems and bacterial genome editing

Bacteria harbor clustered, regularly interspaced, short palindromic repeats (CRISPR, see Glossary) and CRISPR-associated (cas) genes, which constitute an RNA-guided adaptive immune system against invasive genetic elements (1). CRISPR-Cas mediated immunity hinges upon the distinct molecular processes of acquisition, expression, and interference (2). Acquisition occurs via molecular ‘sampling’ of foreign DNA, from which short sequences, termed spacers, are integrated in a polarized manner at the leader end of CRISPR array (1). CRISPR arrays are transcribed constitutively and inducibly as directed by promoter elements in the preceding leader sequence during expression (3-5). The transcript is processed selectively at each repeat sequence, forming mature CRISPR RNAs (crRNAs) that serve as small interfering RNAs. crRNAs guide Cas proteins for sequence-specific recognition and cleavage of target DNA complementary to the spacer to effect interference. CRISPR-Cas systems encode universal cas1 and cas2 genes, and are categorized as Type I, Type II or Type III based on signature genes contributing to the distinct mechanisms by which each system confers interference (6). Type I systems achieve immunity via the CRISPR-associated complex for anti-viral defense (Cascade) through single strand DNA nickase and exonuclease activity, and are defined by the presence of Cas3 (7). Features unique to Type II systems include the signature double-stranded (ds) DNA endonuclease Cas9, the ancillary trans-activating crRNA (tracrRNA), and biogenesis of crRNAs by RNase III (8, 9). Type III systems are marked by the signature gene Cas10, but are mechanistically diverse and less well defined, with some systems even capable of targeting RNA instead of DNA (10).
Delineation of CRISPR-Cas systems into 11 subtypes are similarly based on presence of specific accessory *cas* genes and their respective genetic organization (6).

Cas9 effects interference in Type II systems through sequence-directed endonucleaseolysis at the target locus, achieved by concerted RuvC and HNH nickase activity (11, 12). The streamlined and multifunctional nature of Cas9 is practical for programmable genome editing in diverse organisms, requiring only expression of its cognate tracrRNA and a crRNA corresponding to the target sequence (Figure 2.1). The tipping point for this methodology was the creation of a single guide RNA chimera which combines the functions of the native crRNA and tracrRNA duplex (13). Cas9-mediated genome editing is programmable through design of single guide RNAs (sgRNAs). The specificity of chromosomal cleavage hinges upon selection of a spacer sequence unique to the target allele and is further compounded by the protospacer adjacent motif (PAM), a short conserved sequence that must be proximate to the target proto-spacer (Figure 2.1) (14-16). Cas9 introduces a lethal double-stranded DNA break (DSB) at the target locus, effectively acting as a selection against wild-type sequences during genome editing (11, 12). Pre-existing mutations in the population can be selected for or against (Figure 2.2) but mutations may be introduced subsequent to targeting by host-repair mechanisms (Figure 2.2). Mutations elicited by CRISPR-Cas systems are therefore DNA damage and repair machinery dependent. Heterologous expression of Cas9::sgRNA combinations from *Streptococcus pyogenes* has facilitated high throughput functional genomics in a multitude of eukaryotic organisms and cell lines (17-19). CRISPR-Cas derived genome editing tools have
revolutionized genetic and biological research in model eukaryotic organisms on account of their efficiency, affordability, and accessibility.

Notwithstanding early proof of concept, only three studies have implemented CRISPR-Cas mediated genome editing in bacteria (20-22), making this a relatively unexplored and underrepresented application of CRISPR-Cas systems. Nevertheless, genome editing via CRISPR-Cas constitutes a next-generation method for programmable and high throughput functional genomics in prokaryotic backgrounds. Collectively, these studies substantiate the use of CRISPR-Cas systems as genetic tools in bacteria, and contribute to understanding the fundamental genetic and phenotypic outcomes of targeting bacterial genomes. This review aims to summarize insights from these foundational experiments, highlight considerations for tool development, identify potential biological hurdles, and predict future applications of the technology.

2.3 Lethality of targeting genomes

The lethality of Cas-mediated DNA cleavage was first observed in its natural ecological role of targeting bacteriophages and plasmids (1, 11, 12, 23), but self-targeting events are an evolutionary cost of housing active CRISPR-Cas systems. The observation of self-complementary spacers, at one time constituting up to 22% of known spacer targets in lactic acid bacteria (24), emphasizes the potential selective pressure of self-targeting events. Indeed, identification of self-complementary spacer targets reveals mutations at those chromosomal loci, suggesting that self-targeting events drive mutation or fixation of pre-existing mutations (Figure 2.2). Investigation of spacer acquisition in Streptococcus
thermophilus during exposure to phage led to infrequent observation of chromosomal acquisition events, which correlated with the disappearance of clones containing self-targeting spacers (25). Moreover, several studies in diverse backgrounds have reported the lethality of DNA damage induced by self-targeting CRISPR-Cas systems (20, 22, 26-31). Transformation of plasmids eliciting self-targeting by Cas proteins is cytotoxic as measured by the relative reduction in viable transformants recovered compared to transformation of non-self-targeting plasmids (28). CRISPR-mediated depletion of microbial populations ranges from 3-5 log reductions in populations exhibiting the target sequence, sometimes approaching the transformation efficiency of the respective bacterial background (20, 22, 26-31). DNA cleavage by Cas proteins constitutes a significant threat to the survival and fitness of microorganisms, as evidenced by growth inhibition and aberrant cellular morphology phenotypes consistent with DNA damage observed in Pectobacterium atrosepticum following self-targeting events (20). Active CRISPR-Cas systems cannot coexist in the same cell as the target DNA, which compounds the pressure for mutations to occur, as restoration of the target locus to the wild-type does not circumvent CRISPR targeting (32). Consequently, high fidelity repair mechanisms are not sufficient for survival of self-targeting events. Thus, targeting by CRISPR-Cas systems is a selection against the cell populations exhibiting the target genotype. Selection for pre-existing mutations in genetically heterogeneous cell populations supports CRISPR-Cas directed genome evolution on the population level (Figure 2.2) (20). This phenomenon was demonstrated experimentally by transformation of strain-specific self-targeting plasmids into heterogeneous populations consisting of highly identical Escherichia coli strains (29), in which dose-dependent
depletion of specific population subsets was achieved. The study also reported that lethality was independent of chromosome location, transcriptional activity of the target, strand bias, and coding versus non-coding regions. Collectively, the well-established lethality of self-targeting events substantiates the utility of CRISPR-Cas for mutagenesis in bacterial genomes by selecting for non-wild-type clonal variants. Moreover, they highlight the potential application of Cas cleavage-driven, sequence-specific evolution of bacterial genomes in mixed populations.

2.4 CRISPR-Cas targeting escape strategies

Bacterial cells containing target DNA sequences are efficiently cleared from the population, which is partially due to the low capacity of bacterial DNA repair mechanisms to cope with Cas cleavage. However, some clones are able to maintain wild-type target sequences in the presence of CRISPR-Cas targeting. Genetic analysis of transformants recovered following self-targeting revealed that bacteria can escape targeting by mutation/deletion of the plasmid encoded spacer or chromosomally encoded Cas machinery to effectively preclude self-targeting (Figure 2.3) (20, 22, 33). Mutation of the PAM sequence is a major mechanism by which CRISPR-Cas targeting may be circumvented (Figure 2.3) (15, 20, 21, 33, 34). The seed sequence is comprised of the 8-12 bp most proximate to the PAM and is fundamentally involved in hybridization of crRNA to the cognate target DNA, such that mutations in the seed also abolish targeting (Figure 2.3) (21, 35, 36). Point mutations within the protospacer sequence are relatively well-tolerated and generally do not prevent targeting (20, 21). This is especially true for the PAM-distal spacer
sequence, which is consistent with the removal of 10-11 bp from the 5' end of the spacer during maturation of crRNAs (37). Instead, deletion of the protospacer constitutes a means of evading Cas-mediated cleavage (Figure 2.3). Predictably, mutations that interfere with biogenesis of crRNAs or activity of tracrRNA may also abolish activity, but inactivation of CRISPR-Cas systems constitutes a significant cost to the cell. Bacterial means of escape mirror those of the predominant target of CRISPR-Cas systems, phage populations. In phage, mechanisms facilitating circumvention of CRISPR-Cas targeting typically involve alteration of the PAM, seed or protospacer sequences; achieved through recombination or spontaneous mutation (14, 25, 34). Phage populations are inherently genetically diverse, yet exhibit a high frequency of homologous recombination (HR) events in response to CRISPR-Cas targeting (25). Recombination in bacterial populations is limited by natural barriers preventing accessibility of homologous yet variable DNA segments. However, Jiang and coworkers (22) demonstrated that introduction of exogenous genomic DNA (gDNA) caused recombination-mediated survival within *Streptococcus pneumoniae*, which suggests potential roles of competence and horizontal gene transfer in survival of self-targeting events in bacteria.

2.5 Lessons from bacterial genome editing studies

The lethal effects of CRISPR-Cas self-targeting in bacteria are well reported, but few studies have investigated the molecular outcomes of self-targeting events. Recent work by independent groups have provided invaluable insight into the intersection of genome homeostasis and CRISPR-Cas self-targeting (20-22). These foundational studies pave the way for widespread implementation of CRISPR-Cas technology as a genome editing tool in
bacteria. The experiments revealed the mechanistic underpinnings of CRISPR-Cas targeting, exploited DNA repair/replication pathways for designed genome edits, and delineated the genomic plasticity of bacterial populations using CRISPR-Cas targeting.

2.5.1 Homology directed repair in *S. pneumoniae*

Jiang *et al.* (22) was a landmark study for genome editing using CRISPR-Cas9 in general, and was also the first to demonstrate bacterial genome editing. The study determined that double-crossover HR with a donor template restored CRISPR-Cas9 effected chromosomal injury in *S. pneumoniae*. A prophage served as the target for Cas9 cleavage using two derivative strains of *S. pneumoniae*, differing from the wild-type in an integrated prophage at the *srtA* locus in the chromosome and a prophage integrated strain with a mutated PAM site. Transformation of *S. pneumoniae* cr6 gDNA encoding the prophage targeting Cas9::sgRNA was expected to be lethal to the prophage harboring strain, but not for the strain with the mutated PAM sequence. However, HR at the *srtA* locus was observed, ultimately resulting in the efficient recovery of recombinant clones with deleted prophage genotypes. A similar result was achieved upon co-transformation of a wild-type *srtA* linear editing template with the prophage targeting Cas9::sgRNA (Figure 2.4). The study performed thorough assessments of protospacer and PAM mutations that circumvent Cas9 targeting, providing characterization of the 5'-NGG-3'PAM requirements and seed sequence for the Type II system from *S. pyogenes*. Moreover, the study highlighted the ability to introduce targeted missense mutations as well as whole gene deletions using double-crossover HR in the β-galactosidase encoding gene in *S. pneumoniae*. In order to assess the efficiency of
bacterial genome editing with or without the assistance of Cas9 cleavage, the authors quantified the mutation rate of artificial stop codon in an erythromycin resistance gene with the outcome restoring the EmR phenotype. The experiment revealed a marginal induction of recombination through Cas9 targeting of the stop codon, but even in the absence of Cas9 cleavage, a subpopulation of cells appeared to undergo transformation or recombination at higher frequencies. The study not only established the utility of Cas9-mediated genome editing, but elucidated the molecular underpinnings of the efficiency and limitations of the system.

2.5.2 Efficient and targeted mutagenesis with recombineering in Lactobacillus reuteri

Oh and van Pijkeren employed CRISPR-Cas9 self-targeting in tandem with recombineering for the selection of desired mutations to achieve targeted mutagenesis at nucleotide resolution in L. reuteri (Figure 2.4) (21). Plasmid-based expression of recT, a single-stranded DNA-binding protein, and cas9 was used for single step and dual step strategies by introducing single-stranded oligonucleotides conferring circumvention of Cas9 targeting. The oligonucleotides were designed to harbor a non-targeted PAM sequence to circumvent Cas9 cleavage effectively and avoid competitive Cas9 binding of the oligonucleotide and possible displacement of the RecT protein (38). The study demonstrated that low frequency mutations such as whole gene deletions could be introduced in the cell populations using single-stranded recombineering, and these mutations could be selected for by applying CRISPR-Cas targeting against the wild-type genotype. The study also demonstrates the robustness of the system for mutational biochemical characterization of
proteins through introducing missense mutations using Cas9-assisted codon saturation. Similarly to Jiang et al. (22), the authors address concerns of making mutations at additional non-PAM sites while still using the highly effective PAM-mutation based circumvention. As reported by Oh and van Pijkeren, recombineering represents a precise and efficacious genetic tool in *L. reuteri*, especially when applying CRISPR-Cas assisted selection for desired mutations. However, successful application of this technology may require considerable optimization for use in disparate backgrounds, and the universal efficacy of the method remains undetermined given the significant variation in the capacity of bacteria to carry out recombineering.

2.5.3 Alternative end joining and large deletions in *Pectobacterium atrosepticum*

In contrast to the other bacterial genome editing studies, Vercoe et al. (20) used a natively active Type I-F system, which are less suited for introducing designed mutations, due to the unpredictable and extensive nature of DNA damage caused by Cas3 exonuclease activity. It was demonstrated that lethality from a self-targeting spacer in *P. atrosepticum* was abrogated due to a single PAM sequence mutation, but self-targeting was restored through plasmid-based expression of a programmable repeat-spacer array. In the absence of a donor template for HR, large deletions were achieved through recombination of mobile genetic element features (Figure 2.4). Deletion of the entire pathogenicity island (~98 kbp) occurred reproducibly through recombination of *attL* and *attR* sites flanking the island. Interestingly, the authors reported that the deletion occurred spontaneously at low frequency in wild-type cell populations as detected by PCR amplification of the *attB* excision footprint.
In contrast, variable deletions within the pathogenicity island consistent with A-EJ also occurred through microhomologous sequences. The two mechanisms of mutation observed in this study highlight the ordinal effect of mutation and DNA damage since it can be postulated that deletion of the island was observed through selection of preexisting mutations (Figure 2.2), whereas deletions due to A-EJ likely occurred subsequent to cleavage of the chromosome (Figure 2.2).

2.6 Design considerations

The exceptional range of microbial diversity poses shared challenges for many genetic manipulation strategies. The mechanisms governing DNA homeostasis are highly background specific, thus genetic tool development is limited by factors including transformation efficiency, plasmid replication, capacity for recombination/integration, DNA methylation, and DNA repair pathways. Similarly, application of CRISPR-Cas9 technology for bacterial genome editing hinges upon these molecular processes, but also has specific design considerations for accurate and efficient use of the technology.

2.6.1 Design of spacer sequences

Off target cleavage is expected to occur infrequently in bacterial systems relative to eukaryotes, which can be attributed to lower occurrence of sequences homologous to a given spacer-PAM combination in smaller genomes. Given that single self-targeting events result in a significant reduction in recovery of viable transformants, any off target cleavage leading
to multiple events of CRISPR-Cas induced chromosomal injury in a single cell should compound this reduction in recovery, leading to a decreased incidence of mutation at extraneous loci. However, proper selection of spacer sequences is essential to further prevent unintended cleavage events and to maximize efficiency. To date, two strict criteria for selection and design of spacers are the location of consensus PAM sequences and avoiding incidental sequence identity to extraneous genomic loci. Putative protospacers are constrained by defining the location of putative PAM sequences in the target locus. Since PAM and seed sequences are integral for recognition and activity at the target, spacers must be selected based on uniqueness of these components to prevent off-target cleavage. In Type II-A systems, approximately 10 nt are removed from the 5’ end of the spacer during crRNA maturation, suggesting that they are irrelevant for target specificity (37). Protospacers containing sequences identical to the PAM should not be considered to prevent competitive Cas recruitment that may limit cleavage of the desired locus (38). There is no webtool dedicated to spacer design for bacterial genome editing, and the utility of current eukaryotic tools for designing bacterial spacers is undetermined.

2.6.2 Increasing transformation efficiency for optimized system delivery

Transformation efficiencies are limiting in many backgrounds, but transformation (natural or induced), transduction, and conjugation are all potential avenues for experiments requiring simultaneous co-transformation of editing templates and expression vectors. The lethality of self-targeting compounds the need for high transformation efficiency of CRISPR-Cas components when interference is the direct result of transformation. Limiting
transformation efficiency can be compensated for by designing high frequency mutation strategies facilitating circumvention of CRISPR-Cas targeting, thus increasing recovery of desired genotypes. Development of tightly regulated inducible expression systems bypasses low transformation efficiency of self-targeting plasmids, allowing for induction of self-targeting in highly concentrated cultures. Inducible systems could therefore increase recovery of desired mutations and identification of low frequency mutations. An intriguing observation noted by Jiang et al. (22) suggested that certain bacterial subpopulations were more prone to recombination/transformation. Experiments with multiple rounds of recombination/transformation may therefore constitute short-term directed 'evolution'. This process results in a disproportionate selection of the population with a higher competency phenotype, with potential applications in molecular biology for backgrounds with low transformation or recombination efficiencies.

2.6.3 DNA repair mechanisms in bacteria

Since mutations elicited by CRISPR-Cas systems are both DNA damage and repair machinery dependent, it is prudent to consider the DNA repair pathways present in each microbiological background. The universality of HR, its well-characterized mechanism, and relatively high frequency have led to the widespread development and use of HR-mediated genetic technologies in diverse bacterial backgrounds. Therefore, HR can similarly be considered a practical and applicable repair pathway for introduction of mutations in CRISPR-Cas assisted genome editing. HR-based repair of DSBs and simultaneous generation of desired mutations can be achieved through provision of a homologous editing
template, which affords the potential for both deletions and gene replacements. Recombination can occur between native homologous sequences flanking DSBs in the genome, resulting in deletion of large genomic segments (20). To this end, native repair pathways can either be exploited for generation of desired mutations, or enzymatic machinery can be heterologously expressed to introduce repair pathway platforms for mutagenesis. In particular, recombinant expression of proteins Ku and LigD in bacteria may offer an avenue for high frequency generation and recovery of mutants. In contrast, native pathways may introduce undesired mutations at high frequencies relative to that of designed genetic outcomes, making it necessary to knockout or transcriptionally down-regulate certain pathways to prevent undesired repair from occurring. Conversely, targeting genomes with CRISPR-Cas systems also affords the potential for characterization of native DNA repair pathways in diverse microbial backgrounds.

2.7 Future Applications

CRISPR-Cas selection against target sequences has already been tangibly exploited in a handful of bacteria to elicit genome edits, but further development of the technology has considerable potential for revolutionizing bacterial genetics and genomics.

2.7.1 Cas9 and editing template delivery through transformation of linear nucleic acids

Linear DNA transformation has not been broadly applied as a tool for genetic manipulation in bacteria since replication of DNA requires circularity and host exonuclease
activity causes rapid degradation of linear dsDNA. A few studies have employed linear DNA to elicit gene replacement by double-crossover recombination with success in *E. coli*, *Bordetella pertussis*, *S. thermophilus* and *S. pneumoniae* (22, 39-41). Synthetic DNA molecules are easily designed and affordable for engineered mutagenesis and splicing by overlap extension PCR also allows for generation of editing templates (42). Thus, linear DNA transformation techniques may increasingly be considered a viable option for genome editing when complemented with CRISPR-Cas selection against the wild-type. Transformation of linear RNA is another unexplored avenue for mutagenesis of bacteria, but with the strong selective pressure of Cas9 targeting, RNA-based expression of Cas9 and sgRNAs holds potential for increased throughput in genome editing of bacteria.

### 2.7.2 Exploitation of endogenous and orthogonal systems

CRISPR-Cas systems are found in approximately 46% of bacteria and 84% of archaea (43), highlighting the potential for genome editing applications using endogenously active systems (44). However, there is a general paucity of systems with characterized PAMs and confirmed activity in interference (45). Self-targeting or plasmid interference assays must first be performed to ensure activity and assess PAM recognition. Platforms for genome editing in CRISPR deficient backgrounds may be generated through plasmid or chromosomal expression of vested Cas9s and sgRNAs (46). The use of multiple orthogonal systems offers the advantage of utilizing disparate PAM sequences, thus increasing the range of target sequences without sacrificing specificity (47). Rather, longer PAMs may offer increased specificity and potentially decreased off-target cleavage. Extended PAMs may increase
efficacy of cleavage by further compounding the low frequency at which point mutations effectively confer circumvention of Cas9 recognition. It is noteworthy, however, that recent evidence indicates biochemical recognition of PAMs is not as stringent as bioinformatically determined consensus sequences would suggest (48). Therefore, it may be necessary to empirically determine the specific nucleotides contributing to stringent PAM recognition for each CRISPR-Cas system.

2.7.3 Understanding bacterial genome biology

The high prevalence of mobile genetic elements in bacterial genomes (49) presents a unique challenge for eliciting targeted mutations at these loci. Excision of mobile genetic elements may occur in the face of CRISPR-Cas selective pressure (20), dependent on the frequency of excision relative to that of the desired mutation. Therefore, in silico prediction of mobile genomic segments may be used to identify potentially expendable regions, which can then be experimentally validated with CRISPR-Cas selection. Despite the potential difficulty generating designed mutations in these segments, there are considerable applications for excision of mobile genetic elements, such as defining minimal bacterial genomes and characterization of putative un-annotated proteins and essential genes. CRISPR-Cas selection can therefore be used to screen for clonal subtypes within heterogenous populations, delineating the locus-dependent plasticity of bacterial genomes.
2.8 Concluding remarks

Development of CRISPR-Cas technology in bacteria has yielded applications in typing and strain detection (50, 51), exploitation of natural/engineered immunity against mobile genetic elements (52-54), manipulation of microbial consortia/generation of smart antibiotics (29), and programmable transcriptional regulation (45). However, few studies have focused on the development of CRISPR-Cas genome editing tools in bacteria. This streamlined methodology holds potential for increasing the expediency and efficiency in generation of desired mutations, potentially without the necessity of plasmid integration, extensive screening, or counter-selection. Microbial diversity necessitates development of efficient transformation protocols and genetic tools for bacterial genome editing, but CRISPR-Cas technology opens new avenues in genetic engineering applications.
2.9 References


CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. PLoS Genet 9:e1003454.


development of natural competence in *Streptococcus thermophilus* and *Streptococcus salivarius*. J Bacteriol 192:1444–1454.


**Figure 2.1.** CRISPR-Cas9 targeting of DNA.
Figure 2.2. Manipulation of microbial composition in defined consortia.
Figure 2.3. Bacterial mechanisms of escape from CRISPR-Cas targeting
Figure 2.4. DNA repair outcomes of genome targeting CRISPR-Cas systems.
CHAPTER III

CRISPR-based screening of genomic island excision events in bacteria.

Selle K, Klaenhammer TR, Barrangou R.
3.1 Abstract

Genomic analysis of *Streptococcus thermophilus* revealed that mobile genetic elements (MGEs) likely contributed to gene acquisition and loss during evolutionary adaptation to milk. CRISPR-Cas, the adaptive immune system in bacteria, limits genetic diversity by targeting MGEs including bacteriophages, transposons, and plasmids. CRISPR-Cas systems are widespread in streptococci, suggesting that the interplay between CRISPR-Cas systems and MGEs is one of the driving forces governing genome homeostasis in this genus. To investigate the genetic outcomes resulting from CRISPR-Cas targeting of integrated MGEs, *in silico* prediction revealed 4 genomic islands without essential genes in lengths from 8 to 102 kbp, totaling 7% of the genome. In this study, the endogenous CRISPR3 Type II system was programmed to independently target the four islands through plasmid-based expression of engineered CRISPR arrays. Targeting *lacZ* within the largest 102 kbp genomic island was lethal to wild-type cells and resulted in up to a 2.5-log reduction in the surviving population. Genotyping of Lac− survivors revealed variable deletion events within the flanking IS-elements, all resulting in elimination of the Lac-encoding island. Chimeric insertion sequence footprints were observed at the deletion junctions after targeting all of the four genomic islands, suggesting a common mechanism of deletion via recombination between flanking insertion sequences. These results established that self-targeting CRISPR-Cas systems may direct significant evolution of bacterial genomes on a population level, influencing genome homeostasis and remodeling.
3.2 Introduction

Mobile genetic elements (MGEs) present bacteria with continuous challenges to genomic stability, promoting evolution through horizontal gene transfer. The term MGE encompasses plasmids, bacteriophages, transposable elements, genomic islands, and many other specialized genetic elements (1). MGEs encompass genes conferring high rates of dissemination, adaptive advantages to the host, and genomic stability, leading to their near universal presence in bacterial genomes. To cope with the permanent threat of predatory bacteriophages and selfish genetic elements, bacteria have evolved both innate and adaptive immune systems targeting exogenous genetic elements. Innate immunity includes cell-wall modification, restriction/modification systems, and abortive phage infection (2). Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated genes (Cas) are an adaptive immune system targeted against invasive genetic elements in bacteria (3). CRISPR-Cas mediated immunity relies on distinct molecular processes, categorized as acquisition, expression, and interference (3). Acquisition occurs via molecular ‘sampling’ of foreign genetic elements, from which short sequences, termed spacers, are integrated in a polarized fashion into the CRISPR array (4). Expression of CRISPR arrays is constitutive and inducible by promoter elements within the preceding leader sequence (5, 6). Interference results from a corresponding transcript that is processed selectively at each repeat sequence, forming CRISPR RNAs (crRNAs) that guide Cas proteins for sequence-specific recognition and cleavage of target DNA complementary to the spacer (7). CRISPR-Cas technology has applications in strain typing and detection (8-10), exploitation of natural/engineered immunity against mobile genetic elements (11), programmable genome
editing in diverse backgrounds (12), transcriptional control (13, 14), and manipulation of microbial populations in defined consortia (15).

Although sequence features corresponding to CRISPR arrays were described previously in multiple organisms (16, 17), *Streptococcus thermophilus* was the first microbe where the roles of specific *cas* genes and CRISPR-array components were elucidated (4). *S. thermophilus* is a non-pathogenic thermophilic Gram-positive bacterium used as a starter culture that catabolizes lactose to lactic acid in the syntrophic production of yogurt and various cheeses (18). *S. thermophilus* encodes up to four CRISPR-Cas systems, two of them (CRISPR1 and CRISPR3) are classified as Type II-A systems that are innately active in both acquisition and interference (4, 19). Accordingly, genomic analysis of *S. thermophilus* and its bacteriophages established a likely mechanism of CRISPR-Cas systems for phage/DNA protection. Investigation of CRISPR-Cas systems in *S. thermophilus* led to bioinformatic analysis of spacer origin (4, 20), discovery of the proto-spacer adjacent motif (PAM) sequences (19, 21), understanding of phage-host dynamics (22, 23), demonstration of Cas9 endonuclease activity (7, 24, 25), and recently, determination of the tracrRNA structural motifs governing function and orthogonality of Type II systems (26). Genomic analysis of *S. thermophilus* revealed evolutionary adaptation to milk through loss of carbohydrate catabolism and virulence genes found in pathogenic streptococci (18). *S. thermophilus* also underwent significant acquisition of niche-related genes, such as those encoding including cold-shock proteins, copper resistance proteins, proteinases, bacteriocins, and lactose catabolism proteins (18). Insertion sequences (ISs) are highly prevalent in *S. thermophilus* genomes and contribute to genetic heterogeneity between strains by facilitating dissemination.
of islands associated with dairy adaptation genes (18). The concomitant presence of MGEs and functional CRISPR-Cas systems in \textit{S. thermophilus} suggests that genome homeostasis is governed at least in part by the interplay of these dynamic forces. Thus, \textit{S. thermophilus} constitutes an ideal host for investigating the genetic outcomes of CRISPR-Cas targeting of genomic islands.

CRISPR-Cas systems have recently been the subject of intense research in genome editing applications (12), but the evolutionary roles of most endogenous microbial systems remain unknown (27). Even less is known concerning evolutionary outcomes of housing active CRISPR-Cas systems beyond the prevention of foreign DNA uptake (7), spacer acquisition events (4), and mutation caused by chromosomal self-targeting (28-32). Thus, we sought to determine the outcomes of targeting integrated MGEs with endogenous Type II CRISPR-Cas systems. Four islands were identified in \textit{S. thermophilus} LMD-9, with lengths ranging from 8 to 102 kbp and totaling approximately 132 kbp, or 7% of the genome. In order to target genomic islands, plasmid-based expression of engineered CRISPR arrays with self-targeting spacers were transformed into \textit{S. thermophilus} LMD-9. Collectively, our results elucidate fundamental genetic outcomes of self-targeting events and show that CRISPR-Cas systems can direct genome evolution at the bacterial population level.

### 3.3 Results

#### 3.3.1 Identification of expendable genomic regions

\textit{In silico} prediction of mobile and expendable loci for CRISPR-Cas targeting was performed on the basis of i) location, orientation, and nucleotide identity of IS elements, and
ii) location of essential ORFs. In *Bacillus subtilis*, 271 essential ORFs were identified by determining the lethality of genome-wide gene knockouts (33). The *S. thermophilus* genome was queried for homologues to each essential gene from *B. subtilis* using the BLASTp search tool under the default scoring matrix for amino acid sequences. Homologues to ~239 essential ORFs were identified in *S. thermophilus*, all of which were chromosomally encoded. Proteins involved in conserved cellular processes including DNA replication/homeostasis, translation machinery, and core metabolic pathways were readily identified. No homologues corresponding to cytochrome biosynthesis/respiration were observed, in accordance with the metabolic profile of fermentative bacteria. Each putative essential ORF was mapped to the reference genome using SnapGene software, facilitating visualization of their location and distribution in *S. thermophilus* LMD-9 (Figure 3.1).

IS elements within the *S. thermophilus* genome were grouped by aligning transposon coding sequences using Geneious software (Figure 3.2 and Figure 3.3). Family designations were determined according to BLAST analysis within the IS element database (https://www-is.biotoul.fr//). To predict the potential for recombination-mediated excision of chromosomal segments, the relative location of related IS elements were mapped to the *S. thermophilus* genome (Figure 3.1). The IS1193 and Sth6 families of IS elements appeared most frequently in the genome and are commonly found in *Streptococcus pneumoniae* and *Streptococcus mutans* (34). IS1191 elements were not frequent but exhibited near perfect fidelity between the copies identified in the genome (Figure 3.2). Despite the prevalence of IS1193 elements, many of these loci were shown to be small fragments that exhibited some polymorphism and degeneracy, but there were also several copies present with a high level of sequence identity.
The Sth6 family exhibited considerable polymorphism and high degeneracy, with some copies harboring significant internal deletions (Figure 3.2). IS1167 elements were relatively well conserved (Figure 3.2). IS1167 and IS1191 elements of *S. thermophilus*, and their relative proximity to milk adaptation genes, we postulate that these conserved/high fidelity transposons were recently acquired in the genome.

By combining the location of predicted essential ORFs and IS elements, expendable islands flanked by IS elements of high identity were identified (Figure 3.1) (Table 3.2). The first island contained an operon unique to *S. thermophilus* LMD-9, encoding a putative ATP-dependent oligonucleotide transport system with unknown specificity (Figure 3.4) (35). The second harbors the cell-envelope proteinase PrtS which contributes to the fast-acidification phenotype of *S. thermophilus* (Figure 3.4) (36). Notably, while *prtS* is not ubiquitous in *S. thermophilus* genomes, it has been demonstrated that the genomic island encoding *prtS* is transferable between strains using natural competence (36). The third island contains a putative ATP-dependent copper efflux protein and is present in every sequenced *S. thermophilus* strain (Figure 3.4). The fourth island is the largest by far in terms of length at 102 kbp, and gene content, with 102 predicted ORFs including the *lac* operon (Figure 3.4). This island is found in all strains of *S. thermophilus*, but the specific gene content and length varies among strains. In order to determine the outcome of targeting a large genomic island with both endogenous Type II systems, repeat-spacer arrays were generated for the *lacZ* coding sequence and cloned into pORI28 (Table 3.1). The fourth island was selected for
CRISPR-Cas targeting due to its size, ubiquity in *S. thermophilus* strains, and the ability to screen for *lacZ* mutations on the basis of a β-galactosidase negative phenotype.

### 3.3.2 CRISPR-Cas targeting of *lacZ* selects for large deletion events

In Type II systems, Cas9 interrogates DNA and binds reversibly to PAM sequences with activation of Cas9 at the target occurring via formation of the tracrRNA::crRNA duplex (37), ultimately resulting in dsDNA cleavage (25). Transformation with plasmids eliciting chromosomal self-targeting by CRISPR-Cas systems appeared cytotoxic as measured by the relative reduction in surviving transformants compared to non-self-targeting plasmids (15, 29). Targeting the *lacZ* gene in *S. thermophilus* resulted in a ~2.5-log reduction in recovered transformants (Figure 3.5), approaching the limits of transformation efficiency. Double-stranded DNA breaks (DSBs) constitute a significant threat to the survival of organisms. The corresponding repair pathways often require end resection to repair blunt-ended DNA. Cas9-effected endonucleolysis further exacerbates the pressure for mutations caused by DSBs to occur, as restoration of the target locus to the wild-type does not circumvent subsequent CRISPR targeting. Identification of spacer origins within lactic acid bacteria revealed that 22% of spacers exhibit complementarity to self and that the corresponding genomic loci were altered, likely facilitating survival of naturally occurring self-targeting events (28).

To determine if the target locus was mutated in response to Cas9-induced cleavage, transformants were first screened for loss of β-galactosidase activity. Clones deficient in activity were genotyped at the *lacZ* locus. No mutations due to classical or alternative end joining, nor any spontaneous single nucleotide polymorphisms were observed in any of the
clones sequenced. The absence of single nucleotide polymorphisms may be attributed to a low transformation efficiency compounded by low incidence of point mutations, and the absence of Ku and LigaseIV homologs correlated with an absence of non-homologous end joining (38). PCR screening indicated that the wild-type lacZ was not present, but the PCR amplicons did not correspond to the native lacZ locus; rather, an IS element-flanked sequence at another genomic locus was amplified. To investigate the genotype responsible for the loss of β-galactosidase activity, Single Molecule Real Time sequencing was performed on two clones; one generated from CRISPR3 targeting the 5' end of lacZ, and one generated from CRISPR3 targeting the sequence encoding the ion-binding pocket necessary for β-galactosidase catalysis (Figure 3.6). This sequencing strategy was employed for its long read length to circumvent difficulty in reliably mapping reads to the proper locus, due to the high number of IS elements in the genome (35). Reads were mapped to the reference genome sequence using Geneious software, and revealed the absence of a large segment (~102 kbp) encoding the lacZ open reading frame (Figure 3.6). Both sequenced strains confirmed the reproducibility of the large deletion boundaries, and showed that the deletion occurred independently of the lacZ spacer sequence or CRISPR-Cas system used for targeting. However, the sequencing data did not reliably display the precise junctions of the deletion.

The 102 kbp segments deleted constitute approximately 5.5% of the 1.86 Mbp genome of S. thermophilus. The region contained 102 putative ORFs (STER_1278-1379), encoding ABC transporters, two-component regulatory systems, bacteriocin synthesis genes, phage related proteins, lactose catabolism genes, and several cryptic genes with no annotated function (35). The effect of the deletion on growth phenotype was assessed in broth culture
by measuring OD 600 nm over time (Fig. 3 C). The deletion clones appeared to have a longer lag phase, lower final OD (p<0.01) and exhibited a significantly longer generation time during log phase with an average of 103 min, compared to 62 min for the wild-type (p<0.001). Although the deletion derivatives have 5.5% less of the genome to replicate per generation, and expend no resources in transcription or translation of the eliminated ORFs, no apparent increase in fitness was observed relative to the wild-type. β-galactosidase activity is a hallmark feature for industrial application of lactic acid bacteria and is essential for preservation of food systems through acidification. The capacity of lacZ deficient S. thermophilus strains to acidify milk was therefore assessed by monitoring pH (Figure 3.7). Predictably, the deletion strain failed to acidify milk over the course of the experiment, in sharp contrast to the rapid acidification phenotype observed in the wild-type.

### 3.3.3 Genomic deletions occur through recombination between homologous IS elements

In order to investigate the mechanism of deletion, the nucleotide sequences flanking the segment were determined. The only homologous sequences observed at the junctions were two truncated IS1193 insertion sequences exhibiting 91% nucleotide sequence identity globally over 727 bp. Accordingly, a primer pair flanking the two IS elements was designed to amplify genomic DNA of surviving clones exhibiting the deletion. Each of the deletion strains exhibited a strong band of the predicted size (~1.2 kb), and confirmed the large genomic deletion event (Fig. 4 A). Interestingly, a faint amplicon corresponding to the chromosomal deletion was observed in the wild-type, indicating that this region may naturally excise from the genome at a low rate within wild-type populations. Sequencing of
the junction amplicon was performed for 20 clones generated by chromosomal self-targeting by CRISPR3. Genotyping of the locus revealed the presence of one chimeric IS element in each clone and, furthermore, revealed the transition from the upstream element to the downstream sequence within the chimera for each clone (Figure 3.7). The size of deletions observed ranged from 101,865-102,146 bp. The exact locus of transition was variable, but non-random within the clones, implying the potential bias of the deletion mechanism. *S. thermophilus* harbors typical recombination machinery encoded as RecA (STER_0077), AddAB homologs functioning as dual ATP-dependent DNA exonucleases (STER_ 1681 and STER_ 1682), and a helicase (STER_1742) of the RecD family. The high nucleotide identity between the flanking IS elements and the capacity for *S. thermophilus* to carry out site-specific recombination (4) confirms the potential for RecA-mediated recombination to mediate excision of the genomic segment (Figure 3.7).

It was next hypothesized that CRISPR-Cas targeting could facilitate isolation of deletions for each locus with the same genetic architecture. Thus, three CRISPR3 repeat-spacer arrays, one targeting the oligonucleotide transporter in the first locus, *prtS* from the second locus, and the ATPase copper efflux gene from the third locus were generated and cloned into pORI28 (Table 3.2). In order to screen for deletions, primers flanking the IS elements at each locus were designed to amplify each deletion junction (Figure 3.7). The absence of wild-type loci was also confirmed in each case by designing internal primers for each genomic island (Figure 3.7). Following transformations with the targeting plasmids, deletions at each locus were isolated and the absence of wild-type confirmed. Sequencing of the deletion junction amplicons confirmed that a single chimeric IS element footprint
remained, indicating a common mechanism for deletion at each locus. Interestingly, primers flanking the IS elements also amplified from wild-type gDNA, further suggesting that population heterogeneity naturally occurred at each locus was due to spontaneous genomic deletions. These results imply that sequence-specific Cas9 cleavage selects for the variants lacking protospacer and PAM combinations necessary for targeting. Thus, spontaneous genomic deletions can be isolated using CRISPR-Cas targeting as a strong selection for microbial variants that have already lost those genomic islands.

3.4 Discussion

In this study, native Type IIA systems harbored in *S. thermophilus* were repurposed for defining spontaneous deletions of large genomic islands. By independently targeting four islands in *S. thermophilus*, stable mutants collectively lacking a total of 7% of the genome were generated. Characterization of the deletion junctions suggested that an IS-dependent recombination mechanism contributes to population heterogeneity and revealed deletion events ranging from 8 to 102 kbp. Precise mapping of the chimeric IS elements indicated that natural recombination events are likely responsible for the large chromosomal deletions in *S. thermophilus* and could potentially be exploited for targeted genome editing. Recent landmark studies have highlighted the potential for CRISPR-Cas induced chromosomal deletions and rearrangements in bacteria (29, 30). Jiang and coworkers first reported that sequence-specific Cas9 cleavage selects for pre-existing variants lacking protospacer and PAM combinations necessary for targeting in *Streptococcus pneumoniae*. Similarly, Vercoe and coworkers demonstrated that chromosomal targeting by a Type-IF CRISPR-Cas system
caused elimination of a horizontally acquired pathogenicity island in *Pectobacterium atrosepticum* (30). The concept of sequence-based removal of specific genotypes was further developed as a tool for manipulation of microbial consortia via CRISPR-Cas targeting, resulting in directed genome evolution at the population level (15). In accordance with previous work, our results demonstrate that wild-type clones were removed from the population while mutants without CRISPR-Cas targeted features survived. Thus, adaptive islands were identified and validated, showing that precise targeting by an endogenous Cas9 can be exploited for isolating large deletion variants in mixed populations.

Genome evolution of bacteria occurs through horizontal gene transfer, intrinsic mutation, and genome restructuring. Genome sequencing and comparative analysis of *S. thermophilus* strains has revealed significant genome decay, but also indicates that adaptation to nutrient-rich food environments occurred through niche-specific gene acquisition (18; 35). The presence of MGEs including integrative and conjugative elements, prophages, and IS elements in *S. thermophilus* genomes is indicative of rapid evolution to a dairy environment (39-40). Mobile genetic features facilitate gene acquisition and conversely, inactivation or loss of non-essential sequences. Consequently, MGEs confer genomic plasticity as a means of increasing fitness or changing ecological lifestyles. Our results strongly indicate that CRISPR-Cas targeting of these elements may influence chromosomal rearrangements and homeostasis. This is in contrast to experiments targeting essential features, which resulted in selection of variants with inactivated CRISPR-Cas machinery (41). Mutation of essential ORFs is not a viable avenue for circumvention of CRISPR-Cas targeting, and thus only those clones with inactivated CRISPR-Cas systems remain. By design, targeting genetic elements
predicted to be hypervariable and expendable demonstrated that variants with altered loci were viable, maintaining active CRISPR-Cas systems during self-targeting events.

Despite the near ubiquitous distribution of IS elements in bacterial genomes they remain an enigmatic genetic entity, largely due to their diversity and plasticity in function (34). Our results suggest it is possible to predict recombination between related IS elements by analyzing their location, orientation, and sequence conservation (Fig. S1-S2). CRISPR-Cas targeting can then be employed to empirically validate population heterogeneity at each predicted locus, and simultaneously increase the recovery of low incidence mutants. The high prevalence of MGEs in lactic acid bacteria, and especially _S. thermophilus_, is in accordance with their role in speciation of these hyper-adapted bacteria through genome evolution (39-40). Moreover, recovery of genomic deletion mutants using CRISPR-Cas targeting could facilitate phenotypic characterization of genes with unknown function.

Mutants exhibiting the deletion of the 102 kb island encoding the _lac_ operon had significantly increased generation times relative to the wild-type and achieved a lower final OD. With 102 predicted ORFs therein, it is likely that additional phenotypes are affected and many of the genes do not have annotated functions. Considering the industrial relevance of niche-specific genes such as _prtS_, this method allows for direct assessment of how island-encoded genes contribute to adaption to grow in milk. Moreover, it is in the natural genomic and ecological context of these horizontally acquired traits, since they were likely acquired as discrete islands. These results establish new avenues for the application of self-targeting CRISPR-Cas9 systems in bacteria for investigation of transposition, DNA repair mechanisms, and genome plasticity.
CRISPR-Cas systems generally limit genetic diversity through interference with genetic elements, but acquired MGEs can also provide adaptive advantages to host bacteria. Thus, the benefit of maintaining genomically integrated MGEs despite CRISPR-Cas targeting is an important driver of genome homeostasis. Collectively, our results establish that in silico prediction of GEIs can be coupled with CRISPR-Cas targeting to isolate clones exhibiting large genomic deletions. Chimeric insertion sequence footprints at each deletion junction indicated a common mechanism of deletion for all four islands. The high prevalence of self-targeting spacers exhibiting identity to genomic loci, combined with experimental demonstrations of genomic alterations, suggest that CRISPR-Cas self-targeting may contribute significantly to genome evolution of bacteria (28, 30). Collectively, studies on CRISPR-Cas induced large deletions substantiate this approach as a rapid and effective means to assess the essentiality and functionality of gene clusters devoid of annotation, and define minimal bacterial genomes based on chromosomal deletions occurring through transposable elements.

3.5 Materials and Methods

3.5.1 Bacterial Strains

All bacterial strains are listed in Table 3.3. Bacterial cultures were cryopreserved in an appropriate growth medium with 25% glycerol (vol/vol) and stored at -80°C. *S. thermophilus* was propagated in Elliker media (Difco) supplemented with 1% beef extract (wt/vol) and 1.9% (wt/vol) β-glycerolphosphate (Sigma) broth under static aerobic conditions.
at 37˚C, or on solid medium with 1.5% (wt/vol) agar (Difco), incubated anaerobically at 37˚C for 48 hours. Concentrations of 2 µg/mL of erythromycin (Em) and 5 µg/mL of chloramphenicol (Cm) (Sigma) were used for plasmid selection in *S. thermophilus*, when appropriate. *E. coli* EC1000 was propagated aerobically in Luria-Bertani (Difco) broth at 37˚C, or on brain-heart infusion (BHI) (Difco) solid medium supplemented with 1.5 % agar. Antibiotic selection of *E. coli* was maintained with 40 µg/mL kanamycin (Kn) and 150 µg/mL of Em for recombinant *E. coli*, when appropriate. Screening of *S. thermophilus* derivatives for β-galactosidase activity was assessed qualitatively by supplementing a synthetic Elliker medium with 1% lactose, 1.5% agar, and 0.04% bromo-cresol purple as a pH indicator.

### 3.5.2 DNA Isolation and Cloning

All kits, enzymes, and reagents were used according to the manufacturers' instructions. DNA purification and cloning were performed as described previously (42). Plasmids with *lacZ* targeting arrays were constructed with each consisting sequentially of the (1) native leader sequence specific to CRISPR1 or CRISPR3 (2) native repeats specific to CRISPR 1 or CRISPR 3 (3) spacer sequence specific to the 5’ end of *lacZ* (4) another native repeat. In order to engineer each plasmid, the sequence features listed above were ordered as extended oligomers (Table 3.1), combined using splicing by overlap extension PCR (42) and cloned into pORI28 (Table 3.1).
3.5.3 Selection and design of CRISPR spacers

Putative protospacers were constrained by first defining the location of all putative PAM sequences in the sense and antisense strands of lacZ. Within the 3,081 nt gene, there were 22 CRISPR1 (AGAAW) and 39 CRISPR3 (GGNG) PAM sites that were identical to their bioinformatically derived consensus sequences (21). After potential spacers were identified, the complete proto-spacer, seed, and PAM sequence were subjected to BLAST analysis against the genome of *S. thermophilus* LMD-9 to prevent additional targeting of non-specific loci. The spacers for CRISPR1 and CRISPR3 were disparate in sequence and corresponding PAM sites, but were designed to target the 5’ end of lacZ, resulting in predicted cleavage sites residing 6 nt apart. The leader sequences, repeats, and spacers on each plasmid represented orthogonal features unique to CRISPR1 or CRISPR3, respectively.

To assess target locus-dependent mutations, an additional CRISPR3 plasmid was created with a spacer to the metal cation-binding residue essential for β-galactosidase activity. A CRISPR1 array plasmid containing a non-self-spacer was used as a control to quantify lethality of self-targeting.

3.5.4 Transformation

Plasmids were electroporated into competent *S. thermophilus* containing the temperature-sensitive helper plasmid pTRK669. An overnight culture of *S. thermophilus* was inoculated at 1% (vol/vol) into 50 mL of Elliker medium supplemented with 1% beef extract, 1.9% β-glycerophosphate and Cm selection. When the culture achieved an OD$_{600}$ nm of 0.3, penicillin G was added to achieve a final concentration of 10 µg/mL. Cells were harvested by
centrifugation and washed 3x in 10 mL cold electroporation buffer (1 M sucrose and 3.5 mM MgCl$_2$). The cells were concentrated 100-fold in electroporation buffer and 40 µL of the suspension was aliquoted into 0.1 mm electroporation cuvettes. Each suspension was combined with 700 ng of plasmid. Electroporation conditions were set at 2,500 V, 25 µFd capacitance, and 200 Ohms resistance. Time constants were recorded and ranged from 4.4 to 4.6 ms. The suspensions were immediately combined with 950 µL of recovery medium and incubated for 8 hours at 37˚C. Cell suspensions were plated on selective medium and electroporation cuvettes were washed with medium to ensure recovery of cells.

3.5.5 Growth and activity assessment

Cultures were preconditioned for growth assays by subculturing for 12 generations in a semi-synthethic Elliker medium with glucose as the sole carbohydrate source. Fresh medium was inoculated with an overnight culture at 1% (vol/vol) and incubated at 37˚C statically. OD$_{600}$ monitored hourly until the cultures achieved stationary phase. Acidification of milk was assessed by inoculating skim milk with an overnight culture to a level of $10^8$ cfu/mL and incubating at 42˚C. The pH was subsequently monitored using a Mettler Toledo Seven Easy pH meter and Accumet probe. Skim milk was acquired from the NCSU Dairy plant and Pasteurized for 30 min at 80˚C.
3.6 References


Table 3.1. Primers and oligomers used in this study.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1_N-term_F</td>
<td>CAAGAACAGTTATGATTTTATAATCTACATACTATGTTGGGATGAAAATCTCAAAAATCTGTTTTTGT</td>
<td>Template for SOE-PCR</td>
</tr>
<tr>
<td>C1_N-term_R</td>
<td>AGTTCAAAATACGTCTCCTGGTGAATATAACAGGATCTCTAAACCTCAGAAAATCTTTCAAG</td>
<td>Template for SOE-PCR</td>
</tr>
<tr>
<td>C1(term_R)</td>
<td>GACAATCTCTTAAAGTG</td>
<td>Template for SOE-PCR</td>
</tr>
<tr>
<td>C3_N-site</td>
<td>AGCGGATAACAATTCACGTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC</td>
<td>Template for SOE-PCR</td>
</tr>
<tr>
<td>C3_ABC</td>
<td>ATCCACCCACCGTTTGGAACCATTCGAAACAACACAGCTCTAAAAC</td>
<td>Template for SOE-PCR</td>
</tr>
<tr>
<td>C3_prtS</td>
<td>AGCGGATAACAATTCACGTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC</td>
<td>Template for SOE-PCR</td>
</tr>
<tr>
<td>C3_Cu</td>
<td>AGCGGATAACAATTCACGTTTTTGAACCATTCGAAACAACACAGCTCTAAAAC</td>
<td>Template for SOE-PCR</td>
</tr>
<tr>
<td>C3LF</td>
<td>AGCAGGGATCCTGGTGAATATAAGTATAGATCGTTTG</td>
<td>Amplify Sth3 Leader from gDNA</td>
</tr>
<tr>
<td>C3LR</td>
<td>AGCAGGGATCCTGGTGAATATAAGTATAGATCGTTTG</td>
<td>Amplify Sth3 Leader from gDNA</td>
</tr>
<tr>
<td>C1F</td>
<td>AGCAGGGATCCTGGTGAATATAAGTATAGATCGTTTG</td>
<td>CRISPR1 SOE-PCR Forward</td>
</tr>
<tr>
<td>C3F</td>
<td>AGCAGGGATCCTGGTGAATATAAGTATAGATCGTTTG</td>
<td>CRISPR3 SOE-PCR Forward</td>
</tr>
<tr>
<td>C1C3R</td>
<td>TGCTGGAGCTCGTGAATATAAGTATAGATCGTTTG</td>
<td>CRISPR1 and CRISPR3 SOE-PCR Reverse Deletion junction amplification</td>
</tr>
<tr>
<td>1193F</td>
<td>TTGAACACTAAGACCGCTCATA</td>
<td>Deletion junction amplification</td>
</tr>
<tr>
<td>1193R</td>
<td>CGTAAAGCTTTTGGATGACTCAAG</td>
<td>Deletion junction amplification</td>
</tr>
<tr>
<td>pORI28F</td>
<td>TTGGTTTGATAATGAACTGTGGTG</td>
<td>Sequencing MCS of pORI28</td>
</tr>
<tr>
<td>pORI28R</td>
<td>TTGGTTTGATAATGAACTGTGGTG</td>
<td>Sequencing MCS of pORI28</td>
</tr>
<tr>
<td>GEI</td>
<td>ORF region</td>
<td>Length</td>
</tr>
<tr>
<td>------</td>
<td>------------------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>STER_139-STER_148</td>
<td>8,490</td>
</tr>
<tr>
<td>2</td>
<td>STER_840-STER_848</td>
<td>11,932</td>
</tr>
<tr>
<td>3</td>
<td>STER_881-STER_888</td>
<td>9,891</td>
</tr>
<tr>
<td>4</td>
<td>STER_1277-STER_1380</td>
<td>102,087</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.3 Bacterial Strains and Plasmids used in this study.

<table>
<thead>
<tr>
<th>Strain designation</th>
<th>Description</th>
<th>Original Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli EC1000</td>
<td>Host for pORI plasmids, chromosomal repA&lt;sup&gt;+&lt;/sup&gt; (pWVO1), Km&lt;sup&gt;R&lt;/sup&gt;, Host for pTRK935</td>
<td>Hanahan, 1985</td>
</tr>
<tr>
<td>S. thermophilus LMD-9</td>
<td>Wild-type</td>
<td>Makarova et al., 2009</td>
</tr>
<tr>
<td>S. thermophilus LMD-9 with pTRK669</td>
<td>Wild-type, RepA&lt;sup&gt;+&lt;/sup&gt; and Cm&lt;sup&gt;R&lt;/sup&gt; conferred by pTRK669</td>
<td>This study</td>
</tr>
<tr>
<td>Plasmids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pORI28</td>
<td>Broad range non-replicative vector, Em&lt;sup&gt;B&lt;/sup&gt;</td>
<td>Kok et al., 1984</td>
</tr>
<tr>
<td>pTRK669</td>
<td>Ts-helper plasmid repA&lt;sup&gt;+&lt;/sup&gt;, Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Russell and Klaenhammer, 2001</td>
</tr>
<tr>
<td>pCRISPR1::lacZ</td>
<td>pORI28::CRISPR1-Leader-RSR-lacZ N-terminus spacer</td>
<td>This study</td>
</tr>
<tr>
<td>pCRISPR3::lacZ</td>
<td>pORI28::CRISPR1-Leader-RSR-lacZ active site spacer</td>
<td>This study</td>
</tr>
<tr>
<td>pCRISPR3::ABC</td>
<td>pORI28::CRISPR1-Leader-RSR-ABC spacer</td>
<td>This study</td>
</tr>
<tr>
<td>pCRISPR3::Cu</td>
<td>pORI28::CRISPR1-Leader-RSR-Cu efflux spacer</td>
<td>This study</td>
</tr>
<tr>
<td>pCRISPR3::prts</td>
<td>pORI28::CRISPR1-Leader-RSR-prts spacer</td>
<td>This study</td>
</tr>
<tr>
<td>pCRISPR3::lacZ</td>
<td>pORI28::CRISPR3-Leader-RSR-lacZ N-terminus spacer</td>
<td>This study</td>
</tr>
<tr>
<td>pCRISPR1::Non-self</td>
<td>pORI28::CRISPR1-Leader-RSR-non self spacer</td>
<td>This study</td>
</tr>
</tbody>
</table>
Figure 3.1. Map of essential genes, insertion sequences, and genomic islands in *Streptococcus thermophilus*.
Figure 3.2. Dendrogram of transposon coding sequences distributed throughout the genome of *S. thermophilus* LMD-9.
Figure 3.3. Geneious alignments of transposon coding sequences.
Figure 3.4. Organization of genomic islands and corresponding CRISPR targets.
Figure 3.5. Transformants recovered after electroporation of control and self-targeting plasmids.
Figure 3.6. Genome sequencing and phenotypic analysis of Lac− clones.
Figure 3.7. Depiction of recombination events between ISs.
CHAPTER IV

Removal of lipoteichoic acid impacts the transcriptome and cell surface protein composition of *Lactobacillus acidophilus*
4.1 Abstract

*Lactobacillus acidophilus* NCFM is a well-characterized probiotic microorganism, supported by a decade of genomic and functional phenotypic investigations. *L. acidophilus* deficient in lipoteichoic acid (LTA), a major immunostimulant in Gram-positive bacteria, has been shown to shift immune system responses in animal disease models. However, the pleiotropic effects of removing LTA from the cell surface in lactobacilli are unknown. In this study, we assessed two strains of *L. acidophilus* deficient in LTA for phenotypes associated with LTA function and surveyed their global transcriptional and exoproteome profiles. *L. acidophilus* strains lacking LTA exhibited an elongated cellular morphology and their growth was severely inhibited by elevated manganese concentrations. Differentially expressed genes (24) specific to the LTA-deficient strains included a predicted heavy metal resistance operon and several putative peptidoglycan hydrolases. Cell surface-associated protein surveys revealed distinct changes in the composition and relative abundances of several extracellular proteins and showed a bias of intracellular proteins in LTA-deficient strains of *L. acidophilus*. Taken together, these results elucidate the pleiotropic effects of LTA removal on the transcriptome and cell surface exoproteome of *L. acidophilus* and suggests roles of LTA in cell morphology and ion homeostasis as a structural component of the Gram positive cell wall.

4.2 Introduction

*Lactobacillus acidophilus* NCFM is a widely consumed probiotic strain that has been at the forefront of genomic characterization of probiotic functionality (1, 2). Probiotics are
live microorganisms, which when administered in adequate amounts, confer a health benefit to the host (3). Recently, emphasis has been placed on elucidating the molecular mechanisms and regulatory networks responsible for probiotic activity (4). Unraveling probiotic mechanisms has been greatly enabled by genome sequencing (5) development of genetic tools (6-8), transcriptomics (9), and increasingly, proteomics (10, 11). Many of these approaches have been extensively applied to investigate cell-surface constituents, as they are highly strain-specific and are uniquely positioned to effect host-microbe crosstalk (12, 13).

Predictive and functional genomics studies of *L. acidophilus* NCFM have revealed that surface layer (Slps) and surface layer associated proteins (SLAPs)(10, 12), sortase-dependent proteins (14), and adhesion exoproteins mediate adhesion and immunomodulatory activities of *L. acidophilus* (12). *L. acidophilus* encodes three Slps, the dominant SlpA and minor components SlpB and SlpX (8). As the major component of the S-layer in *L. acidophilus*, SlpA contributes to interaction with intestinal epithelial cells and to modulation of immune responses (12, 15, 16).

Development of functional genomic tools in *L. acidophilus* also facilitated creation of lipoteichoic acid (LTA) deficient strains (17). LTA is a large amphiphilic interfacial polymer consisting of a poly-glycerol or ribitol-phosphate backbone attached to a glycolipid moiety with a hydrocarbon tail embedded in the cell membrane (18). LTA serves pleiotropic roles in Gram-positive physiology and is a major immunomodulatory cell surface component that elicits an inflammatory response from antigen-presenting cells (19). Genetic modification or removal of LTA from probiotic microbes abrogates the expression of pro-inflammatory cytokines and elicits increased levels of regulatory cytokines, conferring an
increased capacity to alleviate mucosal inflammation (17, 20). Indeed, disruption of the gene encoding lipoteichoic acid synthase (*ltaS*) causes a shift in the immune system interaction of *L. acidophilus* towards an IL-10 dependent regulatory phenotype, which has been characterized using murine disease models of colitis and colonic polyposis *in vivo* (17, 21).

To assign immunomodulatory properties specific to SlpA, we created an *L. acidophilus* derivative deficient in LTA and minor S-layer components SlpB and SlpX (16). Administration of live cells of this strain mitigated colitis in a murine model and restored epithelial barrier integrity. Of note, purified SlpA from this strain was also sufficient for ameliorating colitis and restoration of epithelial barrier integrity (16).

Due to *ltaS* pleiotropy, generation of knockout mutants may prove lethal depending on the organism (22). Consequently, few studies detail successful creation and characterization of LTA-deficient strains in either pathogenic or beneficial Gram-positive bacteria (23, 24). These foundational studies established localization and interaction of LTA with cell division machinery, providing important insights into the functions of LTA in the cell wall. Despite the creation and immunological characterization of LTA-deficient strains of *L. acidophilus*, very little is known concerning the overlap and potential redundancy of cell-surface molecules contributing toward fundamental bacterial physiology. We investigated *L. acidophilus* mutants deficient in LTA and minor S-layer components SlpB and SlpX to characterize the impact of LTA on cell division and cell morphology phenotypes, as well as the compensatory transcriptional and exoproteomic changes caused by deletion of LTA.
4.3 Materials and Methods

4.3.1 Bacterial Strains

All bacterial strains and plasmids are listed in Table 4.2. Bacterial cultures were cryopreserved in their respective media with a 15% glycerol concentration (vol/vol) and stored at -80°C. *L. acidophilus* was propagated in de Mann, Rogosa and Sharpe (MRS) (Difco Laboratories, Inc., Detroit, MI) broth under static aerobic conditions at 37°C, or on MRS agar (1.5% wt/vol agar, Difco) incubated anaerobically at 37°C for 48 hours. Concentrations of 2 µg/mL of erythromycin (Em) (Sigma-Aldrich, St. Louis, MO) and 2-5 µg/mL of chloramphenicol (Cm) (Sigma) were used for plasmid selection in *L. acidophilus* NCFM, when appropriate. Selection for 5-fluorouracil resistant *L. acidophilus* was performed by supplementing glucose semi-defined (GSDM) (25) agar with a final concentration of 100 µg/mL of 5-fluorouracil (Sigma) (8). *Escherichia coli* EC1000 was propagated aerobically in Luria-Bertani (Difco) broth at 37°C, or on brain-heart infusion (Difco) solid medium supplemented with 1.5 % agar. Antibiotic selection of *E. coli* was maintained with 40 µg/mL kanamycin (Kn) and 150 µg/mL of Em for recombinant *E. coli*, when appropriate.

4.3.2 DNA Isolation, Manipulation, and Transformation

All kits, enzymes, and reagents were used according to the manufacturers' instructions. DNA purification and cloning were performed as previously described (8). Purification of genomic DNA from *L. acidophilus* employed a ZR Fungal/Bacterial MiniPrep
kit (Zymo Research Corp., Orange, CA). Plasmid DNA was isolated from *E. coli* using Qiagen Spin miniprep kit (Qiagen Inc., Valencia, CA). High fidelity PCR amplification of DNA was performed with PFU HS II DNA polymerase (Stratagene Corp., La Jolla, CA). Routine PCRs were conducted with Choice-Taq Blue polymerase (Denville Scientific Inc., Meutchen, NJ). Primers for PCR amplification were purchased from Integrated DNA Technologies (Coralville, IA). DNA amplicons were separated using 0.8 % agarose gel electrophoresis and stained with ethidium bromide for visualization. DNA extraction from agarose gels was performed with a Zymoclean DNA gel recovery kit. Restriction endonucleases were acquired from Roche Molecular Biochemicals (Indianapolis, IN). Ligations were performed with New England Biolabs (Beverly, MA) quick T4 ligase. Sequencing was performed by Davis Sequencing Inc. (Davis, CA). Rubidium chloride competent *E. coli* cells were prepared as previously described and frozen at -80°C (26). Heat shock transformants of the ligation mixture were subsequently screened by PCR for inserts using primers flanking the multiple cloning site. Plasmids putatively containing inserts were sequenced to ensure fidelity. Newly constructed integration plasmids were electroporated into competent cells containing the temperature-sensitive helper plasmid, pTRK669, according to methods described previously (7). Penicillin G at a concentration of 10 µg/mL was employed in the preparation of the competent cells to promote electroporation efficiency (27).
4.3.3. Construction of an *L. acidophilus* NCFM Δ*slpB* Δ*slpX* Δ*ltAS* deletion mutant (NCK2187)

To generate an *L. acidophilus* NCFM isogenic mutant deficient in SlpB, SlpX and LTA, the *slpB* (LBA0175), *slpX* (LBA0512) and *ltAS* (LBA0447) genes were sequentially deleted in an NCFMΔ*upp* background host (NCK1909) using the *upp*-based counterselective gene replacement system (8). For in-frame deletion of *slpX*, the integration plasmid pTRK956 previously constructed was used (8); the integration plasmids for *slpB* and *ltAS* were constructed as follows: a 1,170-bp in-frame deletion within the 1,374 bp *slpB* gene (85% of the gene eliminated) was constructed by first amplifying a 743-bp and a 757-bp DNA segment flanking the region upstream and downstream of the deletion target, respectively, using *slpB1U*-F/*slpB2U*-R and *slpB3D*-F/*slpB4D*-R primer pairs (Table 4.1). For *ltAS* deletion, a 2022-bp in-frame deletion within the 2061 bp *ltAS* gene (98% of the gene eliminated) was generated by first amplifying a 772-bp and a 758-bp DNA segment flanking the region upstream and downstream of the deletion target, respectively, using *ltAS1U*-F/*ltAS2U*-R and *ltAS3D*-F/*ltAS4D*-R primer pairs. Purified PCR products representing the upstream and downstream regions flanking each deletion target of *slpB* or *ltAS* were fused and amplified to generate copies of Δ*slpB* or Δ*ltAS* alleles via splicing by overlap extension PCR (SOE-PCR) (28), using 10 ng of each PCR product as amplification templates in 50-µl PCR reactions with *slpB1U*-F/*slpB4D*-R or *ltAS1U*-F/*ltAS4D*-R primer pair. The purified SOE-PCR products were digested with *BamHI* and *SacI*, and ligated into the pTRK935 counterselectable integration vector digested with compatible ends. The ligation mixture was transformed into *E.coli* EC1000 (Table 4.2) with selection on Brain Heart Infusion agar
containing kanamycin (40 µg/ml), erythromycin (Em; 150 µg/ml), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and isopropyl-β-D-thiogalactopyranoside (IPTG). The resulting recombinant integration plasmids carrying the ΔslpB (pTRK957) or ΔltaS allele (pTRK1052) were subjected to DNA sequencing to ensure sequence integrity.

For initial construction of the ΔslpB mutant, pTRK957 was electroporated into NCK1909 harboring the pTRK669 helper plasmid (NCK1910; Table 4.2). One EmR and CmR transformant carrying both plasmids were grown overnight in MRS broth containing 2 µg/ml each of Em and Cm and transferred three times (1% inoculum) in MRS broth with Em (ca. 30 generations) in a 42°C water bath. Chromosomal integrants were selected by replica plating onto MRS agar supplemented with 2 µg/ml of Em or 5 µg/ml of Cm. One EmR and Cm-sensitive (CmS) pTRK957 integrant was selected and transferred two to three times (1% inoculum) in MRS broth without Em (ca. 20-30 generations). To isolate plasmid-free double recombinants, the culture was diluted and plated on GSDM (25) supplemented with 100 µg/ml of 5-fluorouracil (5-FU) and incubated at 37°C anaerobically for 48 to 72 h until colonies were visible. Double recombinants with ΔslpB allele were screened by colony PCR using the slpB-up/slpB-dw primer pair (Table 4.1) that specifically anneals to the flanking region of the slpB gene. One ΔslpB mutant was selected and designated as NCK1964, and in-frame deletion and sequence integrity were confirmed by DNA sequencing using the slpB-up/slpB-dw primer pair.

Subsequently, to generate a ΔslpB ΔslpX double mutant, pTRK956 (ΔslpX deletion construct) was electroporated into NCK1964 previously transformed with pTRK669 helper
plasmid. Chromosomal integration and isolation of double recombinants were performed as previously described. PCR screening of ΔslpX double recombinant was performed using slpX-up/slpX-dw primer pair (8). One ΔslpB ΔslpX deletion mutant (NCK2030) was confirmed by DNA sequencing for the in-frame deletion using the slpX-up/slpX-dw primer pair. Finally, to construct the ΔslpB ΔslpX ΔltaS mutant, pTRK1052 was transformed into NCK2030 harboring pTRK669. Chromosomal integration and recovery of double recombinants were performed as previously described. Screening of ΔltaS double recombinant was performed using ltaS-up/ltaS-dw primer pair (Table 4.1). One ΔslpB ΔslpX ΔltaS triple deletion mutant was selected and designated as NCK2187. Confirmation of the in-frame deletion was performed by DNA sequencing using the ltaS-up/ltaS-dw primer pair.

4.3.4 Microscopy and growth curves

Cultures were propagated from frozen stocks and subcultured twice at a 1% (vol/vol) inoculum in MRS under standard conditions and harvested after 16 hr incubation (stationary phase). Cell morphology and chain length was visualized using a Nikon Eclipse E600 phase contrast microscope with a Q-Imaging Micropublisher Camera attachment. At least 40 cells from each of three biological replicates were measured using Image Pro Insight software (Media Cybernetics, Inc., Rockville, MD). Chain lengths were averaged and means were compared using a two-tailed t-test with unequal variance, at a significance threshold of $p < 0.05$. 

94
For comparing the growth profiles of each *L. acidophilus* derivative, cultures were propagated from frozen stocks in MRS and subcultured twice at a 1% (vol/vol) inoculum in GSDM under standard conditions. The growth of each strain in both standard and high Mn medium was monitored in triplicate using a Spec-20 spectrophotometer at 600 nm and reported as an average.

**4.3.5 Transcriptional analysis**

RNA was extracted from log phase cultures (optical density of 0.6-0.8) grown in MRS medium under static conditions at 37°C. Cells were harvested by centrifugation 4000 x g for 10 min and then flash frozen using an ethanol-dry ice bath and stored at -80°C. For RNA isolation, each frozen pellet was resuspended in 1 mL of Tri-reagent (Zymo Direct-zol RNA MiniPrep Kit -Zymo Research, Irvine, CA) and cell lysis was performed using a Mini bead beater set to homogenize for 5 cycles of 1 min beating alternated with 1 min incubation on ice. Total RNA was isolated according to protocols from TRI-reagent and the Direct-zol RNA MiniPrep Kit. Each RNA preparation was quantified with NanoDrop and analyzed for quality using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Preparation of each mRNA library and RNA-sequencing were performed at the High-Throughput Sequencing and Genotyping Unit of the Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign. For each sample, ribosomal RNA was removed with the Ribozero Bacteria Kit (Illumina, San Diego, CA) followed by library preparation with the TruSeq Stranded RNA Sample Prep Kit (Illumina, San Diego, CA). Single-read RNA-sequencing was performed using an Illumina HiSeq 2500 Ultra-High-Throughput
Sequencing system (Illumina, San Diego, CA) with a read length of 180 nucleotides. Raw sequencing reads were quality assessed using FastQC Version 0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and processed using Geneious 8.0.5 (29). Briefly, after adaptor sequences were removed, raw reads were quality trimmed to remove bases with an error probability limit of 0.001 (Phred score of 30) and filtered to remove reads shorter than 20 nt. These quality trimmed and filtered sequences were then mapped to the reference genomes using Bowtie 2 (30) with default settings within Geneious 8.1.7 (29). Two different approaches were employed to assess the RNA-seq data; differential expression (DE)-seq, which uses median of gene expression ratios normalization, and 2-way hierarchical clustering using the trimmed mean of M-values normalization method.

Expression was calculated in Geneious and compared using the median of gene expression ratios method, also known as DE-Seq (31). The cutoff thresholds for differential expression of genes were a minimum of a 2-fold change and a p-value $< 1.36 \times 10^{-5}$, adjusted by the Bonferroni correction for multiple testing, treating each ORF as an independent test. LTA positive strains were compared with LTA negative strains in a pairwise fashion. JMP genomics (Statistical Analysis Software, Cary, NC) was used to construct 2-way hierarchical clustering heat maps with centered rows under the fast-ward algorithm using the log$_2$ transformed trimmed mean of M-values normalization method.
4.3.6 Extraction and identification of extracellular, non-covalently bound cell surface proteins

Non-covalently bound cell surface proteins, including S-layer proteins (SLPs) and S-layer associated proteins (SLAPs) were extracted from the *Lactobacillus* strains using LiCl denaturing salt, as described previously (10). SLP and SLAP pellets were resuspended in 10% (w/v) SDS (Fisher). Proteins were quantified via bicinchoninic acid assay kit (Thermo Scientific) and visualized via SDS-PAGE using precast 4–20% Precise Tris-HEPES protein gels (Thermo Scientific). Gels were stained using AcquaStain (Bulldog Bio) according to the instructions from the manufacturer.

SLAPs extracted from the various *L. acidophilus* strains were identified using LC-MS/MS from the Genome Center Proteomics Core at the University of California, Davis, as described previously (10). For all analyses, total spectral counts were utilized as a semi-quantitative indicator of protein abundance (32). Two-way clustering of total spectral counts was performed using JMP Genomics (version 5, SAS). Protein domains were identified for analysis using the Pfam protein family database (33).

4.4 Results

4.4.1 LTA maintains cell shape and provides resistance to Manganese

A defective cell division phenotype is characteristic of *ltaS* mutants in diverse organisms including *Staphylococcus aureus* (22), *Bacillus subtilis* (23), *Listeria monocytogenes* (24) and *Lactobacillus gasseri* (34). To determine the impact of an *ltaS*
deletion on cellular morphology in *L. acidophilus*, 40 cells from stationary phase cultures were measured and their average length compared in triplicate (Figure 4.1). *L. acidophilus* strains lacking *ItaS* (NCK2025 and NCK2187) exhibited an elongated cellular morphology, exhibiting an increased length of 106-125% compared to the LTA\(^+\) strains (\(p < 0.05\)). Qualitatively, many cells of the NCK2025 culture exhibited bending or curving, but this was less apparent in the NCK2187 culture (Figure 4.1). These results add to the growing body of evidence LTA that influences cell morphology and that LTA removal in rod-shaped bacteria promotes cell elongation.

The proposed role of LTA in maintenance of ion homeostasis has previously led to observation of growth conditions toxic to an \(\Delta ItaS\) *B. subtilis* mutant via increased sensitivity to Mn\(^{2+}\) (23). To assess whether LTA plays a similar role in ion homeostasis in *L. acidophilus*, growth profiles in semi-defined medium with standard Mn\(^{2+}\) concentrations (0.3 mM) and with elevated Mn\(^{2+}\) concentrations (3 mM) were compared (Figure 4.2). Under standard Mn\(^{2+}\) concentration, minimal difference in growth between the LTA\(^+\) and LTA\(^-\) strains was observed. By contrast, both LTA\(^-\) strains displayed severe growth inhibition in the high Mn\(^{2+}\) concentration when compared to LTA\(^+\) strains, irrespective of deletion of SlpB or SlpX. Thus, high concentrations of Mn\(^{2+}\) leads to an LTA-dependent bacteriostatic effect. These results confirm a similar role for LTA in establishing proper ion barrier function in *L. acidophilus* as in *B. subtilis*, even though the standard Mn\(^{2+}\) growth requirements are considerably different for the two species (0.3 mM for *L. acidophilus*; > 0.05 mM for *B. subtilis*). While there is generally a paucity of data on LTA knockouts in bacteria, these
results highlight a conserved function of LTA in diverse firmicutes, even when the ion homeostasis pathways of the two species may be incongruent.

4.4.2 LTA-deficiency elicits compensatory transcriptional changes in *L. acidophilus*

In order to determine whether loss of LTA caused compensatory transcriptional changes, and to identify genes relevant to the function of LTA in *L. acidophilus*, we used RNA-seq to characterize the global transcriptional profile of LTA mutants. Differentially expressed genes specific to LTA deletion were identified by pairwise comparison of each LTA\(^+\) strain (NCK1909 and NCK2030) with each LTA\(^-\) strain (NCK2025 and NCK2187). Genes meeting the cutoffs for differential expression in all four comparisons were considered differentially expressed due to absence of LTA. Each pairwise comparison was visualized as an XY plot of log\(_2\) transformed normalized transcripts per million, with differentially expressed genes marked with enlarged circles (Figure 4.3). Using this conservative approach, a total number of 24 genes were found to be differentially expressed across all pairwise comparisons (Table 4.3), 9 of which were encoded in putative operons. Overall, for differentially expressed genes encoded in putative operons, co-encoded genes followed the same trend despite not being significantly differentially expressed (Table 4.4). In contrast, *ltaS* (LBA0447) was significantly downregulated due to the in-frame DNA deletion, but expression for the rest of the LTA biosynthetic operon was unaffected. Two-way hierarchical clustering was also used to correlate expression patterns with the LTA genotypes (Figure 4.4). Selected portions of the heat map display the genes for which expression most closely
correlated with LTA genotype, with differentially expressed genes identified again being denoted with circles.

Interestingly, LBA0541 and LBA0542, encoded in a putative heavy metal resistance operon, were significantly downregulated (Table 4.3). These genes have predicted functions as ATPase efflux proteins with a broad substrate specificity for metal ions. Moreover, LBA1220 is a putative pyridine mercuric reductase whose function also relates to heavy metal resistance via reduction of Hg\(^{2+}\). The differential expression of genes with predicted functions in heavy metal resistance pathways is striking, given the increased sensitivity of LTA-deficient mutants to elevated Mn\(^{2+}\) concentrations. However, the relationship between LTA removal and the differentially expressed genes is currently unclear and only two of the putative transcriptional regulators related to differentially expressed genes were identified; FlpA and lysine (Table 4.4) (35). FlpA is a repressor predicted to regulate the putative LBA0541-LBA0544 operon through oxygen sensing, and as a transcriptional regulator, lysine influences the diaminopimelic acid biosynthesis pathway, which has implications for cell wall turnover as a peptidoglycan precursor.

Three of the upregulated genes corresponded to N-acetylmuramidases (LBA1140 and LBA1918) and an endopeptidase (LBA1883), which are predicted to function in peptidoglycan turnover (Table 4.3). To survey whether LTA removal impacted collective expression of peptidoglycan hydrolases, we compared expression values across each predicted peptidoglycan hydrolase gene in *L. acidophilus* (Figure 4.5). Interestingly, an endopeptidase (LBA1743), a N-acetylglucosaminidase (LBA0176), and an L-alanine amidase
(LBA0177) exhibited trends of increased expression, although they did not meet the threshold for differential expression. These results suggest general upregulation of genes related to cell wall turnover, but the specific profile of differential expression reveals insights into which peptidoglycan hydrolases may be directly influenced by the absence of LTA among potentially redundant enzymes in the *L. acidophilus* genome. Also striking was the lack predicted function for 10 of the 24 differentially expressed genes, which are putative hypothetical proteins (Table 4.3).

### 4.4.3 LTA mutants exhibit altered cell surface protein profiles

To examine the role of LTA in the presentation of non-covalently attached cell surface proteins, screenings of Slps and SLAPs were performed on each *L. acidophilus* strain. Non-covalently bound extracellular proteins, including Slps and SLAPs were isolated from these strains and visualized (Figure 4.6). The SDS-PAGE image reflects the Slp profile expected for each of the strains; specifically that NCK1909 and NCK2025 are SlpA+SlpB+ SlpX+ whereas NCK2030 and NCK2187 are SlpA+SlpB− SlpX−, although SlpA and SlpB are indistinguishable via SDS-PAGE. Proteins in the SLAP fraction were identified using LC-MS/MS and analyzed compared using two-way clustering based on the similarity of the identified proteins (Figure 4.6). It is clear that the SLAP fraction of NCK2025 is distinct from those found in the other strains, including NCK2187. Notably, although both LTA-deficient strains NCK2025 and NCK2187 cluster together, NCK2025 exhibits a distinctly unique profile in both the SDS-PAGE image and the heat map. These data suggest that the absence of LTA alters non-covalently bound extracellular proteins but also that SlpB and SlpX contribute to the distinction between NCK2025 and NCK2187 in their composition.
However, although NCK2030 is also deficient in SlpB and SlpX, it did not exhibit changes in proteome composition as observed in NCK2187 compared to the NCK1909 parent, strongly implying that the combination of LTA and SlpB SlpX deficiencies are responsible for the unique profile of NCK2187.

To assess differences in intracellular and extracellular protein abundance, the distribution of spectral counts corresponding to proteins with and without a putative secretion signal were compared (Figure 4.7). The LTA− strains exhibited significantly higher distributions of intracellular protein counts relative to the LTA+ strains (p < 0.05). Moreover, the distribution of intracellular protein counts were significantly higher in NCK2025 than in NCK2187 (p < 0.05). As shown previously, detection of intracellular proteins in LiCl protein extracts from LTA+ strains of *L. acidophilus* is negligible save for non-classically secreted proteins (10, 11). The higher distribution of spectra assigned to intracellular proteins in LTA-deficient strains implies either an increased permeability of the cells, or an increased rate of autolysis that released these proteins outside the cell. We then compared spectral counts for selected proteins with secretion signals (Figure 4.8). Spectral counts for each SLAP protein were considered a semi-quantitative measure of relative protein abundance and were compared across the LTA+ and LTA− strains. Five SLAP proteins were present at decreased levels in the LTA-deficient strains, including putative fibronectin-binding protein FbpB (LBA0191), a penicillin-binding protein (LBA0858), and two glycerol-3-phosphate ABC transporters (LBA0585 and LBA1641). Three extracellular proteins were observed at increased levels in LTA-deficient strains, including a protein of unknown function (LBA1497), LysA muramidase (LBA1918), and SlpA (LBA0169). SlpA spectra were
overrepresented in the SLAP fractions of both NCK2025 and NCK2187, but especially in the latter. These results show that removal of LTA from the cell surface increases the presence of intracellular proteins in extracellular preparations, and impacts extracellular protein levels in *L. acidophilus* NCFM.

### 4.5 Discussion

Genetic tools for functional genomics combined with high-throughput data technologies have contributed to 10 years of unraveling host-microbe crosstalk in *L. acidophilus* NCFM (2). These approaches have led to discovery and characterization of genes involved in probiotic activity and identified cell-surface constituents that positively impact human health. LTA is a classical molecule in host-microbe interactions, the deletion of which has proven to shift immunomodulatory responses toward ameliorating inflammation in models of immunological diseases (36). However, comprehensive characterization of LTA functions in firmicutes remains a challenge given the pleiotropy of the *ltAS* gene and difficulty in achieving knockouts. Foundational studies have reported phenotypic characteristics of LTA-deficient bacteria in *B. subtilis*, *S. aureus* and *L. monocytogenes*, and have provided insights into the intermolecular interactions of LTA in the cell wall (22-24). In order to elucidate effects of LTA removal has on bacterial physiology in *L. acidophilus*, we investigated NCK2025 and compared it with the NCK2187 strain deficient in LTA and minor S-layer components SlpB and SlpX.
Our results show ltaS deletion promotes cell elongation in *L. acidophilus*, which is likely mediated through aberrant cell division and that LTA-deficient *L. acidophilus* strains are highly sensitive to elevated Mn$^{2+}$ concentrations. We also demonstrate that LTA mutants exhibit differential expression patterns compared to parent strains. Specifically, we observed 24 genes whose expression significantly correlated with LTA genotypes, 10 of which were without any annotated function. Of those differentially expressed genes with predicted function, several were related to cell wall turnover and to heavy metal resistance. We postulate that differences in expression of these genes may be related to the cell morphology and Mn$^{2+}$ toxicity phenotypes observed, although the direct mechanisms and regulatory networks must be further investigated. We hypothesized that removal of LTA would affect composition of the cell surface proteome due to its capacity to act as a scaffold for extracellular proteins, and were able to identify proteins whose levels were altered in the SLAP fraction of the LTA$^{-}$ strains. Importantly, the LTA$^{-}$ strains exhibited major differences in the levels of intracellular proteins identified from the SLAP fraction, indicating increased permeability or autolysis in the LTA-deficient strains causing liberation of intracellular components.

The effect of LTA removal on cell division has been investigated in a few species, with varying effects depending on growth phase and cell morphology. This can partially be attributed to the role of LTA in regulating autolysins through maintenance of ion homeostasis, but it is also apparent that LTA facilitates assembly or localization of the FtsZ ring during cell division (23). An early report demonstrated that LTA binding of N-acetylmuramoyl-L-alanine amidase inhibited its activity, while the addition of cations...
restored activity (37). Additional studies reported that deletion of \textit{ItaS} correlates with lower levels of peptidoglycan hydrolases in \textit{S. aureus} (38) and substantiated direct association between LTA and autolysins in the cell wall (39). Collectively, these results indicate that LTA has the capacity to regulate autolysins directly through association or exclusion and indirectly through ion homeostasis. It has long been proposed that LTA functions in cation homeostasis due to its negatively charged poly-glycerol/ribitol phosphate backbone (40). In particular, LTA may function in proton and divalent cation sequestration. Experimentally this was substantiated by growing \textit{B. subtilis} and an LTA-deficient derivative under various concentrations of \textit{Mn}^{2+}, from which a sophisticated proposal of LTA-mediated ionic regulation emerged (23). We also observed that LTA-deficient \textit{L. acidophilus} was sensitive to an elevated \textit{Mn}^{2+} concentration. This result provides evidence that even in disparate Gram-positive bacteria, LTA has a conserved role in maintaining the extracellular ionic milieu.

To investigate the compensatory transcriptional responses in LTA-deficient \textit{L. acidophilus} mutants, we used RNA-seq to analyze the transcriptional profiles of four strains. While the focus of our analysis was to identify differentially expressed genes correlated with LTA genotype, the vast majority were not significantly altered. We were able to identify 24 genes expression of which was significantly different in the LTA-deficient strains. Although some of the genes identified were related to predicted functions of LTA, for many their interrelation with LTA was unclear. Some putative peptidoglycan turnover genes in \textit{L. acidophilus} were affected by LTA removal, especially LysA (LBA1918), a muramidase (LBA1140), and an endopeptidase (LBA1883). We postulate that changes in expression of these genes may be compensatory for the loss of autolysin activity at the cell surface. LTA
seems to play a role in facilitating the localization of the FtsZ ring and autolysins during cell division (23, 41). There is also evidence suggesting that LTA excludes autolysins from delocalized peptidoglycan substrates except for the cell septa (39). It may be possible that the delocalized activity of autolysins caused upregulation of peptidoglycan turnover genes to compensate for lack of specific endolysin activity at the cell septa. Some of the genes downregulated in LTA-deficient strains were associated with heavy metal resistance pathways. LBA0541 encodes cadA, a non-specific ATPase efflux protein, LBA0542 also encodes an ATPase efflux protein, and LBA1220 encodes a pyridine mercuric reductase. It is unclear whether differential expression of these genes is related to or partially responsible for the loss of Mn\(^{2+}\) resistance in LTA\(^{-}\) strains, but further investigation may reveal or define their intersection with LTA activity.

Large scale proteomic methods are being increasingly applied to probiotic and fermentative lactic acid bacteria, with particular focus on Slp and SLAP profiling (10, 11). This is a promising approach to defining the molecular basis for probiotic activity since extracellular proteins appear to be strain-specific and are uniquely positioned to interact with the environment, whether it be a food or mucosal niche. Striking differences in the SLAP profiles of the LTA-deficient strains were apparent. Notably, the distribution of spectral counts for intracellular proteins without a predicted secretion signal was higher in both of the LTA-deficient strains, but was remarkable in NCK2025. It has been suggested that LTA acts as a barrier that reduces cell wall permeability, but it may also be possible that the upregulated peptidoglycan hydrolases can cause increased autolysis. Of note, increased SlpA-specific spectral counts were observed in NCK2025 and with an even greater
magnitude in NCK2187. It is tempting to hypothesize that the upregulation of SlpA may have restored some barrier integrity in NCK2187, resulting in less prominence of intracellular proteins. However, since LTA acts as a scaffold for SlpA attachment to the cell wall in *L. acidophilus*, its absence may result in higher recovery of SlpA in the SLAP fraction due to increased dissociation from the cell wall. Until the interaction of LTA with the S-layer is further elucidated, the basis for increased SlpA in the SLAP fractions of LTA mutants remains unclear. Deletion of *ltaS* also impacted two glycerol-3-phosphate ABC transporters, which likely supply the phosphoglycerol transferase with glycerol3-phosphate necessary for polymerization of the polyglycerolphosphate LTA backbone.

Our results indicate functions of LTA in cell division and cell morphology, and have underscored phenotypic results with compensatory transcriptional and extracellular protein changes. This approach has resulted in identification of genes that may be directly related to LTA activity on the cell surface. By determining the consequences of deleting a specific gene/subset of genes on global gene expression and cell surface proteome composition it is possible to uncover regulatory networks and cellular processes relevant to the genotype. These results underscore how deleting *ltaS*, a conserved pleiotropic gene, impacts the transcriptome and exoproteome of *L. acidophilus*, facilitating identification of genes influenced by LTA activity or relevant for its functions.
4.6 References


### Table 4.1. Primers used in this study.

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction of ΔslpB and ΔltaS</strong></td>
<td></td>
</tr>
<tr>
<td>slpB1U-F</td>
<td>GAAATAGGATCCGCTATCATCAGCCTTCAT</td>
</tr>
<tr>
<td>slpB2U-R</td>
<td>AGATACAGCAGAAGCAGCAA</td>
</tr>
<tr>
<td>slpB3D-F</td>
<td>TTGCTGCTTCTGCTGTATCTTTGAGAAGGTTGTT</td>
</tr>
<tr>
<td>slpB4D-R</td>
<td>TAAAGTAGGCCTGATAGGAAAGGCTCAAT</td>
</tr>
<tr>
<td>ltaS1U-F</td>
<td>CAGCAGGATCCAGGGTGATCGATCGACTTCAT</td>
</tr>
<tr>
<td>ltaS2U-R</td>
<td>GGCGATTCCATCGCTCTCT</td>
</tr>
<tr>
<td>ltaS3D-F</td>
<td>AGAGACCGATGGAACGTACACTGATCGCCTGAGTTG</td>
</tr>
<tr>
<td>ltaS4D-R</td>
<td>CTGCTGGAGCTGCCCAGTTGAA</td>
</tr>
<tr>
<td><strong>PCR analysis and DNA sequencing of deletion targets</strong></td>
<td></td>
</tr>
<tr>
<td>slpB-up</td>
<td>TTCGTTGATCGCATAAG</td>
</tr>
<tr>
<td>slpB-dw</td>
<td>GTGTAGTATTGCGGATAACAG</td>
</tr>
<tr>
<td>ltaS-up</td>
<td>GTTCATGGCTTAGTTAC</td>
</tr>
<tr>
<td>ltaS-dw</td>
<td>CATCATCGCTTCTCAT</td>
</tr>
</tbody>
</table>

* restriction enzyme sites, underlined
**Table 4.2. Bacterial strains and Plasmids used in this study.**

<table>
<thead>
<tr>
<th>Strains</th>
<th>Genotype or characteristics</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L. acidophilus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCK1909</td>
<td>NCFM carrying a 315-bp in-frame deletion within the upp gene</td>
<td>Goh et al., 2009</td>
</tr>
<tr>
<td>NCK1910</td>
<td>NCK1909 harboring pTRK669; host for pORI-based counterselective integration vector</td>
<td>Goh et al., 2009</td>
</tr>
<tr>
<td>NCK1964</td>
<td>NCK1909 carrying a 1,170-bp in-frame deletion in the slpB gene</td>
<td>Goh, Klaenhammer, Unpublished</td>
</tr>
<tr>
<td>NCK2025</td>
<td>NCFM carrying a 2,022-bp in-frame deletion in the ltaS gene</td>
<td>Mohamadzadeh et al., 2011</td>
</tr>
<tr>
<td>NCK2030</td>
<td>NCK1964 carrying a 1,356-bp in-frame deletion in the slpX gene</td>
<td>Goh, Klaenhammer, Unpublished</td>
</tr>
<tr>
<td>NCK2187</td>
<td>NCK2030 carrying a 2,022-bp in-frame deletion in the ltaS gene</td>
<td>Lightfoot et al., 2015</td>
</tr>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC1000</td>
<td>RepA+ JM101; Km+; repA from pWV01 integrated in chromosome; host for pORI-based plasmids</td>
<td>Law et al., 1995</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTRK935</td>
<td>3.0 kb; pORI28 with a upp expression cassette and the lacZ′ from pUC19 cloned into BglII/XbaI sites; serves as counterselective integration vector</td>
<td>Goh et al., 2009</td>
</tr>
<tr>
<td>pTRK956</td>
<td>4.5 kb; pTRK935 with a mutated copy of slpX cloned into BamHI/SacI sites</td>
<td>Goh et al., 2009</td>
</tr>
<tr>
<td>pTRK957</td>
<td>4.5 kb; pTRK935 with a mutated copy of slpB cloned into BamHI/SacI sites</td>
<td>Goh, Klaenhammer, Unpublished</td>
</tr>
<tr>
<td>pTRK1052</td>
<td>4.5 kb; pTRK935 with a mutated copy of ltaS cloned into BamHI/SacI sites</td>
<td>Lightfoot et al., 2015</td>
</tr>
</tbody>
</table>
### Table 4.3. List of differentially expressed genes in LTA deficient strains.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Predicted function</th>
<th>Range of log2 change (LTA−/LTA+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downregulated genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LBA0485</td>
<td>Hypothetical protein</td>
<td>-1.5 to -2.0</td>
</tr>
<tr>
<td>LBA0486</td>
<td>Hypothetical protein</td>
<td>-1.9 to -2.2</td>
</tr>
<tr>
<td>LBA0447</td>
<td>Lipoteichoic acid synthase</td>
<td>-2.1 to -2.8</td>
</tr>
<tr>
<td>LBA0541</td>
<td>cadA</td>
<td>-1.3 to -1.8</td>
</tr>
<tr>
<td>LBA0542</td>
<td>Heavy metal-transporting ATPase</td>
<td>-1.7 to -2.6</td>
</tr>
<tr>
<td>LBA0543</td>
<td>Hypothetical protein</td>
<td>-1.7 to -2.5</td>
</tr>
<tr>
<td>LBA0544</td>
<td>Transcriptional regulator</td>
<td>-1.7 to -2.2</td>
</tr>
<tr>
<td>LBA0853</td>
<td>N-acetyldiaminopimelate deacetylase</td>
<td>-1.1 to -1.7</td>
</tr>
<tr>
<td>LBA1220</td>
<td>Pyridine mercuric reductase</td>
<td>-2.2 to -4.9</td>
</tr>
<tr>
<td>LBA1801</td>
<td>Hypothetical protein</td>
<td>-1.1 to -2.2</td>
</tr>
<tr>
<td>Upregulated genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LBA0872</td>
<td>Hypothetical protein</td>
<td>1.1 to 1.7</td>
</tr>
<tr>
<td>LBA0873</td>
<td>Hypothetical protein</td>
<td>1.3 to 1.8</td>
</tr>
<tr>
<td>LBA1045</td>
<td>glutamine ABC transporter ATP-binding protein</td>
<td>1.1 to 2.0</td>
</tr>
<tr>
<td>LBA1140</td>
<td>lysin</td>
<td>1.2 to 1.5</td>
</tr>
<tr>
<td>LBA1184</td>
<td>Hypothetical protein</td>
<td>1.1 to 1.7</td>
</tr>
<tr>
<td>LBA1497</td>
<td>Hypothetical protein</td>
<td>1.1 to 1.3</td>
</tr>
<tr>
<td>LBA1665</td>
<td>oppA</td>
<td>2.1 to 3.4</td>
</tr>
<tr>
<td>LBA1679</td>
<td>ABC transporter permease</td>
<td>1.7 to 2.4</td>
</tr>
<tr>
<td>LBA1680</td>
<td>ABC transporter ATP-binding protein</td>
<td>1.7 to 2.8</td>
</tr>
<tr>
<td>LBA1690</td>
<td>Hypothetical protein</td>
<td>1.0 to 1.2</td>
</tr>
<tr>
<td>LBA1870</td>
<td>Maltose phosphorylase</td>
<td>1.4 to 3.9</td>
</tr>
<tr>
<td>LBA1883</td>
<td>NLP-P60 secreted protein</td>
<td>1.1 to 1.5</td>
</tr>
<tr>
<td>LBA1918</td>
<td>lysA</td>
<td>1.7 to 2.4</td>
</tr>
<tr>
<td>LBA1928</td>
<td>Hypothetical protein</td>
<td>1.0 to 1.8</td>
</tr>
</tbody>
</table>
Table 4.4. Operons containing genes found to be differentially expressed in LTA-deficient strains.

<table>
<thead>
<tr>
<th>Operon</th>
<th>Putative function</th>
<th>Regulator-effector</th>
<th>Gene</th>
<th>Range of log₂ change</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBA0444-LBA0447</td>
<td>Lipoteichoic acid biosynthesis</td>
<td></td>
<td>LBA0444</td>
<td>0.3 to 0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA0445</td>
<td>0.2 to 0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA0446</td>
<td>0.5 to 0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA0447*</td>
<td>-2.1 to -2.8</td>
</tr>
<tr>
<td>LBA0541-LBA0544</td>
<td>Heavy metal resistance</td>
<td>FlpA-Oxygen</td>
<td>LBA0541*</td>
<td>-1.3 to -1.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA0542*</td>
<td>-1.7 to -2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA0543*</td>
<td>-1.7 to -2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA0544*</td>
<td>-1.7 to -2.2</td>
</tr>
<tr>
<td>LBA0852-LBA0857</td>
<td>Lysine biosynthesis</td>
<td>RNA-Lysine</td>
<td>LBA0852</td>
<td>-0.8 to -1.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA0853*</td>
<td>-1.1 to -1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA0854</td>
<td>-0.9 to -1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA0855</td>
<td>-0.3 to -1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA0856</td>
<td>-1.4 to 0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA0857</td>
<td>-1.1 to 0.2</td>
</tr>
<tr>
<td>LBA1042-LBA1046</td>
<td>Glutamine transport</td>
<td></td>
<td>LBA1042</td>
<td>0.2 to 1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA1044</td>
<td>0.8 to 1.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA1045*</td>
<td>1.1 to 2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA1046</td>
<td>0.8 to 1.6</td>
</tr>
<tr>
<td>LBA1184-LBA1189</td>
<td>Antimicrobial export</td>
<td>YhcF</td>
<td>LBA1184*</td>
<td>1.1 to 1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA1186</td>
<td>0.8 to 1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA1187</td>
<td>0.7 to 1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA1188</td>
<td>0.9 to 1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA1189</td>
<td>0.5 to 1.4</td>
</tr>
<tr>
<td>LBA1679-LBA1680</td>
<td>Antimicrobial export</td>
<td>YhcF</td>
<td>LBA1679*</td>
<td>1.7 to 2.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LBA1680*</td>
<td>1.7 to 2.8</td>
</tr>
</tbody>
</table>
Figure 4.1. Microscopy images of *Lactobacillus acidophilus*.
Figure 4.2. Growth profiles of *Lactobacillus acidophilus* NCFM derivatives.
Figure 4.3. Comparison of TPM values.
Figure 4.4. RNA-seq transcriptional profiles of *Lactobacillus acidophilus* derivatives.
Figure 4.5. RNA-seq transcriptional profiles of predicted peptidoglycan turnover proteins in *Lactobacillus acidophilus*. 
Figure 4.6. Extracellular protein profiles of *Lactobacillus acidophilus* LTA derivatives.
Figure 4.7. Distribution of normalized spectral counts for proteins with and without secretion signals.
Figure 4.8. Comparison of normalized spectral counts of select extracellular proteins across *Lactobacillus acidophilus* derivatives.
Chapter V

Conclusions and perspectives
5.1 Conclusions and perspectives

Lactic acid bacteria are a highly diverse clade of bacteria that vary widely in genomic content and phenotypic functionality, necessitating the concert of high-throughput data technologies with reductionist biology to determine the mechanistic underpinning of commercially and medically relevant traits. In this study we employed functional genomic approaches to characterize traits in two highly relevant commercial strains; the starter culture *S. thermophilus* and the probiotic microbe *L. acidophilus*. In *S. thermophilus* we have established a framework for determining the fundamental outcomes of targeting bacterial genomes with CRISPR-Cas systems in the contexts of genetic tool development, sequence-specific removal of population subsets, and genome homestasis. The ability to leverage CRISPR targeting-induced lethality of bacteria to engineer microbial populations offers a next-generation approach to developing sequence-specific programmable antimicrobials. Our results offer insights into the complexities arising from self-targeting by CRISPR-Cas systems and have established models for describing typical genetic outcomes. Moreover, this approach revealed the utility of genome-targeting CRISPR-Cas systems for synthetic biology applications and underscored the potential for repurposing endogenously active systems for genome editing in bacteria. Future research will need to address gaps in understanding of the interplay between genome homestasis, DNA repair pathways and CRISPR-Cas systems.

In *L. acidophilus* we have defined cell morphology and Manganese resistance phenotypes associated with lipoteichoic acid function. Comparative transcriptomic and proteomics analysis of LTA\(^{+}\) and LTA\(^{-}\) strains facilitated identification of gene candidates associated with the contribution of LTA to cell homeostasis and to cell-surface physiology.
High-throughput proteomic investigations reveal cell-surface composition and potential interactions of proteins involved in cell wall homeostasis. These results open new avenues for determining LTA function in the cell surface and demonstrate how comparative transcriptomics of highly related strains can be used to uncover regulatory networks and key genetic determinants involved in the pleiotropy of conserved genes.
Appendix I

CRISPR-based technologies and the Future of Food Science
JFS Special Issue: 75 Years of Advancing Food Science, and Preparing for the Next 75
CRISPR-Based Technologies and the Future of Food Science
Kurt Seiler and Redolphe Barrangou

Abstract: The on-going CRISPR craze is focused on the use of Cas9-based technologies for genome editing applications in eukaryotes, with high potential for translational medicine and next-generation gene therapy. Nevertheless, CRISPR-Cas systems actually provide adaptive immunity in bacteria, and have much promise for various applications in food bacteria that include high-resolution typing of pathogens, vaccination of starter cultures against pests, and the genetic programming of specific antibiotics that can selectively modulate bacterial population composition. Indeed, the molecular machinery from these DNA-encoded, RNA-mediated, DNA-targeting systems can be harnessed in native hosts, or repurposed in engineered systems for a plethora of applications that can be implemented in all organisms relevant to the food chain, including agricultural crops trait-enhancement, livestock breeding, and fermentation-based manufacturing, and for the genesis of next-generation food products with enhanced quality and health-promoting functionalities. CRISPR-based applications are now poised to revolutionize many fields within food science, from farm to fork. In this review, we describe CRISPR-Cas systems and highlight their potential for the development of enhanced foods.

Keywords: CRISPR, genome, microbiology

Introduction
Food science may be generally defined as the application of many scientific fields and disciplines to foods and processing. Although many distinct sciences have historically been applied to the genesis, formulation, processing, storage, enhancement, and enjoyment of food products over time, food science is arguably in a renaissance stage, which is fueled by the availability of key technologies that allow food scientists and engineers to develop health-promoting products. Notwithstanding the many advances we have witnessed over the last century, the field of genetics has contributed critical advances in the recent past. In particular, the use of Cas9 and other CRISPR-DNA editing technologies has profoundly impacted food science, agriculture, ecology, animal husbandry, and medicines. Tangible and impactful improvements in the human condition such as the industrial biosynthesis of vitamins, enzymes, pharmaceuticals, antibiotics, and lysosomal enzymes, have been enabled by advances in genetic methodologies. Forthcoming DNA technologies have accelerated the rate of molecular biology research in diverse backgrounds, where few technologies previously existed or were suboptimal. Specifically, clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated sequences (Cas) are an adaptive immune system against invasive genetic elements in bacteria (Barrangou and Marraffini 2015), which has been co-opted for genome editing in a diverse set of organisms ranging from scientific models to industrial workhorses (Doudna and Charpentier 2014). The ability to cleave and edit DNA with CRISPR-Cas systems has

Revived reductionist biology during a scientific era defined by sequencing, and the state of functional genomics has accelerated to new limits as a result. However, early investigation of CRISPR-Cas systems developed applications outside of genome editing, including manipulation of microbial consortia (Goma and others 2014), designed vaccination of microorganisms against invasive genetic elements (Barrangou and others 2013), and typing of bacterial strains (Barrangou and Horvath 2012). Many of the early applications of CRISPR-Cas systems actually arose from food science-driven research during characterization of industrial starter culture bacteria for improving bulk fermentation processes (Burrangou and others 2007). Food science is a growing field investigating all biological, chemical, and physical processes to improve production of safe and sustainable food for a growing world population. Overall, given the pace at which CRISPR-Cas technology is being developed, one can already envision how applications of CRISPR-Cas systems may further be harnessed to address challenges related to the food and agriculture industries at every level of food manufacturing, from farm to fork.

CRISPR basics and background
Two genetic elements constitute the adaptive immune system in bacteria: CRISPR arrays which confer immunological memory and surveillance, and Cas genes, which encode effector proteins in all stages of immunity (Figure 1). CRISPR-Cas mediated immunity is categorized into three temporally overlapping but mechanistically distinct molecular processes: acquisition, expansion, and interference (Burrangou 2013; Barrangou and Marraffini 2015). Acquisition occurs via sampling of foreign genetic elements by the universal Cas1-Cas2 surveillance complexes, from which short sequences, termed spacers, are integrated in a polarized fashion into the CRISPR array (Barrangou and others 2007). New spacer sequences are added at the leader end of the array, resulting in an
CRISPR applications in Food Science...

Original record of foreign DNA exposure that span from most recent at the 5' end to most ancient at the 3' end. Expression of CRISPR arrays is constitutive under standard conditions, but is also inducible during phage infection (Young and others 2012). The array is transcribed as a long pre-CRISPR RNA (pre-crRNA), and is typically further processed via Cas proteins and host ribonucleases into mature, small interfering crRNAs (Brown and others 2008). Mature crRNAs guide Cas proteins to target DNA via sequence-specific complementarity for recognition and cleavage of target nucleic acids, causing direct interference of homologous DNA elements (Marraffini and Sontheimer 2008; Garneau and others 2010). Various aspects of CRISPR biology, genetics and applications have been extensively covered in comprehensive and focused reviews (Makarova and others 2011; Barrangou and Horvath 2012; Douzou and Charpentier 2014; Barrangou and Marraffini 2015; Barrangou and May 2014; Selle and Barrangou 2015).

CRISPR arrays consist of highly conserved, partially palindromic DNA repeats that alternate with variable short spacer sequences. Repeat sequences, the compelling and defining feature of CRISPR–Cas systems, function in the formation of hairpins necessary for structure-dependent RNA processing during heterogeneity of crRNAs. CRISPR repeat sequences also exhibit partial complementarity to non-activating CRISPR-RNAs (transcrRNA) and leader sequences, suggesting roles in acquisition and in Cas protein recognition (Deltcheva and others 2011). Specific to certain systems, tracrRNAs are noncoding RNAs that direct processing of crRNAs and target DNA recognition and cleavage by Cas proteins (Deltcheva and others 2011; Kareva and others 2013). By contrast, the origin and function of most spacer sequences is mostly unknown, as only a small percentage exhibit similar identity to foreign invasive elements or chromosomal sequences (Horvath and others 2009). This may be due to extinction of genetic elements or the lack of environmental surveillance (sequencing) of the elements to which spacers correspond. Most of the known spacer targets correspond to phages and bacteriophages, but many also appear to target either self or foreign chromosomal sequences (Horvath and others 2009). Although the intuitive function of an adaptive immune system is targeting invasive genetic elements, many spacer sequences exhibit self-complementarity, and many do not appear to be transcribed, suggesting that CRISPR–Cas systems may play additional roles in microbial physiology beyond targeting of foreign genetic elements (Barrangou 2015). An essential feature of type I and type II systems (Makarova and others 2011) is the protospacer adjacent motif (PAM), a short conserved sequence proximate to the spacer sequence in the target DNA (Deveau and others 2008; Horvath and others 2008; Mojica and others 2009).

The PAM governs both the acquisition and interference processes, as it determines viable protospacers in the target sequence, and functions in differentiation of the target from the CRISPR array. PAM sequences vary between CRISPR types and orthogonal systems within CRISPR types, but must be characterized in order to fully exploit functional systems.

Cas proteins are fundamental to each stage of CRISPR-based immunity, as they are responsible for acquisition of new spacers, processing of crRNAs, and recognition and degradation of sequences complementary to crRNAs. However, convergent evolution has resulted in a myriad of DNA recognition and cleavage mechanisms in keeping with microbial diversity, necessitating categorization of CRISPR–Cas based on gene content, operon organization, and distinct clusters of sequence homology (Makarova and others 2011). The most recent definitive description of CRISPR–Cas systems highlights 2 classes based on the composition of the immune effector component (single vs. multi-subunit). The systems are further delineated into several main types, specifically depending on the presence of signature Cas genes (Makarova and others 2011). The hallmark features of type I systems are the large multifunctional endonucleolytic Cas9, the tracrRNA, and ribonuclease III processing of crRNAs (Deltcheva and others 2011). Type I and type III systems encode Cascade, Cas, and Cas proteins that constitute the multi-subunit effector complexes responsible for target nucleic acid recognition (Brown and others 2008). The signature gene of type I systems is cas1, a single stranded nickase with 3'-5' exonuclease activity, which is recruited to the target via the CRISPR-associated complex for antiviral defense (Brown and others 2008; Sinkunas and others 2011). In contrast to other CRISPR systems, type I and type III systems target either or both DNA and RNA, and the signature gene is cas10 (Makarova and others 2011). Estimates indicate 46% of bacterial and 84% of archaeal genomes contain at least one CRISPR–Cas system (Makarova and others 2011). Despite the high distribution of CRISPR–Cas systems in bacteria, the relative youth of the field means that very few of the systems have been characterized for activity in each of the 3 stages of immunity: acquisition, expression and
CRISPR applications in Food Science...

interference. To date, only 14 systems have demonstrated activity in interference, highlighting the need for investigating orthologous systems for holistic understanding of CRISPR–Cas mechanisms and applications (Boudy-Denomy and Davidson 2014).

The study of CRISPR–Cas systems has relatively humble origins in food science-related research, with its biological function being discovered during investigation of phage resistance mechanisms of dairy starter cultures (Barrangou and others 2007). In particular, *Streptococcus thermophilus* notably undergoes attack by predatory bacteriophage during yogurt and cheese fermentation, and is highly enriched for active CRISPR-Cas systems (Barrangou and others 2013). *S. thermophilus* encodes up to 4 CRISPR-Cas systems, 2 of which are inactively active in both acquisition and interference, likely due to the high selective pressure of bacteriophage in dairy processing environments. Accordingly, functional genomic analysis of a food-grade starter culture bacterium and its bacteriophages established the role of CRISPR–Cas systems in plasmid/DNA protection (Barrangou and others 2007). Several seminal contributions to the field were conducted in *S. thermophilus*, because it harbored highly active CRISPR–Cas systems, the genome sequence was available, and there were characterized lytic phages, all of which was related to its widespread industrial use as a starter culture in food fermentations. Basic food science research therefore led to determination of spacer origin (Barrangou and others 2007), inference of PAM sequences (Devos and others 2008; Horvath and others 2008), unraveling of plase-host dynamics (Sun and others 2013; Paz-Espino and others 2015), demonstration of Cas9 endonuclease activity (Garneau and others 2016; Sapranauskas and others 2011; Garneau and others 2012), characterization of CRISPR-Cas structural motif orthology (Selle and others 2011; Briner and others 2014), and recently, designed removal of large genomic islands (Selle and others 2015).

Applications across the food bacterial spectrum

*Listachillus* spp. are scientifically, industrially, and medically relevant microorganisms that are propagated at high levels for fermentation processes or to elicit health benefits as probiotic microorganisms. Thus, as constituents of the human microbiome, or as fermentative bacteria, exposure to bacteriophage is highly likely, suggesting that bacterial plasmid resistance mechanisms would be abundant in these environments. Indeed, *in silico* surveys have revealed that type II systems are disproportionately present in lactobacilli (Horvath and others 2009; Briner and others 2014), making them a reservoir for CRISPR–Cas systems. In this light, the human and food microbiomes are a relatively unexplored trove of new and diverse CRISPR–Cas systems that are potentially suitable for use in food-grade systems. Given that bacteria are ubiquitous throughout the production and consumption of food, CRISPR–Cas technologies have the potential to impact all classes of bacteria across the food spectrum, including pathogenic, commensal, fermentative, probiotic, and spoilage organisms. CRISPR-based technologies with applications in food science include genotyping of bacteria, manipulation of microbial consortia, vaccination against pluses, and genome editing (Figure 2).

CRISPR-based genotyping

Identification and typing of bacterial strains is a considerable challenge due to the inherent diversity of microorganisms and their tendency to undergo horizontal gene transfer. Although genome sequencing can be considered the “gold standard” for identification and typing of bacterial strains in terms of resolution, it is a costly, analytically-challenging and time-intensive process not suitable for high-throughput or rapid applications. The same limitations apply to pulsed field gel electrophoresis, although fingerprinting of restriction digestion profiles is still performed for strain typing during foodborne outbreaks. Recently, repetitive-element PCR-based genotyping using high-resolution microfluidics has proven to be rapid and reliable in strain differentiation, but identification of strains requires a database of fingerprint data for comparison. IS1 sDNA sequencing, although not a typing tool, is relatively fast, and affordable for rough identification of bacterial genus and species, but can be unreliable even for applications requiring resolution down to the species level. By contrast, CRISPR array genotyping offers a rapid, affordable, and high-resolution means of typing bacterial strains within species that carry them (Barrangou and Horvath 2012). Due to the unique polarized nature of spacer acquisition in CRISPR–Cas systems, the highly ordinal composition of CRISPR-Cas arrays provides a means of typing with high resolution at the strain level (Shariat and others 2013), requiring a few PCR amplifications and sequencing of the array to provide a clear comparison of spacer content. Despite this, proof of concept for CRISPR-based typing has only been provided for a limited set of bacteria. In order to effectively be used as a tool for identification and typing, the same CRISPR-array locus must be enriched or ubiquitous within defined taxonomic groups (Genus or species) and the spacer content of the array must be diverse across all strains in a given subset. The presence or absence of a CRISPR array may also be used to differentiate strains, but is only reliable when it correlates with the phylogeny of the organism. In order for adequate spacer comparison to occur, within a given CRISPR locus, the most ancient spacers must share a common origin, which in turn depends on the course of the array. In other words, there must be some shared and some disparate spacers in order to effectively type strains based on array content. Thus, the process is largely contingent on having had active spacer acquisition machinery at some point in evolutionary history, although degeneracy of CRISPR targets can also add to polymorphisms in spacer content. Of course, CRISPR-based typing also depends on the presence of CRISPR–Cas systems in the genomes of genera and species of interest, and these loci have been identified in most archaea and many bacteria, but only documented to occur in approximately 40% of bacterial genomes. To date, CRISPR-based typing schemes have been effectively employed in foodborne pathogens such as *Salmonella* (Shariat and others 2013) and *Escherichia coli* (Toro and others 2013). Yim and others (2013) describe the potential for CRISPR-based typing of *Listachillus* strains (Hodgson and others 2013), and spoligotyping organisms such as *Listachillus italicus* (Briner and Barrangou 2014), illustrating the broad potential of CRISPR-based genotyping across the bacterial spectrum.

Vaccination of industrial microbes

Mobile genetic elements (MGEs) are a class of DNA entities encompassing plasmide, bacteriophages, transposable elements, and integrative and conjugative elements. They are exogenous elements that can transfer and hijack bacterial DNA homeostatic pathways, causing continuous challenges to both population and genetic stability of bacteria. To cope with the persistent threat of predatory bacteriophages and other MGEs, bacteria have evolved both innate and adaptive immune systems targeting exogenous genetic elements. Innate immune mechanisms include cell-wall modification, restriction/modification systems, and adaptive plas...
infection (Labrie and others 2010). In the food industry, predatory bacteriophages constitute a significant threat to efficiency of preservation and continue to be a major source of inconsistent quality or loss in dairy fermentations. Many strategies have thus been developed to combat the ever-present and dynamic phage populations present in processing plant environments. Starter cultures are especially susceptible to lytic phage infection due to a mono-culture population, wherein a single infective phage type can cause the crash of an entire population. Furthermore, the high rate of mutation in phages necessitates the use of multiple resistance mechanisms and control strategies to compensate for their high capacity for adaptation to the host. Specifically, both native biological mechanisms of resistance and environmental control are employed to prevent phage proliferation. Starter culture rotation, growth in the presence of chelators, multistrain starter formulations, and steam sterilization of manufacturing equipment are all means of controlling phage in the dairy processing environment, whereas genetic transfer of plasmids containing native phage resistance mechanisms and/or CRISPR can be used to combat phages in the bacterial population (Hurwitz and Barrangou 2012). CRISPR provides unique advantages in vaccination of starters against predatory bacteriophages (Barrangou and others 2013). Specifically, the process of adding spacers corresponding to phages is iterative, which means that additional spacers can always be acquired to target emerging phages. Moreover, resistance is sequence-specific, which means that the resistance mechanisms can be as broad, or specific as desired, especially if conserved functional sequences are targeted in phage genomes. One spacer may therefore be able to confer resistance to multiple phages if the respective phage genomes contain the same sequence targeted by the spacer. Finally, unlike innate immune mechanisms of phage resistance, the target sequence of the phage must mutate in order to circumvent CRISPR-Cas as a mechanism, which can lead to detrimental mutations in phage machinery (Sun and others 2013; Paez-Espino and others 2015).

Antimicrobials

Self-targeting events of CRISPR-Cas systems are highly lethal, which has been determined experimentally and observed in vivo (Vercoe and others 2013; Beeld and others 2014). The lethality of self-targeting events relates to the nature of DNA destruction induced by CRISPR-Cas interference mechanisms. Type I systems elicit extensive DNA damage through the exonuclease activity of Cas3, introducing deletions that span approximately 40 kb in some experiments (Vercoe and others 2013). DNA damage of this nature is repairable at low frequency ($10^{-5}$), likely by an alternative end-joining mechanism (Vercoe and others 2013). Type II systems elicit double-stranded DNA breaks via Cas9 activity, for which few DNA repair mechanisms exist in bacteria (Garneau and others 2010; Guzman and others 2012; Selle and Barrangou 2015). Bacteria typically use the high-fidelity pathway of homologous recombination to repair double-stranded DNA breaks, but restoration of the target locus to the wild-type does not circumvent targeting by CRISPR-Cas systems (Selle and

![Figure 2: Applications of CRISPR-Cas systems across the food chain. CRISPR-Cas as antimicrobials and typing tools may be applied at every stage of food manufacture, whereas genome editing may be applied to food crops, animal breeds, and industrial microbes. Vaccination of bacteria against mobile genetic elements can be used to protect the population and genetic stability of starter cultures.](image-url)
CRISPR applications in Food Science…

Barrangou 2015). Pathways of dbDNA break repair in bacteria also include the low-fidelity alternative end-joining, and non-homologous end-joining, although both occur at low frequency, and the latter is generally present in spore-forming bacteria and Mycobacterium spp (Arava and Kooiman 2001). Due to a paucity of high-frequency DNA damage repair pathways that can cope with CRISPR-Cas targeting, population-wide depletion of cells exhibiting the target genotype occurs, generally on the order of 3 to 5 logs under experimental conditions. Self-targeting with type I systems is highly efficacious, a lethality is not dependent on chromosomal location, expression level, or strain bias (Gomaa and others 2014). Furthermore, it was demonstrated that proper design of self-targeting spacers could lead to differentiation of highly related strains (Gomaa and others 2014). Several experiments repurposing endogenous or delivered CRISPR-Cas systems have been performed in pathogenic organisms related to the food industry (Betel and others 2014; Citri and others 2014; Birkard and others 2014; Gomaa and others 2014). Control of microbial consortia in processing facilities and in food products is fundamental to protecting the food supply from contamination or proliferation of foodborne pathogens. Self-targeting CRISPR-Cas systems therefore present a novel and high-potential means to deplete microbial populations in a sequence-specific manner, sparing the innocuous native microbiota present in foods. As CRISPR-based microbiome engineering technologies further evolve, we anticipate several application avenues will be generated across the food supply chain, to optimally manage the composition of various microbial populations associated with soil, plants, livestock, manufacturing environments, and the consumer.

Genome editing and remodeling in bacteria

Bacterial genomes exhibit site-specific plasticity that belies the linearity of their interpretation as sequential representations (Barrangou and Leach 2014). The ability to reprogram CRISPR-Cas systems to target any sequence in the genome offers promising applications towards defining minimal bacterial genomes, determining essential genes, and characterizing genetically homogeneous bacterial populations (Jiang and others 2013; Selle and Barrangou 2015). Recently, CRISPR-Cas targeting was used to show that MGEs contribute to genomic plasticity in Staphylococcus thermophilus (Selle and others 2015). Specifically, recombination between insertion sequences of high identity caused spontaneous deletion of large genomic islands, spanning from 8 to 102 kbp in length. Targeting the genomic islands with an endogenous CRISPR-Cas system enabled selection and recovery of naturally occurring mutants lacking genes necessary for acidification and preservation of milk. The approach also confirmed the nonessentiality of genes encoded on the genomic islands, ultimately resulting in excision of 7% of the genome of S. thermophilus. This approach could similarly be applied for removal of genetically encoded MGEs, increasing genome stability. Moreover, CRISPR-based removal of pathogenicity islands and/or virulence factors is an attractive method for neutralizing pathogenic bacteria. Thus, CRISPR-Cas systems facilitate characterization of MGEs and elucidation of bacterial genome plasticity. Similarly, this technology can also be harnessed in combination with single-strand DNA recombination to drive genome editing in probiotic strains such as Lactobacillus reuteri (Oh and van Pijkeren 2014; van Pijkeren and Butturro 2014).

The CRISPR revolution as it applies to the food chain

The use of CRISPR-based technologies has revolutionized the field of genetics in general, and genome editing of eukaryotes in particular. To date, this approach has been successfully employed for targeted mutagenesis of a plethora of genomes including Homo sapiens, Mus musculus, Danio rerio, Drosophila melanogaster, Caenorhabditis elegans, Oryza sativa, and Saccharomyces cerevisiae (Doudna and Charpentier 2014). The streamlined and multifunctional nature of Cas9 from Type II systems is practical for programmable genome editing through precise and directed targeting of chromosomal loci. The tipping point for genome editing was arguably the provision of a synthetic guide molecule that combinerizes the functions of a native crRNA and tracrRNA (Jinek and others 2012) and the development of a corresponding 2-component sgRNA:Cas9 genome editing system. The programmable specificity of exacted chromosomal cleavage is facilitated by selection and design of a spacer sequence unique to the target allele. Specificity is compounded by the PAM, a short conserved sequence that must be proximate to the protospacer in the target sequence (Doyeon and others 2008; Moir and others 2009). Cas9-induced mutagenesis in eukaryotes occurs subsequent to cleavage and is typically mediated through the imperfect DNA repair mechanism of nonhomologous end joining (NHEJ). Following the genesis of double-stranded breaks, NHEJ yields efficient recovery of insertion and deletion knockout donor genotypes.

CRISPR-based genome editing has already been applied to organisms of interest across food science, including yeast, corn, rice, and tomatoes. Genome editing of crops has applications for targeted engineering to improve growth under drought conditions, application of insecticides, low nutritional/fertility conditions, and also to improve the nutrition potential of food crops. Similarly, genome editing can improve yield in animal breeding through desirable alteration and selection of herd genetics. Moreover, there is the potential to increase the disease-resistance of both crops and cattle, but despite the promising outcomes of genome editing using CRISPR-Cas systems, the practical implications of doing so are yet to be universally defined.

One key consideration moving forward is the regulatory status of various CRISPR-derived products. In some cases, the native activity of CRISPR-Cas systems can be harnessed for screening of natural events, such as vaccination against plagues, immunization against plagues, or lethal-self targeting bacterial activity. These would contritute non-GMO processes that would be readily acceptable to regulatory agencies and, by extension, to the public. In contrast, there are many means to exploit engineered CRISPR-Cas systems in heterologous backgrounds that hinge on recombinant DNA technologies that are typically construed as genetic-engineering methods. On-going efforts are focused on exploiting regulatory frameworks already in place for the genesis of genetically-modified variants using other gene editing technologies such as programmable nucleases. Recent advances based on the use of RNA and ribonucleoprotein complexes, as opposed to DNA, have open intriguing avenues based on the process, rather than the outcome.

Conclusions and Perspective

While much of the on-going CRISPR-CRISPR crake (Penning 2013; beadford 2015) has been focused on genome editing applications in human cells, and the potential of Cas9-based gene therapies for clinical applications (Barrangou and May 2015), CRISPR-Cas systems and CRISPR-based technologies hold much promise for a broad range of applications across food science. For applications in food bacteria, native CRISPR-Cas systems present opportunities for genetic manipulation of pathogens, for vaccination of cultures against plagues, and as next-generation antimicrobials. Furthermore,
CRISPR applications in Food Science

Engineered systems can be harnessed for genome-editing applications in crops and livestock for trait enhancement or in next-generation breeding approaches. Building on recent advances in their exploitation in agriculture, livestock, and industrial fermentation, we envision that CRISPR-Cas technologies will drive research and development in many food products, and open new avenues for the future of food science.

Conflicts of Interest

R.B. and K.M.S. are inventors on several patents related to various uses of CRISPR-Cas systems. R.B. is a board member of Carbis Biosciences, and a founder and advisor of Intella Therapeutics, 2 companies that are involved in commercialization and exploitation of CRISPRR applications.

References


Appendix II

Harnessing CRISPR-Cas systems for bacterial genome editing
Harnessing CRISPR–Cas systems for bacterial genome editing

Kurt Selle\textsuperscript{1,2} and Rodolphe Barrangou\textsuperscript{1}

\textsuperscript{1} Department of Food, Bioprocessing, and Nutrition Sciences, North Carolina State University, Raleigh, NC 27606, USA
\textsuperscript{2} Genomic Sciences Graduate Program, North Carolina State University, Raleigh, NC 27695, USA

Manipulation of genomic sequences facilitates the identification and characterization of key genetic determinants in the investigation of biological processes. Genome editing via clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated (Cas) constitutes a next-generation method for programmable and high-throughput functional genomics. CRISPR–Cas systems are readily reprogrammed to induce sequence-specific DNA breaks at target loci, resulting in fixed mutations via host-dependent DNA repair mechanisms. Although bacterial genome editing is a relatively unexplored and underrepresented application of CRISPR–Cas systems, recent studies provide valuable insights for the widespread future implementation of this technology. This review summarizes recent progress in bacterial genome editing and identifies fundamental genetic and phenotypic outcomes of CRISPR targeting in bacteria, in the context of tool development, genome homeostasis, and DNA repair.

CRISPR–Cas systems and bacterial genome editing

Bacteria harbor CRISPR (see Glossary) and cas genes, which constitute an RNA-guided adaptive immune system against invasive genetic elements\textsuperscript{[1]}. CRISPR–Cas-mediated immunity hinges on the distinct molecular processes of acquisition, expression, and interference\textsuperscript{[2]}. Acquisition occurs via molecular ‘sampling’ of foreign DNA, from which short sequences, termed spacers, are integrated in a polarized manner at the leader end of a CRISPR array\textsuperscript{[1]}. CRISPR arrays are transcribed constitutively and inducibly as directed by promoter elements in the preceding leader sequence during expression\textsuperscript{[3–5]}. The transcript is processed selectively at each repeat sequence, forming mature CRISPR RNAs (crRNAs) that serve as small interfering RNAs (siRNAs). crRNA guide Cas proteins for the sequence-specific recognition and cleavage of target DNA complementary to the spacer to effect interference. CRISPR–Cas systems encode universal cas\textsubscript{I} and cas\textsubscript{II} genes and are categorized as Type I, Type II, or Type III based on signature genes contributing to the distinct mechanisms by which each system confers interference\textsuperscript{[6]}. Type I systems achieve immunity via the CRISPR-associated complex for antiviral defense (Cascade) through single-strand DNA nickase and exonuclease activity and are defined by the presence of Cas3\textsuperscript{[7]}. Features unique to Type II systems include the signature double-stranded (ds) DNA endonuclease Cas9, ancillary trans-activating crRNA (tracrRNA), and the biogenesis of crRNAs by RNase III\textsuperscript{[8,9]}. Type III systems are marked by the signature gene Cas10 but are mechanistically diverse and less well defined with some systems even capable of targeting RNA instead of DNA\textsuperscript{[10]}. Delineation of CRISPR–Cas systems into 11 subsystems is similarly based on the presence of specific accessory cas genes and their respective genetic organization\textsuperscript{[6]}. Cas9 effects interference in Type II systems through sequence-directed endonucleolytic at the target locus, achieved by concerted RuvC and RNR nickase activity\textsuperscript{[11,12]}. The streamlined and multifunctional nature of Cas9 is practical for programmable genome editing in diverse organisms, requiring expression of only its cognate tracrRNA and a crRNA corresponding to the target sequence (Figure 1). The tipping point for this methodology was the creation of a single guide RNA (sgRNA) chimeras that combines the functions of the native crRNA and tracrRNA duplexes\textsuperscript{[13]}. Cas9-mediated genome editing is programmable through the design of sgRNAs. The specificity of chromosomal cleavage hinges on the selection of a spacer sequence unique to the target allele and is further

Glossary

Alternative end joining (A-EJ): a 14-16 bp dsDNA repair mechanism resulting from the annealing and ligation of nonhomologous sequences.
Clustered regularly interspaced short palindromic repeats (CRISPR): a DNA locus comprising nearly identical repeats flanked on either side by repetitive sequences of exogenous origin.
CRISPR-associated (Cas) protein: the enzymes that perform the various stages of CRISPR-Cas activity.
Double-stranded DNA break (DSB): DNA damage in which the phosphodiester backbone of both strands is hydrolyzed or cleaved.
Nonhomologous end joining (NHEJ): a ligase-dependent, error-prone DSB repair mechanism in bacteria and eukaryotes.
Potenator: a target sequence in either the host genome or a foreign genetic element, such as a plasmid or phage, that is identified by a CRISPR spacer sequence.
Potenator adjacent motif (PAM): a short (3–4 bp) conserved sequence that must be present in the target protospacer and is essential for acquisition and interference.
Seed: 6–12 bp of spacer proximate to the PAM, directly involved in crRNA-target DNA hybridization.
Single-guide RNA (sgRNA): a single synthetic chimera combining the functions of native crRNA and tracrRNA.
Spacer: short sequences intervening between repeat sequences that correspond to target genomes.
Trans-activating CRISPR RNA (tracrRNA): Type II-specific ancillary RNA that hybridizes to crRNAs and drives Cas9 activity.
compounded by the protospacer adjacent motif (PAM), a short conserved sequence that must be proximate to the target protospacer (Figure 1) [14–16]. Cas9 introduces a lethal double-stranded DNA break (DSB) at the target locus, effectively acting as a selection against wild-type sequences during genome editing [11,12]. Pre-existing mutations in the population can be selected for or against (Figure 2A) but mutations may be introduced subsequent to targeting by host-repair mechanisms (Figure 2B). Mutations elicited by CRISPR-Cas systems are therefore DNA damage and repair machinery dependent. Heterologous expression of Cas9:gRNA combinations from Streptococcus pyogenes has facilitated high-throughput functional genomics in a multitude of eukaryotic organisms and cell lines [17–19]. CRISPR-Cas-derived genome editing tools have revolutionized genetic and biological research in model eukaryotic organisms on account of their efficiency, affordability, and accessibility.

Despite early proof of concept, only three studies have implemented CRISPR-Cas-mediated genome editing in bacteria [20–22], making this a relatively unexplored and underrepresented application of CRISPR-Cas systems. Nevertheless, genome editing via CRISPR-Cas constitutes a next-generation method for programmable and high-throughput functional genomics in prokaryotic backgrounds. Collectively, these studies substantiate the use of CRISPR-Cas systems as genetic tools in bacteria and contribute to our understanding of the fundamental genetic and phenotypic outcomes of targeting bacterial genomes. This review summarizes insights from these foundational experiments, highlights considerations for tool development, identifies potential biological hurdles, and predicts future applications of the technology.

**Lethality of targeting genomes**

The lethality of Cas-mediated DNA cleavage was first observed in its natural ecological role of targeting bacteriophages and plasmids [1,11,12,23], but self-targeting events are an evolutionary cost of housing active CRISPR-Cas systems. The observation of self-complementary spacers, at one time constituting up to 22% of known spacer targets in lactic and bacteria [24], emphasizes the potentially constraining pressure of self-targeting events. Identification of self-complementary spacer targets reveals mutations at those chromosomal loci, suggesting that self-targeting events drive mutation or fixation of pre-existing mutations (Figure 2A). Investigation of spacer acquisition in Streptococcus thermophilus during exposure to phage led to the infrequent observation of chromosomal acquisition events, which correlated with the disappearance of clones containing self-targeting spacers [25]. Moreover, several publications in diverse backgrounds have reported the lethality of DNA damage induced by self-targeting CRISPR-Cas systems [20,22,26–31]. Transformation of plasmids eliciting self-targeting by Cas proteins is cytotoxic as measured by the relative reduction in viable transformants recovered compared with transformation of non-self-targeting plasmids [28]. CRISPR-mediated deletion of microbial populations results in 3–5 log reductions in populations exhibiting the target sequence, sometimes approaching the transformation efficiency of the respective bacterial background [20,22,26–31]. DNA cleavage by Cas proteins constitutes a significant threat to the survival and fitness of microorganisms, as demonstrated by growth inhibition and aberrant cellular morphology phenotypes consistent with DNA damage observed in Pseudomonas aeruginosa following self-targeting events [20]. Active CRISPR-Cas systems cannot exist in the same cell as the target DNA, which compounds the pressure for mutations to occur, as restoration of the target locus to the wild type does not circumvent CRISPR targeting [32]. Consequently, high-fidelity repair mechanisms are not sufficient for the survival of self-targeting events. Thus, targeting by CRISPR-Cas systems is a selection against the cell populations exhibiting the target genotype. Selection for pre-existing mutations in genetically heterogeneous cell populations supports CRISPR-Cas-directed genome evolution at the population level (Figure 2) [20]. This phenomenon was demonstrated experimentally by the transformation of strain-specific self-targeting plasmids into heterogeneous populations comprising highly similar Escherichia coli strains [29], in which dose-dependent depletion of specific population subsets was achieved. This study also reported that lethality was independent of chromosome location, the transcriptional activity of the target, strand bias, and coding versus noncoding regions. Collectively, the well-established lethality of self-targeting events substantiates the utility of CRISPR-Cas for mutagenesis in bacterial genomes by selecting for novel type IIa variants. Moreover, they highlight the potential application of Cas-cleavage-driven, sequence-specific evolution of bacterial genomes in mixed populations.

**CRISPR-Cas targeting escape strategies**

Bacterial cells containing target DNA sequences are efficiently cleared from the population, partially due to the
A major mechanism by which CRISPR-Cas targeting may be circumvented (Figure 3B) [18,20,21,33,34]. The seed sequence comprises the 6-12 bp most proximate to the PAM and is fundamentally involved in the hybridization of crRNA to the cognate target DNA, such that mutations in the seed also abolish targeting (Figure 3B) [21,35,36]. Point mutations within the protospacer sequence are relatively well tolerated and generally do not prevent targeting [20,21]. This is especially true for the PAM-6 to 11 base pair sequence, which is consistent with the removal of 10 or 11 bp from the 5′ end of the spacer during the maturation of crRNAs [37]. Instead, deletion of the protospacer constitutes a means of evading Cas-mediated cleavage (Figure 3D). Predictably, mutations that interfere with the biogenesis of crRNAs or the activity of tracrRNA may also abolish activity, but inactivation of CRISPR-Cas systems constitutes a significant cost to the cell. Bacterial means of escape mirror those of the predominant target of CRISPR-Cas systems, phage populations. In phages, mechanisms facilitating circumvention of CRISPR-Cas targeting typically involve alteration of the PAM, seed, or protospacer sequences, achieved through recombination or spontaneous mutation [14,25,34]. Phage populations are inherently genetically diverse yet exhibit a high frequency of homologous recombination (HR) events in response to CRISPR-Cas targeting [25]. Recombination in bacterial populations is limited by natural barriers preventing access to homologous yet variable DNA segments. However, Jiang and coworkers [22] demonstrated that introduction of exogenous genomic DNA (gDNA) caused recombination-mediated survival in Streptococcus pneumoniae, which suggests potential roles for homologous gene transfer in survival for self-targeting events in bacteria.

Box 1. DNA repair mechanisms in bacteria

**HR**

DNA repair pathways may result in extensive mutations to restore both single-stranded DNA and dsDNA damage caused by Cas-driven cleavage, HR is the most universal and well-characterized DNA repair mechanism in bacteria (Figure 4A) [59]. RecBCD and RecF proteins conduct two HR-mediated DSB repair pathways in Escherichia coli. Induction of a DSB elicits helicase activity from RecD and RecF, of which the latter also acts as an exonuclease. RecA is recruited and causes conformational changes to their DSBs, resulting in nucleoprotein filament formation (Figure 4B) [59]. RecA then unwinds the blunt-ended DNA and RecG nuclease processes a 3′ overhang, allowing RecFOR to recruit RecA to the strand. The remaining stages of strand invasion and resolution are similar to those of the RecBCD pathway.

**A-EJ**

A-EJ is a DSB repair mechanism dependent on RecBCD exonuclease end resection and microhomology-mediated recombination (Figure 4) [60]. RecBCD processivity is thought to facilitate homology searching and leads to the characteristic outcome of variable deletions at the DSB site. In A-EJ, blunt-ended DNA is

resected until the 1-9 nt sites of homologous sequences overlap, facilitating LigA-mediated resolution of chromosomal sites. A-EJ also offers potential for the integration of exogenous and/or unrelated sequence fragments based on microhomologous (1-8 nt) sequences. Due to the partial redundancy in the repair of DSBs and the utilization of the mechanism, the characterization of this pathway is not trivial and thus far its distribution is unknown beyond its function in E. coli. Thus, it remains difficult to predict the potential for A-EJ repair to occur without prior characterization of this repair pathway in the background microorganisms of interest.

**NHEJ**

NHEJ occurs through the blunt-ended DNA binding protein Ku and a novel ligase, LgD (Figure 4D) [60]. End resection of blunt DNA results in variable deletions and insertions during repair. A survey of bacterial genome sequences performed by Alland and Kosak [61] revealed a few bacteria harboring Ku and LgD homologs, including Bacillus subtilis. Ku and LgD in bacteria exhibit some fundamental differences from their eukaryotic counterparts in terms of gene architecture and mechanism of action [62]. Spontaneously, Ku homologs exist as genetically fused heterodimers in bacteria in an operon with LgD, although certain systems in Mycobacterium spp. encode redundant ligases [58]. To date, NHEJ has been best characterized in B. subtilis [61], Pseudomonas aeruginosa [62], and Mycobacterium spp. [63], but no study has yet investigated the utility of CRISPR-Cas self-targeting in tandem with NHEJ for bacterial genome editing.
Lessons from bacterial genome editing studies

The lethal effects of CRISPR-Cas self-targeting in bacteria are well reported, but few studies have investigated the molecular outcomes of self-targeting events. Recent work by separate groups has provided invaluable insights into the intersection of genome homeostasis and CRISPR-Cas self-targeting [20–22]. These foundational studies pave the way for widespread implementation of CRISPR-Cas technology as a genome editing tool in bacteria. The experiments revealed the mechanistic underpinnings of CRISPR-Cas targeting, exploited DNA repair/replication pathways for designed genome edits, and delineated the genomic plasticity of bacterial populations using CRISPR-Cas targeting.

Homology directed repair in S. pneumoniae

Jiang et al. [22] was a landmark study of genome editing using CRISPR-Cas9 in general and was also the first to demonstrate bacterial genome editing. The study determined that double-crossover HR with a donor template restored CRISPR-Cas9-effected chromosomal injury in S. pneumoniae. A prophage served as the target for Cas9 cleavage using two derivative strains of S. pneumoniae differing from the wild type in an integrated prophage at the srtA locus and a prophage integrated strain with a mutated PAM site. Transformation of S. pneumoniae cr6 gDNA encoding the prophage targeting Cas9 sgRNA was expected to be lethal to the prophage-harboring strain but not for the strain with the mutated PAM sequence. However, HR at the srtA locus was observed, ultimately resulting in the efficient recovery of recombinant clones with deleted prophage genes. A similar result was achieved on transformation of a wild type srtA linear editing template with the prophage targeting Cas9 sgRNA (Figure 4A). The study performed thorough assessments of protospacer and PAM mutations that circumvent Cas9 targeting, providing characterization of the 5'-NGG-3' PAM requirements and seed sequence of the Type II system from S. pneumoniae. Moreover, the study highlighted the ability to introduce targeted missense mutations as well as whole-gene deletions using double-crossover HR in the β-galactosidase-encoding gene in S. pneumoniae. To assess the efficiency of bacterial genome editing with or without the assistance of Cas9 cleavage, the authors quantified the mutation rate of an artificial stop codon in an erthromycin resistance gene with the outcome restoring the Kan<sup>R</sup> phenotype. The experiment revealed marginal induction of recombination through Cas9 targeting of the stop codon, but even in the absence of Cas9 cleavage a subpopulation of cells appeared to undergo transformation or recombination at higher frequencies. This study not only established the utility of Cas9-mediated genome editing, but elucidated the molecular underpinnings of the efficiency and limitations of the system.

Efficient and targeted mutagenesis with recombining in Lactobacillus reuteri

Oh and van Pijkeren employed CRISPR-Cas9 self-targeting in tandem with recombining for the selection of desired mutations to achieve targeted mutagenesis at nucleotide resolution in L. reuteri (Figure 4C) [21]. Plasmid-based expression of recT, a single-stranded DNA-binding protein, and cas9 was used for single-step and dual-step strategies by introducing single-stranded oligonucleotides conferring circumvention of Cas9 targeting. The oligonucleotides were designed to harbor a non-targeted PAM sequence to effectively circumvent Cas9 cleavage and avoid competitive Cas9 binding of the oligonucleotide and possible displacement of the RecT protein [38]. This study demonstrated that low-frequency mutations such as whole-gene deletions could be introduced in the cell populations using single-stranded recombining and that these mutations could be selected for by applying CRISPR-Cas targeting against the wild type genotype. The study also demonstrates the robustness of the system for the mutational biochemical characterization of proteins through introducing missense mutations using Cas9-assisted codon saturation. Similarly to Jiang et al. [22], the authors address concerns of making mutations...
at additional non-PAM sites while still using the highly effective PAM-mutation-based circumvention. As reported by Oh and van Pijkeren, recombineering represents a precise and efficacious genetic tool in *L. reuteri*, especially when applying CRISPR-Cas9-assisted selection for desired mutations. However, successful application of this technology may require considerable optimization for use in disparate backgrounds and the universal efficacy of the method remains undetermined given the significant variation in the capacity of bacteria to perform recombineering.

Alternative end joining (A-EJ) and large deletions in *Pectobacterium atrosepticum*.

In contrast to the other bacterial genome editing studies, Veeres et al. [20] used a natively active Type I-E system, which is less suited for introducing desired mutations due to the unpredictable and extensive nature of the DNA damage caused by Cas11 exonuclease activity. It was demonstrated that lethality due to a self-targeting spacer in *P. atrosepticum* was abrogated by a single PAM sequence mutation, but self-targeting was restored through plasmid-based expression of a programmable repeat-spacer array. In the absence of a donor template for HR, large deletions were achieved through recombination of mobile genetic elements (Figure 4A). Deletion of the entire pathogenicity island (~98 kbp) occurred reproducibly through the recombination of attL and attR sites flanking the island. Interestingly, the authors reported that the deletion occurred spontaneously at low frequency in wild type cell populations as detected by PCR amplification of the attB excision footprint. By contrast, variable deletions within the pathogenicity island consistent with A-EJ also occurred through microhomologous sequences. The two mechanisms of mutation observed in this study highlight the ordinal effect of mutation and DNA damage, since it can be postulated that deletion of the island was observed through selection of pre-existing mutations (Figure 2A), whereas deletions due to A-EJ were likely to have occurred subsequent to cleavage of the chromosome (Figure 2B).

**Design considerations.**

The exceptional range of microbial diversity poses shared challenges for many genetic manipulation strategies. The mechanisms governing DNA homoeostasis are highly background-specific; thus, genetic tool development is limited by factors including transformation efficiency, plasmid replication, capacity for recombination/integration, DNA methylation, and DNA repair pathways. Similarly, the application of CRISPR-Cas9 technology for bacterial genome editing hinges on these molecular processes, but there are also specific design considerations for accurate and efficient use of the technology.

**Design of spacer sequences.**

Off-target cleavage is expected to occur infrequently in bacterial systems relative to eukaryotes, which can be attributed to the lower occurrence of sequences homologous to a given spacer–PAM combination in smaller genomes. Given that single self-targeting events result in a significant reduction in the recovery of viable transformants, any off-target cleavage leading to multiple events of CRISPR-Cas9-induced chromosomal injury in a single cell should compound this reduction in recovery, leading to a decreased incidence of mutation at extraneous loci. Proper selection of spacer sequences is essential to further prevent unintended cleavage events and to maximize efficiency. To date, two strict criteria for the selection and design of spacers are the location of consensus PAM sequences and avoiding incidental sequence identity to

---

**Figure 4.** The "compass rose" of bacterial genome editing via clustered regularly interspaced short palindromic repeats (CRISPR-Cas9) systems and mechanisms of DNA repair in bacteria. In the context of CRISPR-Cas9-enhanced genome editing, recombination (A) can occur with a non-Wilson type recombination, or between homologous sequences flanking the target sequence. (B) Recombining strategy for genome editing is achieved through homologous expression of Cas9 and recF followed by transformation of a single-stranded oligomeric DNA with identity to the target sequence. The single-stranded oligomeric DNA anneals to the lagging strand as a primer during discontinuous DNA replication, thus incorporating the desired mutation. The alternative pathway is mediated by the RecBCD complex and UvrD (C), whereas the classical non-homologous end joining pathway (D) is conducted by Ku and LIG. Both result in variable deletions and insertions.
extraneous genomic loci. Putative protospacers are constrained by defining the location of putative PAM sequences in the target locus. Since PAM and seed sequences are integral for recognition and activity at the target, spacers must be selected based on uniqueness of these components to prevent off-target cleavage. In Type II-A systems, approximately 10 nt is removed from the 5’ end of the spacer during crRNA maturation, suggesting that they are irrelevant for target specificity [37]. Protospacers containing sequences identical to the PAM should not be considered, to prevent competitive Cas recruitment that may limit the cleavage of the desired locus [38]. There is no web tool dedicated to spacer design for bacterial genome editing and the utility of current eukaryotic tools for designing bacterial spacers is undetermined.

Increasing transformation efficiency for optimized system delivery
Transformation efficiencies are limiting in many backgrounds, but transformation (natural or induced), transduction, and conjugation are all potential avenues for experiments requiring simultaneous transformation of editing templates and expression vectors. The lethality of self-targeting compounds is the need for high transformation efficiency of CRISPR-Cas components when interference is the direct result of transformation. Limited transformation efficiency can be compensated for by designing high-frequency-mutation strategies facilitating circumvention of CRISPR-Cas targeting, increasing the recovery of desired genotypes. Development of tightly regulated inducible expression systems bypasses the low transformation efficiency that targets linear DNA, allowing the induction of self-targeting in highly concentrated cultures. Inducible systems could therefore increase the recovery of desired mutations and the identification of low-frequency mutations. An intriguing observation noted by Jiang et al. [22] suggested that certain bacterial subpopulations were more prone to recombination/transformation. Experiments with multiple rounds of recombination/transformation may therefore constitute short-term directed evolution. This process results in disproportionate selection of the population with a higher competency phenotype, with potential applications in molecular biology for backgrounds with low transformation or recombination efficiencies.

DNA repair mechanisms in bacteria
Since the mutations elicited by CRISPR-Cas systems are both DNA damage and repair machinery dependent, it is prudent to consider the DNA repair pathways present in each microorganism. The universality of HR and its well-characterized mechanism and relatively high frequency have led to the widespread development and use of HR-mediated genetic technologies in diverse bacterial backgrounds. Therefore, HR can similarly be considered a practical and applicable repair pathway for the introduction of mutations in CRISPR-Cas-assisted genome editing. HR-based repair of DSBs and the simultaneous generation of desired mutations can be achieved through the provision of a homologous editing template, which affords the potential for both deletions and gene replacements. Recombination can occur between native homologous sequences flanking DSBs in the genome, resulting in the deletion of large genomic segments [39]. To this end, native repair pathways can be exploited for the generation of desired mutations or enzymatic machinery can be heterologously expressed to introduce repair pathway platforms for mutagenesis. In particular, recent studies have shown that CRISPR-Cas systems also afford the potential for the characterization of native DNA repair pathways in diverse microbial backgrounds.

Future applications
CRISPR-Cas selection against target sequences has already been tangibly exploited in a few bacteria to elicit genome edits, but further development of the technology has considerable potential for revolutionizing bacterial genetics and genomics.

Cas9 and editing template delivery through transformation of linear nucleic acids
Linear DNA transformation has not been broadly applied as a tool for genetic manipulation in bacteria, since replication of DNA requires circularity and host coenzyme activity causes rapid degradation of linear DNA. A few studies have employed linear DNA to elicit gene replacement by double-crossover recombination, with success in E. coli, Bordetella pertussis, S. thermophilus, and S. pneumoniae [22,39–41]. Synthetic DNA molecules are easily designed and affordable for engineered mutagenesis and splicing by overlap extension PCR also allows the generation of editing templates [42]. Thus, linear DNA transformation techniques may increasingly be considered a viable option for genome editing when complemented with CRISPR-Cas selection against the wild type. Transformation of linear RNA is another unexplored avenue for the mutagenesis of bacteria, but with the strong selective pressure of Cas9 targeting, RNA-based expression of Cas9 and sgRNAs holds potential for increased throughput in genome editing of bacteria.

Exploitation of endogenous and orthogonal systems
CRISPR-Cas systems are found in approximately 46% of bacteria and 84% of Archaea, with CRISPR-Cas recognition sites scattered across the genome. However, CRISPR-Cas systems are also encoded in the genomes of many archaea and bacteria, with CRISPR-Cas recognition sites scattered across the genome. This suggests that CRISPR-Cas systems are not only useful for genetic modification in bacteria, but also for a variety of other applications, including the development of novel diagnostic tools and the study of bacterial evolution.
increasing the range of target sequences without sacrificing specificity [47]. Longer PAMs may offer increased specificity and potentially decreased off-target cleavage. Extended PAMs may increase the efficacy of cleavage by further compounding the low frequency at which point mutations effectively confer circumvention of Cas9 recognition. It is noteworthy, however, that recent evidence indicates that biochemical recognition of PAMs is not as stringent as bioinformatically determined consensus sequences would suggest [48]. Therefore, it may be necessary to empirically determine the specific nucleotides contributing to stringent FAM recognition for each CRISPR-Cas system.

Understanding bacterial genome biology

The high prevalence of mobile genetic elements in bacterial genomes [49] presents a unique challenge for editing targeted mutations at these loci. Excision of mobile genetic elements may occur in the face of CRISPR-Cas selective pressure [50], dependent on the frequency of excision relative to that of the desired mutation. Therefore, in silico prediction of mobile genomic segments may be used to identify potentially excisable regions, which can then be experimentally validated with CRISPR-Cas selection. Despite the potential difficulty of generating designed mutations in these segments, there are considerable applications for the excision of mobile genetic elements, such as defining minimal bacterial genomes and the characterization of putative unannotated proteins and essential genes. CRISPR-Cas selection can therefore be used to screen for clonal subtypes within heterogeneous populations, degrading the locus-dependent plasticity of bacterial genomes.

Concluding remarks

Development of CRISPR-Cas technology in bacteria has yielded applications in typing and strain detection [50,53], the exploitation of natural/engineered immunity against mobile genetic elements [52–54], the manipulation of microbial consortia/generation of smart antibiotics [20], and programmable transcriptional regulation [45]. However, few studies have focused on the development of CRISPR-Cas genome editing tools in bacteria. This streamlined methodology holds potential for increasing expediency and efficiency in the generation of desired mutations, potentially without the necessity of plasmid integration, extensive screening, or counter-selection. Microbial diversity necessitates the development of efficient transformation protocols and genetic tools for bacterial genome editing (Box 2), but CRISPR-Cas technology opens new avenues in genetic engineering applications.

References

1 Barrangou, R. et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712
5 Goh, Y.J. et al. (2011) Specialized adaptation of a lactic acid bacterium to the milk environment: the comparative genomic of Streptococcus thermophilus LMD-1. Microbiol Cell Fact 10, 122
7 Sinkunas, T. et al. (2011) Cas9 is a single-stranded RNA guide and ATP-dependent effector in the CRISPR/Cas immune system. EMBO J 30, 3325–3334
8 Derconsider, E. et al. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607
16 Meiran, F. et al. (2009) Short direct repeats determine the structure of the prokaryotic CRISPR defense system. Microbiology 155, 733–749
20 Veres, R.R. et al. (2013) Cytosine deamination targeting by CRISPR-Cas systems can reshape bacterial genomes and expel or remodel chromosomally integrated plasmids. PLoS Genet 9, e1003351
28 Bikard, D. et al. (2012) CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. Cell Host Microbe 12, 177–186
40 Zelaya, G.R. et al. (1992) Gene replacement in Bordetella pertussis by transformation with linear DNA. Nat. biotechnol. 10, 1029–1032
44 Liu, X.L. et al. (2014) Repurposing endogenous Type I CRISPR-Cas systems for programmable gene repression. Nucleic Acids Res. 43, 672–681
61 Zha, H. and Shuman, S. (2005) Novel 2'-deoxyadenosine and 2'-deoxyguanosine activities of the bacterial non-homologous end-joining protein. DNA Repair 4, 2509–2511
Appendix III

CRISPR-based screening of genomic island excision events in bacteria
CRISPR-based screening of genomic island excision events in bacteria

Kurt Selleab, Todd R. Kleinhamme, and Rodolphe Barrangoua

*Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, NC 27695; and 1Functional Genomics Graduate Program, North Carolina State University, Raleigh, NC 27695

Contributed by Todd R. Kleinhammer, May 19, 2015 (sent for review September 22, 2014)

Genomic analysis of Streptococcus thermophilus revealed that mobile genetic elements (MGEs) likely contributed to gene acquisition and loss during evolutionary adaptation to milk. Clustered regularly interspaced short palindromic repeats (CRISPR)-associated genes (CRISPR-Cas), the adaptive immune system in bacteria, limits genetic diversity by targeting MGEs including bacteriophages, transposons, and plasmids. CRISPR-Cas systems are widespread in streptococci, suggesting that the interplay between CRISPR-Cas systems and MGEs is one of the driving forces governing genome homeostasis in this genus. To investigate the genetic outcomes resulting from CRISPR-Cas targeting of integrated MGEs, in silico prediction revealed four genomic islands without essential genes in lengths from 8 to 102 kbp, totaling 7% of the genome. In this study, the endogenous CRISPR3 type II system was programmed to target the four islands independently through plasmid-based expression of engineered CRISPR arrays. Targeting lacZ within the largest 102-kbp genomic island was lethal to wild-type cells and resulted in a reduction of up to 2.5-log in the surviving population. Genotyping of LacZ survivors revealed variable deletion events between the flanking insertion-sequence elements, all resulting in elimination of the Lac-encoding island. Chimeric insertion sequence footprints were observed at the deletion junctions after targeting all of the four genomic islands, suggesting a common mechanism of deletion via recombination between flanking insertion sequences. These results established that self-targeting CRISPR-Cas systems may direct significant evolution of bacterial genomes on a population level, influencing genome homeostasis and remodeling.

CRISPR | lacZ | acid bacteria | transposons | IS elements | Cas9

Mobile genetic elements (MGEs) present bacteria with continuous challenges to genomic stability, promoting evolution through horizontal gene transfer. The term “MGE” encompasses plasmids, bacteriophages, transposable elements, genomic islands, and many other specialized genetic elements (1). MGEs encompass genes conferring high rates of dissemination, adaptive advantages to the host, and genomic stability, leading to their nearly universal presence in bacterial genomes. To cope with the permanent threat of predatory bacteriophages and selfish genetic elements, bacteria have evolved both innate and adaptive immune systems targeting exogenous genetic elements. Innate immunity includes cell-wall modification, restriction/ modification systems, and abortive phage infection (2). Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated genes (Cas) are an adaptive immune system targeted against invasive genetic elements in bacteria (3). CRISPR-Cas-mediated immunity relies on distinct molecular processes, categorized as acquisition, expression, and interference (3). Acquisition occurs via molecular sampling of foreign genetic elements, from which short sequences, termed “spacers,” are integrated in a polarized fashion into the CRISPR array (4). Expression of CRISPR arrays is constitutive and inducible by promoter elements within the preceding leader sequence (5–6). Interference results from a corresponding transcript that is processed selectively at each repeat sequence, forming CRISPR RNAs (crRNAs) that guide Cas proteins for sequence-specific recognition and cleavage of target DNA complementary to the spacer (7). CRISPR-Cas technology has applications in strain typing and detection (8–10), exploitation of natural/engineered immunity against mobile genetic elements (11), programmable genome editing in diverse backgrounds (12), transcriptional control (13, 14), and manipulation of microbial populations in defined consortia (15).

Although sequence features corresponding to CRISPR arrays were described previously in multiple organisms (16, 17), Streptococcus thermophilus was the first microbe in which the role of specific cas genes and CRISPR array components were elucidated (4). S. thermophilus is a pathogenic, thermophilic Gram-positive bacterium used as a starter culture that catalyzes lactose to lactic acid in the syriphic production of yogurt and various cheeses (18). S. thermophilus encodes up to four CRISPR-Cas systems, two of which (CRISPR1 and CRISPR3) are classified as type II-A systems that are inactive in both acquisition and interference (4, 19). Accordingly, genomic analysis of S. thermophilus and its bacteriophages established a likely mechanism for phage/DNA protection in CRISPR-Cas systems. Investigation of CRISPR-Cas systems in S. thermophilus led to biomorphic analysis of spacer origin (4, 20), discovery of the proto-spacer adjacent motif (PAM) sequences (19, 21), understanding of phage-host dynamics (22, 23), demonstration of Cas9 endonuclease activity (7, 24, 25), and, recently, determination of the transactivating crRNA (tracrRNA) structural motifs governing function and orthogonality of type II systems (26). Genomic analysis of S. thermophilus revealed evolutionary


The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

*To whom correspondence may be addressed. E-mail: kleinhammer@ncsu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1508828112/-/DCSupplemental.

adaptation to milk through the loss of carbohydrate catabolism and virulence genes found in pathogenic streptococci (18). S. thermophilus also underwent significant acquisition of niche-related genes, such as those encoding cold-shock proteins, copper resistance proteins, proteinases, bacteriocins, and lactose catabolism proteins (18). Insertion sequences (ISs) are highly prevalent in S. thermophilus genomes and contribute to genetic heterogeneity among strains by facilitating dissemination of islands associated with dairy adaptation genes (18). The concomitant presence of MGEs and functional CRISPR-Cas systems in S. thermophilus suggests that genome homeostasis is governed at least in part by the interplay of these dynamic forces. Thus, S. thermophilus constitutes an ideal host for investigating the genetic outcomes of CRISPR-Cas targeting of genomic islands.

CRISPR-Cas systems recently have been the subject of intense research in genome editing applications (12), but the evolutionary roles of most endogenous microbial systems remain unknown (27). Even less is known concerning evolutionary outcomes of housing active CRISPR-Cas systems beyond the prevention of foreign DNA uptake (7), spacer acquisition events (4), and mutation caused by chromosomal self-targeting (38–32). Thus, we sought to determine the outcomes of targeting integrated MGEs with endogenous type II CRISPR-Cas systems. Four islands were identified in S. thermophilus LMD-9, with lengths ranging from 8 to 102 kbp and totaling ~132 kbp, or 7% of the genome. To target genomic islands, plasmid-based expressions of engineered CRISPR arrays with self-targeting spacers were transformed into S. thermophilus LMD-9. Collectively, our results elucidate fundamental genetic outcomes of self-targeting events and show that CRISPR-Cas systems can direct genome evolution at the bacterial population level.

Results
Identification of Expandable Genomic Regions. In silico prediction of mobile and expendable loci for CRISPR-Cas targeting was performed on the basis of (i) the location, orientation, and nucleotide identity of IS elements, and (ii) the location of essential ORFs. In Bacillus subtilis, 271 essential ORFs were identified by determining the lethality of genome-wide gene knockouts (33). The S. thermophilus genome was queried for homologs to each essential gene from B. subtilis using the BLASTp search tool under the default scoring matrix for amino acid sequences. Homologs to ~239 essential ORFs were identified in S. thermophilus, all of which were chromosomally encoded (Table S1). Proteins involved in conserved cellular processes including DNA replication/homeostasis, translation machinery, and core metabolic pathways were readily identified. No homologs corresponding to cytochrome biosynthesis/respiration were observed, in accordance with the metabolic profile of fermentative bacteria. Each putative essential ORF was mapped to the reference genome using SnapGene software, facilitating visualization of their location and distribution in S. thermophilus LMD-9 (Fig. 1).

IS elements within the S. thermophilus genome were grouped by aligning transposon coding sequences using Geneious software (Fig. S1). Family designations were determined according to BLAST analysis within the IS element database (https://www.isbiotech.fr/). To predict the potential for recombination-mediated excision of chromosomal segments, the relative locations of related IS elements were mapped to the S. thermophilus genome (Fig. 1). The IS1193 and IS80 families of IS elements appeared most frequently in the genome and are commonly found in Streptococcus pneumoniae and Streptococcus mitis (54) (S1193) and (IS80) elements were not frequent but exhibited nearly perfect identity between the copies identified in the genome (Figs. S1A and S2A). Despite the prevalence of IS1193 elements, many of these loci were shown to be small fragments that exhibited some polymorphism and degeneracy, but several copies with a high level of sequence identity were present also (Figs. S1B and S2B). The IS80 family exhibited considerable polymorphism and high degeneracy, with some copies harboring significant internal deletions (Figs. S1C and S2C). IS1167 elements were well conserved (Figs. S1D and S2D). Based on the conservation of length and sequence of the IS1167 and IS1193 elements of S. thermophilus and their relative proximity to milk adaptation genes, we postulate that these conserved/high-fidelity transposons were acquired in the genome recently.

By combining the location of predicted essential ORFs and highly similar IS elements, expendable islands were identified (Fig. 1 and Table 1). The first island contained an operon unique to S. thermophilus LMD-9, encoding a putative ATP-dependent oligosaccharide transport system with unknown specificity (Fig. S3A) (33). The second harbors the cell-envelope protein PrsA, which contributes to the fast-acidification phenotype of S. thermophilus (Fig. S3B) (36). Notably, although prsA is not ubiquitous in S. thermophilus genomes, it has been demonstrated that the genomic island encoding prsA is transferable between strains using natural competence (36). The third island contains a putative ATP-dependent copper efflux protein and is present in every sequenced S. thermophilus strain (Fig. S3C). The fourth island is the largest by far in terms of length, at 102 kbp, and gene content, with 15 predicted ORFs including the lac operon (Fig. S3D). This island is found in all strains of S. thermophilus, but the specific gene content and length vary among strains. To determine the outcome of targeting a large genomic island with both endogenous type II systems, repeat-spacer arrays were generated for the lacZ coding sequence (Fig. S3D) and were cloned into pORI28 (Table S2). The fourth island was selected for CRISPR-Cas targeting because of its size, ubiquity in S. thermophilus strains, and the ability to screen for lacZ mutations on the basis of a β-galactosidase-negative phenotype.

CRISPR-Cas Targeting of lacZ Selects for Large Deletion Events. In type II systems, Cas9 interrogates DNA and binds reversibly to PAM sequences with activation of Cas9 at the target occurring via formation of the tracrRNA:crRNA duplex (37), ultimately resulting in dsDNA cleavage (Fig. S4A and B) (25). Transformation with plasmids eliciting chromosomal self-targeting by CRISPR-Cas systems appeared cytotoxic as measured by the relative reduction in surviving transformants compared with non-self-targeting plasmids (15, 29). Targeting the lacZ gene in S. thermophilus resulted in an ~2.5-log reduction in recovered transformants (Fig. 2), approaching the limits of transformation efficiency. Double-strand breaks (DSBs) constitute a significant threat to the survival of organisms. The corresponding repair pathways often require end resection to repair blunt-ended DNA. Cas9–effected endonucleolyis further exacerbates
Table 1. Genomic island characteristics and CRISPR targets

<table>
<thead>
<tr>
<th>GEl</th>
<th>ORF region</th>
<th>Length, bp</th>
<th>IS family</th>
<th>Target gene</th>
<th>CRISPR system (PAM)</th>
<th>Spacer</th>
<th>PAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>STER_1390-STER_148</td>
<td>8,460</td>
<td>IS6</td>
<td>Oligopeptide transporters</td>
<td>3 (GGG)</td>
<td>GGCGCTGGG</td>
<td>GGG</td>
</tr>
<tr>
<td>2</td>
<td>STER_840-STER_884</td>
<td>11,932</td>
<td>IS51/111/IS167</td>
<td>Potentiation pf5</td>
<td>3 (GGG)</td>
<td>GGCGCTGGG</td>
<td>GGG</td>
</tr>
<tr>
<td>3</td>
<td>STER_10-STER_888</td>
<td>9,891</td>
<td>IS1191</td>
<td>Copper efflux operon</td>
<td>3 (GGG)</td>
<td>GGCGCTGGG</td>
<td>GGG</td>
</tr>
<tr>
<td>4</td>
<td>STER_1277-STER_1380</td>
<td>102,087</td>
<td>IS1193</td>
<td>102 ORFs</td>
<td>3 (GGG)</td>
<td>GGCGCTGGG</td>
<td>GGG</td>
</tr>
</tbody>
</table>

GEl, genomic island.

the pressure for mutations caused by DSBs to occur, because restoration of the target locus to the wild type does not circumvent subsequent CRISPR targeting. Identification of spacer origins within lactic acid bacteria revealed that 22% of spacers exhibit complementarity to self and that the corresponding genomic loci were altered, likely facilitating the survival of naturally occurring self-targeting events (28).

To determine if the target locus was mutated in response to Cas9-induced cleavage, transformants first were screened for loss of β-galactosidase activity. Clones deficient in activity were genotyped at the lacZ locus. No mutations caused by classical or alternative end joining and no spontaneous SNPs were observed in any of the clones sequenced. The absence of SNPs may be attributed to a lower transformation efficiency compounded by a low incidence of point mutations, and the absence of Ku and ligase IV homologs correlated with an absence of nonhomologous end joining (28). PCR screening indicated that wild-type lacZ was present, but the PCR amplicons did not correspond to the native lacZ locus; rather, an IS element-flanked sequence was amplified. To investigate the genotype responsible for the loss of β-galactosidase activity, single-molecule real-time sequencing was performed on two clones, one generated from CRISPR3 targeting the 5' end of lacZ, and the other generated from CRISPR3 targeting the sequence encoding the intergenic region. A distance of 3,400 bp was revealed by the absence of a large segment (~102 kbp) encoding the lacZ ORF (Fig. 3A and B). Both sequenced strains confirmed the reproducibility of the large deletion in the genomic DNA sequence and revealed the absence of a large segment (~102 kbp) encoding the lacZ ORF (Fig. 3A and B). Both sequenced strains confirmed the reproducibility of the large deletion boundaries and showed that the deletion occurred independently of the lacZ spacer sequence used for targeting.

The 102-kbp segments deleted constitute ~5.5% of the 1,866-Mbp genome of S. thermophilus. The region contained 102 putative ORFs (STER_1278-1370), encoding ATP-binding cassette (ABC) transporters, two-component regulatory systems, bacteriocin synthesis genes, flagellar genes, and lactose catabolism genes, and several cryptic genes with no annotated function (35). The exact structure of the strain in the deletion was determined in broth culture by measuring OD at 600 nm over time (Fig. 3C).

The deletion clones appeared to have a longer lag phase and lower final OD (P < 0.01) and exhibited a significantly shorter generation time during log phase (average of 103 min compared to 62 min for the wild type; P < 0.001). Although the deletion derivatives have 5.5% less of the genome to replicate per generation and expend no resources in transcription or translation of the eliminated ORFs, no apparent increase in fitness was observed relative to the wild type. β-Galactosidase activity is a hallmark feature for industrial applications of lactic acid bacteria and is essential for preservation of food systems through acidification. The capacity of lacZ-deficient S. thermophilus strains to acidify milk therefore was assessed by monitoring pH (Fig. 4D). Predictably, the deletion strain failed to acidify milk over the course of the experiment, in sharp contrast to the rapid acidification phenotype observed in the wild type.

Genomic Deletions Occur Through Recombination between Homologous IS Elements. To investigate the mechanism of deletion, the nucleotide sequences flanking the target were determined. The only homologous sequences observed at the junctions were two truncated IS1193 insertion sequences exhibiting 91% nucleotide sequence identity globally over 727 bp. Accordingly, a primer pair flanking the two IS elements was designed to amplify genomic DNA of surviving clones exhibiting the deletion. Each of the deletion strains exhibited a strong band of the predicted size (~12 kbp) and confirmed the large genomic deletion event (Fig. 4A). Interestingly, a faint band corresponding to the chromosomal deletion was observed in the wild type, indicating that this region may exist naturally within the genome at a low rate within wild-type populations. Sequencing of the junction amplicon was performed for 20 clones generated by chromosomal self-targeting by CRISPR3. Genotyping of the junction revealed the presence of one chimeric IS element in each clone, and furthermore revealed the transition from the upstream element to the downstream sequence within the chimera for each clone (Fig. 4B). The size of deletions observed ranged from 101.965 to 102.146 bp. The exact locus of transition was variable but was not randomized within the clones, implying the potential bias of the deletion mechanism. S. thermophilus harbors atypical recombination machinery encoded as RecA.
Fig. 3. Genome sequencing and phenotypic analysis of lacZ clones (A and B). Sequence data revealed an absence of the chromosomal segment encoding lacZ in two mutants independently created by targeting the 5′- and 3′- and central-binding region (B) coding sequences of lacZ using the CRISPR-Cas system. The size of the deletions ranged from 90.86 to 100.16 bp in length, constituting 1.5% of the genome of S. thermophilus. (C) Growth of large-deletion strains generated by CRISp spacer 1 (circles) and CRISp spacer 2 (diamonds) compared with wild type (square) in semisynthetic EiEi medium (replicated as mean ± SD). Shown is OD at 600 nm of three independent biological replicates. (D) Acidification capacity of wild-type S. thermophilus (squares) and large-deletion strains (circles/diamonds) in skim milk relative to an unmodified control (triangles).

(STER_0977). AddAB homologues functioning as dual ATP-dependent DNA exonucleases (STER_1681 and STER_1682), and a Helicase (STER_1742) of the RecD family. The high nucleotide identity between the flanking IS elements and the capacity for S. thermophilus to carry out site-specific recombination (4) confirms the potential for RecA-mediated recombination to mediate excision of the genomic segment (Fig. 4C).

It was hypothesized that CRISPR-Cas targeting could facilitate isolation of deletions for each locus with the same genetic architecture. Thus, three CRISPR3 repeat-spacer arrays, one targeting the oligomucleotide transporter in the first locus, one targeting proS from the second locus, and one targeting the ATPase copper efflux gene from the third locus, were generated and cloned into pORL2 (Table S2). To screen for deletions, primers flanking the IS elements at each locus were designed to amplify each deletion junction (Fig. 4D). The absence of wild-type loc was confirmed in each case by designing internal primers for each genomic island (Fig. 4E). After transformations with the targeting plasmids, deletions at each locus were isolated, and the absence of wild type was confirmed. Sequencing of the deletion junction amplicons confirmed that a unique chimeric IS element footprint remained, indicating a common mechanism for deletion at each locus. Interestingly, primers flanking the IS elements also amplified from wild-type gDNA, further suggesting that population heterogeneity that naturally occurred at each locus resulted from spontaneous genomic deletions. These results imply that sequence-specific Cas9 cleavage selects for the variants lacking protospacer and PAM combinations necessary for targeting. Thus, spontaneous genomic deletions can be isolated using CRISPR-Cas targeting as a strong selection for microbial variants that already have lost those genomic islands.

Discussion

In this study, native type IIIA systems harbored in S. thermophilus were repurposed for defining spontaneous deletions of large genomic islands. By independently targeting four islands in S. thermophilus, stable mutants collectively losing a total of 7% of the genome were generated. Characterization of the deletion junctions suggested that an IS-dependent recombination mechanism contributes to population heterogeneity and revealed deletion events ranging from 8 to 102 kbp. Precise mapping of the chimeric IS elements indicated that natural recombination events are likely to be responsible for the large chromosomal deletions in S. thermophilus and potentially could be exploited for targeted genome editing. Recent landmark studies have highlighted the potential for CRISPR-Cas-induced chromosomal deletions and rearrangements in bacteria (29, 30). Jiang et al. (29) first reported that sequence-specific Cas9 cleavage selects for preexisting variants lacking protospacer and PAM combinations necessary for targeting in S. pneumoniae. Similarly, Vercoe et al. (30) demonstrated that chromosomal targeting by a type-IF CRISPR-Cas system caused deletion of a horizontally acquired pathogenicity island in Proteus mirabilis (30). The concept of sequence-based removal of specific genotypes was developed further as a tool for manipulation of microbial consortia via CRISPR-Cas targeting, resulting in directed genome evolution at the population level (15). In accordance with previous work, our results demonstrate that wild-type clones were removed from the population, but mutants without CRISPR-Cas-targeted features survived. Thus, adaptive islands were identified and validated, showing that precise targeting by an endogenous Cas9 can be exploited for isolating large deletion variants in mixed populations.

Genome evolution of bacteria occurs through horizontal gene transfer, intrinsic mutation, and genome restructuring. Genome sequencing and comparative analysis of S. thermophilus strains
have revealed significant genome decay but also indicate that adaptation to nutrient-rich food environments occurred through niche-specific gene acquisition (18, 25). The presence of MGEs including integrative and conjugative elements, prophages, and IS elements in S. thermophilus genomes is indicative of rapid evolution to a dairy environment (39, 40). Mobile genetic features facilitate gene acquisition and, conversely, inactivation or loss of nonessential sequences. Consequently, MGEs confer genomic plasticity as a means of increasing fitness or changing ecological lifestyles. Our results strongly indicate that CRISPR-Cas targeting of these elements may influence chromosomal rearrangements and homoeostasis. This finding is in contrast to experiments targeting essential features, which resulted in the selection of variants with inactivated CRISPR-Cas machinery (41). Mutation of essential ORFs is not a viable avenue for circumvention of CRISPR-Cas targeting, and thus only those clones with inactivated CRISPR-Cas systems remain. By design, targeting genetic elements predicted to be hypervariable and expendable demonstrated that variants with altered loci were viable, maintaining active CRISPR-Cas systems during self-targeting events.

Despite the nearly ubiquitous distribution of IS elements in bacterial genomes, they remain an enigmatic genetic entity, largely because of their diversity and plasticity in function (34). Our results suggest it is possible to predict recombination between related IS elements by analyzing their location, orientation, and sequence conservation (Figs. S1 and S2). CRISPR-Cas targeting them can be used to validate population heterogeneity empirically at each predicted locus and simultaneously increase the recovery of low-incidence mutants. The high prevalence of MGEs in lactic acid bacteria, and especially S. thermophilus, is in accordance with their role in speciation of these hyper-adapted bacteria through genome evolution (39, 40). Moreover, recovery of genomic deletion mutants using CRISPR-Cas targeting could facilitate phenotypic characterization of genes with unknown functions. Mutants exhibiting the deletion of the 102-kb island encoding the lac operon had significantly increased generation times relative to the wild type and achieved a lower final OD. With 102 predicted ORFs therein, it is likely that additional phenotypes are affected, and many of the genes do not have annotated functions. CRISPR-Cas targeting allows direct assessment of how island-encoded genes contribute to adaptation to grow in milk; this understanding is important, given the industrial relevance of niche-specific genes such as prs. Moreover, it is in the natural genomic and ecological context of these horizontally acquired traits, because they likely were acquired as discrete islands. These results establish avenues for the application of self-targeting CRISPR-Cas9 systems in bacteria to investigate transposition, DNA repair mechanisms, and genome plasticity.

CRISPR-Cas systems generally limit genetic diversity through interference with genetic elements, but acquired MGEs also can provide adaptive advantages to host bacteria. Thus, the benefit of maintaining genetically integrated MGEs despite CRISPR-Cas targeting is an important driver of genome homeostasis. Collectively, our results establish that in silico prediction of GEIs can be coupled with CRISPR-Cas targeting to isolate clones exhibiting large genomic deletions. Chimeric insertion sequence footprints at each deletion junction indicated a common mechanism of deletion for all four islands. The high prevalence of self-targeting spacers exhibiting identity to genomic loci, combined with experimental demonstrations of genomic alterations, suggests that CRISPR-Cas self-targeting may contribute significantly to genome evolution of bacteria (28, 30). Collectively, studies on CRISPR-Cas-induced large deletions substantiate this approach as a rapid and effective means to assess the essentiality and functionality of gene clusters devoid of annotation and to define minimal bacterial genomes based on chromosomal deletions occurring through transposable elements.

Materials and Methods

Bacterial Strains. All bacterial strains are listed in Table S2. Bacterial cultures were grown/kept in an appropriate growth medium with 25% glycerol (v/vol) and stored at -80 °C. E. thermophilus was propagated in EiIker medium (Difco) supplemented with 1% beef extract (wt/vol) and 1.9% (wt/vol) d-glucosylphosphate (Sigma) broth under static aerobic conditions at 37 °C on an orbital shaker at 150 rpm. Antibiotic selection of E. coli was maintained with 40 μg/ml kanamycin and 150 μg/ml of gentamycin.

DNA Isolation and Cloning. All kits, enzymes, and reagents were used according to the manufacturers’ instructions. DNA purification and cloning were performed as described previously (42). Plasmids with lacZ targeting arrays were constructed with each consisting sequentially of the (i) native
leader sequence specific to CRISPR1 or (ii) CRISPR3 native repeat specific to CRISPR1 or (iii) CRISPR3 spacer sequence specific to the 5' end of laiz, another native repeat. To engineer each plasmid, the sequence features listed above were ordered as extended oligonucleotides (Table S3), combined using splicing by overlap extension PCR (42), and cloned into pORI28 (Table S3).

Selection and Design of CRISPR Spacers. Putative protospacers were conserved by first defining the location of all putative PAM sequences in the same and adjacent frames of laiz. Within the 3,081-bp genome, there were 22 CRISPR1 (AGAAGAW) and 39 CRISPR3 (GGGAG) PAM sites that were identical to their biologically derived consensus sequences (71). After potential spacers were collected, the complete protospacer, seed, and PAM sequence were subjected to BLAST analysis against the genome of S. thermophilus LMD-9 to prevent additional targeting of romsppecic loci. The spacers for CRISPR1 and CRISPR3 were disqualified in sequence and corresponding PAM sites but were designed to target the 5' end of laiz, resulting in predicted cleavage sites residing 6 bp apart. The leader sequences, repeats, and spacers on each plasmid represented orthogonal features unique to CRISPR1 or CRISPR3, respectively. To assess target locus-dependent mutations, an additional CRISPR1/CRISPR3 plasmid was created with a spacer to the metal- 

Transformation. Plasmids were electroprogrammed into competent S. thermophilus containing the temperature-sensitive helper plasmid pTMA80. An overnight culture of S. thermophilus was inoculated at 1% (v/v) into 50 mL of B伊利克 medium supplemented with 1% beef extract, 1.9% (g/l) glycerol-phosphate, and 0.1% yeast extract. After the culture achieved an OD600 of 0.3, penicillin G was added to achieve a final concentration of 100 μg/mL.

Cells were harvested by centrifugation and washed three times in 10 mM cold 

electroprogrammable buffer (1 M sucrose and 3.5 mM MgCl2). The cells were con- 

centrated 100-fold in electroporation buffer, and 45 μL of the suspension was aliquoted into 0.1 mL electroporation cuvettes. Each suspension was combi- 

ned with 700 ng of plasmid. Electroporation conditions were set at 2,500 V, 25 μF capacitance, and 200 Ω resistance. Time constants were recorded and ranged from 4.4 to 4.5 ms. The suspensions were combined immediately with 950 μL of recovery medium and were incubated for 8 h at 37 °C. Cell sus- 

pensions were plated on solid medium, and electroporation efficiencies were washed with medium to ensure recovery of cells.

Growth and Activity Assessment. Cultures were preconditioned for growth assay by subculturing for 12 generations in a semi-synthetic Eikler medium with 0.1% (v/v) glycerol. Cultures were assayed with an overnight culture at 1% (v/v) and inoculated at 37 °C statically. OD600 was monitored hourly until the cultures achieved stationary phase. 

Acknowledgment. We thank Alele Briner, Madeleine Shoup, Chelse Beisel, Yong Jun-Goh, Sarah O'Callahan, and Robert Conner for helpful discussions, and the North Carolina State University (CDSU) Dairy Plant for providing skim milk. This work was supported by NSF grant number 1207336, the North Carolina Agriculture Foundation, and DuPont Nutrition and Health. K.S. was supported in part by the 2013-2014 Daniels Program in Food Safety of The Edward D. Martin Company, Inc. during the completion of this work.

References

Appendix IV

Programmable removal of bacterial strains by used of genome-targeting CRISPR-Cas systems
Programmable Removal of Bacterial Strains by Use of Genome-Targeting CRISPR-Cas Systems

Ahmed A. Gomaa, Heidi E. Klumpe, Michelle L. Luo, Kurt Selle, Rodolphe Barrangou, Chase L. Beisel
Department of Chemistry and Biochemistry, North Carolina State University, Raleigh, North Carolina, USA; Department of Food, Bioprocessing, and Nutritional Sciences, North Carolina State University, Raleigh, North Carolina, USA

ABSTRACT CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) systems in bacteria and archaea employ CRISPR RNAs to specifically recognize the complementary DNA of foreign invaders, leading to sequence-specific cleavage or degradation of the target DNA. Recent work has shown that the accidental or intentional targeting of the bacterial genome is cytotoxic and can lead to cell death. Here, we have demonstrated that genome targeting with CRISPR-Cas systems can be employed for the sequence-specific and titratable removal of individual bacterial strains and species. Using the type I-E CRISPR-Cas system in Escherichia coli as a model, we found that this effect could be elicited using native or imported systems and was similarly potent regardless of the genomic location, strand, or transcriptional activity of the target sequence. Furthermore, the specificity of targeting with CRISPR RNAs could readily distinguish between even highly similar strains in pure or mixed cultures. Finally, varying the collection of delivered CRISPR RNAs could quantitatively control the relative number of individual strains within a mixed culture. Critically, the observed selectivity and programmability of bacterial removal would be virtually impossible with traditional antibiotics, bacteriophages, selectable markers, or tailored growth conditions. Once delivery challenges are addressed, we envision that this approach could offer a novel means to quantitatively control the composition of environmental and industrial microbial consortia and may open new avenues for the development of "smart" antibiotics that circumvent multidrug resistance and differentiate between pathogenic and beneficial microorganisms.

IMPORTANCE Controlling the composition of microbial populations is a critical aspect in medicine, biotechnology, and environmental cycles. While different antimicrobial strategies, such as antibiotics, antimicrobial peptides, and lytic bacteriophages, offer partial solutions, what remains elusive is a generalized and programmable strategy that can distinguish between even closely related microorganisms and that allows for fine control over the composition of a microbial population. This study demonstrates that RNA-directed immune systems in bacteria and archaea called CRISPR-Cas systems can provide such a strategy. These systems can be employed to selectively and quantitatively remove individual bacterial strains based purely on sequence information, creating opportunities in the treatment of multidrug-resistant infections, the control of industrial fermentations, and the study of microbial consortia.

Microorganisms play critical roles in human health and environmental nutrient cycles and are regularly employed in diverse industrial processes. Within each context, a central challenge is controlling the specific composition of a mixed population. A few strategies can remove some microorganisms but not others (Fig. 1): defined growth conditions, conventional antibiotics, and antimicrobial peptides with some strain specificity, lytic bacteriophages, and the expression of antibiotic resistance genes, auxotrophic markers, or toxins under unique expression systems (1). Unfortunately, most of these approaches offer constrained opportunities to selectively remove individual bacterial strains (e.g., antibiotics and antimicrobial peptides) or require detailed knowledge of the genetics, metabolism, and physiology of each constituent of the population (e.g., selective growth conditions). Lytic bacteriophages often offer exquisite specificity (2). However, individual bacteriophages must be isolated against each strain and would require additional screening to determine the degree of specificity. Furthermore, lytic bacteriophages replicate as part of the infection cycle, eventually wiping out the entire target population or breeding resistance. What remains elusive is a generalized and programmable strategy that can distinguish between even closely related microorganisms and that allows for fine control over the composition of a microbial population. We propose that CRISPR-Cas systems could provide such a strategy.

CRISPR-Cas systems are RNA-directed adaptive immune systems in many bacteria and most archaea that recognize nucleic acids of invading plasmids and viruses (3, 4). Recognition is directed by CRISPR RNAs that are processed from transcribed arrays of alternating target-specific "spacer" sequences and identical "repeat" sequences. The spacer region of each CRISPR RNA base
purs with complementary nucleic acids, driving cleavage or degradation by the Cas proteins within minutes of invasion (5–7).

Three types of CRISPR-Cas systems, which vary in their specificity and mode of action, have been defined. Type I systems cleave and degrade DNA, type II systems cleave DNA, and type III systems cleave DNA or RNA (8). Type I and II systems require two principal factors to effectively target DNA: (i) complementarity between the CRISPR RNA spacer and the target “protospacer” sequence and (ii) a protospacer-adjacent motif (PAM) specific to each CRISPR-Cas system flank the protospacer (9–11). Effective targeting can occur even for multiple mismatches between the CRISPR RNA and the protospacer, although mismatches within the “seed” region flanking the PAM are more disruptive (9, 12). Similar factors are required for DNA targeting by type III systems, where these systems evaluate base pairing between the target sequence and the region flanking the protospacer (13).

While these factors help safeguard against accidental targeting of genomic sequences, they provide a simple set of design rules to achieve DNA targeting. This has primarily been exploited with type II systems for genome editing, whereby cleavage is followed by DNA repair through non-homologous and joining (NHEJ) or homologous recombination (11, 14, 15). However, within microorganisms with poor or absent NHEJ, genome targeting can be lethal. For instance, natural systems that acquired genome-targeting spacers appear to possess inactive systems or mutated target loci (16), potentially explaining the evolution of Pseudomonas aeruginosa harboring a type I-F CRISPR-Cas system (17). In the industrial bacterium Streptococcus thermophilus, cultures under biocatalyst attack rarely integrate genome-targeting spacers and, in such an event, rapidly disappear from the population (18).

In the hyperthermophilic archaeon Sulfolobus solfataricus harboring a type III-A system, infecting cells with viral particles encoding a genome-targeting spacer slowed the growth of the culture under selecting conditions and led to recombination between the virally encoded spacer and the endogenous CRISPR array (19). In the bacterium Eshetrichia coli expressing a type I-E or II-A system in trans, transformation of a plasmid with spacers targeting endogenous genes or a lysogenized bacteriophage led to extremely low recovery of viable transformants (20–23). Similar results were obtained in the bacterial pathogen Streplococcus pneumoniae expressing a type II system in trans, wherein viable transformants contained mutations or deletions that inactivated CRISPR-Cas-mediated targeting (11, 24). Finally, in the bacterial phytopathogen Pseudomonas atrovirens harboring a native type I-F system, induction of genome-targeting spacers from a tightly regulated plasmid was cytotoxic within a few hours of induction and led to extensive deletions in the target loci (25). This extensive evidence presents an opportunity for the sequence-specific removal of microorganisms by reprogramming CRISPR-Cas systems.

Here, we investigated the potential of CRISPR-Cas systems for the sequence-specific targeting and selective removal of individual strains of bacteria. Using the E. coli type I system as a model, we found that targeting the E. coli genome led to partial removal of cells as long as the target sequence contained a PAM and was complementary to the spacer. In contrast to targeting of bacteriophages and plasmids, genome targeting accommodated multiple mutations in the seed region. Furthermore, this selectivity was effective regardless of genomic location, strain, or transcriptional activity. Finally, using genomic sequence information, we could selectively remove closely related bacterial strains whether in pure or mixed cultures. The extent of removal could even be modulated by mixing targeting and non-targeting plasmids, opening the possibility of quantitatively controlling the composition of microbial consortia and selectively treating multidrug-resistant infections, particularly with ongoing advances in the delivery of nucleic acids to microorganisms.

RESULTS

Genome targeting with the type I-E CRISPR-Cas system in E. coli. We first evaluated the impact of targeting a natural genomic locus with the type I-E CRISPR-Cas system from Escherichia coli K-12, one of the best-characterized CRISPR-Cas systems to date. This system encodes six cas genes in two operons (casABCDEF and cas3) required for CRISPR RNA processing and the cleavage and degradation of target DNA (26). Because the casABCDEF operon is repressed by H-NS in E. coli K-12 under normal growth conditions (27), we used a previously developed system consisting of two plasmids (pCasA-E and pCas3) (see Fig. S1 in the supplemental material) that inductively express all six cas genes (26). In addition, we generated a third plasmid, encoding an altered version of the endogenous CRISPR array in E. coli K-12 that accommodates the sequential insertion of engineered spacer sequences (pCRISPR) (see Fig. S1 and S2). pCRISPR plasmids encoding engineered, genome-targeting spacers were transformed into E. coli K-12 strain BW25113 cells equipped with inducible expression of the T7 polymerase (BW25113 T7) and harboring the two cas-expressing plasmids (pCasA-E and pCas3) (see Fig. S1). As part of the assay, we measured the transformation efficiency, a proxy for removal of strains in pure cultures, by dividing the number of viable transformants for each genome-targeting CRISPR plasmid by the number of viable transformants for the original pCRISPR plasmid. Lower ratio or transformation efficiencies indicate a greater extent of removal.

We began with a spacer that is complementary to the template strand of the essential ftsK gene, involved in cell division (Fig. 2A). The selected protospacer was immediately downstream of AAG, one of the four PAMs for this CRISPR-Cas system (9). The resulting anti-ftsK (as-ftsK) plasmid exhibited transformation efficiency ~10^{-4} fold lower than that of the original pCRISPR plasmid (Fig. 2B), paralleling the transformation efficiency of plasmids encoding prophage-targeting CRISPR RNAs (21, 22). In the absence of the casABCDEF operon, the as-ftsK plasmid and the original pCRISPR plasmid yielded similar transformation efficiencies (Fig. 2B), ruling out transformation issues and confirming the role of the casABCDEF operon. Forced expression of the chromosomally encoded ftsK gene through deletion of the hns gene also resulted in a low transformation efficiency (see Fig. S3A in the
the resistance cassette into the *E. coli* genome. Finally, we sequenced the CRISPR locus on the α-fosA plasmid, which revealed various deletions that removed the α-fosA spacer (see Fig. 5A in the supplemental material). In total, alterations to the CRISPR RNAs appear to principally account for surviving colonies, at least within our experimental setup.

We next evaluated the sequence specificity of CRISPR-Cas-mediated removal. We introduced different point mutations into the seed region of the α-fosA spacer (Fig. 2B), where the seed region for the type I-E CRISPR-Cas system is conventionally identified as the first through fifth, seventh, and eighth nucleotides flanking the PAM (9, 12). Prior work demonstrated that single point mutations within this region of the spacer for type I CRISPR-Cas systems disrupts immunity against bacteriophages (9). However, we found that point mutations to the third (m5), fifth (m7), or seventh (m8) nucleotide of the wild-type (WT) α-fosA spacer only marginally disrupted removal (Fig. 2B). Pairing point mutations (m2,5; m7,9; m5,7) further disrupted removal, while only the combination of all three point mutations (m2,5,7) fully disrupted removal.

To further probe the specificity of removal, we introduced compensatory mutations within the native fosA gene (m2,5,7') (see Fig. S5 in the supplemental material). The matched pairing of the m2,5,7' spacer and the m2,5,7' strain resulted in a large extent of removal, albeit less than that seen with the pairing of the WT spacer and the WT strain (Fig. 2B). Separately, the mismatched pairing of the WT spacer and the m2,5,7' strain exhibited negligible removal (Fig. 2B), excluding the possibility of unintended targeting at other genomic loci.

**Potent removal by targeting diverse locations throughout the genome.** Programming of CRISPR-Cas systems to remove individual strains would greatly benefit from the ability to readily target any PAM-flanking sequence throughout a genome. Previous examples of genome targeting successfully targeted different genes on both strands of the genome (11, 19, 20, 25). However, a comprehensive and quantitative investigation of genome targeting has not been conducted. Toward this goal, we designed 10 additional spacers that target different protospacers flanked by a PAM throughout the *E. coli* K-12 genome (Fig. 3A). The corresponding protospacers covered a diverse range of locations, including the positive and negative strands of the genome, template strands and non-template strands of genes, and within transcribed regions. Furthermore, we targeted both essential and non-essential genes because of their relative capacities to tolerate mutations or deletions. In all cases, the extent of removal was statistically similar to that of the original α-fosA spacer (P values between 0.05 and 0.88) (Fig. 3B), suggesting that removal is based on chemoosmotic injury rather than on perturbing the natural functions of the target locus. Furthermore, in the absence of the casABCDE operon, each plasmid and the original CRISPR plasmid yielded similar transformation efficiencies (see Fig. S6 in the supplemental material). The PAM was an essential feature, similar to findings of previous studies (11, 20, 25), since targeting a separate site within the fosA gene lacking a PAM resulted in negligible removal (Fig. 3B). Based on these results, we conclude that potent removal can be achieved by targeting diverse locations throughout the genome as long as a PAM is present. Interestingly, the simultaneous targeting of multiple locations (ααd, mbA, fosA, and musB) exhibited extents of removal similar to those with targeting of only one of the locations (fosA) (P = 0.48) (see Fig. S7).
FIG 3 Similar efficiencies when targeting diverse locations throughout the genome. (A) Protoparson locations in the E. coli K-12 genome. Dark circles within and outside the circle reflect spacers designed to base pair with the negative (−) or positive (+) strand of the chromosome, respectively. Dark circles reflect protospacers flanked by 20 bp PAM (white), on the template strand (blue), or on the non-template strand (green) of coding regions or in non-transcribed regions (purple). (B) Transformation efficiency for CRISPR plasmids encoding spacers targeting the sites shown in panel A to BW25113 (T) harboring pCas9 and pCas8. See the legend for Fig. 2B for an explanation of the transformation efficiency. Values represent the geometric means and SEM of data from three independent experiments.

To explore the broad potential of our approach through native CRISPR-Cas systems outside of E. coli, we explored the impact of genome targeting in the Gram-positive bacterium Streptococcus thermophilus. In particular, we assessed genome targeting through the two native type II CRISPR-Cas systems (CRISPR1 and CRISPR2) previously shown to be active under normal growth conditions (4, 29). The transformation efficiencies of plasmids encoding CRISPR1 and CRISPR3 RNAs targeting the lacZ gene were near the limit of detection (10⁻¹-fold lower than that of the empty plasmid) (see Fig. S1B and Table S1 in the supplemental material). Therefore, potent removal can be achieved through different native CRISPR-Cas systems.

Sequence-specific removal of individual strains. The flexibility and sequence specificity of genome targeting open the intriguing possibility of using CRISPR-Cas systems to specifically remove individual microbial species and strains. To begin exploring this possibility, we focused on two substrains of E. coli K-12 (BW25113-T7) and E. coli B (BL21(DE3)) (Fig. 4A). Because the genomes of these strains bear more than 99% sequence homology and almost all cellular processes are identical (30), selectively removing one of the strains would be extremely difficult with antimicrobial agents or under defined growth conditions. However, the distinguishing sequences afford ample opportunities to selectively target either strain with programmed CRISPR-Cas systems. Using in silico genomic analyses, we identified one PAM-flanking sequence unique to E. coli K-12 (within the fopB gene, involved in the transport of L-fucose), one PAM-flanking sequence unique to E. coli B (within the ogr gene, located within the P2 prophage), and one PAM-flanking sequence shared by both strains (within the groI gene, involved in protein folding). We subsequently designed CRISPR spacers that recognize PAM-adjacent protospacers in each genome and measured removal in pure cultures harboring pCasA-E and pCas3 (see Fig. S1 in the supplemental material). As expected, targeting fopB removed only the K-12 strain, targeting ogr removed only the B strain, and targeting groI removed both strains (Fig. 4B).

One potential application of programmable removal with CRISPR-Cas systems is targeting pathogenic bacteria while sparing commensal bacteria. Toward this goal, we focused on E. coli K-12 (BW25113-T7), a derivative of commensal E. coli that naturally inhabits the human digestive tract, and on Salmonella enterica (SBS000A1, a derivative of LT2), a major food pathogen. Both species are Gram-negative enterobacteria, and they share ~71% sequence homology (Fig. 4A) (31). Using genomic analyses, we designed CRISPR spacers targeting a PAM-flanking sequence unique to E. coli (within the araB gene, involved in the regulation of acetyl-coenzyme A [CoA] biosynthesis), a PAM-flanking sequence unique to S. enterica (within the mvM gene, encoding a putative virulence factor), and a shared PAM-flanking sequence (within the groI gene). The resulting plasmids were transformed into pure cultures harboring pCasA-E and pCas3 (see Fig. S1 in the supplemental material). As expected, targeting araB removed only E. coli, targeting mvM removed only S. enterica, and targeting groI removed both strains (Fig. 4B).

Selective and titratable removal of individual strains in mixed cultures. We next proceeded from pure cultures to mixed
cultures in order to evaluate the selective removal of target strains. We repeated the transformation experiments with *E. coli* B [BL21(DE3)] and *E. coli* K-12 (BW25133-T7), except that both strains were cocultured and plated on agar with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and isopropyl-β-D-thiogalactopyranoside (IPTG). Under these plating conditions, BL21(DE3) yields blue colonies, whereas BW25133-T7 yields white colonies (see Fig. S8 in the supplemental material). Similar to the experiments with pure cultures, targeting the PAM-flanking sequence within the ocr gene selectively removed BL21(DE3), targeting the PAM-flanking sequence within the fucP gene selectively removed BW25133-T7, and targeting with the original pCRISPR plasmid maintained both strains (Fig. S5A). Furthermore, in the absence of the casABCD operon, the two strains were maintained in similar ratios regardless of the transformed plasmid (Fig. S5A). These results demonstrate that CRISPR-Cas systems can be employed for the selective removal of bacterial strains in mixed cultures.

The above mixed-culture experiments utilized single plasmids to either remove or maintain individual strains. We hypothesized that transforming combinations of targeting and nontargeting plasmids could remove a portion of targeted cells, conferring control over the composition of the population. To test this hypothesis, we transformed different amounts of the pCRISPR plasmid and the BL21(DE3)-targeting plasmid (total of 100 ng) and then quantitated the ratio of blue and white colonies. Remarkably, the fraction of the BL21(DE3)-targeting plasmid strongly correlated with the selective removal of BL21(DE3) (Fig. S5B). The almost-perfect linear correlation ($R^2 = 1.00$) further suggests that almost all transformed cells received a single plasmid that either removed or sustained *E. coli* B. We thus conclude that CRISPR-Cas systems can be reprogrammed to quantitatively modulate the composition of a mixed population.

**DISCUSSION**

We demonstrated the sequence-specific removal of individual strains using CRISPR-Cas systems. While the extent of removal was extremely high (>$99.99\%$ for the *fusR* plasmid [Fig. 2B]), a fraction of the transformed cells consistently survived targeting. Sequencing survivors revealed consistent loss of the genome-targeting spacer in the transformed pCRISPR plasmid (see Fig. S4 in the supplemental material), likely through recombination between the identical repeats. This insight is consistent with findings of a recent study showing loss or inactivation of CRISPR elements under evolutionary pressure (28). This insight also suggests one potential countermeasure against surviving colonies: reducing the number of repeats within the CRISPR array. The array could even be reduced to a single repeat-spacer, paralleling the engineering of single-guide RNAs for use with Cas9 (32). Other potential strategies include expressing multiple CRISPR-Cas systems or diminishing CRISPR-encoding plasmids that underwent recombination. Targeting multiple sites at one time did not appear to be an effective strategy (see Fig. S7), perceptibly due to rearrangement of the CRISPR-encoding plasmid (see Fig. S4).

We also demonstrated that potental removal could be achieved using type I and type II CRISPR-Cas systems. An interesting distinction between these systems is that type I systems cleave and degrade DNA through the action of a 5'-to-3' exonuclease, whereas type II systems only cleave DNA (33). The additional effect of DNA degradation by type I systems may further improve
the potency of genome targeting by preventing DNA repair, although a direct comparison between type I and type II systems would be difficult to directly evaluate this potential contribution.

We found that multiple mismatches within the seed region were required to fully disrupt targeting by the type I-E system from E. coli (Fig. 2). This may explain the absence of mutations within the protospacer of surviving colonies (see Fig. S4 in the supplemental material). Separately, the number of required mismatches contrasts with the single mismatches that disrupted im- position of the S. aureus phage (9). This discrepancy is intriguing considering that the same Cas-encoding plasmids were used in these studies. One possibility is that the seed region can accommodate different numbers of mismatches when targeting genomic DNA or when targeting invader DNA. Such differences may help explain emerging reports of elevated off-target effects associated with genome editing (34–37).

Delivery arguably poses the most immediate challenge to the downstream use of CRISPR-Cas systems for the selective and titratable removal of microorganisms. However, opportunities in nanoparticle development and the engineering of bacteriophages present potential solutions. Nanoparticles have been used to deliver nucleic acids to bacteria (38), but little subsequent work has been done; the delivery of CRISPR RNAs may provide the impetus to further investigate nanoparticles as vehicles of delivery to microorganisms. Separately, bacteriophages have been widely used for heterologous protein expression, gene delivery, and the treat- ment of bacterial infections (39, 40). Lyssogenic bacteriophages or phage with a broad host range would be particularly beneficial for the delivery of CRISPR-Cas encoding constructs (41–43). While silver nanoparticles also could be used to remove bacteria (44, 45), they lack the specificity or the programmability offered by genome-targeting CRISPR-Cas systems and cannot be easily dosed to quantitatively control the compo- nent of a microbial consortium.

Once delivery challenges are overcome, we foresee CRISPR-Cas systems being exploited to control bacterial populations in diverse ecological niches and scientific fields. In biotechnology, CRISPR-Ca systems could be used to selectively clear contaminating microorganisms or to quantitatively control the composition of microbial consortia in industrial processes or in environmental samples. In medicine, CRISPR-Cas systems could be used to control the composition of the gut flora or as "smart" antibiotics that circumvent commonly transmitted modes of antibiotic resistance and distinguish between beneficial and pathogenic bac- teria. For applications that require removal of more than one strain, multiple spacers that target shared or unique sequences could be encoded in a single CRISPR array. The arrays could also be combinatorially with a set of genes to instigate removal of strains lacking functional CRISPR-Cas systems (15, 46). Because of the sequence specificity of targeting, CRISPR-Cas systems could be used to distinguish strains separated by only a few base pairs. The use of CRISPR-Cas systems would require detailed knowledge of the genomic sequences of the bacterial population, although the dwindling cost and increasing speed of high-throughput sequencing along with powerful metagenomics tools would alleviate this challenge. Overall, CRISPR-Cas systems offer a unique opportunity for the selective and titratable removal of microorganisms for industrial and medical purposes, which can be added to the ever-expanding applications of this versatile immune system (11, 20, 25, 47, 48).

MATERIALS AND METHODS

Strains and plasmid construction. See Table S2 in the supplemental ma- terial for a list of all strains used in this work. E. coli K12 strain BW25113 T7 was generated by transforming mcrE17.7-RNAp-wt-d from Y35163 to BW25113 by PI transduction. Successful transduction was verified by PCR. BW25113-T7m5.57 (Fig. 2A) was generated using three rounds of oligonucleotide-mediated recombination with mcrA-mcrB-5 spacer-cropc and the pKD4-topol plasmid encoding the A2 red recombination gene (49, 50). The oligonucleotide contained two phosphorothioate link- ages at each end to improve the recombination efficiency (51, 52). Succes- sive recombinants were verified by PCR and by sequencing.

See Table S2 in the supplemental material for a list of all plasmids used in this work. The origins of replication for the pCAGEG, pCAG-E, and pCRISPR plasmids used with E. coli and S. enterica belong to different incompatibility groups (26, 53). In the pCAG plasmid, lacking the conA/CABDE operon (pCAG E), pCAG E was digested with NotI/ NsiI, blunt ended using T4 DNA polymerase, and ligated into the pCRISPR plasmid digested with NsiI and XhoI.

To generate the pCRISPR plasmid, the pBAD18 plasmid (53) was linearized with XbaI and amplified by PCR using primers pBAD18 faithful pBAD18 rev. A chemically synthesized gBlock (IDT) was then inserted downstream of the Pmcr promoter by Gibson assembly (54). The gBlock encoded four repeats and three intervening spacers from the endogenous CRISPR locus in E. coli K-12 MG1655 (see Table S3 in the supplemental material, where the first spacer was modified to include a KpnI restriction site and an XhoI restriction site. These restriction sites allow the sequential insertion of engineered repeat-spacer pairs (see Fig. S2). Each pair was chemically synthesized as two oligonucleotides (IDT), phosphorylated with T4 polynucleotide kinase, annealed, and ligated into the pCRISPR plasmid digested with KpnI and XhoI.

The pBAD18 aaxnbelAaxnbelB cyclin was constructed in a manner similar to that for the pCAG plasmid, wherein a chemically synthe- sized gBlock (IDT) was inserted downstream of the Pmcr promoter of the linearized pBAD18 plasmid by Gibson assembly (54). The gBlock encoded the first repeat-spacer sequence from the endogenous E. coli CRISPR locus, followed by four repeats and four intervening spacers tar- geting four different locations in E. coli K1225113 (aax, mcrA, frdA, and mcrD) (see Table S2 in the supplemental material).

To generate pCRD8 (55) with engineered spacers, pCRD8 and each insert generated through PCR assembly were digested with BamHI and SacI and ligated together. To generate the insert encoding the lacZ1 spacer, template-free PCR was conducted with C1 lacZ1 fwd/C1 lacZ1 rev, followed by using the resulting product in a subsequent PCR with C1 BamHI_fwd/C1 SacI rev. To generate the inserts encoding the lacZ1 and lacZ3 spacers, first the CRISPR leader region was amplified by PCR from LMD-9 genomic DNA with C3 leader_fwd/C3 leader_rev. Next, the resulting product was used as the template in a subsequent PCR with C3 leader_fwd/lacZ2_rev or C3 leader_fwd/lacZ3 rev. Finally, each PCR product was used as the template in a final round of PCR with C3 leader_fwd/C3C4 SacI rev. All oligonucleotides and enzymes were purchased from IDT and NEB, respectively. All cloned plasmids were verified by sequencing.

Growth conditions. All E. coli and Salmonella strains were cultured in LB medium (10 g/Liter tryptone, 5 g/Liter yeast extract, and 10 g/Liter sodium chloride) at 37°C and 250 rpm with appropriate antibiotics. The same strains were plated on LB agar (LB medium with 1.5% agar) supple- mented with appropriate inactin and incubated at 37°C. S. thermophilus LMD-9 was cultured in Elliker broth (Elliker medium [Difco] supple- mented with 1% beef extract) and plated on Elliker broth (Elliker broth without the 1.5% agar). Both culturing and plating of LMD-9 were con- ducted at 37°C. Antibiotics were administered at the following final concentrations: 50 µg/ml streptomycin, 50 µg/ml kanamycin, 50 µg/ml am- plicillin, 2 µg/ml chloramphenicol, and 2 µg/ml erythromycin. Inducers were administered at the following final concentrations: 0.1 mM IPTG and 0.002% L-arabinose.
Design of native CRISPR RNAs. An overview of the approach to design and insert spacer sequences into the pCRISPR array within the pCRISPR plasmid is shown in Fig. S2 in the supplemental material. The spacers were designed to identify one of the known PAMs for the type I-E CRISPR-Cas system in E. coli (AAG, GAG, GAT, and ATG). The downstream 32 nucleotides (nt) were then used as the spacer within the engineered repeat spacer pair. Note that the last two nucleotides of the spacer are fixed as GCG because of the adopted designing strategy (see Fig. S2). However, these nucleotides fall well outside the seed region and therefore are expected to have a negligible effect on targeting.

The spacers for S. thermophilus were designed by identifying a known PAM for CRISPR (NNRNAAG) or for CRISPR (NGGNGC). The sequence of the 31 nt upstream of each was integrated into oligonucleotides that were used to generate a leader region followed by a single repeat-spacer repeat that was subsequently cloned into pCRISPR. This construct relies on processing through the native trans-activating RNase E and RNase III.

Transformation assay. Frozen stocks of S. aureus and Salmonella strains harboring pCas3 and pCasA-E (or pCasA-E') were streaked to isolation on LB agar. Individual colonies were inoculated into 3 mL of LB medium and shaken overnight at 37°C. The cultures were then back diluted into 25 mL of LB medium and grown to an A600 of 0.6 to 0.8, which was measured on a NanoDrop 2000 spectrophotometer (Thermo Scientific). The cells were then pelleted and washed with ice cold 10% glycerol 2 times before being resuspended in 150 to 350 mL of 10% glycerol. The suspended cells (50 mL) were transformed with 50 ng of pCRISPR or pCRISPR encoding the indicated spacer, using an MicroPulsar electroporator (Bio-Tad) and recovered in 300 mL of SOC medium (Quality Biological) for 1 h (E. coli) or for 2 h (Salmonella). After the recovery period, 200 mL of different dilutions of the cells were plated on LB agar with indicated antibiotic efficiency was calculated by dividing the number of transformants for the tested plasmid by the number of transformants for the original pCRISPR plasmid. To normalize for experimental variability in transformation efficiency, the same batch of cells prepared for electroporation was transformed with each tested plasmid and the pCRISPR plasmid.

A. thermophilus strain LMD-1 harboring pTRK69 was grown in 50 mL of Tiller broth and prepared for electroporation as described previously, which concentrated the culture 100-fold (57). The resuspended cells (50 mL) were transformed with 1 pg of the pCRISPR control plasmid or pCRISPR containing the indicated spacer. Transformants were collected in 950 mL of Tiller broth overnight and plated on Tiller agar. Plates were then incubated for 48 h in a Coy anaerobe chamber with a gas mixture of 10% hydrogen, 5% carbon dioxide, and 85% nitrogen before the colonies were counted. The transformation efficiency was calculated by dividing the number of transformants for the tested plasmid by the number of transformants for the pCRISPR control plasmid.

The average limit of detection of the killing assay, calculated as 1/n (no. of transformants for the control plasmid) was 7 x 10^8 for E. coli. 4 x 10^7 for Salmonella, and 2 x 10^8 for S. Thermophilus. The high transformation efficiency for Salmonella was achieved by purifying the pCRISPR plasmids, the pCas3 plasmid, and the pCasA-E plasmid individually from SB3000 (57).

Mixed culture transformation assay. The transformation assay for mixed cultures resembled that for the pure culture with a few notable differences. Cultures of E. coli K-12 and E. coli B strains harboring pCas3 and pCasA-E (or pCasA-E') were grown separately to an A600 of ~0.8, and then equal numbers of cells were mixed from the back dilution prior to preparing the culture for electroporation. An aliquot of the resuspended cell mixture (50 mL) was then transformed with the pCRISPR plasmid, pCRISPR encoding the indicated spacer, or a defined mixture of both plasmids for a total of 100 ng. The transformed cells were recovered in 300 mL of SOC medium for 90 min. After the recovery period, 200 mL of different dilutions of the cells were plated on LB agar with indicated antibiotics. The ratio of blue (E. coli R12) to white (E. coli K-12) colonies on the sample plate was divided by the same ratio on the pCRISPR plate, yielding the normalized ratio. To normalize for experimental variability in transformation efficiency, the same batch of cells prepared for electroporation was transformed with each tested plasmid mixture and the pCRISPR control plasmid.

Analysis of escape mutants. Colonies from the transformation assay with the α 63a plasmid (pCR804) were inoculated into 5 mL of LB medium with appropriate antibiotics and inducers. Growth was assessed on the plates based on the A600 after 13.5 h of growth. Cultures exhibiting measurable growth (A600 > 0.01) were stored as glycerol stocks. Plasmids were then isolated from each escape mutant, and equal amounts of DNA were resolved by agarose gel electrophoresis. Each isolated set of plasmids was also transformed into E. coli K-12 and plated on LB agar containing one of the three antibiotics. Finally, the plasmid mixture from each escape mutant was sequenced using primers that specifically bind within the FadR promoter or the double terminator of the α 63a plasmid. To analyze the protospeakers, approximately 400 bp surrounding the protospeaker within the fadA gene of the escape mutant was PCR amplified and subjected to sequencing.

Statistical analyses. All P values were calculated using the Student t test, assuming log normal distributions, two tails, and unequal variances.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mbio.00928-13/DCSupplemental.

Figures: S1, EPS file, 0.7 MB; S2, EPS file, 0.6 MB; S3, EPS file, 0.5 MB; S4, EPS file, 0.6 MB; S5, EPS file, 0.5 MB; S6, EPS file, 0.5 MB; S7, EPS file, 0.5 MB; S8, EPS file, 3.4 MB; Table S1, DOC file, 0.1 MB; Table S2, DOC file, 0.2 MB.

ACKNOWLEDGMENTS

We thank G. Seelig for initial discussions, R. H. Kelly for E. coli BL21 (DE3), D. M. Dowans for Salmonella SB3000, and U. Qamar for E. coli DB163 and plasmids pYU1397 and pWJ4000. This work was supported by university start-up funds and the Faculty Research and Professional Development program [to C.L.B.] and a graduate fellowship through the North Carolina Agriculture Foundation and through the Dannon Company, Inc. (to K.S.).

REFERENCES


Appendix V

Guide RNA modules direct Cas9 activity and orthogonality
Guide RNA Functional Modules
Direct Cas9 Activity and Orthogonality

Alexandra E. Briner,1,5 Paul D. Donohoe,3,5 Ahmed A. Gomaa,3,5,6 Kurt Selte,1 Euan M. Slorach,3 Christopher H. Nye,2 Rachel E. Haurwitz,2 Chase L. Reisel,2 Andrew P. May,1,7 and Rodolphe Barrangou1,6

1Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, NC 27695, USA
2Caribou Biosciences, Inc., Berkeley, CA 94710, USA
3Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695, USA
4Chemical Engineering Department, Faculty of Engineering, Cairo University, Giza 12613, Egypt
5Co-first authors
6Correspondence: cbisein@ncsu.edu (C.L.B), arty.may@cariboubio.com (A.P.N.), rbarrangou@ncsu.edu (R.B.)
7Co-first authors
http://dx.doi.org/10.1016/j.molcell.2014.09.019

SUMMARY

The RNA-guided Cas9 endonuclease specifically targets and cleaves DNA in a sequence-dependent manner and has been widely used for programmable genome editing. Cas9 activity is dependent on interactions with guide RNAs, and evolutionarily divergent Cas9 nucleases have been shown to work orthogonally. However, the molecular basis of selective Cas9-gRNA interactions is poorly understood. Here, we identify and characterize six conserved modules within native crRNA tracrRNA duplexes and single guide RNAs (sgRNAs) that direct Cas9 endonuclease activity. We show that the bulge and nusus are necessary for DNA cleavage and demonstrate that the nusus and hairpins are instrumental in defining orthogonality between systems. In contrast, the crRNA tracrRNA complementary region can be modified or partially removed. Collectively, our results establish guide RNA features that drive DNA targeting by Cas9 and open new design and engineering avenues for CRISPR technologies.

INTRODUCTION

Clustered regularly interspaced short palindromic repeats (CRISPR) and associated Cas proteins constitute the CRISPR-Cas system, which provides adaptive immunity against invasive genetic elements in bacteria and archaea (Eckert et al., 2007; Garneau et al., 2010; Makarova et al., 2011; Sparras, 2013). In Type II CRISPR-Cas systems, the signature RNA-guided endonuclease Cas9 specifically targets sequences complementary to CRISPR spacers and generates double-stranded DNA breaks (DSBs) using two nickase domains (Garneau et al., 2010; Gasson et al., 2012; Jinek et al., 2012; Makarova et al., 2011; Sparras et al., 2013). Any DNA sequence may be targeted, as long as it is flankered by a Cas9-specific protospacer-adjacent motif (PAM) (Garneau et al., 2010; Gasson et al., 2012; Jinek et al., 2012; Sparras et al., 2013; Sternberg et al., 2014). Targeting and cleavage by Cas9 systems rely on a RNA-duplex consisting of CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA) (Deltcheva et al., 2011). This native complex can be replaced by a synthetic single guide RNA (sgRNA) chimera that mimics the crRNA tracrRNA duplex (Jinek et al., 2012). sgRNAs in combination with Cas9 make convenient, compact, and portable sequence-specific targeting systems that are amenable to engineering and heterologous transfer into a variety of model systems of industrial and translational interest. Accordingly, the Cas9-sgRNA technology, which provides a practical means to generate DSBs, has revolutionized genome editing (Cong et al., 2013; Jiang et al., 2013; Mali et al., 2013; Sander and Joung, 2014), opened new avenues for high-throughput genome-wide genetic screens (Shalem et al., 2014; Wang et al., 2014), and expanded the toolbox for transcriptional control (Gilbert et al., 2015; Qi et al., 2013).

Furthermore, the absence of cross-interactions between evolutionarily distant Cas9 sgRNAs (Chylinski et al., 2013; Ettle et al., 2013; Fontana et al., 2014) has allowed multiple, independent targeting to be achieved within a cell when coexisting functional Type II CRISPR-Cas systems operate concurrently (Barrangou et al., 2007; Honsho et al., 2008). Despite the widespread use of these molecular machines, the critical features of sgRNAs and their involvement in defining functionally orthologous Cas9 endonucleases remain to be characterized. Indeed, early attention on Cas9 targeting and cleavage focused on spacer-target complementarity and PAM sequence sensitivity, while there remains a paucity of information defining the elements that drive Cas9-sgRNA interactions and that dictate orthogonality between Type I CRISPR-Cas systems. Therefore, we set out to identify and characterize features within sgRNAs that impart Cas9 targeting and cleavage specificity to open new engineering avenues for CRISPR technologies.

RESULTS

Identification of sgRNA Functional Modules

Recent structural studies have provided insights into the Streptococcus pyogenes Cas9 (SpCas9) interaction with a sgRNA and complementary strand target DNA (Jinek et al., 2012; Nishimasu et al., 2014). Distinct portions of the sgRNA are predicted to form various features that interact with Cas9 and/or the DNA target. However, the boundaries and respective roles of these
portions remain to be determined. To address this, we generated a series of sgRNA variants in which we modified portions of the sequence and/or structure (Table S1 available online. Supplemental Experimental Procedures). We then assessed the ability of the sgRNA variants to support Cas9 cleavage of an AV51 DNA sequence both biochemically, using a double-stranded DNA cleavage assay, and in human HEK293 cells (Supplemental Experimental Procedures) using a T7E1 assay. sgRNA variants that supported Cas9-dependent cleavage biochemically were generally also active in cells, although cell-based activity was not measurable for some variants exhibiting weak cleavage activity biochemically (Figures 1A and S1). Similar data were obtained from a subset of sgRNA variants for a second spacer targeting a sequence from the VEGFA gene (Figure S1, Table S1). Results established six distinct structural modules that constitute sgRNAs: the spacer, the lower stem, the upper stem, the bulge, the nexus, and the hairpins (Figure 1A). It has been established previously that the spacer sequence determines the location of Cas9 endonuclease cleavage (Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012; Sapanovas et al., 2011). Therefore, we focused on the five sgRNA modules located downstream of the spacer sequence, and tested 77 unique sgRNA variants with modified sequences (Table S1). Mutations in both the lower and upper stems were relatively tolerant to sequence variations, including nucleotide substitutions, insertions and deletions and even predicted structural disruptions (Figure 1, variants v1 through v6 and Figure S1). GVs 1-5, 10-12, 26, 33, 34). Abrogation of cleavage in cells required incorpora-

tion of two concurrent mismatches within the lower stem (Figure 1, variant v4; Figure S1, variant G24). Remarkably, complete removal of the hairpins was required to fully abolish cleavage in cell-based assays (Figures 1, variant v20, and S1; GVs 63–66). In contrast, the bulge was less tolerant to mutation, as alteration was functionally disruptive; a sgRNA variant with a perfectly base-paired duplex sequence at the site of the bulge is inactive (Figure 1, variant v6), as are those in which purine nucleotides are swapped (A-G, S-A) (Figure S1). Interestingly, however, a sgRNA variant in which the upper stem is removed or the bulge is replaced by a tetranucleotide retained activity even in HEK293 cells (Figure 1, variants v11 and v12), suggesting that the upper stem is dispensable. In virtually all Type II crRNA sequences, the nucleotide immediately 3′ of the spacer sequence at the base of the lower stem is a conserved guanine nucleotide. This nucleotide pairs with a similarly conserved uracil nucleotide at the base of the lower stem (Nishimura et al., 2011) but can be mutated without loss of function (Figure 1, variants v5 and v6). This suggests that the sequence conservation at this location is not critical for Cas9 nucleosome activity and is perhaps required for another function of the CRISPR repeat, such as spacer acquisition.

Immediately downstream of the lower stem is an additional stem loop that we have named the nexus; this feature exhibits sequence and structural features important for cleavage (Figures 1, variants v13 through v18, and S1; Table S1). The first two adenine nucleotides are highly conserved among tracrRNAs of Type II systems, and introduction of additional nucleotides between the two adenine nucleotides and the stem loop reduced activity (Figure 1, variant v13). The base of the stem loop is formed by a pair of guanine nucleotides (G53 and G54) base paired to two cytosine nucleotides (C80 and C81). Substitution of G53 and G54 for cytosines abrogated activity (Figure 1, variant v14). However, substitution of G-C pairs with A-U pairs is
tolerated at both positions, and G-U and G-A pairings at the base of the nuclease are tolerated in biochemical assays (Fig. S1; Table S1; GV-61 and GV-62), suggesting that there is some conformational flexibility at the nuclease interaction site. Conversely, increases in the length of the nuclease significantly reduced biochemical activity and abolished cell-based activity (Figures 1, variant v17, and S1; GV 68–71, 73, and 74). Downstream of the nuclease stem loop, substitutions along the length of the extended region between the nuclease and the first harpin (Harpin1) are tolerated (Figure S1; GV75). Harpin1 can be removed from the sgRNA backbone as long as it is replaced by two nucleotides, presumably to enable appropriate spacing between the nuclease and harpin 2 (Figure S1; GV 79–81). Based on the sgRNA variants tested, we speculate that differences between observed biochemical and cell-based cleavage activity are due to RNA loding, stability, complex formation, and/or stoichiometry differences in live cells.

These results are consistent with recent structural and biochemical data on the mechanism of DNA recognition and cleavage by Cas9 (Jinek et al., 2012; 2014; Nishimura et al., 2014; Anders et al., 2016). Notably, the upper and lower stem interact with Cas9 mainly through sequence-independent interactions with the phosphate backbone (Jinek et al., 2014; Nishimura et al., 2014). In contrast, the bulge participates in specific side-chain interactions with the Rec1 domain. The bulgebase of U44 interacts with the side chains of Tyr 325 and His 326, while G43 interacts with Tyr 329. The nuclease forms the core of the sgRNA:Cas9 interactions and lies at the intersection between the sgRNA and both Cas9 and the target DNA. The nucleobases of AS1 and AS2 interact with the side chain of Phe 1105; U56 with Arg 457 and Asn 459; the nucleobase of U59 inserts into a hydrophobic pocket defined by side chains of Arg 74, Asn 77, Pro 475, Leu 455, Phe 446, and Ile 448; C60 interacts with Leu 455, Asa 456, and Asn 459, and C61 interacts with the side chain of Arg 70, which in turn interacts with C15.

Overall, our results demonstrate that discrete sequences and structures such as the bulge and nuclease are critical for sgRNA composition.

Conservation of Modules in Type II-A CRISPR-Cas Systems

These findings establish important modules in sgRNA that are required to support SpyCas9 activity. However, while used widely for genome editing, SpyCas9 is merely one of many Cas9 orthologs found naturally (Chylinski et al., 2013; Fontana et al., 2014). We therefore investigated whether the same sgRNA sequence features also occur in other CRISPR-Cas systems, focusing on Type II-A systems. We sampled 41 Cas9 sequences from Streptococcus and Lactobacillus genomes, in which Type II systems preferentially occur (Makarova et al., 2011), and identified their corresponding CRISPR repeat and predicted tracrRNA sequences. The Cas9 protein sequences clustered into three main sequence groups (Figure S3). Similar grouping was observed when clustering was carried out using either CRISPR repeat or predicted tracrRNA sequences (Figures 2A and S2), as anticipated, given the presence of an anti-CRISPR repeat within the tracrRNA and the intimate molecular relationship between Cas9 and crRNA:tracrRNA pairs (Deltcheva et al., 2011; Fontana et al., 2014; Makarova et al., 2011).

Within the tracrRNA sequences, we consistently observed the functional modules identified for SpyCas9 (Figure 2B), with conservation of the overall sgRNA/crRNA:tracrRNA structure between families, and high levels of sequence conservation within clusters. The presence of a bulge with a directional kink between the lower stem and the upper stem was observed consistently across a diversity of systems (Figure 2B). The length of the lower stem was highly conserved within and variable between families. Interestingly, the highest level of conservation was observed for the nuclease sequences (Figures 2B and S2). The general nuclease shape with a GC-rich stem and an offset uracil was shared between the two Streptococcus families. In contrast, the di-synthetic double stem nuclease (Figure 2B) was unique to, and ubiquitous in, Lactobacillus systems. Remarkably, some bases within the nuclease were strictly conserved even between distinct families (Figure S2), including AS2 and CS5, further highlighting the critical role of this module. Based on the SpyCas9 crystal structure (Nishimura et al., 2014), AS2 interacts with the backbone of residues 1,103–1,107 close to the 3′ end of the target strand, suggesting that the interaction of the nuclease with the protein backbone may be required for PAM binding. Indeed, the structure of SpyCas9 with target DNA shows that this region of the protein directly interacts with the PAM in duplex form and suggests a role for the nuclease in presenting this region of the protein in an appropriate conformation to engage double-stranded DNA (Anders et al., 2014).

Crossing Cas9:sgRNA Orthogonality Boundaries

These findings suggest a potential relationship between the structure and sequence of the sgRNA and the diversity of Cas9 proteins. This prompted us to determine the sgRNA modules that define Cas9 groups that are orthogonal—groups between which crRNAs are not cross-functional (Barrangou et al., 2007; Horvath et al., 2008; Esseltine et al., 2013; Fontana et al., 2014). In particular, we selected the endonucleases from the two naturally coexisting orthogonal S. thermophilus Type II-A systems, Sh10Cas9 and Sh30Cas9 (Horvath et al., 2009), to investigate the link between sgRNA composition and Cas9 orthogonality.

We designed a series of experiments based on self-targeting in Escherichia coli (see Supplemental Information; Figure 3B) to test whether specific mutations in a sgRNA could facilitate DNA targeting with a previously orthologous Cas9. We identified a region within the E. coli genome that contained overlapping Cas9 target sites for the Sh10Cas9 and Sh30Cas9 systems to minimize any quantitative differences between targeting sites (Figure 3B). We generated chimeric versions of the two sgRNA backbones and interchanged the spacer, lower stem–bulge, upper stem, nuclease (Figure 3C), and tested their ability to drive self-targeting (Gouna et al., 2014) by either Sh10Cas9 or Sh30Cas9. First, we confirmed that each sgRNA drives targeting only with its cognate Cas9 (Figure 3C, top panel). Next, we swapped the spacer sequences to match the PAM with the other Cas9. This swap was insufficient to confer activity with the non-cognate Cas9, demonstrating the importance of other modules. Interestingly, the CRISPR1 sgRNA with the CRISPR3 spacer still conferred activity with the Sh30Cas9 (Figure 3C, top-left panel), suggesting flexibility in the spacing requirement between the
PAM and the protospacer for this Cas9, supporting previous observations (Chan et al., 2014). We then swapped the nuclease and hairpins between the sgRNAs. Remarkably, this swap resulted in sgRNAs that guided the orthologous Cas9 but not the cognate Cas9 (Figure 3C, lower panel), demonstrating that the orthogonality barrier between these proteins had been crossed. Swapping individual portions of the nuclease or hairpins was insufficient to guide the opposite Cas9 (Figure 3C, lower panel), suggesting that both modules are important for Cas9 specificity. The importance of the nuclease and hairpin modules aligns with the cell-based cleavage results using SpyCas9 (Figures 1 and S1) and is consistent with sequence and structural differences observed between the nuclease and hairpin modules in the CRISPR1 and CRISPR3 systems. In addition, the significance of the nuclease and hairpin modules, as opposed to the flexibility of the crRNA-tracrRNA complementary sequences, challenges the canonical view that the CRISPR-repeat sequence plays a key role in defining orthogonal CRISPR-Cas systems (Torzilari et al., 2014; Esvelt et al., 2013; Chylinski et al., 2013). Altogether, these results show that the tracrRNA nuclease and hairpins are critical for Cas9 pairing and can be swapped to cross orthogonality barriers separating disparate Cas9 proteins (Esvelt et al., 2013), which is
instrumental for further harnessing of orthogonal Cas9 proteins associated with different PAMs.

**DISCUSSION**

Recent structural and biochemical data have begun to shed light on the mechanism of DNA recognition and cleavage by Cas9 (Jinek et al., 2012, 2014; Nishimasu et al., 2014; Anders et al., 2014). Electron micrographs of the apo-, RNA-bound, and protein/RNA/DNA complexes indicated that upon binding guide RNA, Cas9 undergoes a dramatic conformational change to facilitate target DNA binding and cleavage (Jinek et al., 2014). Crystal structures show that, consistent with images from the electron microscope, the SpyCas9-sgRNA/DNA complex and apo-SpyCas9 occupy significantly different conformations, with substantial rearrangement of RNA- and DNA-binding domains taking place between the two structures (Jinek et al., 2014; Nishimasu et al., 2014). The nexus occupies a critical position in the SpyCas9-sgRNA/DNA complex, coordinating a number of key components of the protein and sgRNA, positioning both protein and RNA appropriately to receive target DNA duplexes for cleavage. Upon binding sgRNA/DNA, the arginine-rich bridge helix binds to the base of the nexus and to the lower stem. Additionally, the nexus interacts with two small regions (which we propose to establish as Nexus Interacting Region 1 [NIR1] 446–497 and Nexus Interacting Region 2 [NIR2] 1,105–1,138) from the two loops of SpyCas9. Both of these regions are disordered in the apoSpyCas9 structure and notably contain two tryptophan residues identified as being important in PAM recognition (Jinek et al., 2014). NIR2 also interacts directly with the lower stem, and the face opposite the nexus-binding site lies in close proximity to the 3’ end of the target stand, suggesting that interaction with the nexus may be required to order the PAM recognition site. Notably, in the Actinomycetes naeslundii Cas9 (AnaCas9) apo-structure (Jinek et al., 2014), NIR2 is ordered and contains a ~30 aa insertion. Because NIR2 is positioned to interact with the PAM duplex in the SpyCas9-sgRNA/DNA complex structure, it is tempting to speculate that AnaCas9 may recognize a larger nexus and potentially an extended accompanying PAM sequence.

Altogether, these results reveal six distinct features within guide RNAs and establish the bulge and nexus as the most critical features for Cas9 targeting. This challenges the canonical view that the crRNA CRISPR repeat/tracrRNA anti-repeat is the driver of Cas9 functionality and orthogonality (Eswell et al., 2013; Fontana et al., 2014). This also provides a basis for optimization of sgRNA composition and design. For instance, the dispensability of the upper stem, hairpin 1, and the sequence flexibility of the lower stem suggest the design of guide RNAs that are more compact and conformationally stable than traditional sgRNAs. This may open new opportunities to append RNA features to the CRISPR guides. Furthermore, the ability to dictate Cas9 orthogonality using chimeric sgRNAs with altered nexus and hairpin sequences opens new avenues for the exploitation of novel Cas9 proteins, with the potential to harness the diversity of natural Cas9 orthologs, including short Cas9 variants for convenient packaging and delivery. This also expands multiplexing opportunities by using a single Cas9 with various chimeric guides or by concurrently using orthogonal systems with different combinations of standard or chimeric sgRNAs. Collectively, our findings open up new avenues for Cas9-dependent DNA targeting and set the stage for the development of next-generation CRISPR-based technologies.
Molecular Cell
Features of Cas9 Guide RNAs

EXPERIMENTAL PROCEDURES

sgRNA Engineering and DNA Cleavage
GguiR RNAs were produced by in vitro transcription from plasmid template incorporating a T7 promoter, with templates assembled by PCR using internal assembly oligonucleotides containing the specific variant sequences and universal primer sequences corresponding to the T7 promoter (forward) and the 5’ end of the guideRNA (reverse). Double-stranded DNA target regions for biotinylated assays were amplified by PCR from HOK93 (ATCC) genomic DNA (gDNA) isolated using QuickExtract (Epicenter). Cas9 was produced by the MacLab at UC Berkeley as previously described (Jinek et al., 2012). Each sgRNA-Cas9 combination was incubated at 37°C for 10 min, and cleavage reactions were initiated by the addition of target DNA to a final concentration of 12.5 nM before incubation for 15 min at 37°C. The appearance of DNA bands at 300 bp and 1400 bp (indicated cleavage).

Percentage cleavage was calculated using the program FIJ (v. 0.6.0.0-β; http://fiji.sc/FIJ) by measuring the ratio between the sum of area under peaks for cleavage bands and the sum of area under both the parental and cleavage bands.

C9-1 expressing cells (HOK93 sgCas9) were generated by transfecting HOK93 cells with a linearized plasmid containing a Cas9-GFP fusion gene under the control of the CMV promoter and a neomycin resistance gene. gDNA was isolated from HOK93 sgCas9 control cells without guide RNA transcription. DNA for T7F assays was generated by PCR amplification of the target Avsf1 locus. Further details about material and procedures used for guide RNA generation, DNA template generation, in vitro transcription, target dsDNA generation, Cas9 protein production, DNA cleavage assay, cell culture, and cell line generation, cell transfection, target dsDNA generation, and T7F assay are available in the Supplemental Experimental Procedures section.

Orthogonal sgRNA/Cas9 System Engineering
The sgRNA from the CRISPR locus and the CRISPR locus were PCR amplified from genomic DNA of S. thermophilis MBL-2 and cloned into pC9-Cas9 (Addgene 44259; Doudna et al., 2013). To construct the sgRNA-expressing plasmid, the Spel restriction site in the pC9-Cas9 plasmid (Addgene 44259; Doudna et al., 2013) was removed and a B(3)EII (BstY) polymerase and T7F (T7F) cleavage to guide RNA transcription. DNA for T7F assays was generated by PCR amplification of the target Avsf1 locus, and transformation efficiency was calculated by dividing the number of transformants for the target sgRNA-Cas9 control plasmid, as described previously (Gomaa et al., 2014). Additional details about material and procedures used for plasmid construction, strain and growth conditions, and bacterial transformation assays are available in the Supplemental Experimental Procedures section, as well as in Table S5C.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, two tables, and the Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2013.04.018.

AUTHOR CONTRIBUTIONS

A.E.B., P.D.O., A.A.G., A.P.M., C.L.B., and R.B. conceived and designed the study. A.E.B. and K.M.S. performed computational analyses. P.D.O., E.M.S., C.H.L., and A.A.G. performed experiments. R.E.H., C.L.B., A.P.M., and R.B. analyzed data and wrote the manuscript, with contributing input from all authors. A.P.M., P.D.O., E.M.S., C.H.L., and R.E.H. declare competing financial interests as either employees or directors at and stockholders of Carlsberg Bioclinics, Inc. Patent applications have been filed associated with this work.

ACKNOWLEDGMENTS

We thank all members of Carlsberg Bioclinics, Inc. and of the CRISPR laboratory at NC State for constructive insights. We specifically thank Rebecca Stotz for technical assistance with cloning. This work was supported by startup funds from North Carolina State University, a grant from the National Science Foundation (CAREER-1400135) to C.L.B. and R.B., and a grant from the Kaiser Institute of Engineering, Technology & Science to C.L.B. We thank Jennifer Doudna and Martin Jinek for critical reading of the manuscript and helpful discussions. The authors would like to thank their many colleagues and collaborators in the CRISPR field for their insights into these fascinating molecular systems.

REFERENCES

Molecular Cell
Features of Cas9 Guide RNAs


Appendix VI

Occurrence and activity of a type CRISPR-Cas system in *Lactobacillus gasseri*
Occurrence and activity of a type II CRISPR-Cas system in Lactobacillus gasseri

Rosemary Sanzok-Dawes,† Kurt Selle,† Sarah O‘Flaherty,† Todd Klaenhammer,‡ and Rodolphe Barrangou‡

1Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, NC, USA
2Functional Genomics Program, North Carolina State University, Raleigh, NC, USA

Bacteria encode clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated genes (cas), which collectively form an RNA-guided adaptive immune system against invasive genetic elements. In silico surveys have revealed that lactic acid bacteria harbour prolific and diverse set of CRISPR-Cas systems. Thus, the natural evolutionary role of CRISPR-Cas systems may be investigated in these ecologically, industrially, scientifically and medically important microbes. In this study, 17 Lactobacillus gasseri strains were investigated and 6 harboured a type II-A CRISPR-Cas system, with considerable diversity in array size and spacer content. Several of the spacers showed similarity to plasmid and plasmid sequences, which are typical targets of CRISPR-Cas immune systems. Aligning the protospacers facilitated inference of the protospacer adjacent motif sequence, determined to be 5’-NTA-3’ flanking the 3’ end of the protospacer. The system in L. gasseri JV W90 and NCD 1342 interfered with transforming plasmids containing sequences matching the most recently acquired CRISPR spacers in each strain. We report the distribution and function of a native type II-A CRISPR-Cas system in the commensal species L. gasseri. Collectively, these results open avenues for applications for bacteriophage protection and genome modification in L. gasseri, and contribute to the fundamental understanding of CRISPR-Cas systems in bacteria.

INTRODUCTION

Bacteria encode clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated genes (cas), which collectively form an RNA-guided adaptive immune system against invasive genetic elements (Barrangou et al., 2007). CRISPR-Cas systems are highly prevalent, appearing in approximately 46% of bacterial and 84% of archaical genomes (Grissa et al., 2007). CRISPR-Cas-mediated immunity hinges upon the distinct molecular processes of acquisition, expression and interference (Barrangou & Marraffini, 2014). Acquisition occurs via molecular ‘sampling’ of foreign DNA, from which short sequences termed spacers are integrated in a polarized manner at the leader end of the CRISPR array.

These authors contributed equally to this work.

Abbreviations: CRISPR, clustered regularly interspaced short palindromic repeat; crRNA, CRISPR RNA; PAM, protospacer adjacent motif; Rep-PCR, repetitive-element PCR; tracrRNA, trans-activating CRISPR RNA.

Two supplementary tables are available with the online Supplementary Material.

(Barrangou et al., 2007). CRISPR arrays are transcribed constitutively and indirectly, directed by promoter elements in the preceding leader sequence during expression (Brouns et al., 2008; Young et al., 2012). The transcript is processed selectively at each repeat sequence, forming small interfering CRISPR RNAs (crRNAs) that function to guide Cas proteins. Interference is carried out through sequence-specific recognition and cleavage of target nucleic acid complementary to the spacer. CRISPR-Cas systems contain universal cas1 and cas2 genes, and are categorized as type I, type II or type III based on signature genes, namely cas9, cas9 and cas10, contributing to the distinct mechanisms by which each system confers immunity (Makarova et al., 2011).

Despite the high prevalence of CRISPR-Cas in bacteria, relatively few systems have been experimentally characterized for their functional activity and ecological role as adaptive immune systems targeting invasive genetic elements in vivo (Hondy-Denomy & Davidsohn, 2014). Intensive study of Streptococcus thermophilus has provided valuable insights into how CRISPR-Cas systems shape the dynamic interplay between plages and their bacterial hosts (Barrangou et al., 2007).
Levin et al., 2013; Paco-Espino et al., 2013; Sun et al., 2013). Little is known about the mechanisms or evolutionary impact of CRISPR-Cas loci outside of a few systems, which limits repurposing of endogenous systems for genome editing or transcriptional control in their native backgrounds (Liu et al., 2015; Sestra & Charpentier, 2014; Selle & Barrangou, 2015). It is essential to characterize systems from diverse bacteria to expand the ecological understanding of CRISPR-Cas and facilitate development of new CRISPR-Cas-based genetic tools derived from orthogonal systems. In silico surveys have revealed that type II systems are disproportionately present in lactic acid bacteria (Chylinski et al., 2014; Horvath et al., 2009), making them a reservoir for orthologous CRISPR-Cas systems. Hallmark features of type II systems are the signature DNA endonuclease Cas9 (Garneau et al., 2010; Supranansayakul et al., 2011), the trans-activating CRISPR RNA (tracrRNA) (Delcheva et al., 2011) and maturation of crRNAs by RNAse III (Chylinski et al., 2014). Cas9 targets DNA for double-stranded cleavage, which occurs through dual HNH and RuvC nuclease domains (Garneau et al., 2010; Gasiunas et al., 2012; Sサプリ華縛等も、2011; Jinek et al., 2012).

*Lactobacillus gasseri* is a commensal lactic acid bacterium frequently isolated from human mucosal tissues (Azcarate-Peril et al., 2008). It is considered an autotrophic micro-organism, colonizing niches such as the oral cavity, vagina and gastrointestinal tract in healthy individuals (Delgado et al., 2007, Rodrigues da Cunha et al., 2012). *L. gasseri* has a long history of safe human consumption and experimentally substantiated health benefits, leading to use of well-characterized strains as probiotics (Selle & Klaenhammer, 2013). *L. gasseri* strains exhibit high intra-species diversity (Azcarate-Peril et al., 2008), in contrast to

<table>
<thead>
<tr>
<th>Table 1. Strains and plasmids used in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
</tr>
<tr>
<td>L. gasseri</td>
</tr>
<tr>
<td>NCK 99 (ATCC 199907)</td>
</tr>
<tr>
<td>NCK 334 (ATCC 33123)</td>
</tr>
<tr>
<td>NCK 1340 (AM1)</td>
</tr>
<tr>
<td>NCK 1341 (TR12)</td>
</tr>
<tr>
<td>NCK 1342 (SH10)</td>
</tr>
<tr>
<td>NCK 1344 (NR2)</td>
</tr>
<tr>
<td>NCK 1348 (RF14)</td>
</tr>
<tr>
<td>NCK 1349 (RF81)</td>
</tr>
<tr>
<td>NCK 1557</td>
</tr>
<tr>
<td>NCK 2140</td>
</tr>
<tr>
<td>NCK 2141</td>
</tr>
<tr>
<td>CRISPR-Cas-containing <em>L. gasseri</em> strains</td>
</tr>
<tr>
<td>NCK 1336 (WD19)</td>
</tr>
<tr>
<td>NCK 1343 (SH15)</td>
</tr>
<tr>
<td>NCK 1344 (FR4)</td>
</tr>
<tr>
<td>NCK 1346 (MH1)</td>
</tr>
<tr>
<td>NCK 1347 (ML3)</td>
</tr>
<tr>
<td>J5/V03</td>
</tr>
<tr>
<td>E. coli</td>
</tr>
<tr>
<td>MC1061</td>
</tr>
<tr>
<td>NCK 2346</td>
</tr>
<tr>
<td>NCK 2347</td>
</tr>
<tr>
<td>NCK 2348</td>
</tr>
<tr>
<td>NCK 2349</td>
</tr>
<tr>
<td>NCK 2350</td>
</tr>
<tr>
<td>Plasmids</td>
</tr>
<tr>
<td>pGK12</td>
</tr>
<tr>
<td>pTRK 1090</td>
</tr>
<tr>
<td>pTRK 1091</td>
</tr>
<tr>
<td>pTRK 1092</td>
</tr>
<tr>
<td>pTRK 1093</td>
</tr>
<tr>
<td>pTRK 1094</td>
</tr>
</tbody>
</table>

ATCC, American Type Culture Collection.
Table 2. Spacer sequences and BLAST matches

<table>
<thead>
<tr>
<th>Strain</th>
<th>Spacer no.</th>
<th>Spacer/protospacer match</th>
<th>3’ Flank</th>
<th>Identity/accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCK 1346</td>
<td>10</td>
<td>AGGAAAGTTCTACAGTTGATACGG</td>
<td>GGTAAGA</td>
<td>Phi ADH phage A11319.1</td>
</tr>
<tr>
<td>NCK 1342</td>
<td>11</td>
<td>AGGAAAGTTCTACAGTTGATACGG</td>
<td>GGTAAGA</td>
<td>Prophage LMP 1865</td>
</tr>
<tr>
<td>NCK 1342</td>
<td>8</td>
<td>GGAATTCCTACAGTTGATACGG</td>
<td>GGTAAGA</td>
<td>P1 Phage KCS 4X</td>
</tr>
<tr>
<td>IV-V03</td>
<td>2</td>
<td>GGAATTCCTACAGTTGATACGG</td>
<td>GGTAAGA</td>
<td>P1 Phage KCS 4X</td>
</tr>
<tr>
<td>IV-V03</td>
<td>12</td>
<td>GGAATTCCTACAGTTGATACGG</td>
<td>GGTAAGA</td>
<td>P1 Phage KCS 4X</td>
</tr>
<tr>
<td>IV-V03</td>
<td>18</td>
<td>GGAATTCCTACAGTTGATACGG</td>
<td>GGTAAGA</td>
<td>P1 Phage KCS 4X</td>
</tr>
<tr>
<td>K7</td>
<td>6</td>
<td>GGAATTCCTACAGTTGATACGG</td>
<td>GGTAAGA</td>
<td>P1 Phage KCS 4X</td>
</tr>
<tr>
<td>NCK 1345</td>
<td>29</td>
<td>GGAATTCCTACAGTTGATACGG</td>
<td>GGTAAGA</td>
<td>P1 Phage KCS 4X</td>
</tr>
<tr>
<td>NCK 1345</td>
<td>21</td>
<td>GGAATTCCTACAGTTGATACGG</td>
<td>GGTAAGA</td>
<td>P1 Phage KCS 4X</td>
</tr>
<tr>
<td>NCK 1347</td>
<td>17</td>
<td>GGAATTCCTACAGTTGATACGG</td>
<td>GGTAAGA</td>
<td>P1 Phage KCS 4X</td>
</tr>
</tbody>
</table>

to the closely related and highly conserved *Lactobacillus acidophilus* species. The evolutionary interaction of bacteriophages and lactic acid bacteria is clearly highlighted by the multitude of phage-resistance mechanisms employed by these bacteria, including CRISPR-Cas systems (Barrangou et al., 2007; Horvath et al., 2009; Coste & Ross, 2002). Indeed, various temperate bacteriophages specific to *L. gasseri* have been sequenced or characterized (Raya et al., 1989; Ismail et al., 2009; Baugh et al., 2014). We sought to survey and investigate one type II system in *L. gasseri* for its potential activity, based on key characteristics alluding to its functionality. The system exhibited maintenance of CRISPR arrays, diversity of spacer content, spacers targeting invasive genetic elements and intra-species/system synthesis of Cas operon (Bowory-Denovoy & Davidovs, 2014). In this study, the distribution of a type II CRISPR-Cas system amongst *L. gasseri* strains was determined, and its functionality vivo was demonstrated in *L. gasseri IV-V03 and NCK 1342.

METHODS

**Bacterial strains and growth conditions.** The bacterial strains and plasmids used in this study are shown in Table 1. *Escherichia coli* MC1061 was grown aerobically, in LB broth (Difco) at 37°C. *E. coli* MC1061 transformants were selected on brain heart infusion agar (1.5%, w/v; Difco) with 150 μg erythromycin (Erm) ml^{-1}. The *L. gasseri* strains were propagated statically in MRS de Man, Rogosa and Sharpe broth (Difco) in loosely capped tubes at 37°C anaerobic conditions and 37°C, and transformants were selected in the presence of 5 μg Erm ml^{-1} (Fisher Scientific).

**Repetitive-element PCR (Rep-PCR) strain genotyping.** *L. gasseri* strains were grown in MRS, and genomic DNA extracted using an Ultraclean microbial DNA isolation kit (MoBio) and quantified using a Nanodrop 1000 spectrophotometer (Thermo Scientific). After normalization to 20 ng μl^{-1} with UltraPure distilled water (Invitrogen), Twenty-three microinlurities Master Mix from the Rep-PCR Lactobacillus based Diversity Kit (BioMérieux) was added to a tube with 2 μl of genomic DNA. DNA amplification was performed in a MyCycler thermal cycler (Bio-Rad), programmed for 2 min at 94°C (initial denaturation), 35 cycles of 70 s at 98°C (denaturation), 90 s at 70°C (annealing) and 90 s at 70°C (extension), and 3 min at 70°C (final extension), using AmpliTaq DNA polymerase from Applied Biosystems. The reaction product was then added to the DiversityLab DNA reagents and supplies (BioMérieux), according to the manufacturer’s protocol. The chip samples were analysed using DiversityLab software version 3.1 and similarity of strains was determined by comparing the resultant electropherograms/barcodes.

**CRISPR strain genotyping.** The CRISPR database, CRISPRdb (Grissa et al., 2007a), and CRISPRFinder (Grissa et al., 2007b) were used to identify a putative CRISPR locus in the draft genome of *L. gasseri* IV-V03. (GenBank accession no. AGCQ00000000). *L. gasseri IV-V03* was obtained through BEI Resources (Biodefense and Emerging Infections Research Resources). National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA, as part of the Human Microbiome Project. *L. gasseri* EM 104. Planners were designed to amplify the CRISPR-Cas locus from the draft genome sequence of *L. gasseri IV-V03*. The National Center for Biotechnology Information (NCBI) Basic Local Alignment Sequence Tool (BLAST) (Altschul et al., 1990) was used to determine *L. gasseri* spacer matches (Table 2).
DNA manipulations. All L. gasseri genomic DNA was extracted using an UltraClean microbial DNA isolation kit (Mo Bio). Plasmid DNA from E. coli was obtained using a QIAprep Spin miniprep kit (Qiagen). PCR primers and interference inserts (Table S1, available in the online Supplementary Material) were synthesized by Integrated DNA Technologies. PCR amplifications for cloning and screening were generated using standard protocols and Choice-Taq blue DNA polymerase (Denville Scientific). PCR products were analyzed in 1% agarose gel using 1 kb Plus ladder (Invitrogen) as the molecular mass ladder and purified using a QIAquick gel extraction kit (Qiagen). DNA sequencing was performed by Dovis Sequencing and Eton Bioscience.

Construction of interference plasmids. Interference plasmids were constructed by ligation using T4 DNA ligase (New England BioLabs) and protospacer/protospacer adjacent motif (PAM) inserts (Table S1) with EcoRI-digested (Promega) pRK22. The constructs were transformed into rubidium chloride competent E. coli MC1061 (Hartman, 1985). The resulting interference plasmids were isolated from E. coli transformants. PCR screened for the presence of the insert and sequenced across the multiple cloning sites to confirm insert sequence accuracy prior to transformation. Transformation of L. gasseri NGX 1342, and IV-VO with the interference plasmids was performed as described previously (Walker et al., 1986).

RESULTS

Rep-PCR genotyping of L. gasseri strains
Genotyping by Rep-PCR was performed to assess the diversity of a set of 17 L. gasseri strains. Rep-PCR hinges on the...
random distribution of genus-specific repetitive elements in bacterial genomes (Healy et al., 2005). PCR amplification was performed using primers specific to the repetitive element, resulting in non-specific amplification of intervening sequences. Analysis of the Rep-PCR fingerprint or barcode using microfluidics allows for differentiation of strains at high resolution. Considerable genomic diversity was observed within the *L. gasseri* species (Fig. 1).

**Distribution and spacer content of CRISPR-Cas systems in *L. gasseri***

A putative type II-A system was previously identified in *L. gasseri* JV-V03, based on a highly conserved 36 nt repeat, the presence of the universal genes *cas1* and *cas2*, the signature gene *cas9*, a putative tracrRNA and *cas10* (Fig. 2) (Chytilska et al., 2013; Briner et al., 2014). Since CRISPR array/Cas operon maintenance is an indication of function, 16 *L. gasseri* strains (excluding JV-V03) were surveyed to determine the distribution and conservation of the type II-A system present in *L. gasseri* JV-V03. Genomic DNA from each strain was assessed for the signature *cas9* gene using internal primers (in1cas9, in2cas9) and for CRISPR arrays using primers designed to flank the loci (Hgfl F2, Hgfl R; core 1, core 4 and core 6 F and R) (Table S1). Of the 17 isolates, type II-A CRISPR-Cas systems were identified in 6 strains by PCR analysis, with amplicons ranging from 1.2 to 2.2 kb (Fig. 3). CRISPR arrays were variable in the number of spacers, ranging from 11 to 30 (Fig. 4). Each of the array amplicons was sequenced to determine the spacer content of the strains, which provided a record of acquisition events that are highly ordinal and sequence-specific. Bacteria inhabiting the gastrointestinal and genitourinary tracts are exposed to many foreign DNA elements, mainly bacteriophages and plasmids (Breitbart et al., 2008; Minot et al., 2011). In an attempt to specifically identify and document exposure to foreign DNA, we matched spacer sequence to known plasmid and bacteriophage protectors from the NCBI database using BLAST. The results in Table 2 show commonalities in environmental exposures and presumed acquisition of adaptive immunity in their evolutionary history. Of the 83 unique spacer sequences analysed from *L. gasseri* arrays, only 12 revealed BLAST hits with high identity (>85%) to invasive genetic elements. Alignment of flanking sequences at the 3' end of the high fidelity protospacer matches facilitated inference of the PAM sequence, which was hypothesized to be 5'-NTAA-3' (Table 2). The PAM includes at least one non-specific nucleotide at the 3' end of the protospacer, is 4 bp long and is purine-rich. The motif is consistent with previously characterized PAMs, although the presence of the thymine nucleotide is a rarity (Fowell et al., 2013). However, the absence of any guanosines correlates with the low G+C content of *L. gasseri* phage genomes (Baugh et al., 2014). The repeat sequences within each array were highly conserved, although some repeat degeneracy was observed (Table S2).

Comparative analysis of CRISPR spacer composition is a powerful tool for determining relatedness of strains through shared spacers, while allowing for differentiation based on unique spacers (Horvath et al., 2009; Barrangou & Horvath, 2012; Sharanat et al., 2013; Briner & Barrangou, 2014). Analysis of the spacer content for the type II-A system in *L. gasseri* revealed considerable diversity (Fig. 4), which was true even for the most ancestral spacers, as only two pairs of strains exhibited any shared spacers throughout the CRISPR array. The lack of shared spacers suggests considerable genetic diversity of *L. gasseri* strains, which mirrors the diversity observed using the Rep-PCR method, and from FGE genotyping studies in *L. gasseri* (Crowell, 1998). Two strains, NCK 1345 and 1347, appeared to be identical based on Rep-PCR, but differences in spacer content in NCK 1345 and 1347 were apparent. The differences were due to the internal deletion of several 30 bp spacers, offering a distinction not detectable in Rep-PCR, suggesting that comparative CRISPR array analysis can be used to differentiate the two strains (Fig. 4).

**Type II-A system in *L. gasseri* prevents plasmid uptake**

The presence of cas genes and all the necessary genetic elements, along with diverse repeat-spacer arrays, implies that these loci are active, but interference assays are essential for directly testing whether systems are capable of
targeting DNA elements. In this study, we employed a plasmid interference assay, which relies on native expression of the endogenous tracrRNA, crRNA and Cas9 in vivo. Two of the strains that contain complete CRISPR-Cas systems, but different spacer compositions, were selected for the interference experiments. L. gasseri NCK 1342, an isolate from a renal patient’s endoscopy, and JV-V03, a female genital-urinary tract isolate that was part of the Human Microbiome Project, were used (Table 1). Protospeacer/PAM sequences corresponding to the leader proximate spacer in each strain were cloned separately into the shuttle vector pGK12. In this scenario, the protospeacer/PAM : : plasmid is analogous to previously encountered DNA, and the strain’s native CRISPR spacer, being part of the 'memory' complex, confers immunity by interference. Thus, a population targeting the protospeacer/PAM : : plasmid would exhibit a lower transformation efficiency than the pGK12 vector control due to plasmid cleavage. Indeed, consistent with other studies (Marraffini & Sontheimer, 2008; Gemia et al., 2014), targeting of plasmids exhibiting perfect spacer : : PAM combinations manifested 3.13 and 3.89 log reductions in transformants of L. gasseri NCK 1342 and JV-V03, respectively (Fig. 5). The reduction occurred in both NCK 1342 and JV-V03 with different protospeacer-containing plasmids, showing that efficiency of interference by the system was similar for each strain/spacer combination used. However, targeting of the plasmids was highly dependent on the PAM and seed sequences (Fig. 5). Introducing an incorrect PAM sequence (5'-GCIC-3') or a single nt deletion at position 30 within the JV-V03 spacer abolished targeting of the pGK12 : : protospeacer/PAM (Fig. 5). This is in accordance with the well-established roles of the PAM and seed sequences in CRISPR-Cas activity (Deveau et al., 2008; Horvath et al., 2008; Mozic et al., 2009; Semenov et al., 2011; Wiedenheft et al., 2011). The plasmid interference data validated the bioinformatically inferred 5'-NTAA-3' PAM sequence and provided concomitant evidence of expression of CRISPR-Cas components at a level sufficient for interference activity.

**Mechanisms of escape from the CRISPR-Cas system**

Despite the efficacious targeting of protospeacer/PAM : : plasmids, transformants were recovered, indicating evasion of targeting or inactivation of the CRISPR-Cas system. Therefore, the recovered transformants were genotyped at the CRISPR array level and the plasmids were analysed for loss of the insert. PCR amplification across the cloning site of pGK12 did not reveal any protospeacer/PAM deletions. Next, 15 transformants were selected and screened for loss of spacers from the CRISPR array by PCR amplification. The results showed that 9/15 transformants exhibited deletions within the CRISPR array (Fig. 6), implying...
that recombination occurred between repeats flanking the spacer responsible for interference of the plasmid. Sequencing confirmed the loss of the spacers, and showed that although variable deletions occurred, there was bias in the transformants toward recombination at certain repeat sequences. Other possible escape scenarios could be mutations in the crRNA coding sequences. PCR amplification of crRNA from genomic DNA of some remaining IV-V03 escape clones revealed several smaller amplicons relative to the wild-type, suggesting that deletions occurred (data not shown).

**DISCUSSION**

CRISPR-Cas systems are present in approximately 46% of bacterial and 84% of archaeal genomes, but there are significant and outstanding gaps in determining the ecological or evolutionary roles of the majority of endogenous microbial systems. Investigation of the natural evolutionary roles of CRISPR-Cas systems and their application in the development of genetic tools requires mechanistic and functional examination. A report recently highlighted that ~14 CRISPR-Cas systems had been vetted for their in vivo interference activity, of which only 5 were type II systems (Bondy-Denomy & Davidson, 2014). Here we submit the first report, to our knowledge, of a native, functional type II-A CRISPR-Cas system in the Lactobacillus genus, known to harbour numerous uncharacterized type II systems of a diverse nature, for which no functional work currently exists (Briner et al., 2014). Moreover, the system was characterized in *L. gasseri*, a microbe that impacts human health through autochthonously inhabiting the human gastrointestinal and genito-urinary tracts.

Considering that *L. gasseri* is commonly isolated from human mucosal niches and is used in commercial dairy processes, it could be hypothesized that possession of an immune system in these bacteriophage-enriched environments may contribute to environmental and/or in vivo persistence. However, the sporadic distribution (6/17) of the CRISPR-Cas system in *L. gasseri* implies an accessory ecological role in the highly diverse species. Interestingly, inferring the origin of spacer sequences from *L. gasseri* CRISPR arrays revealed sequence complementarity to bacteriophage and plasmid targets (Table 2), implying the system's ability to acquire spacers targeting prophages. Unlike bacterial species that undergo continuous attack from lytic bacteriophages, only temperate bacteriophages have been observed in *L. gasseri*, which represent several of the spacer hits reported here (Baugher et al., 2014; Raya et al., 1989). It is still currently unknown how CRISPR-Cas systems interact with prophages or how they may provide immunity when phage infection does not apply immense selective pressure. Therefore, *L. gasseri* strains containing this CRISPR-Cas system may serve as a platform for elucidating lyssogenic prophage interactions. It is notable that many of the spacer matches corresponded to plasmids (Table 2), as only two plasmids have been characterized in *L. gasseri*, one of which spacer 21 from NCK 1345 matched.

Importantly, we validated a novel PAM sequence and established plasmid interference using the system, which
suggests the native functionality of the system in *L. gasseri*, both transcriptionally and biochemically. Targeting efficacy of the system appeared equivalent to other type II systems such as CRISPR1 and CRISPR3 from *S. thermophila* (Barrangou et al., 2007). In agreement with previous studies, it was observed that spacer loss, likely through recombination between CRISPR repeats, is a major mechanism of preventing CRISPR-Cas interference (Jiang et al., 2013). Transforms exhibiting deletion of spacers likely incur a lower adaptive cost to the cell relative to mutations inactivating the CRISPR-Cas system as a whole. Cells employing this mechanism of preventing CRISPR targeting not only maintain CRISPR-Cas activity, but allow for uptake of horizontally acquired DNA in a subset of the bacterial population. Thus, repeat-mediated recombination may allow for acquisition of adaptive DNA, even when limiting genetic diversity typically constitutes the evolutionary cost of housing an active CRISPR-Cas system.

Characterization and exploitation of CRISPR-Cas systems in bacteria have led to effective typing and strain detection (Horvath et al., 2008; Barrangou & Horvath, 2012; Shariat et al., 2013; Briner & Barrangou, 2014), engineered immunity against mobile genetic elements (Garneau et al., 2010; Barrangou & Horvath, 2012), sequence-specific endogenous killing of bacteria (Gomma et al., 2014) and programmable transcriptional regulation (Bilardi et al., 2013; Luo et al., 2015). In silico surveys of lactic acid bacterial genomes suggest that these microorganisms are a significant reservoir for orthogonal type II CRISPR-Cas systems (Chylinski et al., 2013; Briner
et al., 2014). Characterization of the active type II-A system in L. gasseri facilitates its use as a platform for genome editing: programmable transcriptional repression and investigation of bacteriophage dynamics in L. gasseri. Moreover, the L. gasseri Cas9, with its cognate tracrRNA and PAM sequence, may be considered for genome editing purposes in food-grade systems and human gene therapy, since these applications require Cas9s from non-pathogenic sources. Collectively, our results open avenues for similar applications in L. gasseri and contribute to the fundamental understanding of CRISPR-Cas systems in bacteria.

ACKNOWLEDGEMENTS

This work was supported by the North Carolina Agriculture Foundation and DuPont Nutrition and Health. We thank Alexandra Fast and Evelyn Durman for technical assistance, Dr. Yong Fan for insightful discussions and use of colony to Dr. Philippe Horvath for providing the CRISPR spacer Excel Macro K.S. and R.B. are co-investigators on patent applications related to various uses of CRISPR-Cas systems.

REFERENCES


Clark, R. H. (2001). Distribution and strain polymorphisms in probiotic lactobacilli. MS thesis, North Carolina State University, Raleigh, NC, USA.


180


Edited by: T. Mradek.
SIGNR3-dependent immune regulation by *Lactobacillus acidophilus* surface layer protein A in colitis
SIGNR3-dependent immune regulation by *Lactobacillus acidophilus* surface layer protein A in colitis

Yalina L Lightfoot, Kurt Selle, Tao Yang, Yong Jun Goh, Bikash Sahay, Mojgan Zadeh, Jennifer L Owen, Natacha Colliou, Eric Li, Timo Johannsen, Bernd Lepenies, Todd R Klaenhammer, Mansour Mohamadzadeh

**Abstract**

Intestinal immune regulatory signals govern gut homeostasis. Breakdown of such regulatory mechanisms may result in inflammatory bowel disease (IBD). *Lactobacillus acidophilus* contains unique surface layer proteins (Slp), including SlpA, SlpB, SlpX, and lipoteichoic acid (LTA), which interact with pattern recognition receptors to mobilize immune responses. Here, to elucidate the role of SlpA in protective immune regulation, the NCK2387 strain, which solely expresses SlpA, was generated. NCK2387 and its purified SlpA bind to the C-type lectin SIGNR3 to exert regulatory signals that result in mitigation of colitis, maintenance of healthy gastrointestinal microbiota, and protected gut mucosal barrier function. However, such protection was not observed in SIGNR3−/− mice, suggesting that the SlpA/SIGNR3 interaction plays a key regulatory role in colitis. Our work presents critical insights into SlpA/SIGNR3-induced responses that are integral to the potential development of novel biological therapies for autoimmune inflammatory diseases, including IBD.

**Keywords:** colitis; immune regulation; *Lactobacillus acidophilus*; SIGNR3

**Subject Categories:** Microbiology, Virology & Host Pathogen Interaction

DOI: 10.1002/emb.14902960966

**Introduction**

The gastrointestinal (GI) microbiota plays a critical role in determining the immunologic outcome of various signaling events in host cells via their gene products, exceeding the human genome by a 100-fold (Ley et al., 2006; Qin et al., 2010). As such, the composition of the GI microbiota and host immunity are mutually and continuously influence each other (Maslowski & Mackay, 2011; McDermott & Heffragle, 2014). Therefore, it is inevitable that intestinal homeostasis is tightly controlled by regulatory immune mechanisms, which are established by the interactions of the trillions of microbes and their gene products with numerous pattern recognition receptors (PRRs), including C-type lectin receptors (CLRs), such as the specific intercellular adhesion molecule-3-grabbing non-integrin-hemorhag-related 1 (SIGNR3) (Konstantinov et al., 2008; Oort & Reis e Sousa, 2011). Disruption of this delicate balance by immunologic signals has devastating consequences that may result in intestinal disorders, including inflammatory bowel disease (IBD). When this occurs, highly activated innate cells trigger innate-infiltrating pathogenic T-cell subsets (e.g. Th1, Th17), and even regulatory T cells (Tregs) with proinflammatory characteristics (Kharazie et al., 2012; Geren et al., 2014; Neutrophils, 2015) that ultimately drive tissue destruction and intestinal disease progression. Innate cells (e.g. dendritic cells, macrophages) are the initial targets of the culpable microbes and their gene products, which subsequently affect the regulation/stimulation of intestinal immunity (Ivanov & Henda, 2013; Atarashi et al., 2014). Given these intertwined relationships, it is not surprising that microbial products have been linked to the pathology of intestinal autoinflammation (Nicholson et al., 2012). The underlying associations between gut microbes and inflammatory diseases (e.g. IBD) have already been well documented; however, the cellular and molecular mechanisms by which intestinal commensal gut product(s) and their molecular receptor(s) impact immune responses remain unclear.

Information regarding the immunobiologic functions of *Lactobacillus acidophilus* surface layer proteins (Slps) is relatively limited. Slps are parasympathetic (glyco) protein arrays that are abundant on the cell surfaces of few bacteria and archaea, including
L. acidophilus (Johnson et al., 2013). The S-layer of L. acidophilus NCFM is composed of three Slp-encoding genes: slpa (LBA0195), slpb (LBA0175), and slpc (LBA0212) (Goth et al., 2009). Diverse functional roles have been proposed for Slp, including cell shape determinants, molecular sieves, protective layers against viral infection, anchoring sites for surface-associated enzymes, and facilitators of cellular adhesion through PRRs, including C-type lectins (CLECs) (Konstantinov et al., 2008).

CLECs recognize carbohydrate structures on self and nonself antigens (Engert, et al., 2002; Ovino & Reis e Sousa, 2011). Twenty-nine CLECs, including DC-specific ICAM-3-grabbing non-integrin (DC-SIGN), have been identified on dendritic cells (DCs) and macrophages (Mφs) thus far (van Kooyk & Geijtenbeek, 2003; Ehlers, 2010; Sacchi & Reis e Sousa, 2013). DC-SIGN, which was previously shown to bind L. acidophilus SlpA in vitro (Konstantinov et al., 2008), is a calcium-dependent carbohydrate-binding protein with specificity for the mannose-containing glycans of microbial surface components and hormone-containing Lewis antigens (Ehlers, 2010). Of the eight murine homologs of DC-SIGN, SIGN3R (C2D9RD) exhibits the most biochemical similarity to human DC-SIGN (Penningsland et al., 2006). SIGN3R contains a carbohydrate-recognition domain (C RID) and signals through a semi-immune receptor tyrosine-based activation motif (hITAM) (Tanne et al., 2009). This signaling potentially downregulates the ubiquitously expressed immunoreceptor Tyrosine-based Amino-terminal Motif (Itam) (Robins et al., 2010) that catalyzes proinflammatory leukotriene B₄ (LTB₄) synthesis from LTA₄ (Snegrova et al., 2010), which subsequently activates interleukin (IL)-1β production. Here, we identify Sipa as a key effector molecule expressed by L. acidophilus and demonstrate its in vitro protective role in mucin colitis models. Moreover, we provide evidence that protection by L. acidophilus Sipa is conferred via signaling through a single CLR, namely SIGN3R.

Results

NCK2187 promotes intestinal immune regulation in steady state

Recently, we showed that transient colonization of the colon with NCK2025 (LTA⁺) significantly mitigated chemical and T-cell-mediated colitis (Mohammadzadeh et al., 2011). Additionally, NCK2025 significantly ablated inflammation-promoting polyposis in our nilex (Apoe⁻/⁻) Tg4-cre mouse model, where protection correlated with the regulation of innate and T-cell-induced inflammation (Khazzie et al., 2012). We therefore hypothesized that the controlled inflammation resulted from cross talk between NCK2025 Sipa with intestinal cells. To test this hypothesis, the sipp-counterselective gene replacement strategy (Gon et al., 2009) was used to generate in-frame deletions in the slpa and slpb genes of NCK2030. The LTA⁻ derivative was created by a deletion of the phosphoglycerol transferase gene (Mohammadzadeh et al., 2011), resulting in NCK2187, which expresses only Sipa (Fig 1A and B). To demonstrate that the newly generated NCK2187 transiently colonizes the gut, the clearance kinetics of both the

![Figure 1. Lactobacillus acidophilus NCK2187 strain development and characteristics.](image-url)

**Figure 1. Lactobacillus acidophilus NCK2187 strain development and characteristics.**

A. Agarose gel image illustrating PCR amplification of slpa (LBA0447), slpb, and slpc deletions in NCK2187.
B. SDS-PAGE gel of S. mansfieldi purified SipA (A) and SipB (B) from the parental strain NC as well as NCK2187 and NCK2030 (LTA⁺, Sipa⁺, SipB⁺, SipX⁺) and NCK2187 (LTA⁻, Sipa⁺, SipB⁻, SipX⁻).
C. Be mice were orally gavaged with 10⁶ CFU cryoprophylacterin-resistant NCK2187. Total colonic Sipa and SipB were assessed using the Foss of cryoprophylacterin-resistant strain, n = 3. Tissue pellets were collected daily and tested for the presence of cryoprophylacterin-resistant strain, n = 3, using the Foss 3-plex. Data are representative of five independent experiments and are shown as mean ± SEM and are representative of two experiments performed in triplicates. *P < 0.05, **P < 0.01, ***P < 0.001.
erythromycin-resistant NCK56 and NCK2187 strains were determined in C57BL/6 (B6) mice that were orally treated once with 10^9 CFU/mouse. Data showed that mice cleared both NCK56 and NCK2187 after 3 days, indicating that the deletion of LTA, SipA, and SipK in NCK2187 did not alter its transient passage through the GI tract when compared to its WT parent (Fig 1C).

To investigate the activation of colonic DCs when co-cultured with NCK56 or NCK2187, colonic cells were obtained from naïve B6 mice. While such intestinal cell-bacterial co-cultures did not significantly change the expression of DC co-stimulatory molecules (e.g., CD80) (not shown) or IL-10, only NCK56 deactivated the levels of IL-18, IL-6, IL-12, and TNF-α (Fig 1D). Next, naïve B6 mice were orally gavaged with NCK56 or NCK2187, and colonic immune responses were analyzed. Treatment with NCK2187 significantly increased the frequency of colonic Foxp3^- Tregs when compared to both untreated (PBS) and NCK56-treated mice (Fig 2A). Moreover, IL-17A^- and IFN-γ^- CD4^- T cells were significantly reduced by NCK2187 treatment (Fig 2A). NCK2187^-treated B6 Foxp3-GFP mice also exhibited higher numbers of colonic IL-10^- TGF-β^- Tregs than did NCK56^-treated and untreated mice (Fig 2B and C). Collectively, oral treatment with this novel L. acidophilus strain induced colonic regulatory immune responses.

Protective properties of NCK2187 and its SipA against intestinal inflammation and dysbiosis

To clarify the consequences of the immunoregulatory responses observed above during inflammation, B6 Rag2^-/- mice adaptively transferred with CD45Rb^-/- CD4^- T cells were orally treated with NCK56, NCK2187, its purified SipA, or PBS (Fig 3). Untreated (PBS) and NCK56-treated mice with adoptively transferred T cells developed severe colitis as demonstrated by weight loss, bloody diarrhea, shortening of the colon, and increased damage of the colon (Fig 3A-C, Supplementary Fig S1A). Furthermore, the levels of systemically induced proinflammatory IL-1β, IL-6, TNF-α, IL-17A, G-CSF, and macrophage inflammatory protein (MIP)-1α were significantly enhanced in the sera of these groups of mice (Fig 3D). In contrast, similar to the Treg co-transferred mice, NCK2187 and its purified SipA significantly protected Rag2^-/- mice from T-cell-induced colitis (Supplementary Tables S1, S2 and S3). NCK2187^- and SipA^-treated mice gained weight throughout the course of the study and did not develop bloody diarrhea in the way that the PBS and NCK56 groups did (Fig 3A). Furthermore, rectal and colonic atrophy due to pathogenic inflammation was not observed in these mice, as the tissue destruction and immune cell infiltration associated

Figure 2. Lactobacillus acidophilus NCK2187 promotes intestinal regulation in steady state.

A. B6 mice were orally gavaged with 10^9 CFU NCK56 (blue) or NCK2187 (green) on days 0, 3, 5, and 7, or left untreated, and immune responses in the colon were analyzed at day 3 by flow cytometry.

B, C. B6 Foxp3-GFP mice were treated and evaluated as in A. Regulatory cytokine production in Foxp3^- GFP^- (green dotted line) versus Foxp3^- GFP^- (white bar) cells was measured by intracellular staining and FACS analysis (C).

Data information: for all data shown, n = 5 mice/group. Data represent four individual experiments and are shown as mean ± SEM. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. Black asterisks compare NCK2187 to untreated (PBS) mice, and red asterisks to NCK56^-treated mice.
erythromycin-resistant NCK56 and NCK2187 strains were determined in C57BL/6 (B6) mice that were orally treated once with 10⁶ CFU/mouse. Data show that mice cleared both NCK56 and NCK2187 after 3 days, indicating that the deletion of LTA, SlpA, and SlpK in NCK2187 did not alter its transit passage through the GI tract when compared to its WT parent (Fig 1C).

To investigate the activation of colonic DCs when co-cultured with NCK56 or NCK2187, colonic cells were obtained from naïve B6 mice. While such intestinal cell-bacterial co-cultures did not significantly change the expression of DC co-stimulatory molecules (e.g., CD80) (not shown) or IL-10, only NCK56 elevated the levels of IL-18, IL-6, IL-12, and TNF-α (Fig 1D). Next, naïve B6 mice were orally gavaged with NCK56 or NCK2187, and colonic immune responses were analyzed. Treatment with NCK2187 significantly increased the frequency of colonic FoxP3+ Tregs when compared to both untreated (PBS) and NCK56-treated mice (Fig 2A). Moreover, IL-17A+ and IFN-γ+ CD4+ T cells were significantly reduced by NCK2187 treatment (Fig 2A). NCK2187-treated B6 FoxP3-GFP mice also exhibited higher numbers of colonic IL-10+ TGF-β1+ Tregs than did NCK56-treated and untreated mice (Fig 2B and C). Collectively, oral treatment with this novel L. acidophilus strain induced colonic regulatory immune responses.

Protective properties of NCK2187 and its SlpA against intestinal inflammation and dysbiosis

To clarify the consequences of the immunoregulatory responses observed above during inflammation, B6 Rag1-/- mice were adoptively transferred with CD45.1+/CD4+ T cells were orally treated with NCK56, NCK2187, its purified SlpA, or PBS (Fig 3). Untreated (PBS) and NCK56-treated mice with adoptively transferred T cells developed severe colitis as demonstrated by weight loss, bloody diarrhea, shortening of the colon, and increased damage of the colon (Fig 3A-C, Supplementary Fig S1A). Furthermore, the levels of systemically induced proinflammatory IL-1β, IL-6, TNF-α, IFN-γ, G-CSF, and macrophage inflammatory protein (MIP)-1α were significantly enhanced in the sera of these groups of mice (Fig 3D). In contrast, similar to the Treg co-transferred mice, NCK2187 and its purified SlpA significantly protected Rag1-/- mice from T-cell-induced colitis (Supplementary Tables S1, S2 and S3). NCK2187- and SlpA-treated mice gained weight throughout the course of the study and did not develop bloody diarrhea in the way that the PBS and NCK56 groups did (Fig 3A). Furthermore, rectal and colonic atrophy due to pathogenic inflammation was not observed in these mice, as the tissue destruction and immune cell infiltration associated

---

**Figure 2. Lactobacillus acidophilus NCK2187 promotes intestinal regulation in steady state.**

A. B6 mice were orally gavaged with 10⁶ CFU NCK56 (blue) or NCK2187 (green) on days 0, 3, 6, and 9. Bone marrow immune response in FoxP3-GFP (green dotted line) versus FoxP3-GFP (white bar) cells was measured by intracellular staining and FACS analysis (C).

B, C. B6 FoxP3-GFP mice were treated and evaluated as in (A). Regulatory cytokine production in FoxP3-GFP (green dotted line) versus FoxP3-GFP (white bar) cells was measured by intracellular staining and FACS analysis (C).

Data information: *P = 0.05; **P < 0.005, ****P < 0.0001. Black asterisks compare NCK2187 to untreated (PBS) mice, and red asterisks to NCK56-treated mice.
Figure 3.
with T-cell-induced colitis were significantly abrogated in NCK2187- and SipA-treated groups (Fig 3B and C). Additionally, systemic inflammation was significantly reduced in these groups of mice (Fig 3D). Interestingly, the genes encoding the receptors for Ltb4, Ltb4-1 and Ltb4-2, were significantly downregulated in the colons of NCK2187- and SipA-treated mice (Supplementary Fig S1B), which may have contributed to the reduced expression of colonic Ltb4 (Supplementary Fig S1B).

T-cell-induced colitis resulted in intestinal epithelial mucus degeneration in mice that did not receive NCK2187 or its purified SipA (Fig 3C). Indeed, the expression of right junction-associated genes was significantly downregulated in PBS- and NCK56-treated Ragn1-/- mice (Fig 4A). Furthermore, FITC-dextran permeability assays confirmed that these mice were suffering from dysregulation of intestinal barrier function (Fig 4B). Accordingly, NCK2187 and SipA significantly protected barrier integrity and function (Fig 4A and B). An immunologically and anatomically weakened intestinal epithelial barrier during intestinal inflammation allows luminal bacteria to interact with the intestinal mucosa and the infiltrating immune cells, initiating inflammatory responses directed against the gut commensal bacteria and introducing dysbiosis. To determine the status and composition of the microbiota in the T-cell-induced colitis model (week 7), we analyzed the microbial communities in the colons of the different experimental groups and found that the severity of colitis was associated with significant changes in the microbiota (Fig 4C and D). Furthermore, the microbial communities in PBS- and NCK56-treated Ragn1-/- mice were more closely related to the same type of bacteria but with lower diversity in the intestinal luminal milieu and the communities observed in the intestinal mucosa, which may promote the growth of previously overrepresented microbial communities so that they dominate the population. Alterations in the microbial composition were also manifested at lower taxonomic levels NCK2187, SipA, and TcgA-treated groups once again showed similar relative abundance and distribution of several unclassified genera (Fig 4E).

To further investigate the regulatory role of L. acidophilus SipA, we then selected an experimentally infectious model using Cimobacter rodentium that results in a breach of the intestinal epithelial barrier, potentially orchestrated by uncontrolled proinflammatory immune responses (Leib et al., 2009). Accordingly, data demonstrate that treatment with NCK2187 and its purified SipA significantly accelerated pathogen clearance (Supplementary Fig S2A), resulting in reduced size of the draining lymph nodes (Supplementary Fig S2B), and decreased colonic IL-1β expression (Supplementary Fig S2C). Conversely, this trend was not observed in C. rodentium-inoculated mice that were treated with NCK56 or PBS, suggesting that L. acidophilus SipA regulates induced proinflammation (e.g., IL-1β), which results in less colonic damage and bloody diarrhea (Supplementary Fig S2D and E). Histologic analyses of colonic mucosal damage (e.g., goblet cell loss, abnormal crypts) and inflammation with C. rodentium plus NCK56 revealed increased lymphoplasmacytic infiltrates with lesser but mildly increased neutrophilic infiltrates within the lamina propria and colonic submucosa, which were decreased in L. acidophilus SipA-treated groups (Supplementary Fig S2D). Future studies are warranted to elucidate the cellular and molecular mechanisms involved in such regulatory protective intestinal responses upon GI pathogen challenge.

**Lactobacillus acidophilus SipA binding to SIGNR2 promotes colonic regulatory immune responses**

Symbiotic bacteria and their gene products dictate the nature of innate responses via their sensing receptors (Ivano & Hond, 2012; Yang et al., 2014); however, such stimulatory signals must be regulated by other receptors to avoid intestinal inflammation. As stated earlier, SIGNR2 exhibits the most biochemical similarity to human DC-SIGN. We screened all known murine SIGNR1-8 and found the Signr1 and Signr3 genes to be differentially activated in the colonic tissue of mice orally treated with NCK2187 (Fig 5A), prompting us to evaluate the binding of SipA to SIGNR1 and SIGNR3. Subsequently, the corresponding extracellular fragments of SIGNR1 and SIGNR3 were fused to the Fc portion of human IgG (SIGNR1-Fc, SIGNR3-Fc) and then transiently expressed in Chinese hamster ovary (CHO)-S cells (Erksson et al., 2013) (Supplementary Fig S3). Data demonstrate that while expressed SIGNR1-Fc bound to purified SipA co-acted on charged fermont heads, SIGNR1-Fc, Ctrl-CFC (control proteins tagged with Fc), and the secondary rat anti-human Fc antibody alone did not, suggesting SipA-binding specificity to SIGNR1 (Fig 5B). Additionally, we observed the binding of L. acidophilus SipA to full-length SIGNR3 but not SIGNR1 expressed on CHO-S cells, indicating again receptor-binding specificity to this bacterial protein (Fig 5C). Moreover, given that SIGNR3 was found to be the mouse ortholog of DC-SIGN, most closely resembling its human homolog in terms of ligand binding (Powlesland et al., 2006), we in vitro binding assays confirmed that purified SipA binds to DC-SIGN-Fc (Fig 5D), as well as onto CHO-S cells expressing DC-SIGN (Fig 5E).

To clarify the role of NCK2187 SipA/SIGNR3 binding and signaling in vivo, we first orally treated WT B6 and B6 Signr3-/- ICR mice...
A. Relative expression of tight junction-associated genes Cldn3 and Ocln, determined by RT-PCR relative to 385 hRNA (A), as well as passive transepithelial resistance (PCT) of Caco-2 monolayers (B) at 7 days post-transfection. Data are shown as means ± SEM. *p < 0.05, **p < 0.01. The asterisk indicates the significance level.

B. FITC fluorescence intensity of CD45R0+ T cells in the small intestine. Data are shown as means ± SEM. *p < 0.05, **p < 0.01. The asterisk indicates the significance level.

C. PC2 scatter plot showing the relative abundance of different phyla in the gut microbiota. The colors represent the treatment groups. The size of the circles indicates the relative abundance of each phylum. The dots represent individual samples.

D. Heat map showing the relative abundance of different taxa at the genus level. The colors represent the treatment groups. The size of the squares indicates the relative abundance of each taxon. The dots represent individual samples.

E. Scatter plot showing the relative abundance of different taxa at the species level. The colors represent the treatment groups. The size of the circles indicates the relative abundance of each species. The dots represent individual samples.

Figure 4. Lactobacillus acidophilus NCK2187 and its SlpA protect intestinal barrier function and prevent dysbiosis in pathogenic T-cell-induced colitis.

Colitis was induced in B6 RAG-2−/− mice as described in Fig 1. A.-E. Relative expression of tight junction-associated genes Cldn3 and Ocln, determined by RT-PCR relative to 385 hRNA (A), as well as passive transepithelial resistance (PCT) of Caco-2 monolayers (B) at 7 days post-transfection. Data are shown as means ± SEM. *p < 0.05, **p < 0.01. The asterisk indicates the significance level. B. FITC fluorescence intensity of CD45R0+ T cells in the small intestine. Data are shown as means ± SEM. *p < 0.05, **p < 0.01. The asterisk indicates the significance level. C. PC2 scatter plot showing the relative abundance of different phyla in the gut microbiota. The colors represent the treatment groups. The size of the circles indicates the relative abundance of each phylum. The dots represent individual samples. D. Heat map showing the relative abundance of different taxa at the genus level. The colors represent the treatment groups. The size of the squares indicates the relative abundance of each taxon. The dots represent individual samples. E. Scatter plot showing the relative abundance of different taxa at the species level. The colors represent the treatment groups. The size of the circles indicates the relative abundance of each species. The dots represent individual samples.

with our bacterial strains and analysed the immunologic responses induced in steady state. While NCK2187 treatment led to reduced IL-1β production in both conventional (not shown) and germ-free (GF) B6 mice (Fig SF), no anti-inflammatory effects were observed in NCK2187-treated KO mice (Fig SG). Furthermore, the Treg-promoting properties of NCK2187 (Fig 2A and B) were abrogated in
Figure 5. Lactobacillus acidophilus NCC2187 and its SlpA bind to murine SIGN3R2 to induce regulatory signals.

A. I26 mice were orally gavaged with 3.5×10^6 CFU NOK56 or NOK2187, and the sIgM gene expression of C-type lectin receptors were measured by RT-PCR. Each box represents an individual mouse; n = 4. Data represent three independent experiments.

B. Binding of SlpA to various mAbs fusion proteins was analyzed by flow cytometry. Gray filled line = SlpA-coated beads; orange = SlpA-coated beads + secondary antibody; green = SlpA-coated beads + control fusion protein; blue = SlpA-coated beads + SIGNR1+Fc; red = SlpA-coated beads + SIGNR3-Fc. Binding assay results were confirmed five independent times.

C. Binding of SlpA to SIGN3R2, but not SIGNR1, was observed by CHO-S cells. Gray filled line = untransfected CHO-S cells; blue = untransfected CHO-S cells labeled SlpA; red = SIGNR3-transfected CHO-S cells labeled SlpA; green = SIGNR1-transfected CHO-S cells labeled SlpA. Binding assays in CHO-S cells were performed three independent times.

D. Binding of SlpA to recombinant human DC-SIGN5000-Fc chimera was analyzed by flow cytometry. Gray filled line = SlpA-coated beads + secondary antibody; green = SlpA-coated beads + control fusion protein; red = SlpA-coated beads + DC-SIGN-Fc. Binding assay results were confirmed five independent times.

E. Binding of SlpA to DC-SIGN expressed by CHO-S cells. Gray filled line = untransfected CHO-S cells; blue = untransfected CHO-S cells labeled SlpA; red = DC-SIGN-transfected CHO-S cells labeled SlpA. Binding assays in CHO-S cells were performed three independent times.

F, G. IL-1β production by colonic DCs of naïve WT B6 (gray) or SlpA (blue) were determined by flow cytometry. n = 5 mice/group. Data represent four independent experiments and are shown as mean ± SEM. *p < 0.05. Black asterisks compare NCC2187 to untreated (PBS) mice, and red asterisks to NOK56-treated mice.

H. Frequency of colonic Foxp3+ T cells in KO mice treated with NOK56 or NOK2187 was measured by flow cytometry. n = 6 mice/group. Data represent four independent experiments and are shown as mean ± SEM.
Figure 6. Lactobacillus acidophilus NCK2187 and its SipA do not protect against DSS-induced colitis in Sig{sup}-/- mice. WT (n = 8) or Sig{sup}-/- (n = 3) mice were orally gavaged with NC56, NCK2187, or SipA on days -3 and -1, and 3% DSS was given in deionized water. Mice were gavaged with bacteria or purified SigA every other day for an additional three times and monitored for disease progression.

A. Colitis severity was determined in part by body weight loss, x = SipA group.
B. C. Colitis score based on histopathology and gross morphology of the colons were also used as measures of disease. Scale bars = 30 μm, x = SipA group.
D. E. F. G. Colitis score was used to calculate distance between the microbial communities of the different samples (day 10), and then the distance matrix was visualized using PCoA. Light grey: WT, red: SipA, brown: WT + DSS, blue: WT + DSS + SipA, aqua: KO + DSS, red: KO + DSS + NC56, yellow: KO + SipA, purple: KO + DSS + SipA, n = 5 mice.

H. Species richness and microbial diversity at DSS-treated mice. Top: The Chao richness index was used as a measure of species richness. Bottom: The Shannon diversity index was used to estimate microbial diversity in each group. n = 5 mice. Color coding is as in (B).

Data information: Data are shown as mean ± SEM and are representative of three different experiments. *p < 0.05, **p < 0.01, ***p < 0.001.

K0 mice (Fig 5F). These data strongly suggest that NCK2187 delivers immunoregulatory signals via its interaction with SIGN3.

NCK2187 and its SipA do not prevent dextran sulfate sodium (DSS)-induced colitis in the absence of SIGN3 signaling

Previous reports have highlighted the role of specific CLRs in experimental colitis. For instance, mice lacking Sig1r1 expression are less susceptible to induced colitis (Saunders et al., 2010), while mice deficient in (dextran) and Sig1r1 exhibit exacerbated disease (Iber et al., 2012; Eriksson et al., 2013). To further investigate SipA/SIGN3 signaling in disease conditions, DSS-treated WT and Sig1r1−/− mice were orally gavaged with NC56, NCK2187, or SipA, and monitored for disease progression. Consistent with our T-cell-induced colitis model, disease progression and severity were significantly reduced in WT mice orally treated with NCK2187 or purified SipA (Fig 6, Supplementary Table S5); however, NCK2187 and SipA did not confer any protection in K0 mice (Fig 6). Measurements included weight loss, histopathology-based colitis scores, evaluation of gross mucosal damage, and immune cell recruitment and activation (Figs 6 and 7).

Disease progression and inflammation were associated with alterations in intestinal barrier integrity and the composition of the gut microbiota (Fig 6H). Several gut permeability markers were evaluated by RT-PCR. The restoration of claudins (claudin-3, claudin-12, and claudin-5) in WT mice treated with NCK2187 or SipA indicates that NCK2187 and SipA could be capable of promoting intestinal tight junctions (Fig 6).

Accordingly, only NCK2187 and SipA treatments of WT mice prevented increased fecal albumin levels after DSS treatment (Fig 6G). The other hand, no positive effects by NCK2187 or SipA on barrier integrity were noted in K0 mice (data not shown). In terms of microbiota composition, protected WT mice (NCK2187 and SipA treated) clustered together in UniFrac analyses, while diseased untreated and NCK56-treated WT mice clustered separately (Fig 6G). Conversely, the microbial communities of all DSS-treated Sig1r1−/− mice formed a single cluster, indicating that dysbiosis was uniformly distributed independent of treatment group (Fig 6G). Furthermore, richness and diversity, analyzed by the Chao richness index and Shannon diversity index, respectively, were maintained in NCK2187- and SipA-treated WT mice, while no such effects were observed in K0 mice (Fig 6H).

We then analyzed induced immune responses in the colon of DSS-treated mice to determine the differences, if any, among our treatment groups. Neutrophil infiltration within the colon of NCK2187- and SipA-treated SIGN3-sufficient mice that were given DSS decreased to nearly PBS-treated control mice levels; while, in contrast, an even higher frequency of infiltrating neutrophils was detected in Sig1r1−/− mice after the induction of colitis, irrespective of the treatment group analyzed (Fig 7A, Supplementary Table S6). Similarly, the number of IL-1β-producing DCs and macrophages was significantly decreased with NCK2187 and SipA treatment in WT mice; however, no changes were observed among the different treatment groups in the absence of SipA/SIGN3 signaling (Fig 7B). We have previously shown, as have others, that pathogenic inflammation can result in proinflammatory Foxp3− RORγt− Tregs (Bhattacharyya et al., 2011). However, no significant change in the total number of Foxp3− Tregs was measured among the WT and K0 groups, we found that the quality of these Tregs was significantly altered. A large number of Foxp3+, IL-10+, and TNF-α were repressed only in NCK2187- and SipA-treated WT mice, but not in K0 mice (Fig 7C). Correspondingly, the levels of circulating cytokines in the sera (IL-1β, IL-10, IL-17A, IFN-γ, and TNF-α) were restored only in NCK2187- and SipA-treated WT mice, but not in Sig1r1−/− mice (Supplementary Fig S4). Collectively, these clinical and immunologic data provide robust evidence in support of an immunoregulatory role for L. acidophilus SipA that is highly dependent on intact SIGN3-mediated binding.

Discussion

The human GI tract harbors trillions of microbes, most of which are bacteria (Qin et al., 2010) and are critical determinants of the health of the host (Nicholson et al., 2012; Subramanian et al., 2014). This is especially true in the case of IBD, given the intimate association of the gut microbiota and their gene products with the adhesion colonic tissue (Hold et al., 2014; Hutenhaver et al., 2014). Early experiments provided evidence suggesting that susceptibility to pathogenic intestinal inflammation in experimental colitis was dependent upon the presence of enteric antigens (Kudóm et al., 1991), and were later supported by human studies, which demonstrated that an imbalance in the commensal bacterial composition, termed "disbiosis," is a defining characteristic of patients suffering from IBD (Sisoo et al., 2006; Frank et al., 2007). Accordingly, a major focus in
the field has been the identification of effector bacterial strains that influence the immune system (Ahern et al., 2014), and thus may be employed to reprogram undesired immune responses, both locally and systemically.

As reviewed elsewhere (Ahern et al., 2014), it is important that the search for microbes with immunoregulatory properties accounts for differences at the strain level and not merely at the species level. Our goal has been to take this concept one step further and to
Figure 7. Lactobacillus acidophilus NCC2187 and its SipA do not prevent immune infiltration and activation during DSS-induced colitis in Sigm3−/− mice. 
Sign3+/+ (WT) or Sign3−/− (KO) mice were orally gavaged with NCC2187 or SipA on days −3 and −1, and DSS was given in the drinking water. Mice were gavaged with bacteria or purified SipA every other day for an additional three times, and immunity was analyzed by flow cytometry at day 10. 
A. Representative plots indicate the frequency of mononuclear cells from untreated or DSS-treated WT (left) and KO mice (right). Empty bars = WT; hatched bars = KO; white bars = untreated; purple bars = DSS; red bars = DSS + NOD58; green bars = DSS + NOD2187; blue bars = DSS + SipA. 
B, C. Colon, DCs and macrophages (Mφs) were analyzed by flow cytometry for the production of IgG1 and IgA. wounded tissues from mice were evaluated for expression of +RORγT Cq n = 5 mice/group. Gray hatched line = isotype control; black = untreated; purple = DSS; red = DSS + NOD58; green = DSS + NOD2187; blue = DSS + SipA. 
Data information: Data are shown as mean ± SEM and are representative of three different experiments. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

identify specific bacterial molecule host receptor interactions that may account for the responses induced by effector bacterial strains. For instance, we previously demonstrated that oral treatment using a L. acidophilus strain lacking the gene responsible for LTA biosynthesis significantly reduced pathogenic inflammation in the GI tract, thereby promoting the mitigation of induced colitis (Mohammadzadeh et al., 2012). Yet, the question that remained unaddressed was “What molecule(s) of this L. acidophilus strain deficient in LTA significantly diminish the induced pathogenic inflammation that can result in intestinal disorders?” In addition to LTA, L. acidophilus is composed of various Sips, and several important roles have been attributed to the Sips, including immune effector properties (Konstantinov et al., 2008; Lightfoot & Mohammadzadeh, 2013).
To address this question, we systematically deleted genes using innovative strategies that allow the direct selection of double recombinants, thereby facilitating the construction of novel bacterial strains, such as NCK2187, and the assignment of roles to SIp candidate genes that are responsible for SIpA, SIpB, and SIpX protein expression (Goh et al., 2009). This sophisticated molecular approach to targeting genes in L. acidophilus defined the functional role of SIpA and demonstrated how this bacterial product affects intestinal immune cells and conventional T-cell subset activation, including Tregs, in steady state and murine colitis models. This is the first thorough study that has been carried out to rigorously examine the role of L. acidophilus SIpA in intestinal disorders using immunomodulated, infective, and transgene-induced colitis models. As seen in Figures 3 and 4, NCK2187 and its purified SIpA not only mitigated T-cell–induced colitis by significantly reducing proinflammation, but also protected the composition of the microbiota and intestinal barrier function. Additionally, systemic immune responses were also altered, whereas the levels of proinflammatory cytokines, including IL-1β, whose detrimental role in IBD was recently demonstrated (Cuevas et al., 2012), decreased significantly. These data implicate the IL-1β signaling axis in our model of intestinal protection. Importantly, a similar regulatory trend was also observed in C. rodentium–infected colitis, wherein C. rodentium–infected mice fed NCK2187 or its SIpA exhibited reduced colitis symptoms, including a reduction in diarrhea, colorectal atrophy, and inflammation. These data inspire the search to further elucidate molecular mechanisms that could potentially be involved, including the regulation of Wnt11 receptors. Accordingly, gene-screens results, along with SIpA binding to SIGIRR, clearly highlight the involvement of SIGIRR in the process of tempering highly activated gut immune responses. Additional data regarding SIGIRR engagement using SigIRR+ mice clarified the role of this signaling molecule in induced immune regulation, as was also documented in the Lethal enteritis infantum murine model (Raffels et al., 2013). Our data support the notion that the SIpA/SIGIRR interaction significantly reduces the high-affinity receptors for LTBR (in T-cell–transferred Rag2−/− mice). Downregulation of LTBR, and/or its receptors is critical in preventing inflammation activation, which otherwise results in increased IL-1β (Raffels et al., 2013). Interestingly, interrupting the interaction between SIpA and SIGIRR resulted in hypoimmune immunity and the production of IL-1β in KO mice under inflammatory conditions. Such dysregulated immune responses in KO mice promoted neuropathic infiltration and significantly affected the function of colonic FoxP3+ Tregs, which were converted to proinflammatory FoxP3+ RORγ+ Tregs, all of which significantly contributed to pathogenic inflammation. A condition seen in IBD progression to constant, balanced immunity was restored in WT mice that were treated with NCK2187 or SIpA. Induced colon inflammation in WT mice that were treated with NCK2186, but not in NCK2187– or SIpX–treated mice, and in SIpX+ mice, regardless of treatment, resulted in microbial dysbiosis and barrier dysfunction, another hallmark of IBD.

In conclusion, we have shown that the interaction of SIpA with SIGIRR can impact the status of innate and T-cell polarization in induced colitis. Our findings suggest that effective modulation of these cellular and molecular factors might significantly modify pathogenic inflammation that results in colitis and would therefore restore intestinal homeostasis by rebalancing deteriorated immunity, the composition of the gut microbiota, and mucosal barrier function. More importantly, our data support the translational implications of SIpA/DC-SIGN interaction and the involved regulatory signals, which may advance the development of potential therapeutic approaches for inflammatory diseases in the future, including chronic colitis. If hyperactivated innate and T-cell subsets play critical roles in IBD progression, our data provide clues concerning intestinal cellular mechanisms, and the properties of SIpA that may be used for immune therapeutic approaches in future clinical applications.

Materials and Methods

Mice

C57BL/6 [B6] and B6 recombination-activating gene 1-deficient (Rag2−/−) mice were purchased from Jackson Laboratories (Bar Harbor, ME, USA). Germ-free (GF) B6 mice were obtained from the National Institute of Child Health and Human Development (NIH) and housed in the Laboratory of Immunology and Infectious Diseases (LIID). The mouse strain B6.Cg-Tg(TCR5)1HjEc (B6.Cg-Tg(TCR5)1HjEc) was provided by the NIH-sponsored Mice Disease Model Regional Resource Center (MIDRCC) National System and was backcrossed at the Mouse Genetics unit at the National Institutes of Health, Bethesda, MD, USA. The mouse strain B6.Cg-Tg(TCR5)1HjEc was purchased as a standard stock from the Jackson Laboratory. Mice were bred in-house in the animal facility at the College of Veterinary Medicine, UH. Mice were maintained under specific pathogen-free, allergen-free conditions and sex-matched controls used at 6–8 weeks of age in accordance with the Animal Welfare Act and the Public Health Policy on Humane Care. Mice were randomized into the described treatment groups and disease severity scored blindly. Procedures were approved by the Institutional Animal Care and Use Committee (IACUC), protocol number 2014–6559.

Bacterial strains

The app-counterselective gene replacement strategy (Goh et al., 2009) was used to generate an in-frame deletion of the phosphoglycerol transferase gene within NCK2030 (LTA− SIpB− SIpX− SIpA−) to generate NCK2187 (LTA− SIpB− SIpX− SIpA−). Wild-type L. acidophilus NCFM (NCC56) and NCK2187 were propagated anaerobically in MRS broth (Difco, BD, Franklin Lakes, NJ, USA) at 37°C for 18 h. In preparation for oral treatment, bacteria were washed twice with sterile PBS, and the number of colony-forming units (CFU) was estimated by measuring the optical density at 600 nm. The concentrations of each L. acidophilus strain was accordingly adjusted to the desired final concentration of 1 × 10⁸ CFU/100 μl. To determine the clearance kinetics of the different L. acidophilus strains, groups of mice (n = 3) were orally gavaged with erythromycin-resistant (EmR) NCC56 or NCK2187 (1 × 10⁸ CFU/100 μl/mouse). Fece pellets were collected from cage every day thereafter for up to 7 days. Each fecal pellet was then reconstituted in 10% MRS (0.2 g/2 ml). The homogenized material was serially diluted and plated...
onto MRS agar containing Em (2 μg/ml). The daily average excreted L. acidophilus was quantified. For the oral gavage of mice, each mouse received colonizer 1 × 10^7 CFU of NCK56 or NCK2187 to 100 μl of PBS. Mice enrolled in steady-state studies were orally gavaged every 3 days for a total of four times, and immune changes were analyzed at day 14. The gavage schedule was determined based on the clearance kinetics of the bacterial strains in Balb/c mice, as all mice used in our studies are the Balb/c background.

**Surface layer protein A isolation**

SipA was purified from NCK2187 with LCL. Cultures from 18-h-grown NCK2187 were pelleted at 5,000 × g for 10 min. Bacterial pellets were washed with cold PBS and re-pelleted before extractants. Pellets were resuspended in 5 M LiCl (Sigma-Aldrich, St. Louis, MO, USA) and gently stirred for 30 min, and the bacteria were removed by centrifugation (20,000 × g, 15 min). Supernatants were dialyzed against distilled water using a dialysis bag with a cut-off molecular weight of 3 kDa for salt removal. The protein precipitate was dissolved in 1 M LiCl and pelleted at 20,000 × g for 30 min. The SipA protein preparation was washed with water a minimum of three times before use. SDS-PAGE gels and proteomics analyses were used to confirm SipA purity. For oral gavage, mice were given 150 μg of SipA in 500 μl PBS.

**Lamina propria leukocyte (LPL) preparation**

Freshly isolated colons were cut into 0.5 cm sections, and intramural lymphocytes were removed with a digestion buffer consisting of HBSS (GIBCO, Life Technologies, Grand Island, NY, USA) containing 5 mM EDTA (GIBCO, Life Technologies) and 10 mM HEPES (GIBCO, Life Technologies), for 20 min at 37°C. Remaining colonic tissues were digested in DMDM (GIBCO, Life Technologies), supplemented with 0.25 mg/ml Collagenase Type VII (Sigma-Aldrich), 0.125 U/ml Liberase TM Research Grade (Roche Applied Science, Indianapolis, IN, USA), 10 mM HEPES, 0.1 M CaCl₂ (Sigma-Aldrich), and 5% FBS (GIBCO, Life Technologies). Three digestions of 10 min each at 37°C were performed. Single-cell suspensions obtained were combined and stained for flow cytometry-based analyses or used for ex vivo studies.

**Ex vivo stimulation of colonic LPLs**

Isolated colonic LPLs were co-cultured with NCK56 or NCK2187 for 12 h at 37°C. Supernatants were then collected and stored at 80°C for later cytokine analyses using Bio-Plex Pro Mouse Cytokine Immunoassay kits (Bio-Rad, Hercules, CA, USA). Activation phenotypes of DCs were analyzed by flow cytometry using the appropriate antibodies to quantitate expression levels of MHC-II molecules and co-stimulatory markers.

**Flow cytometry and antibodies**

Colonic LPLs were stained with LIVE/DEAD Aqua Dead Cell Stain kit (Molecular Probes, Life Technologies). Washed cells were incubated with Mouse Fc Blocking Reagent (Miltenyi Biotec, Auburn, CA, USA) as per the manufacturer’s instructions before staining with combinations of the following antibodies or their corresponding isotype controls: CD45 (30-F11), CD11c (N418), CD11b (M1/70), F4/80 (BM8), GR1 (RB6-8C5), I-A/I-E (MLR20), CD3 (14-2C11), CD4 (R4-55), CD8 (53-67), PDL-1 (N93), H-2Kd, CD69, CD44 (12-M5H14), and IgM, IgG1, II, and III. For detection of intracellular cytokines, cells were fixed and permeabilized with BD Cytofix/PermFix (BD Biosciences). Cytokine T cells were stimulated with phorbol 12-myristate 13-acetate (PMA; 50 ng/ml) and ionomycin (2.5 μg/ml) in the presence of brefeldin A (BioLegend) for 2.5 h. The Transcription Factor Fixation/Permeabilization kit from eBioscience (San Diego, CA, USA) was used for FoxP3 staining. After staining, a BD LSRII Fortessa (BD Biosciences) cell analyzer was used to acquire fixed cells. Data were analyzed with FlowJo software (Tree Star, Ashland, OR, USA).

**T-cell-induced colitis**

In preparation for the adoptive transfer of CD435^+ CD4^+ T cells into B6 Rag1^−/− mice, spleen and mesenteric lymph node (MLN) single cell suspensions obtained from healthy B6 mice were pooled and incubated in AntigenPure goat anti-mouse IgG (H+L)-coated cell culture plates (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA, USA) at 37°C for 1 h. CD4^+ T cells were isolated from non-adherent cells using the CD4^+ T cell Isolation Kit II (MACS, Miltenyi Biotec, San Diego, CA, USA), and CD3^+ CD4^− T cells were then depleted by positive selection (MACS, Miltenyi Biotec). Bound CD3^+ CD4^+ T cells were collected and injected into the regulatory T cells (Treg) group. The resulting cell suspensions after negative and positive selection were consistently comprised of >98% CD3^+ CD435^+ CD4^+ T cells. Because the T-cell transfer model of colitis is a chronic inflammatory model, which requires several weeks for gross inflammation to manifest, B6 Rag1^−/− mice were orally gavaged once with NCK56, NCK2187, or SipA prior to the adoptive transfer of T cells by intraperitoneal injection (i.p.); 1 day later, the mice were orally gavaged once more, and once a week for four consecutive weeks thereafter. Colitis progression was monitored by determining mouse weight loss, fur coat development, and focal occult blood (FOB) presence throughout the study. Stool consistency was scored as follows: 0 = normal, 2 = pasty, and 4 = watery with perianal staining.

**Murine Citrobacter rodentium infection**

Citrobacter rodentium infection in mice causes pathology similar to that seen with colitis induced by the human pathogens, enteropathogenic *Escherichia coli* and enterohemorrhagic *E. coli* (Lappin et al., 2010). These Gram-negative bacteria induce attaching and effacing (AE) lesions on the intestinal epithelium (Raper et al., 2004), which initiate pathological inflammation resulting in intestinal colonization (Mundy et al., 2005; Lappin et al., 2007; Stecher et al., 2007). Briefly, B6 mice were gavaged with either NCK56, NCK2187, SipA, or PBS at days 3 and 1. Subsequently, these groups of mice were orally infected with *C. rodentium* (10^8 CFU/mouse). Mice were then treated with *L. acidophilus* strains, SipA, or PBS every other day until day 14 (Supplementary...
DSS-induced colitis

WT and KO mice were treated with 3% DSS in the drinking water for 5 days (made fresh every 2–3 days) to induce colitis. Mice were monitored for disease progression through day 10 after treatment as described above. For prevention studies, mice were orally gavaged with NRS56, NC2187, or SigA at days −3 and −1 and then every other day after 3% DSS treatment, for a total of five gavages (two before, and 3 after 3% DSS). Due to the acute nature of this model, and the fact that our bacterial strains are cleared from B6 mice within 3 days, we employed this treatment regimen to ensure the continuous presence of the bacteria during the acute inflammatory window.

Histopathology

Colitis severity in T-cell-, DSS- and C. rodentium infection-induced colitis was determined by histopathology. Tissues were fixed, sectioned, and stained with hematoxylin and eosin (Histology Tech Services, Gainesville, FL, USA). Stained sections were blindly evaluated by a board-cred veterinary pathologist. Colitis was graded based on seven parameters (0–7), as previously described (Cheng et al., 2014).

FITC-dextran intestinal permeability assay

Passive transmural absorption of FITC-labeled dextran (Sigma-Aldrich) in vivo was used to determine intestinal barrier function as previously described (Nappo et al., 1996). Mice were gavaged with FITC-dextran, MW: 40,000 (60 mg/100 g body weight). Blood was collected retro-orbitally after proper anesthesia; mice were sacrificed after blood collection. Fluorescence intensity in the serum was measured with a fluorimeter (485 nm excitation, 519 nm emission). FITC-dextran concentrations in the mouse sera were determined from standard curves generated by serial dilution of FITC-dextran using blank subtraction in the test samples using sera from mice that were not gavaged with the permeability tracer.

Colonoscopy of DSS- and T-cell-induced colitis mice

Macropscopic damage in the colon of B6 Rag-1−/−, WT, and KO mice was visualized with a Multi-Purpose Rigid Telescope attached to a TELEPACK X (Karl Storz Endoscopy, Germany). Mice were killed for 4–6 h, and subsequently, the colons of the living subjects were imaged under appropriate anaesthetic conditions.

Real-time PCR and 16S ribosomal DNA sequencing

Colon tissues from B6 Rag-1−/−, WT, and KO mice were isolated and processed for changes in gene expression as previously described (Lightfoot et al., 2014). Microbiota analyses were performed on the Illumina MiSeq (Illumina, Inc., San Diego) as outlined previously (Lightfoot et al., 2014). Primers used, as well as their sequences, are listed in Supplementary Table S7.

SIGNR1, SIGIRR, and DC-SIGN binding assays

C-type lectin receptors, SIGNR1 and SIGIRR, were fused to the Fc part of human IgG1 (SIGNR1-Fc and SIGIRR-Fc) as previously described (Eriksrud et al., 2013). Brieﬂy, the extracellular regions of murine SIGNR1 and SIGIRR were ampliﬁed and ligated into the expression vector pHLUG-myc-Fc2 (Introgen, Toulouse, France) for expression in myeloma-free CHO-K1 cells (ATCC). Expression in CHO-S cells was driven by an HIV-1LTR promoter, and secretion into the culture supernatant was mediated by an external IL2 signal sequence (IL2s) (Eriksrud et al., 2013). Fusion proteins were incubated with SigA-coated beads and subsequently with anti-Fc antibodies for detection of binding to beads. Additionally, CHO-S cells were transfected with SIGNR1, SIGIRR, or DC-SIGN-expressing commercial vectors (Supplementary Fig S1) and 48 h post-transfection, cells were incubated with beads coated with labeled SigA for 30 min. The recombinant human DC-SIGN/CDC209-Fc chimera was purchased from R&D Systems. Binding of SigA-coated beads (Dynabeads MyOne Carboxylic Acid, Life Technologies) to fusion proteins or CHO-S cells was analyzed by flow cytometry.

Statistical analyses

Representative data indicate mean ± SEM. Signiﬁcance was determined by two-tailed unpaired t-tests for two-group comparisons (GraphPad Prism v6.0f for Mac OS X, La Jolla, CA, USA). Statistical signiﬁcance for differences in weight, diarrhea score, and FOB score was calculated using multiple unpaired t-tests correcting for multiple comparisons with the Holm-Sidak method in Prism v6.0f.

Supplementary information for this article is available online at http://embj.oxfordjournals.org

Acknowledgements

This work was supported by NIH R01, AI069870, the Department of Defense CA111002, the NHICCR Clinical and Translational Science Award, the Ocala Royal Dames, Gatorade Foundation, the Florida Breast Cancer Foundation (UCF), the University of Florida (L.R.; RR11999), NIAID NIH FR2 DK101367 (Y.L.), and the North Carolina Agricultural Foundation. B.L. acknowledges funding from the German Federal Ministry of Education and Research (Flz 0315464) and the Collaborative Research Center (SFB) 765. The authors wish to acknowledge the NIH-sponsored Mouse Model Regional Resource Center (MMRRC) National System as the source of genetically altered materials for use in this study, strain CD1 09-UCR. The materials were produced and deposited to the MMRRC by the Consortium for Functional Dysbiotics supported by the National Institute of General Medical Sciences (EAM2319). We thank Ms. Ashleigh Mica (L.R.); Dr. April English (L.R.); and Mr. Brandon Johnson (N.P.) for technical support.

Author contributions

All authors contributed extensively to the work presented here and discussed data at all stages of the manuscript, which contributed to the writing of the manuscript. All authors read and approved the final manuscript. Y.L. designed and performed experiments, analyzed data, and wrote the paper; T.V. performed
microscopy and RT-PCR experiments and analyzed data. YC and KS constructed the NCC217R strain. K5 purified L. acidophilus SP4. KS purified L. acidophilus SP4. performed experiments and analyzed data. NZ designed and performed experiments. JL evaluated and scored all colon tissue sections. NC designed and performed experiments. CL performed all histological analyses for interpretation of microscope data. TI and IB, contributed reagents. TRK conceived and implemented the technology for developing L. acidophilus mutant strains. NM conceived the project, supervised the immunologic and bacteriologic aspects of the work, designed experiments, and wrote the paper.

Conflict of interest
The authors declare that they have no conflict of interest.

References

Lupp C, Robertson ML, Wickham ML, Sierou KN, Champion OL, Caynor EC, Pauley W (2013) Suppressive TGF-β1 mediated inflammation disrupts the intestinal microbiota and promotes the outgrowth of enterobacteria. Cell Host Microbe 20: 204
T, Kleinhammer TR (2011) Regulation of induced colonic inflammation by Lactobacillus acidophilus deficient in lipoteichoic acid. Proc Natl Acad Sci USA 108(Suppl. 3): 14623–14630


Oseofo F, Reis e Sousa C (2011) Myeloid C-type lectin receptors in pathogen recognition and host defense. Immunity 34: 651–664


