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

Rosemary Sanozky-Dawes,1† Kurt Selle,1,2† Sarah O'Flaherty,1 Todd Klaenhammer1,2 and Rodolphe Barrangou1,2

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 a 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 phage 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’-NTAA-3’ flanking the 3’ end of the protospacer. The system in *L. gasseri JV-V03* and NCK 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 archaeal genomes (Grissa et al., 2007a). 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 (Barrangou et al., 2007). CRISPR arrays are transcribed constitutively and inductively, 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 cas3, 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* (Bondy-Denomy & Davidson, 2014). Intensive study of *Streptococcus thermophilus* has provided valuable insights into how CRISPR-Cas systems shape the dynamic interplay between phages and their bacterial hosts (Barrangou et al., 2007;
Levin et al., 2013; Paez-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 (Luo et al., 2015; Doudna & 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., 2013; 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; Sapranauskas et al., 2011), the trans-activating CRISPR RNA (tracrRNA) (Deltcheva et al., 2011) and maturation of crRNAs by RNase III (Chylinski et al., 2014). Cas9 targets DNA for doubled-stranded cleavage, which occurs through dual HNH and RuvC nickase domains (Garneau et al., 2010; Gasiunas et al., 2012; Sapranauskas et al., 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 autochthonous 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

<table>
<thead>
<tr>
<th>Strains</th>
<th>Relevant characteristic/source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L. gasseri</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCK 99 (ADH, ATCC 19992)</td>
<td>Human faecal (strain ADH)</td>
<td>ATCC</td>
</tr>
<tr>
<td>NCK 334 (ATCC 33323)</td>
<td>Human isolate, neotype</td>
<td>ATCC</td>
</tr>
<tr>
<td>NCK 1340 (AM1)</td>
<td>Patient endoscopy</td>
<td>Kullen et al. (2000)</td>
</tr>
<tr>
<td>NCK 1341 (JK12)</td>
<td>Healthy faecal</td>
<td>Kullen et al. (2000)</td>
</tr>
<tr>
<td>NCK 1343 (SD10)</td>
<td>Healthy faecal</td>
<td>Kullen et al. (2000)</td>
</tr>
<tr>
<td>NCK 1344 (FR2)</td>
<td>Healthy endoscopy</td>
<td>Kullen et al. (2000)</td>
</tr>
<tr>
<td>NCK 1348 (RF14)</td>
<td>Healthy faecal</td>
<td>Kullen et al. (2000)</td>
</tr>
<tr>
<td>NCK 1349 (RF81)</td>
<td>Healthy faecal</td>
<td>Kullen et al. (2000)</td>
</tr>
<tr>
<td>NCK 1557</td>
<td>Frozen yogurt</td>
<td>Clark (2001)</td>
</tr>
<tr>
<td>NCK 2140</td>
<td>Newborn faecal</td>
<td>Rodrigues da Cunha et al. (2012)</td>
</tr>
<tr>
<td>NCK 2141</td>
<td>Newborn faecal</td>
<td>Rodrigues da Cunha et al. (2012)</td>
</tr>
<tr>
<td><strong>CRISPR-Cas-containing L. gasseri strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCK 1338 (WD19)</td>
<td>Patient endoscopy</td>
<td>Kullen et al. (2000)</td>
</tr>
<tr>
<td>NCK 1342 (JG141)</td>
<td>Patient endoscopy</td>
<td>Kullen et al. (2000)</td>
</tr>
<tr>
<td>NCK 1345 (FR4)</td>
<td>Healthy endoscopy</td>
<td>Kullen et al. (2000)</td>
</tr>
<tr>
<td>NCK 1346 (ML1)</td>
<td>Healthy endoscopy</td>
<td>Kullen et al. (2000)</td>
</tr>
<tr>
<td>NCK 1347 (ML3)</td>
<td>Healthy endoscopy</td>
<td>Kullen et al. (2000)</td>
</tr>
<tr>
<td>JV-V03</td>
<td>Female genito-urinary tract</td>
<td>BEI resources</td>
</tr>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCI061</td>
<td>Transformation host</td>
<td>Casadaban &amp; Cohen (1980)</td>
</tr>
<tr>
<td>NCK 2346</td>
<td>MCI061 transformant with pTRK 1090</td>
<td>This study</td>
</tr>
<tr>
<td>NCK 2347</td>
<td>MCI061 transformant with pTRK 1091</td>
<td>This study</td>
</tr>
<tr>
<td>NCK 2348</td>
<td>MCI061 transformant with pTRK 1092</td>
<td>This study</td>
</tr>
<tr>
<td>NCK 2349</td>
<td>MCI061 transformant with pTRK 1093</td>
<td>This study</td>
</tr>
<tr>
<td>NCK 2350</td>
<td>MCI061 transformant with pTRK 1094</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pGK12</td>
<td>ori (pWV01), Em R, Cm R RepA +</td>
<td>Kok et al. (1984)</td>
</tr>
<tr>
<td>pTRK 1090</td>
<td>pGK12::protospacer/PAM 42i</td>
<td>This study</td>
</tr>
<tr>
<td>pTRK 1091</td>
<td>pGK12::protospacer/PAM 42NP</td>
<td>This study</td>
</tr>
<tr>
<td>pTRK 1092</td>
<td>pGK12::protospacer/PAM JV1</td>
<td>This study</td>
</tr>
<tr>
<td>pTRK 1093</td>
<td>pGK12::protospacer/PAM JVNP</td>
<td>This study</td>
</tr>
<tr>
<td>pTRK 1094</td>
<td>pGK12::protospacer/PAM JV37i</td>
<td>This study</td>
</tr>
</tbody>
</table>

ATCC, American Type Culture Collection.
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; Coffey & Ross, 2002). Indeed, various temperate bacteriophages specific to *L. gasseri* have been sequenced or characterized (Raya et al., 2007; Ismail et al., 2009; Baugher 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 function.

The system exhibited maintenance of CRISPR arrays, activity, based on key characteristics alluding to its function—et al. Indeed, various temperate bacteriophages specific to *L. gasseri* have been sequenced or characterized (Raya et al., 2007; Ismail et al., 2009; Baugher 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 function.

### 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 ml⁻¹ of gentamycin (Em) ml⁻¹. The *L. gasseri* strains were propagated statically in MRS, and genomic DNA extracted using an UltraClean microbial DNA isolation kit (Mo Bio) and quantified with UltraPure distilled water (Invitrogen). Twenty-three microlitres Master Mix from the Rep-PCR Lactobacillus-based DiversiLab kit (bioMérieux) was added to a tube with 2 μl 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 30 s at 94 °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 DiversiLab system chip along with the DiversiLab DNA reagents and supplies (bioMérieux), according to the manufacturer’s protocol. The chip samples were analysed using DiversiLab software version 3.4 and similarity of strains was determined by comparing the resultant electropherogram/barcodes.

### CRISPR strain genotyping.

The CRISPR database, CRISPR db (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* JV-V03 (GenBank accession no. ACGO00000000). *L. gasseri* JV-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*, HM-104. Primers were designed to amplify the CRISPR-cas locus from the draft genome sequence of *L. gasseri* JV-V03. The National Center for Biotechnology Information (NCBI) Basic Local Alignment Sequence Tool (BLASTN; Altschul et al., 1990) was used to determine *L. gasseri* spacer matches (Table 2).

### 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 (Mo Bio) and quantified using a Nanodrop 1000 spectrophotometer (Thermo-Scientific), then normalized to 20 ng μl⁻¹ with UltraPure distilled water (Invitrogen). Twenty-three microlitres Master Mix from the Rep-PCR Lacto- bacillus-based DiversiLab kit (bioMérieux) was added to a tube with 2 μl 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 30 s at 94 °C (denaturation), 30 s at 55 °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 DiversiLab system chip along with the DiversiLab DNA reagents and supplies (bio- Mérieux), according to the manufacturer’s protocol. The chip samples were analysed using DiversiLab software version 3.4 and similarity of strains was determined by comparing the resultant electropherogram/barcodes.

### Table 2. Spacer sequences and BLAST matches

<table>
<thead>
<tr>
<th>Strain</th>
<th>Spacer no.</th>
<th>Spacer/protospacer match</th>
<th>Flank</th>
<th>Identity/accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCK 1346</td>
<td>10</td>
<td>AGCAAGATTTCACAATTAGCTTAA</td>
<td>GTTCAA</td>
<td>Phi-ADH phage AJ131519.1</td>
</tr>
<tr>
<td>NCK 1342</td>
<td>11</td>
<td>CCTAGATGATCCTGCTATTAGATA</td>
<td>GTAAAT</td>
<td>Prophage Lj965</td>
</tr>
<tr>
<td>NCK 1342</td>
<td>8</td>
<td>GCAATATCAATGTCGCTTTAGTAA</td>
<td>ATGAAA</td>
<td>Phage KC5a</td>
</tr>
<tr>
<td>JV-V03</td>
<td>2</td>
<td>AATTCTTGTTAATGACGGTCAATT</td>
<td>ACTAAG</td>
<td>Plasmid pWCFS103</td>
</tr>
<tr>
<td>JV-V03</td>
<td>12</td>
<td>ATTCTTGTTAATGACGGTCAATT</td>
<td>CTAATG</td>
<td>Phage TD1</td>
</tr>
<tr>
<td>JV-V03</td>
<td>18</td>
<td>TATGCTTCGAAAGTTAATTAGGTTA</td>
<td>TTTGTA</td>
<td>Plasmid pHN1</td>
</tr>
<tr>
<td>K7</td>
<td>6</td>
<td>CAAGATTTAAACAGATATGTTAGGC</td>
<td>ATGAAA</td>
<td>Plasmid pWN2</td>
</tr>
<tr>
<td>K7</td>
<td>8</td>
<td>GACGTCTTTAAATAGGTGTTGTTG</td>
<td>ACTAAG</td>
<td>Plasmid pWN2</td>
</tr>
<tr>
<td>NCK 1345</td>
<td>29</td>
<td>TCTCTCGAAAAAAGTTGTTGCTT</td>
<td>GTAAA</td>
<td>phi jbl1</td>
</tr>
<tr>
<td>NCK 1345</td>
<td>21</td>
<td>TTCCTTCGAAAAAGTTGTTGCTT</td>
<td>ATAGAA</td>
<td>Plasmid pLgLA39</td>
</tr>
<tr>
<td>NCK 1347</td>
<td>22</td>
<td>TCGTTTGAAGAGAAAGTTGTTGCT</td>
<td>GTAAA</td>
<td>phi jbl1</td>
</tr>
<tr>
<td>NCK 1347</td>
<td>17</td>
<td>CATCAGTACAGCAAGCTTACAGTT</td>
<td>ATAGAA</td>
<td>Plasmid pWCFS103</td>
</tr>
</tbody>
</table>
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 amplicons for cloning and screening were generated using standard protocols and Choice-Taq blue DNA polymerase (Denville Scientific). PCR products were analysed in 1 % agarose gels 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 Davis 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 Eco1CR1-digested (Promega) pGK12. The constructs were transformed into rubidium chloride competent *E. coli* MC1061 (Hanahan, 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 site to confirm insert sequence accuracy prior to transformation. Transformation of *L. gasseri* NCK 1342, and JV-V03 with the interference plasmids was performed as described previously (Walker et al., 1996).

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 csn2 (Fig. 2) (Chylinski et al., 2013; Briner et al., 2014). Since CRISPR array/Cas operon maintenance is an indication of function, 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 (mrcasF, mrcasR) and for CRISPR arrays using primers designed to flank the locus (hgflk F2, hgflk 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 analyses, 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 genito-urinary 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 protospacers 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 (Esvelt et al., 2013). However, the absence of any guanosines correlates with the low G+C content of L. gasseri phage genomes (Baugher 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; Shariat 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 divergence of L. gasseri strains, which mirrors the diversity observed using the Rep-PCR method, and from PFGE 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 \textit{in vivo}. Two of the strains that contain complete CRISPR-Cas systems, but different spacer compositions, were selected for the interference experiments. \textit{L. gasseri} NCK 1342, an isolate from a renal patient’s endoscopy, and JV-V03, a female genitourinary tract isolate that was part of the Human Microbiome Project, were used (Table 1). Protospacer/PAM sequences corresponding to the leader proximate spacer in each strain were cloned separately into the shuttle vector pGK12. In this scenario, the protospacer/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 protospacer/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; Gomaa \textit{et al.}, 2014), targeting of plasmids exhibiting perfect spacer :: PAM combinations manifested 3.13 and 3.89 log reductions in transformants of \textit{L. gasseri} NCK 1342 and JV-V03, respectively (Fig. 5). The reduction occurred in both NCK 1342 and JV-V03 with different protospacer-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’-GCTC-3’) or a single nt deletion at position 30 within the JV-V03 spacer abolished targeting of the pGK12 :: protospacer/PAM (Fig. 5). This is in accordance with the well-established roles of the PAM and seed sequences in CRISPR-Cas activity (Deveau \textit{et al.}, 2008; Horvath \textit{et al.}, 2008; Mojica \textit{et al.}, 2009; Semenova \textit{et al.}, 2011; Wiedenheft \textit{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 protospacer/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 protospacer/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
sequences. Other possible escape scenarios could be mutations in the cas coding sequences. PCR amplification of cas9 from genomic DNA of some remaining JV-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 lysogen–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. thermophilus* (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., 2009; 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 (Gomaa et al., 2014) and programmable transcriptional regulation (Bikard et al., 2013; Luo et al., 2015). *In silico* surveys of lactic acid bacterial genomes suggest that these micro-organisms are a significant reservoir for orthogonal type II CRISPR-Cas systems (Chylinski et al., 2013; Briner et al., 2013).

**Fig. 6.** Analysis of escape mechanisms from the CRISPR-Cas system. Amplification of spacer arrays from NCK 1342 and JV-V03 transformants (Tf) that escaped the CRISPR-Cas system shows smaller repeat-spacer arrays. Sequencing of those amplicons revealed which spacers had been lost. (a) NCK 1342 agarose gel labels indicate the escapee transformant designations. Escapees Tf 1 and Tf 2 (top schematic) are missing the most recent spacer used to create the NCK 1342 interference plasmid (pTRK1090) and escapee Tf 5 (lower schematic) has lost the most recent ten spacers and repeat sequences. (b) JV-V03 agarose gel labels indicate the escapee transformant designations. Escapees Tf 1, Tf 5 and Tf 9 (top schematic) have lost the most recent three spacers and escapees Tf 3, Tf 4 and Tf 6 (lower schematic) are missing the most recent nine spacers.
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 bacterium–phage 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 Briner and Evelyn Durmaz for technical assistance, Dr Yong Jun Goh for insightful discussions and are grateful to Dr Philippe Horvath for providing the CRISPR spacer Excel Macro. K.S. and R.B. are co-inventors on patent applications related to various uses of CRISPR-Cas systems.

REFERENCES


