Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium

Michael E. Pyne, Mark R. Bruder, Murray Moo-Young, Duane A. Chung & C. Perry Chou

Application of CRISPR-Cas9 systems has revolutionized genome editing across all domains of life. Here we report implementation of the heterologous Type II CRISPR-Cas9 system in Clostridium pasteurianum for markerless genome editing. Since 74% of species harbor CRISPR-Cas loci in Clostridium, we also explored the prospect of co-opting host-encoded CRISPR-Cas machinery for genome editing. Motivation for this work was bolstered from the observation that plasmids expressing heterologous cas9 result in poor transformation of Clostridium. To address this barrier and establish proof-of-concept, we focus on characterization and exploitation of the C. pasteurianum Type I-B CRISPR-Cas system.

In silico spacer analysis and in vivo interference assays revealed three protospacer adjacent motif (PAM) sequences required for site-specific nucleolytic attack. Introduction of a synthetic CRISPR array and cpaAIR gene deletion template yielded an editing efficiency of 100%. In contrast, the heterologous Type II CRISPR-Cas9 system generated only 25% of the total yield of edited cells, suggesting that native machinery provides a superior foundation for genome editing by precluding expression of cas9 in trans. To broaden our approach, we also identified putative PAM sequences in three key species of Clostridium. This is the first report of genome editing through harnessing native CRISPR-Cas machinery in Clostridium.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) and CRISPR-associated (Cas) proteins comprise the basis of adaptive immunity in bacteria and archaea. CRISPR-Cas systems are currently grouped into six broad types, designated Type I through VI. CRISPR-Cas Types I, II, and III, the most prevalent systems in both archaea and bacteria, are differentiated by the presence of cas3, cas9, or cas10 signature genes, respectively. Based on the composition and arrangement of cas gene operons, CRISPR-Cas systems are further divided into 16 distinct subtypes. Type I systems, comprised of six distinct subtypes (I-A to I-F), exhibit the greatest diversity and subtype I-B is the most abundant CRISPR-Cas system represented in nature. CRISPR-Cas loci have been identified in 45% of bacteria and 84% of archaea due to widespread horizontal transfer of CRISPR-Cas loci within the prokaryotes.

CRISPR-based immunity encompasses three distinct processes, termed adaptation, expression, and interference. Adaptation involves the acquisition of specific nucleotide sequence tags, referred to as protospacers in their native context within invading genetic elements, particularly bacteriophages (phages) and plasmids. During periods of predation, protospacers are rapidly acquired and incorporated into the host genome, where they are subsequently referred to as spacers. Cas1 and Cas2, which form a complex that mediates acquisition of new spacers, are the only proteins conserved between all CRISPR-Cas subtypes. Chromosomally-encoded spacers are flanked by 24–48 bp partially-palindromic direct repeat sequences, iterations of which constitute CRISPR arrays. Up to 587 spacers have been identified within a single CRISPR array, exemplifying the exceptional level of attack experienced by many microorganisms in nature. During the expression phase of CRISPR immunity, acquired spacer sequences are expressed and, in conjunction with Cas proteins, provide resistance against invading genetic elements.
elements. CRISPR arrays are first transcribed into a single large precursor crRNA by a promoter located within the CRISPR leader (lead). The resulting transcript is cleaved and processed into individual mature crRNAs by the Cas6 endonuclease (Type I systems) or the ubiquitous RNase III enzyme (Type II systems). Processing is mediated by characteristic secondary structures (hairpins) formed by Type I pre-crRNAs or by a trans-activating RNA (tracrRNA; brown) possessing homology to direct repeat sequences in Type II systems. A single synthetic guide RNA (gRNA) can replace the dual crRNA-tracrRNA interaction (not shown). Mature crRNAs are guided to invading nucleic acids through homology between crRNAs and the corresponding invader protospacer sequence. Type I interference requires the multiprotein Cascade complex (comprised of cas6-cas8b-cas7-cas5 in Clostridium difficile and C. pasteurianum), encoded downstream of the Type I CRISPR array. Type I and II interference mechanisms require recognition of one of multiple protospacer adjacent motif (PAM) sequences, which collectively comprise the consensus PAM element (red). The location of the PAM and the site of nucleolytic attack relative to the protospacer sequence differs between Type I and II CRISPR-Cas systems. Representative PAM sequences from C. difficile (Type I-B) and Streptococcus pyogenes (Type II) CRISPR-Cas loci are shown. Nucleolytic attack by Cas3 or Cas9 results in a DNA nick (DN) or blunt double-stranded DNA break (DB), respectively. Both CRISPR-Cas loci contain cas1 and cas2 genes (not shown), while the Type I and II loci also contain cas4 and csn2 genes, respectively (not shown).

Figure 1. Comparison of Type I (left) and Type II (right) CRISPR-Cas interference mechanisms. CRISPR arrays, comprised of direct repeats (DRs; royal blue and dark green) and spacer tags (light blue and light green) are first transcribed into a single large pre-crRNA by a promoter located within the CRISPR leader (lead). The resulting transcript is cleaved and processed into individual mature crRNAs by the Cas6 endonuclease (Type I systems) or the ubiquitous RNase III enzyme (Type II systems). Processing is mediated by characteristic secondary structures (hairpins) formed by Type I pre-crRNAs or by a trans-activating RNA (tracrRNA; brown) possessing homology to direct repeat sequences in Type II systems. A single synthetic guide RNA (gRNA) can replace the dual crRNA-tracrRNA interaction (not shown). Mature crRNAs are guided to invading nucleic acids through homology between crRNAs and the corresponding invader protospacer sequence. Type I interference requires the multiprotein Cascade complex (comprised of cas6-cas8b-cas7-cas5 in Clostridium difficile and C. pasteurianum), encoded downstream of the Type I CRISPR array. Type I and II interference mechanisms require recognition of one of multiple protospacer adjacent motif (PAM) sequences, which collectively comprise the consensus PAM element (red). The location of the PAM and the site of nucleolytic attack relative to the protospacer sequence differs between Type I and II CRISPR-Cas systems. Representative PAM sequences from C. difficile (Type I-B) and Streptococcus pyogenes (Type II) CRISPR-Cas loci are shown. Nucleolytic attack by Cas3 or Cas9 results in a DNA nick (DN) or blunt double-stranded DNA break (DB), respectively. Both CRISPR-Cas loci contain cas1 and cas2 genes (not shown), while the Type I and II loci also contain cas4 and csn2 genes, respectively (not shown).
Cas9 cleaves DNA three nucleotides upstream of the PAM element\textsuperscript{21,22}, while Cas3 nicks the PAM-complementary strand outside of the area of interaction with crRNA\textsuperscript{20}.

Owing to the simplicity of CRISPR-Cas9 interference in Type II systems, the \textit{S. pyogenes} CRISPR-Cas9 machinery has recently been implemented for extensive genome editing in a wide range of organisms, such as \textit{E. coli}\textsuperscript{27-29}, yeasts\textsuperscript{30,31}, mice\textsuperscript{32}, zebrafish\textsuperscript{33}, plants\textsuperscript{34}, and human cells\textsuperscript{35,36}. In bacteria, CRISPR-based methods of genome editing signify a critical divergence from traditional techniques of genetic manipulation involving the use of chromosomally-encoded antibiotic resistance markers, which must be excised and recycled following each successive round of integration\textsuperscript{37}. Within \textit{Clostridium}, a genus with immense importance to medical and industrial biotechnology\textsuperscript{38,39}, as well as human disease\textsuperscript{40}, genetic engineering technologies are notoriously immature, as the genus suffers from overall low transformation efficiencies and poor homologous recombination\textsuperscript{41}. Existing clostridial genome engineering methods, based on mobile group II introns, antibiotic resistance determinants, and counter-selectable markers, are laborious, technically challenging, and often ineffective\textsuperscript{42-44}. In contrast, CRISPR-based methodologies provide a powerful means of selecting rare recombination events, even in strains suffering from poor homologous recombination. Such strategies have been shown to be highly robust, frequently generating editing efficiencies up to 100\%\textsuperscript{27,29,45}. Accordingly, the \textit{S. pyogenes} Type II CRISPR-Cas9 system has recently been adapted for use in \textit{C. beijerinckii}\textsuperscript{46} and \textit{C. cellulolyticum}\textsuperscript{47}, facilitating highly precise genetic modification of clostridial genomes and paving the way for robust genome editing in industrial and pathogenic clostridia.

Here we report development of broadly applicable strategies of markerless genome editing based on exploitation of both heterologous (Type II) and endogenous (Type I) bacterial CRISPR-Cas systems in \textit{C. pasteurianum}, an organism possessing substantial biotechnological potential for conversion of waste glycerol to butanol as a prospective biofuel\textsuperscript{48}. While various tools for genetic manipulation of \textit{C. pasteurianum} are under active development recently\textsuperscript{49,50}, effective site-specific genome editing for this organism is lacking. In this study, we demonstrate the first implementation of \textit{S. pyogenes} Type II CRISPR-Cas9 machinery for markerless and site-specific genome editing in \textit{C. pasteurianum}. Recently, we sequenced the \textit{C. pasteurianum} genome\textsuperscript{51} and identified a central Type I-B CRISPR-Cas locus, which we exploit here as a chassis for genome editing based on earlier successes harnessing endogenous CRISPR-Cas loci in other bacteria\textsuperscript{52,53}. Our strategy encompasses plasmid-borne expression of a synthetic Type I-B CRISPR array that can be site-specifically programmed to any gene within the organism’s genome. Providing an editing template designed to delete the chromosomal protospacer and adjacent PAM yields an editing efficiency of 100\% based on screening of 10 representative colonies. To our knowledge, the approach described here is the first report of genome editing in \textit{Clostridium} by co-opting native CRISPR-Cas machinery. Importantly, our strategy is broadly applicable to any bacterium or archaeon that encodes a functional CRISPR-Cas locus and appears to yield more edited cells compared to the commonly employed heterologous Type II CRISPR-Cas9 system.

Results

Implementation of the Type II CRISPR-Cas9 system for genome editing in \textit{C. pasteurianum}.

Recently, two groups reported a CRISPR-based methodology employing the Type II system from \textit{S. pyogenes} for use in genome editing of \textit{C. beijerinckii} and \textit{C. cellulolyticum}\textsuperscript{46,47}. This system requires expression of the \textit{cas9} endonuclease gene in \textit{trans}, in addition to a chimeric guide RNA (gRNA) containing a programmable RNA spacer. To determine if the \textit{S. pyogenes} machinery could also function for genome editing in \textit{C. pasteurianum}, we constructed a Type II CRISPR-Cas9 vector by placing \textit{cas9} under constitutive control of the \textit{C. pasteurianum} thiolase (\textit{thl}) gene promoter and designing a synthetic gRNA expressed from the \textit{C. beijerinckii} \textit{sCbe1}_5830 small RNA promoter\textsuperscript{46}. We selected the \textit{cpaAIR} gene as a target double-stranded DB site through the use of a 20 nt spacer located within the \textit{cpaAIR} coding sequence, as this gene has been previously disrupted in \textit{C. pasteurianum}\textsuperscript{50}. An \textit{S. pyogenes} Type II PM sequence (5’-NGG-3’), required for recognition and subsequent cleavage by \textit{Cas9}\textsuperscript{27}, is located at the 3’ end of the \textit{cpaAIR} protospacer sequence within the genome of \textit{C. pasteurianum} (Fig. 2a). Transformation of \textit{C. pasteurianum} with the resulting vector, designated pCas9gRNA-cpaAIR, yielded an average transformation efficiency of 0.03 colony-forming units (CFU) \(\mu\text{g}^{-1}\) DNA (Fig. 2b). Only one out of five attempts at transfer of pCas9gRNA-cpaAIR produced a single transformant, indicating efficient Cas9-mediated killing of host cells. To demonstrate genome editing using this system, we constructed pCas9gRNA-delcpaAIR through introduction of a \textit{cpaAIR} gene deletion editing cassette into plasmid pCas9gRNA-cpaAIR. The editing cassette was designed to contain 1,029 bp and 1,057 bp homology regions to the \textit{cpaAIR} locus, which together flank the putative \textit{cpaAIR} double-stranded DB site. Homologous recombination between the plasmid-borne editing cassette and the \textit{C. pasteurianum} chromosome is expected to result in a \textit{cpaAIR} gene deletion comprising 567 bp of the \textit{cpaAIR} coding sequence, including the protospacer and associated PAM element required for \textit{Cas9} attack, and 19 bp of the upstream \textit{cpaAIR} gene region, including the putative \textit{cpaAIR} gene promoter (Fig. 2a). Compared to the lethal pCas9gRNA-cpaAIR vector, introduction of pCas9gRNA-delcpaAIR established transformation. A transformation efficiency of 2.6 \(\text{CFU} \mu\text{g}^{-1}\) DNA was obtained using pCas9gRNA-delcpaAIR, an 87-fold increase compared to pCas9gRNA-cpaAIR (Fig. 2b). Genotyping of 10 pCas9gRNA-delcpaAIR transformants generated the expected PCR product corresponding to \textit{cpaAIR} gene deletion, resulting in an editing efficiency of 100\% (Fig. 2c). Sanger sequencing of a single pCas9gRNA-delcpaAIR transformant confirmed successful deletion of a 762 bp region of the \textit{cpaAIR} coding sequence (data not shown).

Despite an editing efficiency of 100\% using heterologous Type II CRISPR-Cas9 machinery, an average of only 47 total CFU were obtained by introducing 15–25 \(\mu\text{g}\) of pCas9gRNA-delcpaAIR plasmid DNA (2.6 \(\text{CFU} \mu\text{g}^{-1}\) DNA). Such a low transformation efficiency may impede more ambitious genome editing strategies, such as integration of large DNA constructs and multiplexed editing. Since expression of the \textit{Cas9} endonuclease has been shown to be moderately toxic in a multitude of organisms [e.g. \textit{mycobacteria}, yeast, algae, and \textit{mice}\textsuperscript{54-56}], even in the absence of a targeting gRNA, we prepared various \textit{cas9}-expressing plasmid constructs to determine
if expression of cas9 leads to reduced levels of transformation. Introduction of a cas9 expression cassette lacking a gRNA into plasmid pMTL85141 (transformation efficiency of $6.3 \times 10^8$ CFU µg$^{-1}$ DNA), generating p85Cas9, resulted in a reduction in transformation efficiency of more than two orders of magnitude (26 CFU CFU µg$^{-1}$ DNA) (Fig. 2b). Modifying the pIM13 replication module of p85Cas9 to one based on pCB102 in plasmid p83Cas9 further reduced transformation to barely detectable levels (0.7 CFU µg$^{-1}$ DNA). Importantly, transformation of C. pasteurianum with p85delCas9, generated through deletion of the putative cas9 gene promoter in p85Cas9, restored transformation to typical levels ($2.2 \times 10^3$ CFU µg$^{-1}$ DNA). Collectively these data demonstrate that expression of Cas9 in the absence of a gRNA significantly reduces transformation of C. pasteurianum. It is noteworthy that we also observed a dramatically reduced level of transformation of Clostridium acetobutylicum.

**Figure 2.** Genome editing in C. pasteurianum using the heterologous S. pyogenes Type II CRISPR-Cas9 system. (a) cpaAIR gene deletion strategy using Type II CRISPR-Cas9. Introduction of a double-stranded DB to the cpaAIR locus was achieved by programming a gRNA spacer sequence (green) and expressing heterologous cas9 within plasmid pCas9gRNA-cpaAIR. cpaAIR-targeted gRNA, containing cas9 binding handle (orange), is directed to the chromosomal cpaAIR gene through base-pairing to the protospacer sequence and Cas9-recognition of the S. pyogenes PAM element (5'-NGG-3'; red). Insertion of a cpaAIR gene editing cassette in pCas9gRNA-cpaAIR, generating pCas9gRNA-delcpaAIR, leads to homologous recombination and deletion of a portion of the cpaAIR coding sequence, including the protospacer and PAM elements. Unmodified cells are selected against by Cas9 cleavage, while edited cells possessing a partial cpaAIR deletion are able evade attack. Genes, genomic regions, and plasmids are not depicted to scale. (b) Transformation efficiency corresponding to Type II CRISPR-Cas9 vectors (pCas9gRNA-cpaAIR and pCas9gRNA-delcpaAIR) and various cas9 expression derivatives and control constructs (pMTL85141, p85Cas9, p83Cas9, p85delCas9). Transformation efficiency is reported as the number of CFU generated per µg of plasmid DNA. Data shown are averages resulting from at least two independent experiments and error bars depict standard deviation. (c) Colony PCR genotyping of pCas9gRNA-delcpaAIR transformants. Primers cpaAIR.S and cpaAIR.AS were utilized in colony PCR to screen 10 colonies harboring pCas9gRNA-delcpaAIR. Expected product sizes are shown corresponding to the wild-type (2,913 bp) and the cpaAIR deletion mutant (2,151 bp) strains of C. pasteurianum. Lane 1: linear DNA marker; lane 2: no colony control; lanes 3: wild-type colony; 4: colony harboring pCas9gRNA-cpaAIR; lanes 5–14: colonies harboring pCas9gRNA-delcpaAIR.
using plasmid p85Cas9, which could also be rescued through deletion of the cas9 gene promoter in p85delCas9 (data not shown).

**Analysis of the **C. pasteurianum** Type I-B CRISPR-Cas system and identification of putative protospacer matches to host-specified spacers.** Due to the inhibitory effect of cas9 expression on transformation, we reasoned that the S. pyogenes Type II CRISPR-Cas9 system imposes significant limitations on genome editing in Clostridium, as the clostridia are transformed at substantially lower levels compared to most bacteria. To evade poor transformation of cas9-encoded plasmids, we investigated the prospect of genome editing using endogenous CRISPR-Cas machinery. We recently sequenced the genome of C. pasteurianum and unveiled a CRISPR-Cas system comprised of a 37-spacer CRISPR array upstream of a core cas gene operon (cas6-cas8b-cas7-cas5-cas3-cas4-cas1-cas2) (abbreviated cas68b753412). A promoter within the putative leader sequence (lead) drives transcription of the CRISPR array.

We used BLAST and PHAST to analyze all 45 spacer tags specified in the C. pasteurianum genome in an attempt to identify protospacer matches from invading nucleic acid elements, including phages, prophages, plasmids, and transposons. Since seed sequences, rather than full-length protospacers, have been shown to guide CRISPR interference, mismatches in the PAM-distal region of protospacer were permitted, while spacer-protospacer matches possessing more than one mismatch in 7 nt of PAM-proximal seed sequence were omitted. Although no perfect spacer-protospacer matches were identified, several hits were revealed possessing 2–7 mismatches to full-length C. pasteurianum spacers (Table 1). All protospacer hits identified were represented by spacers 18, 24, and 30 from the central C. pasteurianum Type I-B CRISPR array, whereby multiple protospacer hits were obtained using spacers 24 and 30. Importantly, protospacer matches were derived from predicted Clostridium and Bacillus phage and prophage elements.

**Probing the **C. pasteurianum** Type I-B CRISPR-Cas system using in vivo interference assays and elucidation of protospacer adjacent motif (PAM) sequences.** We selected the best protospacer hits,
Spacer number | Spacer-protospacer match | Invading element | Mismatches | Putative PAM sequence
--- | --- | --- | --- | ---
18 | GTAAAATTTGATTGTCCTCATTGCGATGAAGAAA ATAAAATTTGATTGCCCTCACTGTGATGAAGAAA | Clostridium pasteurianum BC1 (vicinity of phage genes) | 4 | 5’-TTTCA-3’
24 | TTGCAATAGAATGTGATAAAGACCATACACATATGT TTGCAATAGAATGTGATAAAGACCATACACATATGT | Clostridium phage ϕCD111 | 2 | 5’-AATTG-3’
TTGCAATAGAATGTGATAAAGACCATACATATGT TTGCAATAGAATGTGATAAAGACCATACATATGT | Clostridium acidurici 9a (transposase) | 4 | 5’-AATTA-3’
TTGCAATAGAATGTGATAAAGACCATACATATGT TTGCAATAGAATGTGATAAAGACCATACATATGT | Clostridium acetica strain DSM 1496 plasmid CACET_5p (transposase) | 7 | 5’-AATTTC-3’
30 | ATATATGGATTGAAAGGTTTCAGAAGTAAAATA ATATATGGATTGAAAGGTTTCAGAAGTAAAATA | Clostridium botulinum CDC_297 (intact prophage) | 3 | 5’-TATCT-3’
ATATATGGATTGAAAGGTTTCAGAAGTAAAATA ATATATGGATTGAAAGGTTTCAGAAGTAAAATA | Clostridium pasteurianum NRRL B-598 (intact prophage) | 3 | 5’-TTTCT-3’
ATATATGGATTGAAAGGTTTCAGAAGTAAAATA ATATATGGATTGAAAGGTTTCAGAAGTAAAATA | Bacillus licheniformis ATCC 14580 (phage terminase) | 4 | 5’-TCTCA-3’
ATATATGGATTGAAAGGTTTCAGAAGTAAAATA ATATATGGATTGAAAGGTTTCAGAAGTAAAATA | Bacillus pumilus strain NJ-V2 (phage terminase) | 4 | 5’-TTCG-3’
ATATATGGATTGAAAGGTTTCAGAAGTAAAATA ATATATGGATTGAAAGGTTTCAGAAGTAAAATA | Bacillus subtilis strain SG6 (intact prophage) | 5 | 5’-TTTCA-3’

Table 1. Putative protospacer matches identified through in silico analysis of C. pasteurianum CRISPR spacers. *Spacer-protospacer mismatches are underlined. †For hits found within bacterial genomes, the location of the protospacer sequence relative to prophage regions and mobile genetic elements is provided in parentheses. ‡5 nt of adjacent sequence is provided. PAM sequences corresponding to the top protospacer hit from each spacer (bolded) were selected for in vivo interference assays.

possessing 2–4 nt mismatches to C. pasteurianum spacers 18, 24, and 30 (Table 1), for further characterization. Previous analyses of Type I CRISPR-Cas systems have employed a 5 nt mismatch threshold for identifying putative spacer-protospacer hits26,61, as imperfect pairing affords flexibility in host recognition of invading elements or indicates evolution of invading protospacer sequences as a means of evading CRISPR attack60. While the top spacer 30 hit was found to possess homology to an intact prophage from C. botulinum, the best spacer 24 match was predicted to target clostridial phage ϕCD111, a member of the Siphoviridae phage family. C. pasteurianum has recently been shown to harbor an intact and excisable temperate prophage from the same phage family, further supporting the notion that spacer 24 targets phage ϕCD111. The single protospacer match to spacer 18 was found to possess homology to a partial prophage region within the genome of C. botulinum, which is opposite the end of protospacers in Type I CRISPR-Cas systems, which is opposite to the arrangement of protospacer-adjacent sequence; 2) 5 nt of 5′ protospacer-adjacent sequence; and 4) 5 nt of 5′ and 3′ protospacer-adjacent sequence (Fig. 3b). Although the PAM element is typically located at the 5′ end of protospacers in Type I CRISPR-Cas systems, which is opposite to the arrangement of protospacer-adjacent sequence; 3) 5 nt of 3′ protospacer-adjacent sequence; and 4) 5 nt of 5′ and 3′ protospacer-adjacent sequence (Fig. 3b).
to the arrangement observed in Type II systems (Fig. 1), we elected to assay both 5′ and 3′ protospacer-adjacent sequences in the event that the _C. pasteurianum _Type I-B machinery exhibits atypical PAM recognition. Protospacer derivatives were synthesized as complementary single-stranded oligonucleotides, which were annealed and inserted into plasmid pMCTL85141. Interestingly, all three protospacers triggered an interference response from _C. pasteurianum_ when a suitable protospacer-adjacent sequence was provided (Fig. 3b). Plasmids devoid of 5′ protospacer-adjacent sequence (pSpacer18, pSpacer24, pSpacer30, pSpacer18-3′, pSpacer24-3′, and pSpacer30-3′), efficiently transformed _C. pasteurianum_ (1.0–2.4 × 10⁶ CFU µg⁻¹ DNA) (Fig. 3b). Conversely, plasmids containing 5′ protospacer-adjacent sequence (pSpacer18-5′, pSpacer24-5′, pSpacer30-5′, pSpacer18-flank, pSpacer24-flank, and pSpacer30-flank), were unable to transform _C. pasteurianum_ (Fig. 3b). These data indicate that _C. pasteurianum_ expresses Cas proteins that recognize specific PAM sequences encompassed within 5 nt at the 5′ end of protospacers. Interference by host Cas proteins was found to be robust and highly specific.

We analyzed the 5′-adjacent sequences corresponding to protospacers 18, 24, and 30, resulting in three functional PAM sequences represented by 5′-T TTCA-3′, 5′-AATTG-3′, and 5′-TA TCT-3′, respectively (Fig. 3b and Table 1). Due to the promiscuity of most PAM elements, the identified PAM sequences presumably represent only a small subset of sequences that together constitute the consensus recognized by _C. pasteurianum_. It is noteworthy, however, that the third nucleotide of all three functional PAM sequences, as well as six additional sequences that were not assayed _in vivo_ (Table 1), represents a conserved thymine (T) residue, which may be essential for recognition of invading determinants by _C. pasteurianum_ Cas proteins. Within protospacer constructs lacking 5′ adjacent sequence, namely pSpacer18, pSpacer24, pSpacer30, pSpacer18-3′, pSpacer24-3′, and pSpacer30-3′, protospacers are preceded by the sequence 5′-CCCGG-3′ or 5′-GCCGG-3′, encompassing the partial SacII cloning site. It is evident that this sequence does not constitute a PAM sequence recognized by _C. pasteurianum_. CRISPR-Cas machine (Fig. 3b). Similarly, in their native context of _C. pasteurianum_, protospacers 18, 24, and 30 are preceded by the sequence 5′-TAAAT-3′, which is also not recognized by host Cas proteins in order to avoid self attack. Although this sequence resembles the three functional PAM sequences identified through interference assays, particularly 5′-T A TCT-3′, the central conserved T nucleotide is lacking, further supporting the importance of this residue in self and non-self distinction by _C. pasteurianum_.

By assuming the PAM sequence recognized by _C. pasteurianum_ is 5 nt in length and based on a _C. pasteurianum_ chromosomal GC content of 30%, it is possible to calculate the frequency that each PAM sequence occurs within the genome of _C. pasteurianum_. All three 5 nt _C. pasteurianum_ PAM sequences are comprised of four A/T residues and one G/C residue, indicating that all PAM sequences should occur at the same frequency within the _C. pasteurianum_ chromosome. Since the probability of an A or T nucleotide occurring in the genome is 0.35 and the probability of a C or G nucleotide is 0.15, the frequency of each PAM sequence within either strand of the _C. pasteurianum_ genome is 1 ÷ [(0.35)³(0.15)(2 strands)] = 222 bp. More importantly, the overall PAM frequency is only 74 bp, indicating that one of the three functional PAM sequences is expected to occur every 74 bp within the genome of _C. pasteurianum_. This frequency is further reduced to 27 bp if the true PAM recognized by _C. pasteurianum_ is represented by 3 nt, which is a common feature of Type I-B PAMs. In comparison, the Type II CRISPR-Cas9 system from _S. pyogenes_ recognizes a 5′-NGG-3′ consensus, which is expected to occur every 22 bp in the genome of _C. pasteurianum_.

**Repurposing the endogenous Type I-B CRISPR-Cas system for markerless genome editing.**

The high frequency of functional PAM sequences within the genome of _C. pasteurianum_ suggests that the endogenous Type I-B CRISPR-Cas system could be co-opted to attack any site within the organism’s chromosome and, therefore, provide selection against unmodified host cells. To first assess self-targeting of the _C. pasteurianum_ CRISPR-Cas system, we again selected the _cpaAIR_ gene as a target. The 891 bp _cpaAIR_ gene was found to possess a total of 19 potential PAM sequences (5′-T TTCA-3′, 5′-AATTG-3′, and 5′- TA TCT-3′), which is more than the 12 PAM sequences expected based on a genomic frequency of 74 bp. We selected one PAM sequence (5′-AATTG-3′) within the coding region of the _cpaAIR_ gene as the target site for _C. pasteurianum_ self-cleavage, whereby sequence immediately downstream embodies the target protospacer. Analysis of the core 37 spacers encoded by _C. pasteurianum_ revealed minimal variation in spacer length (34–37 nt; mean of 36 nt), while GC content was found to vary dramatically (17–44%). Subsequently, we generated a synthetic _cpaAIR_ spacer by selecting 36 nt immediately downstream of the designated PAM sequence, which was found to possess a GC content of 28%. A CRISPR expression cassette was designed by mimicking the sequence and arrangement of the native Type I-B CRISPR array present in the _C. pasteurianum_ genome (Figure S1B). Specifically, a 243 bp CRISPR leader was utilized to drive transcription of the synthetic _cpaAIR_ CRISPR array, comprised of the 36 nt _cpaAIR_ spacer flanked by 30 nt direct repeats. The synthetic array was followed by 298 bp of sequence located at the 3′ end of the endogenous chromosomal CRISPR array. The resulting cassette was synthesized and inserted into plasmid pMCTL85141, generating pCParray-cpaAIR (Fig. 4a). While several attempts at transformation of _C. pasteurianum_ using pCParray-cpaAIR failed to generate transformants, an overall transformation efficiency of 0.6 CFU µg⁻¹ DNA was obtained (Fig. 4b), compared to 6.3 × 10⁹ CFU µg⁻¹ DNA for the pMCTL85141 parental plasmid, a difference of more than four orders of magnitude. We reasoned that the synthetic _cpaAIR_ spacer triggered self-attack of _C. pasteurianum_ through introduction of a DN and subsequent strand degradation by Cas3. To verify the location of the DN site within the _cpaAIR_ target gene and, more importantly, demonstrate manipulation of the Type I-B CRISPR-Cas system for genome editing, we introduced the aforementioned _cpaAIR_ editing cassette utilized for cas9-mediated genome editing (from plasmid pCas9gRNA-delcpAIR) into plasmid pCParray-cpaAIR (Fig. 4a). Transformation of _C. pasteurianum_ with the resulting plasmid, pCParray-delcpAIR, produced an abundance of transformants, yielding a transformation efficiency of 9.5 CFU µg⁻¹ DNA, an increase of more than an order of magnitude compared to pCParray-cpaAIR lacking an editing cassette (Fig. 4b). Despite a low-level of background resulting from transformation with pCParray-cpaAIR, genotyping of 10 pCParray-delcpAIR transformants generated a PCR product corresponding to _cpaAIR_ gene deletion in all colonies screened, yielding an
editing efficiency of 100% (Fig. 4c). Sanger sequencing of a single pCParray-delcpaAIR transformant confirmed successful deletion of a 762 bp region of the \textit{cpaAIR} coding sequence (data not shown). Importantly, this outcome is consistent with localization of the DN within the \textit{cpaAIR} locus, as well as provides proof-of-principle repurposing of the host Type I-B CRISPR-Cas machinery for efficient markerless genome editing.

Identification of putative PAM sequences in industrial and pathogenic clostridia. As the first step towards expanding our CRISPR-Cas hijacking strategy to other prokaryotes, we surveyed the clostridia for...
species harboring putative CRISPR-Cas loci. One cellulytic and one acetogenic species, namely *Clostridium thermocellum* and *Clostridium autoethanogenum*, respectively, in addition to *Clostridium tetani*, a human pathogen, were selected. Like *C. pasteurianum*, all three species encode putative Type I-B systems, while *C. tetani* and *C. thermocellum* harbor an additional Type I-A or Type III locus, respectively. Only spacers associated with Type I-B loci were analyzed, corresponding to 98, 31, and 169 spacers from *C. autoethanogenum*, *C. tetani*, and *C. thermocellum*, respectively. In *silico* analysis of clostralid spacers against firmicute genomes, phages, and plasmids yielded putative protospacer matches from all three clostridial Type I-B CRISPR-Cas loci analyzed (Table 2). In total 10 promising protospacer hits were obtained, which were found to target phages (2 hits), plasmids (2 hits), and regions of bacterial genomes in the vicinity of phage and/or transposase genes (2 hits). Six spacers were found to target clostridial genomes and clostridial phage and prophage elements. Interestingly, spacers from the *C. autoethanogenum* Type I-B locus were analyzed in an earlier report and no putative protospacer matches were identified23, which we unveiled as our probable protospacer hits, including the only perfect spacer-protospacer match identified in this study. Overall, putative protospacer matches contained 0–8 mismatches when aligned with clostridial spacers. Analysis of clostralid 5′-protospacer-adjacent sequences revealed a number of conserved sequences (Table 2). Interestingly, all 10 putative PAM sequences were found to possess a conserved A residue in the immediate 5′-protospacer-adjacent position. Based on a 3 nt consensus, prospective PAMs of 5′-TAA-3′ were predicted for the Type I-B CRISPR-Cas loci of *C. autoethanogenum*, *C. tetani*, and *C. thermocellum*, respectively.

**Discussion**

This work details the development of a genome editing methodology allowing efficient introduction of precise chromosomal modifications through harnessing an endogenous CRISPR-Cas system. Our strategy leverages the widespread abundance of prokaryotic CRISPR-Cas machinery, which have been identified in 45% of bacteria, including 74% of *clostridia*. An exceptional abundance of CRISPR-Cas loci, coupled with an overall lack of sophisticated genetic engineering technologies and tremendous biotechnological potential, provides the rationale for our proposed genome editing strategy in *Clostridium*. We selected *C. pasteurianum* for proof-of-concept CRISPR-Cas repurposing due to the presence of a Type I-B CRISPR-Cas locus (Fig. 3a) and established industrial relevance for biofuel production46,49. Analysis of *C. pasteurianum* CRISPR tags led to elucidation of the probable origins of three spacer sequences, all of which returned protospacer matches from clostralid phage and prophage determinants (Table 1). *C. pasteurianum* CRISPR-Cas proteins proved to be functional and highly active against plasmid-borne protospacers possessing a 5′-adjacent PAM sequence, as no interference response was generated from protospacers harboring 3′-adjacent sequence in the absence of a 5′ PAM sequence (Fig. 3b). This finding is consistent with other Type I-CRISPR-Cas systems, in which the PAM positioned 5′ to the protospacer is essential for interference by host cells and contrasts Type II CRISPR-Cas9 systems, whereby the PAM is recognized at the 3′ end of protospacers14,25,26. Following elucidation of functional PAM sequences, we developed a genome editing strategy encompassing expression of a synthetic programmable Type I-B CRISPR array that guides site-specific nucleolytic attack of the *C. pasteurianum* chromosome by co-opting the organism’s native

**Table 2. Putative protospacer matches identified through in silico analysis of clostralid CRISPR spacers.**

| Organism (CRISPR-Cas subtype) | Spacer-protospacer match | Invading element | Mismatches | Putative PAM sequence |
|-------------------------------|--------------------------|------------------|------------|------------------------|
| *C. autoethanogenum* DSM 10061 (Type I-B) | AAGGTTGATCGATTCCTTATCCTTATAGTACCTTGAAGGTC | *Clostridium ljungdahlii* DSM 13528 (incomplete prophage) | 0 | 5′-ATTAA-3′ |
| | TAGACCACTTTTAAAGTCAATTTTGATGCTG | *Clostridium phage vB_CpeS-CP51 | 4 | 5′-ACTAA-3′ |
| | AATACAGTTTTAATGTAACGATTAGGGG | Bacillus thuringiensis HD-789 plasmid pBD789R-3 | 4 | 5′-AGAA-3′ |
| | GCACACTCAACGAGCAAACTAAACGGCAGGAAAGTTGATT | Enterococcus durans strain KLD6 6.0930 (vicinity of transposase and phage genes) | 8 | 5′-ATCAA-3′ |
| | GGACCTTCAGGAGCAAACTAAACGGCAGGAAAGTTGATT | *Clostridium botulinum* A2 str. Kyoto (intact prophage) | 3 | 5′-TTTTA-3′ |
| | AAGAGCAGGCGAGGAGATCTTCAAAATGAA | *Clostridium botulinum* F str. 230613 (intact prophage) | 3 | 5′-TATAA-3′ |
| | AACGAAGCGGCAGGAGGATCTTCAAAATGAA | Bacillus sp. FJAT-4402 (questionable prophage) | 7 | 5′-CATCA-3′ |
| *C. thermocellum* ATCC 27405 (Type I-B) | ATCTTTCTTTTTTTACTCTAAATGGCAATGGAAGGTCG | *Clostridium stercorarium* subs. stercorarium DSM 8532 (intact prophage) | 2 | 5′-TTTCA-3′ |
| | TGATGGATGGATCAGGTAAGTCTCCAGATGGAAGGTCG | *Clostridium clariflavum* DSM 19732 (vicinity of transposase) | 2 | 5′-GGACA-3′ |
| | ACGAGGCGGTTGTTACCTTATGGAGATGGAAGGTCG | Staphylococcus phage vB_SauM_Remus | 6 | 5′-AATCA-3′ |

*Spacers-protospacer mismatches are underlined. In instances where multiple protospacer hits were obtained from a single spacer query, the top hit is provided. Generally, PAM sequences were found to be identical between multiple protospacer hits from a single spacer sequence. For hits found within bacterial genomes, the location of the protospacer sequence relative to prophage regions and mobile genetic elements is provided in parentheses. 5′ nt of adjacent sequence is provided. Potential conserved residues are bolded.*
Cas proteins. Cas3-mediated DNA attack affords selection against unmodified host cells, whereby edited cells are efficiently obtained through co-introduction of an editing template (Fig. 4a,b). We have demonstrated 100% editing efficiency (10/10 correct colonies) by targeting the cpaAIR locus in combination with introduction of a cpaAIR gene deletion cassette (Fig. 4c).

Our native CRISPR-Cas repurposing methodology contrasts current approaches of CRISPR-mediated genome editing in bacteria, which rely on the widely-employed Type II CRISPR-Cas9 system from S. pyogenes. In Clostridium, such heterologous CRISPR-Cas9 genome editing strategies have recently been implemented in C. beijerinckii and C. cellulolyticum. While editing efficiencies >95% were reported using C. cellulolyticum, no efficiency was provided for CRISPR-based editing in C. pasteurianum, which involves the use of a phenotypic screen to identify mutated cells. Although we have shown 100% editing efficiency in C. pasteurianum through application of the same S. pyogenes CRISPR-Cas9 machinery (Fig. 2a,c), the total yield of edited cells was only 25% compared to the endogenous Type I-B CRISPR-Cas approach (Figs 2b and 4b). By assessing transformation of various cas9 expression constructs, we ascribe this outcome to poor transformation of vectors expressing cas9 in trans (Fig. 2b). A low to moderate level of Cas9 toxicity has been documented in a diverse range of organisms, including prototaoza, Drosophila, yeast, and human cells, and likely results from the generation of lethal ectopic chromosomal DNA breaks. We have also observed reduced transformation of E. coli ER1821 in this study using plasmids expressing heterologous cas9 (data not shown). In more dramatic instances, for example in mycobactaria and the algae Chlamydomonas reinhardtii, toxicity leads to erratic Cas9 expression and overall poor genome editing outcomes. Such reports emphasize the importance of mitigating Cas9 toxicity or developing alternative methodologies facilitating efficient genome editing.

Many authors have described the relatively high efficiency of plasmid transfer to C. pasteurianum, success in obtaining targeted mutants using constitutive expression of heterologous cas9. For key organisms lacking endogenous CRISPR-Cas loci, such as C. acetobutylicum, C. ljungdahlii, and C. tetani, cas9 expression, as observed in this study. Hence, for key organisms lacking endogenous CRISPR-Cas loci, such as C. acetobutylicum and C. ljungdahlii, in which the heterologous Type II system is obligatory for genome editing, we recommend inducible expression of cas9. For this purpose, several clostridial inducible gene expression systems have recently been characterized. Our success in obtaining targeted mutants using constitutive expression of heterologous cas9 potentially results from the relatively high efficiency of plasmid transfer to C. pasteurianum (up to 10^4 CFU μg −1 DNA). It is probable that Cas9-mediated genome editing efforts could be impeded in species that are poorly transformed, rendering endogenous CRISPR-Cas machinery the preferred platform for genome editing. Furthermore, since linear DNA is a poor substrate for transformation of Clostridium and because it is generally unfeasible to co-transfer two DNA substrates to Clostridium due to poor transformation, all of the genetic components required for Type I-B or Type II CRISPR-Cas functionality in this study were expressed from single vectors. This shortcoming exposes an additional advantage of our endogenous CRISPR-Cas hijacking strategy, as only a small CRISPR array (0.6 kb) and editing template are required for genome editing, resulting in a compact 5.7 kb editing vector (pCParray-delcpaAIR). On the other hand, editing using the heterologous Type II system requires expression of the large 4.2 kb cas9 gene, in addition to a 0.4 kb gRNA cassette and editing template. The large size of the resulting pCas9gRNA-delcpaAIR editing vector (9.7 kb) not only limits transformation but also places significant constraints on multiplexed editing strategies involving multiple gRNAs and editing templates. Owing to overall low rates of homologous recombination in Clostridium, such ambitious genome editing strategies could be enhanced through coupling of native or heterologous CRISPR-Cas machinery to highly recombinogenic phage activities.

In this context, one functional clostridial phage recombinase has been characterized to date. To initiate efforts aimed at co-opting Type I CRISPR-Cas machinery in other key species, we examined CRISPR spacer tags from one acetogenic (C. autoethanogenum), one cellulolytic (C. thermocellum), and one pathogenic (C. tetani) species (Table 2). Subsequent in silico analysis of clostridial spacer tags, coupled with our experimental validation of C. pasteurianum PAM sequences and a recent report detailing characterization of the C. difficile Type I-B CRISPR-Cas locus, provide an in depth glimpse into clostridial CRISPR-Cas defence mechanisms (Table 3). Overall, clostridial Type I-B PAM sequences are characterized by a notable lack of guanine (G) residues. Additionally, several PAM sequences unveiled in this study are recognized across multiple species of Clostridium, such as 5′-TCA-3′ by C. pasteurianum, C. tetani, and C. thermocellum, and 5′-TAA-3′ by C. autoethanogenum and C. tetani, which suggests horizontal transfer of CRISPR-Cas loci between these organisms. Indeed, C. tetani harbors 7 distinct Type I-B CRISPR arrays, of which employ the same direct repeat sequence utilized by the C. pasteurianum Type I-B system. Since PAM sequences determined in this study are highly similar between C. pasteurianum (5’-TCA-3′, 5’-TTG-3′, 5’-TCT-3′) and C. tetani (5′-TCA-3′, 5′-TTA-3′, 5′-TAA-3′), it is plausible that these organisms recognize the same PAM consensus. More broadly, clostridial Type I-B PAM sequences bear a striking overall resemblance to sequences recognized by the Type I-B system from the distant archaeon Haloferax volcanii (5′-ACT-3′, 5′-TTG-3′, 5′-TAG-3′, and 5′-CAC-3′), which are also distinguished by an overall low frequency of G residues. Collectively these data suggest that many PAM sequences are common amongst Type I-B CRISPR-Cas systems, even in evolutionarily distant species, such as the case of Haloferax and Clostridium. In this context, we posit that empirical elucidation of PAMs is unnecessary, as highly pervasive PAM sequences (e.g., 5′-TCA-3′ and 5′-TAA-3′) or validated sequences from closely-related species can easily be assessed for functionality in a target host strain. This consequence simplifies our proposed CRISPR-Cas repurposing approach, as a functional PAM sequence and a procedure for plasmid transformation are the only prerequisite criteria for implementing our methodology in any target organism harboring active Type I CRISPR-Cas machinery.

Genome editing strategies based on the S. pyogenes Type II system reported previously and the CRISPR-Cas hijacking approach detailed in this study, represent a key divergence from earlier methods of gene disruption and integration in Clostridium. Currently, the only procedures validated for modifying the genome of C. pasteurianum involve the use of a programmable group II intron and heterologous counter-selectable mazF marker. Whereas group II introns are limited to gene disruption, as deletion and replacement are not possible,
Figure S1. Oligonucleotides and synthetic DNA constructs were purchased from Integrated DNA Technologies (IDT; Coralville, Iowa). A promoter from the spacer sequence (ctgatgaagctaatacagat), which was expressed from the sCbei_5830 small CPA AIR C. beijerinckii gene by specifying a 20 nt PAM sequence, was utilized to fuse 1,028 bp and 1,057 bp CPA AIR C. pasteurianum strain engineering technologies. Although our work here focused on C. pasteurianum, repurposing of endogenous CRISPR-Cas machinery is readily adaptable to most of the genus Clostridium, including many species of immense relevance to medicine, energy, and biotechnology, as well as half of all bacteria and most archaea.

### Materials and Methods

#### Strains, plasmids, and oligonucleotides.
Strains and plasmids employed in this study are listed in Table 4. Clostridium pasteurianum ATCC 6013 was obtained from the American Type Culture Collection (ATCC; Manassas, VA) and propagated and maintained according to previous methods. Escherichia coli strains DH5α and ER1821 (New England Biolabs; Ipswich, MA) were employed for plasmid construction and plasmid maintenance, respectively. Recombinant strains of C. pasteurianum were selected using 10 μg ml^−1^ thiamphenicol and 30 μg ml^−1^ kanamycin or 30 μg ml^−1^ chloramphenicol. Antibiotic concentrations were reduced by 50% for selection of double plasmid recombinant cells. Desalted oligonucleotides and synthetic DNA constructs were purchased from Integrated DNA Technologies (IDT; Coralville, IA). Oligonucleotides utilized in this study are listed in Table S1 and synthetic DNA constructs are detailed in Figure S1.

### DNA manipulation, plasmid construction, and transformation.
A cas9 E. coli-Clostridium expression vector, p85Cas9, was constructed through amplification of a cas9 gene cassette from pCas927 using primers cas9+SacI+ and cas9+Xhol+AS and insertion into the corresponding sites of pMTL85145. To construct an E. coli-C. pasteurianum Type II CRISPR-Cas9 plasmid (pCas9gRNA-cpaAIR) based on the S. pyogenes CRISPR-Cas9 system, we designed a synthetic gRNA targeted to the C. pasteurianum CPA AIR gene by specifying a 20 nt cpaAIR spacer sequence (ctgatgaagctaatacagat), which was expressed from the C. beijerinckii S.Cbei_5830 small RNA promoter. A promoter from the C. pasteurianum thiolase gene was included for expression of cas9. The resulting 521 bp DNA fragment (Figure S1A) was synthesized and inserted into the SacII and BstZ17I sites of p85Cas9. To modify pCas9gRNA-cpaAIR for genome editing via deletion of cpaAIR, splicing by overlap extension (SOE) PCR was utilized to fuse 1,028 bp and 1,057 bp cpaAIR homology regions generated using the primer sets delcpaAIR.PvuI+ and delcpaAIR.SacI+AS and delcpaAIR.SacI+S+delcpaAIR.PvuI+. Respectively. The resulting PvuI-digested product was cloned into the PvuI site of pCas9gRNA-cpaAIR, yielding pCas9gRNA-delcpaAIR. Plasmid p83Cas9, a p85Cas9 derivative containing the pCB102 replication module, was constructed by amplifying cas9 from pCas927 using primers cas9+SacI+ and cas9+Xhol+AS and inserting the resulting product into the corresponding sites of pMTL83151. A promoterless cas9 derivative of p85Cas9, designated p85delCas9, was derived by amplification of a partial promoterless cas9 fragment from pCas9gRNA-cpaAIR using primers −cas9. SacI+S+cas9+BstZ17I+AS and cloning of the resulting product into the SacII + BstZ17I sites of p85Cas9.

### Table 4: Summary of clostridial Type I-B CRISPR-Cas loci analyzed to date.

| Species                        | Number of spacers (total)^a | PAM sequences^b | PAM^c | Reference |
|-------------------------------|-----------------------------|-----------------|-------|-----------|
| C. pasteurianum DSM 10061     | 22, 43, 33 (98)             | 5′-TAA-3′ 5′-TAA-3′ 5′-GAA-3′ | 5′-NAA-3′ | This study^d |
| C. difficile 630/R20291        | 1, 2, 1, 1, 4, 2, 3, 2, 14, 11, 4, 5, 4, 14, 9, 26, 9 (116) | 5′-CCCA-3′ 5′-CCT-3′ | 5′-CCW-3′ | 62,7 |
| C. pasteurianum ATCC 6013     | 37, 8 (45)                  | 5′-TCA-3′ 5′-TTG-3′ 5′-TCT-3′ | ND^d | This study^d |
| C. tetani 12124569            | 22, 3, 4, 2, 4, 5, 10, 3 (53) | 5′-TAA-3′ 5′-TTA-3′ 5′-TCA-3′ | 5′-TNA-3′ | This study^d |
| C. thermocellum ATCC 57405    | 51, 96, 169, 78, 42 (436)  | 5′-TCA-3′ 5′-TCA-3′ 5′-ACA-3′ | 5′-NCA-3′ | This study^d |

Table 3. Summary of clostridial Type I-B CRISPR-Cas loci analyzed to date. ^aSpacers corresponding to Type I-B CRISPR-Cas loci analyzed in this study are bolded. ^b3 nt PAM and PAM sequences are shown. Experimentally-verified motifs are bolded. ^cW = weak (A or T). ^dND = not determined due to highly varied PAM sequences.

Techniques based on homologous recombination using antibiotic resistance determinants and counter-selectable markers, such as pyrE/pyrF, codA, and mazF,^24,37,48, are technically-challenging and laborious due to a requirement for excision and recycling of markers. In general, these strategies do not provide adequate selection against unmodified cells, necessitating subsequent rounds of enrichment and selection.^2,43,76,77 Thus, both native and heterologous CRISPR-Cas machineries offer more robust platforms for genome modification of C. pasteurianum and related clostridia.

Currently, endogenous CRISPR-Cas systems have been harnessed in only a few prokaryotes, namely E. coli^24,79, Pectobacterium atrosepticum^80, Streptococcus thermophilus^78, and two species of archaea^52,81. In conjunction with these reports, our success in co-opting the chief C. pasteurianum CRISPR-Cas locus contributes to a growing motivation towards harnessing host CRISPR-Cas machinery in a plethora of prokaryotes. The general rationale of endogenous CRISPR-Cas repurposing is not limited to genome editing, as a range of applications can be envisioned. In a recent example, Luo et al.^79 deleted the native cas3 endonuclease gene from E. coli, effectively converting the host Type I-E CRISPR-Cas immune system into a robust transcriptional regulator for gene silencing. Such applications dramatically extend the existing molecular genetic toolbox and pave the way to advanced strain engineering technologies. Although our work here focused on C. pasteurianum, repurposing of endogenous CRISPRS-Cas loci is readily adaptable to most of the genus Clostridium, including many species of immense relevance to medicine, energy, and biotechnology, as well as half of all bacteria and most archaea.
C. *pasteurianum* protospacer constructs lacking protospacer-adjacent sequences were derived by annealing oligos spacer18.AatII.S + spacer18.SacII.AS (pSpacer18), spacer24.AatII.S + spacer24.SacII.AS (pSpacer24), and spacer30.AatII.S + spacer30.SacII.AS (pSpacer30). The pSpacer constructs were then cloned into pCas9 and pCas9gRNA vectors to generate pSpacerCas9 and pSpacerCas9gRNA vectors, respectively. These vectors were then used to transform *Escherichia coli* DH5α and *Escherichia coli* ER1821 cells to create *C. pasteurianum* protospacer expression vectors. The vectors were subsequently used to transform *C. pasteurianum* ATCC 6013 and Δ*cpaAIR* markerless deletion mutant cells to create *C. pasteurianum* protospacer expression vectors. The resulting cells were then used to generate *C. pasteurianum* protospacer expression vectors for use in subsequent experiments.

| Strain | Relevant characteristics | Source or reference |
|--------|--------------------------|---------------------|
| *Escherichia coli* DH5α | F′ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR supG4 thyA169 Δ(lacZΔM15-argF)169 BfdR17 (erm39 cym39) Δλλ | Lab stock |
| *Escherichia coli* ER1821 | F′ endA1 glnV44 thi-1 relA1 (mcrA) rfbD1 (mcrC) spoT1 Δ(mcrC-mrr)114Δ IS10 | Lab stock; New England Biolabs |
| *Clostridium pasteurianum* ATCC 6013 | Wild-type | American Type Culture Collection |
| *Clostridium pasteurianum* Δ*cpaAIR* | Markerless *cpaAIR* deletion mutant | This study |

| Plasmid | Relevant characteristics | Source or reference |
|---------|--------------------------|---------------------|
| pFnuDIIIMKn | M.FnuDII methyltransferase plasmid for methylation of *E. coli-C. pasteurianum* shuttle vectors (KmR; p15A ori) | 49 |
| pMTL83151 | *E. coli-Clostridium* shuttle vector (CmR; CoE1 ori; pCB102 ori) | 57 |
| pMTL85141 | *E. coli-Clostridium* shuttle vector (CmR; CoE1 ori; pCB101 ori) | 57 |
| pCas9 | *E. coli* cas9 and tracrRNA expression vector (CmR; p15A ori) | 27 |
| pCas9gRNA-cpaAIR | Type II CRISPR expression vector containing *cas9* and gRNA targeted to the *C. pasteurianum* *cpaAIR* gene | This study |
| pCas9gRNA-delcpaAIR | Type II CRISPR genome editing vector derived by inserting a *cpaAIR* deletion editing cassette into pCas9gRNA-cpaAIR | This study |
| p85Cas9 | *cas9* expression vector derived by inserting *cas9* with its native promoter from pCas9 into pMTL85141 | This study |
| p83Cas9 | *cas9* expression vector derived by inserting *cas9* and the tracrRNA from pCas9 into pMTL83151 | This study |
| p85delCas9 | Derived by deleting the *cas9* promoter from p85Cas9 | This study |
| pSpacer18 | *C. pasteurianum* protospacer 18 construct lacking flanking sequences | This study |
| pSpacer18-5′ | *C. pasteurianum* protospacer 18 construct including 5′ protospacer-adjacent sequence | This study |
| pSpacer18-3′ | *C. pasteurianum* protospacer 18 construct including 3′ protospacer-adjacent sequence | This study |
| pSpacer18-flank | *C. pasteurianum* protospacer 18 construct including flanking protospacer-adjacent sequence | This study |
| pSpacer24 | *C. pasteurianum* protospacer 24 construct lacking flanking sequences | This study |
| pSpacer24-5′ | *C. pasteurianum* protospacer 24 construct including 5′ protospacer-adjacent sequence | This study |
| pSpacer24-3′ | *C. pasteurianum* protospacer 24 construct including 3′ protospacer-adjacent sequence | This study |
| pSpacer24-flank | *C. pasteurianum* protospacer 24 construct including flanking protospacer-adjacent sequence | This study |
| pSpacer30 | *C. pasteurianum* protospacer 30 construct lacking flanking sequences | This study |
| pSpacer30-5′ | *C. pasteurianum* protospacer 30 construct including 5′ protospacer-adjacent sequence | This study |
| pSpacer30-3′ | *C. pasteurianum* protospacer 30 construct including 3′ protospacer-adjacent sequence | This study |
| pSpacer30-flank | *C. pasteurianum* protospacer 30 construct including flanking protospacer-adjacent sequence | This study |
| pCPArray-cpaAIR | Type I-B CRISPR expression vector containing a synthetic CRISPR array targeted to the *C. pasteurianum* *cpaAIR* gene | This study |
| pCPArray-delcpaAIR | Type I-B CRISPR genome editing vector derived by inserting a *cpaAIR* deletion editing cassette into pCPArray-cpaAIR | This study |

Table 4. Strains and plasmids employed in this study.
Protospacer constructs possessing 5′ or 3′ protospacer-adjacent sequences were prepared by annealing oligos spacer 18′-3′. SacII.AS (pSpacer 18′-3′), spacer 18′-3′. SacII.AS (pSpacer 18′-3′), spacer 24′-3′. SacII.AS + pSpacer 24′-3′, spacer 30′-3′. SacII.AS (pSpacer 30′-3′), or spacer 30′-3′. SacII.AS + pSpacer 30′-3′. SacII.AS (pSpacer 30′-3′). Protospacer constructs possessing 5′ and 3′ flanking protospacer-adjacent sequence were prepared by annealing oligos spacer 18-flank.SacII.AS (pSpacer 18-flank), spacer 24-flank.SacII.AS + pSpacer 24-flank, or spacer 30-flank.SacII.AS + pSpacer 30-flank.SacII.AS (pSpacer 30′-flank). In all instances protospacer oligos were designed such that annealing generated AatII and SacII cohesive ends for ligation with AatII-+SacII-digested pMPL85141.

To construct the endogenous CRISPR array vector, pCP array-cpaAIR, a synthetic CRISPR array was designed containing a 243 bp CRISPR leader sequence and a 37 nt crpAIR spacer flanked by 30 nt direct repeat sequences. The synthetic array was followed by 298 bp of sequence found downstream of the endogenous CRISPR array in the chromosome of C. pasteurianum to ensure design of the synthetic array mimics that of the native sequence. The resulting 667 bp fragment (Figure S1B) was synthesized and cloned into the SacI site of pMTL85141. A genome editing derivative of pCP-Parray-cpaAIR for deletion of crpAIR was derived by subcloning the PvuII-flanked deletion cassette from pCas9gRNA-delcpaAIR into pCP-Parray-cpaAIR, yielding pCP-DelpcpaAIR.

DNA manipulation was performed according to established methods62. Commercial kits for DNA purification and agarose gel extraction were obtained from Bio Basic Inc. (Markham, ON). Plasmids were introduced to C. pasteurianum μg of plasmid DNA. and are expressed as colony-forming units (CFU) per μg of plasmid DNA.

Identification of putative protospacer matches to clostridial spacers. Clostridial spacers were utilized to query firmicute genomes, phages, transposons, and plasmids using BLAST. Parameters were optimized for somewhat similar sequences (BlastN)64. Putative protospacer hits were assessed based on the number and location of mismatches, whereby multiple PAM-distal mutations were tolerated, while protospacers containing more than one mismatch within 7 nt of PAM-proximal seed sequence were rejected60. Firmicute genomes possessing putative protospacer hits were analyzed for prophage content using PHAST59 and surrounding sequences were inspected for elements indicative of DNA mobility and invasion, such as transposons, transposases, integrase, and terminases.

References
1. Barrangou, R. & Marraffini, L. A. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol. Cell 54, 234–244 (2014).
2. Sorek, R., Lawrence, C. M. & Wiedenheft, B. CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu. Rev. Biochem. 82, 257–266 (2013).
3. Makarova, K. S. et al. An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. 13, 722–736 (2015).
4. Shmakov, S. et al. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. Mol. Cell 60, 385–397 (2015).
5. Makarova, K. S. et al. An updated classification of the CRISPR–Cas systems. Nat. Rev. Microbiol. 9, 467–477 (2011).
6. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR-associated (Cas) protein families and multiple Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol 1, e60 (2010).
24. Deveau, H. et al. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J. Bacteriol. 190, 1390–1400 (2008).
25. Mojica, F. Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733–740 (2009).
26. Shah, S. A., Erdmann, S., Mojica, F. J. & Garrett, R. A. Protospacer recognition motifs: mixed identities and functional diversity. RNA biology 10, 891–899 (2013).
27. Wang, W. Y., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31, 233–239 (2013).
28. Jiang, Y. et al. Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. Appl. Environ. Microbiol. 81, 2506–2514 (2015).
29. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Coupling the CRISPR/Cas9 system with lambda Red recombineering enables simplified chromosomal gene replacement in Escherichia coli. Appl. Environ. Microbiol. 81, 5103–5114 (2015).
30. DiCarlo, J. E. et al. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas9 systems. Nucleic Acids Res. 41, 4336–4343 (2013).
31. Horwitz, A. A. et al. Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. Cell Systems 1, 88–96 (2015).
32. Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR-Cas-mediated genome engineering. Cell 153, 910–918 (2013).
33. Hwang, W. Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas9 system. Nat. Biotechnol. 31, 227–229 (2013).
34. Shan, Q. et al. Targeted genome modification of crop plants using a CRISPR-Cas9 system. Nat. Biotechnol. 31, 686–688 (2013).
35. Cong, L. et al. Multiplex genome engineering using CRISPR-Cas9 systems. Science 339, 819–823 (2013).
36. Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823–826 (2013).
37. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. USA 97, 6640–6645 (2000).
38. Tracy, B. P., Jones, S. W., Fast, A. G., Indurthi, D. C. & Papoutsakis, E. T. Clostridia: The importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. Curr. Opin. Biotechnol. 23, 364–381 (2012).
39. Van Mellaert, L., Barbe, S. & Anne, J. Clostridium spores as anti-tumour agents. Trends Microbiol. 14, 190–196 (2006).
40. Halleroy, C. L. Toxigenic clostridia. Clin. Microbiol. Rev. 3, 66–98 (1990).
41. Pyne, M. E., Bruder, M., Moo-Young, M., Chung, D. A. & Chou, C. P. Technical guide for genetic advancement of underdeveloped and intractable Clostridium. Biotechnol. Adv. 32, 623–641 (2014).
42. Al-Hinai, M. A., Fast, A. G. & Papoutsakis, E. T. Clostridia: The importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. Curr. Opin. Biotechnol. 23, 364–381 (2012).
43. Heap, J. T. et al. Sequence analysis of whole genomes of Trypanosoma cruzi and Trypanosoma brucei. Nat. Biotechnol. 27, 233–239 (2009).
44. Heap, J. T. et al. Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of Clostridium autoethanogenum and analysis of CRISPR systems in industrial relevant Clostridia. Biotechnol. Biofuels 7, 40, doi:10.1186/1754-6834-7-40 (2014).
45. Yazdani, S. S. & Gonzalez, R. Anaerobic fermentation of glycerol: A path to economic viability for the biofuels industry. Curr. Opin. Biotechnol. 18, 213–219 (2007).
46. Peng, D., Kurup, S. P., Yao, P. Y., Minning, T. A. & Tarleton, R. L. CRISPR-Cas9-mediated single-gene and gene family disruption in Trypanosoma cruzi. mBio 6, e02097–e02014, doi:10.1128/mBio.02097-14 (2015).
47. Gratz, S. J. et al. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics 196, 961–971 (2014).
69. Sebo, Z. L., Lee, H. B., Peng, Y. & Guo, Y. A simplified and efficient germline-specific CRISPR/Cas9 system for Drosophila genomic engineering. *Fly* 8, 52–57 (2014).

70. Charpentier, E. & Doudna, J. A. Biotechnology: Rewriting a genome. *Nature* 495, 50–51 (2013).

71. Datta, S., Costantino, N., Zhou, K. M. & Court, D. L. Identification and analysis of recombinering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc. Natl. Acad. Sci. USA* 105, 1626–1631 (2008).

72. Dong, H., Tao, W., Gong, F., Li, Y. & Zhang, Y. A functional recT gene for recombinering of *Clostridium*. *J. Biotechnol.* 173, 65–67 (2014).

73. Hartman, A. H., Liu, H. L. & Melville, S. B. Construction and characterization of a lactose-inducible promoter system for controlled gene expression in *Clostridium perfringens*. *Appl. Environ. Microbiol.* 77, 471–478 (2011).

74. Dong, H. J., Tao, W. W., Zhang, Y. P. & Li, Y. Development of an anhydrotetracycline-inducible gene expression system for solvent-producing *Clostridium acetobutylicum*: A useful tool for strain engineering. *Metab. Eng.* 8, 52–57 (2014).

75. Charpentier, E. & Doudna, J. A. Biotechnology: Rewriting a genome. *Nature* 495, 50–51 (2013).

76. Dong, H., Tao, W., Gong, F., Li, Y. & Zhang, Y. A functional recT gene for recombinering of *Clostridium*. *J. Biotechnol.* 173, 65–67 (2014).

77. Sandoval, N. R., Venkataramanan, K. P., Groth, T. S. & Papoutsakis, E. T. Whole-genome sequence of an evolved *Clostridium pasteurianum* strain reveals Spo0A deficiency responsible for increased butanol production and superior growth. *Biotechnol. Biofuels* 8, 227, doi: 10.1186/s13068-015-0408-7 (2015).

78. Olson, D. G. & Lynd, L. R. Transformation of *Clostridium thermocellum* by electroporation. *Methods Enzymol.* 510, 317–330 (2012).

79. Luo, M. L., Mullis, A. S., Leenay, R. T. & Beisel, C. L. Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.* 43, 674–681 (2015).

80. Vercoe, R. B. et al. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *Plos Genet* 9, e1003454, doi: 10.1371/journal.pgen.1003454 (2013).

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

82. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning*. Vol. 2 (Cold spring harbor laboratory press New York, 1989).

Acknowledgements
We thank N. Minton and L. Marraffini for kindly providing pMTL8000-series and pCas9 vectors, respectively.

Author Contributions
M.E.P. helped conceive of the study, participated in its design and coordination, carried out the experiments, and drafted the manuscript. M.R.B. participated in coordination of the study and assisted with experimental procedures. M.M.Y. participated in the study design and coordination. D.A.C. and C.P.C. helped conceive of the study, participated in its design and coordination, and helped draft the manuscript. All authors read and approved the final manuscript.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: D.A.C. is a founder and employee of Neemo Inc., at which M.E.P. has also been employed. Neemo Inc. has a financial interest in the production of biofuels using *Clostridium*. The remaining authors declare no competing interests.

How to cite this article: Pyne, M. E. et al. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. *Sci. Rep.* 6, 25666; doi: 10.1038/srep25666 (2016).