FEMS Microbiology Letters, 366, 2019, fnz059

https://doi.org/10.1093/femsle/fnz059
Advance Access Publication Date: 15 March 2019
Research Letter

RESEARCH LETTER – Biotechnology & Synthetic Biology

CRISPR-Cas, a highly effective tool for genome editing in *Clostridium saccharoperbutylacetonicum* N1-4(HMT)

Aretha N. Atmadjaja[1,2,‡], Verity Holby[1,‡], Amanda J. Harding[1], Preben Krabben[1,†], Holly K. Smith[1] and Elizabeth R. Jenkinson[1,*]

[1]Green Biologics Ltd, R&D labs, 154AH Brook Drive, Milton Park, Abingdon OX14 4SD, UK and [2]School of Life Sciences, University of Nottingham, Nottingham NG7 2UH, UK

*Corresponding author: Green Biologics Ltd, 154AH Brook Drive, Milton Park, Abingdon OX14 4SD, UK. Tel: +44 1235 435710; Email: liz.jenkinson@greenbiologics.com

†Present Address: Microbesphere Ltd., 9 Marsh Lane, Didcot, Oxfordshire, UK. OX11 8FD.

One sentence summary: This research establishes a genome editing tool using the endogenous CRISPR-Cas mechanism of solventogenic clostridia.

‡These authors contributed equally to this work.

Editor: Richard van Kranenburg

ABSTRACT

The solventogenic clostridia have long been known for their ability to convert sugars from complex feedstocks into commercially important solvents. Although the acetone-butanol-ethanol process fell out of favour decades ago, renewed interest in sustainability and 'green' chemistry has re-established our appetite for reviving technologies such as these, albeit with 21st century improvements. As CRISPR-Cas genome editing tools are being developed and applied to the solventogenic clostridia, their industrial potential is growing. Through integration of new pathways, the beneficial traits and historical track record of clostridial fermentation can be exploited to generate a much wider range of industrially relevant products. Here we show the application of genome editing using the endogenous CRISPR-Cas mechanism of *Clostridium saccharoperbutylacetonicum* N1-4(HMT), to generate a deletion, SNP and to integrate new DNA into the genome. These technological advancements pave the way for application of clostridial species to the production of an array of products.

Keywords: *Clostridium saccharoperbutylacetonicum* N1-4(HMT); genome editing; endogenous CRISPR-Cas; industrial biotechnology; sporulation; green chemistry

Abbreviations

ABE:	Acetone, Butanol and Ethanol
CGM:	Clostridial Growth Media
DR:	Direct Repeat
gDNA:	genomic DNA
HR:	Homology Region
HRM:	High Resolution Melt
Ldr:	Leader sequence
PAM:	Protospacer Adjacent Motif
RCM:	Reinforced Clostridial Media
SNP:	Single Nucleotide Polymorphism
Spc:	Spacer

INTRODUCTION

The use of microbial engineering to generate strains that can produce high value and bulk chemicals through fermentation of renewable and sustainable feedstocks has been gaining

Received: 26 September 2018; **Accepted:** 14 March 2019

ever more attention as environmental considerations, improved purity of fermentation-derived products, and the desire for secure, alternative energy sources has begun to impact on the appetite for sustaining oil-derived chemical processes (Woodley, Breuer and Mink 2013). The solventogenic clostridia are one class of microbes that hold promise for exploitation in this field of industrial biotechnology, in particular for the commercial production of acetone, butanol and ethanol (ABE) (Jones and Woods 1986).

To date, optimisation of strains to improve beneficial phenotypes, for example high productivities and altered solvent ratios, has used technologies such as adaptive lab evolution (Lin and Blaschek 1983; Liu, Gu and Yu 2012; Barrick and Lenski 2013), random mutagenesis and targeted genetic modifications (Cooksley *et al.* 2012; Croux *et al.* 2016). Significant recent advances in technologies such as gene synthesis and high throughput screening mean these can now be joined by the more sophisticated tools of synthetic biology (Gyulev *et al.* 2018), to tune and tweak pathways precisely towards the production of desired chemicals. In order to fully exploit the potential of these tools, more reliable and efficient methods for editing the genomes of clostridia need to be established.

Previously, the most significant breakthroughs have been made with *Clostridium acetobutylicum* for which reproducible knock out and knock in technologies have been available and widely used (Heap *et al.* 2010; Heap *et al.* 2012). However, these methods have their drawbacks, and with the discovery and subsequent advent of CRISPR-Cas genome editing, the application of this method to modify clostridial genomes is gaining popularity. CRISPR-Cas mediated modifications can be very specific and use the same basic protocol. A single nucleotide polymorphism (SNP), an in-frame deletion or an insertion can be made without leaving scars in the genome. The ability to make precise deletions that are not reliant on insertional inactivation (for example those generated by ClosTron, Heap *et al.* 2010) mitigate polar effects, enabling fine tuning of pathways in a more predictable manner. The technology does not overcome the inherent inefficiency of homologous recombination in clostridia (Wasels *et al.* 2017) but it does significantly reduce the time needed for identification of modified colonies without the need for counter-selection markers (Li *et al.* 2016).

To date Type II cas9-based technologies have been successfully applied in a number of clostridial strains, including *C. acetobutylicum* (Li *et al.* 2016), *C. beijerinckii* (Wang *et al.* 2015; Li *et al.* 2016), *C. pasteurianum* (Pyne *et al.* 2016) and *C. saccharoperbutylacetonicum* (Wang *et al.* 2017), with varying efficiencies. As methods have evolved, a two-step (Wasels *et al.* 2017) or inducible-*cas9* approach (Wang *et al.* 2016) has been shown to be more efficient than a single-step approach in which *Cas9* and gRNA are expressed simultaneously, and a cas9 with nickase activity (Cas9n) has demonstrated higher editing efficiency than the wild type cas9 in *C. acetobutylicum* and *C. beijerinckii* (Li *et al.* 2016). Despite this progress, cas9 editing in clostridia still has some disadvantages. Pyne *et al* and Zhang *et al* have shown that transformation of the cas9 endonuclease can significantly reduce transformation efficiencies for *C. pasteurianum*, *C. acetobutylicum* and *C. tyrobutyricum*. They subsequently identified the key features of the endogenous Type IB CRISPR-Cas machinery and successfully co-opted it for use as an efficient genome editing tool in both *C. pasteurianum* (Pyne *et al.* 2016) and *C. tyrobutyricum* (Zhang *et al.* 2018). The same technology has been developed for use in *C. saccharoperbutylacetonicum* (Jenkinson and Krabben 2016).

Here, we demonstrate for the first time the application of the endogenous CRISPR-Cas system from *C. saccharoperbutylacetonicum* N1-4(HMT) to generate strains carrying precise deletions, single base changes (SNPs) and integration of exogenous DNA. The versatility of this editing method enables us to explore the potential of clostridia as platform hosts for producing a myriad of commercially valuable biochemicals.

METHODS

Bacterial strains, plasmids and culture conditions

The strain used in this work was *Clostridium saccharoperbutylacetonicum* N1-4(HMT), DSM 14 923. Plasmids were based on either pMTL82154 or pMTL83251 (Heap *et al.* 2009) (Table 1). All relevant primer sequences are shown in Table 1.

Clostridial cultures were grown anaerobically in either reinforced clostridial media (RCM) (Oxoid), clostridial growth media (CGM) (5 g L^{-1} yeast extract, 0.75 g L^{-1} K_2HPO_4, 0.75 g L^{-1} KH_2PO_4, 0.4 g L^{-1} $MgSO_4$, 0.01 g L^{-1} $FeSO_4$, 0.01 g L^{-1} $MnSO_4$, 1 g L^{-1} NaCl, 2 g L^{-1} $(NH_4)_2SO_4$, 2 g L^{-1} asparagine, pH 6.6) or TYIR (2.5 g L^{-1} yeast extract, 2.5 g L^{-1} tryptone, 0.025 g L^{-1} $FeSO_4$, 0.5 g L^{-1} $(NH_4)_2SO_4$, 10 mM MES) with glucose added at 1%–5%. Recombinant clostridial strains were selected on 75 μg ml^{-1} of thiamphenicol or 40 μg ml^{-1} erythromycin. *E. coli* strains used in the construction of shuttle vectors were DH5α (NEB) or Turbo (NEB), and were selected on chloramphenicol (25 μg ml^{-1}) or erythromycin (500 μg ml^{-1}).

Identification of CRISPR-Cas features

Direct repeat (DR) and spacer (spc) sequences were identified in the genome sequence of *C. saccharoperbutylacetonicum* N1-4(HMT) (NC_020291.1) using the CRISPRFinder programme (Grissa, Vergnaud and Pourcel 2007). Putative protospacer adjacent motif (PAM) sites were identified by nucleotide BLAST analysis of the spc sequences in the *C. saccharoperbutylacetonicum* N1-4(HMT) CRISPR-Cas array (Altschul *et al.* 1990) and analysis of the flanking nucleotides.

Confirmation of PAM site

To assess efficiency of the putative PAMs, oligonucleotides were designed to match the sequence of spc 53 from the *C. saccharoperbutylacetonicum* N1-4(HMT) CRISPR-Cas array with a 5' CCA, CCT or CCC (Table 1). When annealed the primer pairs give a 5' blunt end and a 3' HindIII compatible overhang (Table 1, ref 1–6). 10 pmol of each primer were suspended in 100 μl molecular biology grade water (Sigma), and annealed: Primer mix was incubated at 95°C for 5 min and then gradually cooled, 1°C/min, from 70°C to 30°C. The final step was a temperature decrease from 30°C to 4°C over 5 min. Annealed primers were used as the 'insert' in a ligation reaction using pMTL83251 cut with SmaI and HindIII (or ZraI/HindIII for the no PAM control). Colonies were screened by PCR and sequenced before transforming into *C. saccharoperbutylacetonicum* N1-4(HMT).

Electroporation

Plasmids were transformed into *C. saccharoperbutylacetonicum* N1-4(HMT) using electroporation. Cultures were grown in 60 ml CGM with 5% glucose in anaerobic cabinets (Don Whitley) until

Table 1. Plasmids and primers used in this study. Upper case letters in primer sequences refer to restriction endonuclease or PAM sites incorporated into the sequence.

	Plasmid	Description	Reference
	pMTL83251	Backbone for homologous recombination vectors	Heap *et al.* 2009
	pMTL82154	Backbone for CRISPR-Cas targeting vectors	Heap *et al.* 2009
	pMTL8325_Spc53_CCA	Confirmation of PAM sequence	This work
	pMTL8325_Spc53_CCT		
	pMTL83251_Spc53_CCC		
	pMTL83251_Spc53		
	pMTL82154_Spo0A_SNP	Recombination vector for making Spo0A SNP	This work
	pMTL83251_Ldr_Spo0A_spc	Targeting vector for selecting Spo0A mutants, both SNP and deletion	This work
	pMTL82154_Spo0A_del	Recombination vector for making Spo0A deletion	This work
	pMTL82154_int_1kb	Recombination vector for making 1 kb integration	This work
	pMTL82154_int_3kb	Recombination vector for making 3 kb integration	This work
	pMTL82154_int_5kb	Recombination vector for making 5 kb integration	This work
	pMTL83251_Ldr_Spc_int	Targeting vector for selecting integrated mutants	This work
Ref	**Primer name**	**Sequence**	**Notes**
1	Spc_53_PAM_T_F	CCTattactagaatgggaggaaaaagcaaattaagaaacA	Anneals with Spc_53_PAM_T_R to make spc region 53 with CCT PAM
2	Spc_53_PAM_T_R	AGCTTgtttcttaatttgctttttcctcccattctagtaatAGG	Pair with Spc_53_PAM_T_F
3	Spc_53_PAM_A_F	CCAattactagaatgggaggaaaaagcaaattaagaaacA	Anneals with Spc_53_PAM_A_R to make spc region 53 with CCA PAM
4	Spc_53_PAM_A_R	AGCTTgtttcttaatttgctttttcctcccattctagtaatTGG	Pair with Spc_53_PAM_A_F
5	Spc_53_PAM_C_F	attactagaatgggaggaaaaagcaaattaagaaacA	Anneals with Spc_53_PAM_C_R to make spc region 53. Clone into SmaI site to make CCC PAM
6	Spc_53_PAM_C_R	AGCTTgtttcttaatttgctttttcctcccattctagtaat	Pair with Spc_53_PAM_C_F
7	spo0A_SNP_HR1_F	agggaaaattatggggtaag	Amplifies Spo0A 857 bp upstream of SNP. Pair with Spo0A_SNP_HR2_R
8	Spo0A_SNP_HR2_R	gttaaatttaaagtttcatttagtctc	Amplifies Spo0A 680 bp downstream of SNP.
9	SpoHR1_F	gaatggacaaaaggagagaaagaaagatgg	Amplifies spo0A SNP region. Anneals in HR region.
10	spo0A_seq_R	aggtaagttgctggtaacaactcacc	Pairs with spoHR1 _F Will only anneal in the chromosome.
11	spo0A_SNP_HRM_F	ggatatactgttcatacagaaaaagg	HRM primer for nested PCR
12	spo0A_SNP_HRM_R	ctaaataagtctgtcattagacctag	Pair with spo0A_SNP_HRM_F
13	spo0A_del_HR1_F	atggacctcaagaatttgatatag	Amplifies HR1 for Spo0A deletion
14	spo0A_del_HR1_R	cttatcagcgatatcatcggcaataagtacag	Pair with spo0A_del_HR1_F
15	spo0A_del_HR2_F	cttattgccgatgatatcgctgataagcttcgtttg	Amplifies HR2 for Spo0A deletion
16	spo0A_del_HR2_R	taaacaactctgtctcttctac	Pair with spo0A_del_HR2_F
17	intvectorHR1_F	GCGGCCgctaaatcacttttatatgcaattaatgg	Amplifies HR1 for integration
18	intvectorHR1_R	ACGCGTcaagttacaatttacaattacttatg	Pair with intvectorHR1_F
19	intvectorHR2_F	CTCGAGgtaaaacatttacaatagaaatc	Amplifies HR2 for integration
20	intvectorHR2_R	AGGCCTctttaatgaaattactatttttttgatac	Pair with intvectorHR2_F
21	LP_5kb_F1	ATTATAGagctccaaacgtcagcgagctggttg	
22	LP_1kb_R1	ctggcctattaccacaaagccattcc	Pair with LP_5kb_F1 to amplify 1 kb fragment from λDNA
23	LP_3kb_R1	ccagcgatggcggttatggtttg	Pair with LP_5kb_F1 to amplify 3 kb fragment from λDNA
24	LP_5kb_R1	cgatgacgcctgtacgcatttgg	Pair with LP_5kb_F1 to amplify 5 kb fragment from λDNA
25	Int_genSeqF2	tcttggaacttcagcagaacc	Binds upstream of integration site. Will only anneal in the chromosome
26	LP_seq_R1	ggttcagggatcgcctcacc	Reverse primer out of lambda DNA fragment

mid-exponential phase. The media was dispensed into an Erlenmeyer flask and left in the anaerobic cabinet overnight to equilibrate. Cells were pelleted at 4000 x *g* for 10 min at 4°C and washed in EPB_S buffer (300 mM sucrose, 0.6 mM Na_2HPO_4, 4.4 mM NaH_2PO_4, 10 mM $MgCl_2$), re-centrifuged and resuspended in 2 ml of EPB_NS buffer (as for EPB_S but without $MgCl_2$). 500–1000 ng DNA, was added to 200 µl cells in 2 mm electroporation cuvettes (Cell Projects) and incubated on ice for 5 min before pulsing at 1.5 kV (BioRad). All steps, except for centrifugation, were carried out in the anaerobic cabinet. Cells were plated on CGM agar containing 5% glucose and relevant antibiotic.

CRISPR-CAS GENOME EDITING

Recombination plasmids

Spo0A SNP

The *spo0A* region (approx. 1.5 kb) was amplified from genomic DNA (gDNA) and blunt cloned into the pMTL82154 SmaI site (Table 1, ref 7–8). Required mutations were incorporated through site directed mutagenesis following kit instructions (Quikchange SDM, Agilent).

Spo0A Deletion

The deletion vector was constructed using overlapping PCR. The homology regions (HR1 and HR2, both 760 bp) were PCR amplified from gDNA. The 'Spo0A_del_HR1_R' and 'Spo0A_del_HR2_F' primers (Table 1, ref 14 and 15) had a region of homology so the HR1 and HR2 PCR products could be used in a second round to prime off of each other before carrying out the final round using 'Spo0A_del_HR1_F' and 'Spo0A_del_HR2_R' (Table 1, ref 13 and 16) to obtain the full length 'deletion' cassette. This was then blunt cloned into the SmaI site of pMTL82154.

Integration

The integration site is between two transcriptional terminators of the formate acetyltransferase (Cspa_c10940) and rubredoxin (Cspa_c10950) genes. The vector was constructed through sequential cloning into pMTL82154. First the HR1 region (approx. 950 bp) was PCR amplified from gDNA (Table 1, ref 17–18) and cloned into the NotI/MluI sites followed by addition of 370 bp which consisted of a short cloning site and 358 bp promoter sequence from Cspa_c25620. The second transcriptional terminator site (found 76 bp upstream of Cspa_c10950) was synthesised (Life Technologies) with a modification to remove the PAM site and cloned into the NcoI/XhoI sites. Finally the HR2 region (approx. 950 bp) was PCR amplified from gDNA (Table 1, ref 19–20) and cloned into the XhoI/PvuII sites.

The λDNA fragments were PCR amplified (DNA from NEB) using primers shown in Table 1, ref 21–24, and ligated into the SacI and StuI sites of the integration vector multiple cloning site.

All recombination vectors were checked by PCR and Sanger sequencing before transforming into *C. saccharoperbutylacetonicum* N1-4(HMT).

Targeting spc plasmids

Spo0A targeting spc

The leader (Ldr) sequence was identified as the region between the end of cas 2 and the first DR (Fig. 1A). This 181 bp sequence was synthesised (Life Technologies) and cloned into pMTL83251 using BamHI/StuI sites. The targeting spc flanked by DRs 'DR_Spo0Aspc_DR' was also synthesised and cloned downstream of the Ldr using StuI/HindIII sites.

Integration targeting spc

The 'DR_IntSpc_DR' targeting construct was synthesised (Life Technologies) and cloned under the Ldr sequence in the same way.

To validate the 'killing efficiency' of the targeting vectors, they were transformed into *C. saccharoperbutylacetonicum* N1-4(HMT) and colony numbers compared to an empty pMTL83251 plasmid.

Genome editing

Recombination vectors were transformed into *C. saccharoperbutylacetonicum* N1-4(HMT) first; following recovery the transformants were plated on selective agar (thiamphenicol) and incubated in an anaerobic cabinet at 32°C. After approximately 24–48 h, the transformants were restreaked onto a fresh plate and the presence of the recombination plasmid confirmed by colony PCR.

Homologous recombination at the target site was promoted by subculturing several times in 10 ml RCM containing thiamphenicol. The final subculture was inoculated into 60 ml CGM and a second electroporation carried out as described previously to transform the targeting spc plasmid into the cells. Transformants were plated on selective agar (erythromycin). Any transformants that were recovered at this stage were checked for the presence of the desired modification as described below.

Identification of edited mutants

Cell lysis

Clostridial colonies were streaked onto CGM agar plates with a small amount of colony being resuspended in a PCR tube containing 5 μg proteinase K (NEB) in 25 μl TE pH 8.0. Tubes were incubated at 55°C for 15 min followed by 80°C for 15 min. 1 μl was used as the template for PCR.

SNPs

High Resolution Melt (HRM) curve analysis was used to identify a single base pair change. In a two-step process, primers were designed to amplify a 1 kb region of the genome containing the SNP, with at least one primer annealing to a sequence out-with the homology arms (Table 1, ref 9–10). HRM works best with short amplicons so the resultant PCR product was then diluted to a copy number of 1×10^4 and used as the template to amplify 50–150 bp in a second round of PCR: 5 μl Precision Melt Supermix (Bio-Rad), 0.2 μM each primer (Table 1, ref 11–12), 1 μl template, dH_20 to 10 μl for each sample. The following controls were included: WT template, SNP template (from a plasmid, 1 ng), no template control and a negative control (from the first round PCR, diluted 1/5000). Amplification was carried out using the CFX96 real-time PCR detection system (Bio-Rad): 40 cycles of 95°C for 10 s, 59.9°C for 30 s and 72°C for 30 s followed by 95°C for 30 s and a final hold at 60°C for 1 min before performing the end point HRM melt curve. This was set at 60°C to 95°C with 0.2°C increments for 10 s. The melt curve was analysed using Precision Melt Analysis software (Bio-Rad). The melting points of the double stranded amplicon for each transformant were compared against the controls and the colony that grouped with the 'SNP template' was confirmed by Sanger sequencing.

Deletions

The reaction was prepared using Illustra PureTaq RTG beads (GE Healthcare) with 1 μl template and primers 'SpoHR1_F' and 'spo0A_seq_R' (Table 1, ref 9–10). An initial denature step at 95°C for 5 min was followed by 35 cycles of 95°C for 30 s, 64.6°C for 30 s, 72°C for 2 min and a final extension step of 72°C for 10 min.

Integrations

Amplification was carried out using Phusion polymerase (NEB) using 1 μl template and primers 'Int_genSeqF2' and 'LP_seq_R1' (Table 1, ref 25–26). An initial denature step at 95°C for 30 s was followed by 35 cycles of 98°C for 10 s, 62.3°C for 30 s, 72°C for 1 min and a final extension step of 72°C for 10 min.

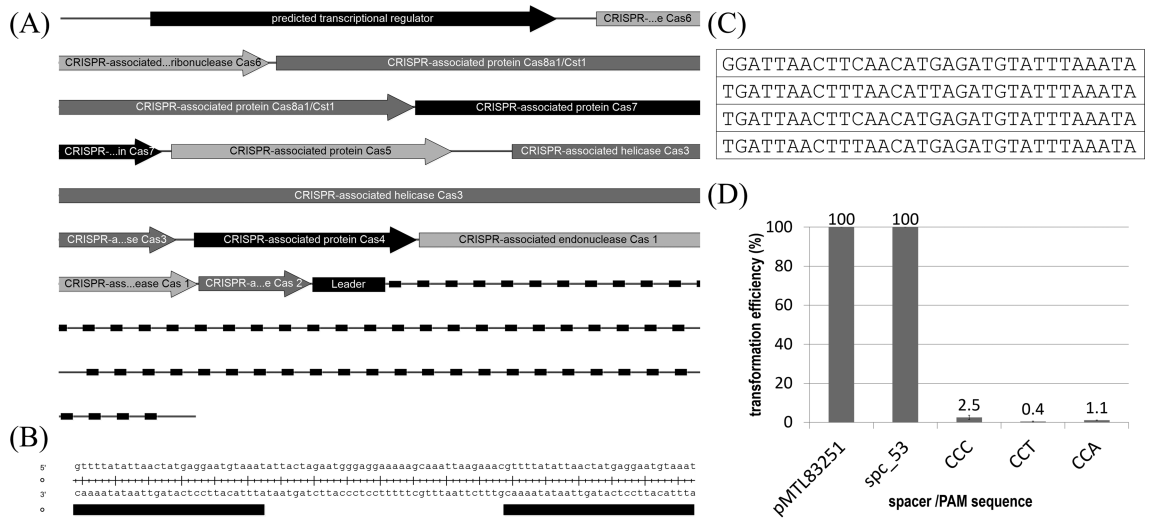


Figure 1. Characterisation of *C. saccharoperbutylacetonicum* N1-4(HMT) CRISPR-Cas features. (A) The CRISPR-Cas gene cluster. (B) Sequence of the main cluster DR with spc 53. (C) Sequence of DRs in putative second cluster. (D) Plasmids carrying the spc 53 sequence with putative PAM sequences (CCC, CCT or CCA) were transformed and number of colonies compared to empty vector (pMTL83251) and a 'no PAM' control (spc_53).

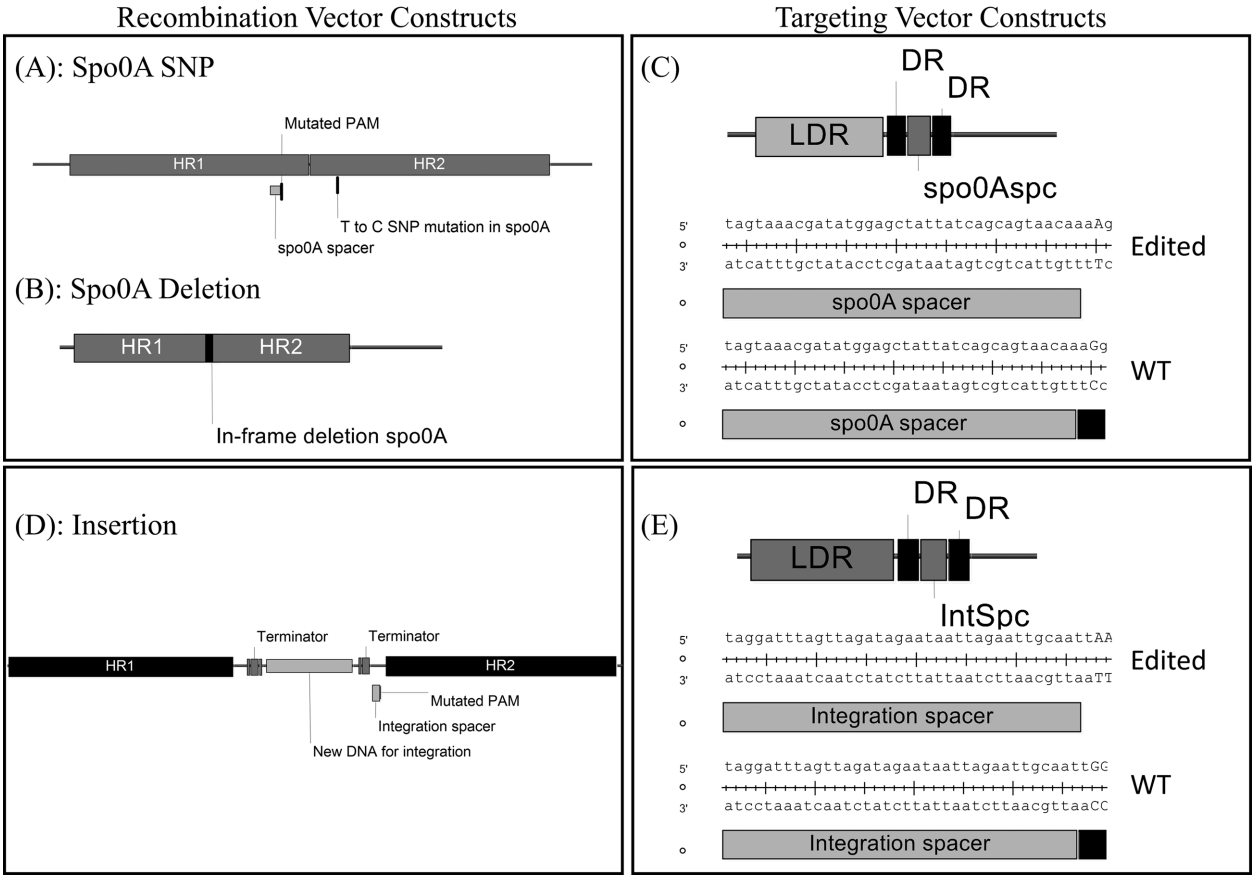


Figure 2. Overview of recombination and targeting cassettes for 'hijacking' endogenous CRISPR-Cas for genome editing. (A) Recombination vector for generating the *spo0A* SNP carries two mutations: one for generating the T to C SNP, the second for modifying the PAM site. (B) Recombination vector for generating the in-frame *spo0A* deletion. The resultant construct no longer carries the PAM_protospacer sequence. (C) Targeting construct for selecting mutant *spo0A* strains (SNP or deletion) over the WT sequence. The Ldr_DR_Spo0Aspc_DR cassette will recognise the WT sequence as it still has a functional PAM (CCT). (D) Recombination vector for generating the integration strains. The PAM has been mutated in this construct. (E) Targeting construct for selecting integrants over the WT sequence. The Ldr_DR_IntSpc_DR cassette will recognise the WT sequence as it still has a functional PAM (CCA).

Bottle screens

Cultures were grown anaerobically in RCM in serum bottles overnight at 32°C to an OD_{600nm} > 2.0 and pH 5.0–5.5. From this starter culture, 25 ml TYIR was inoculated using a 15% inoculum. The subculture was grown anaerobically at 32°C for 3–5 h, until it reached OD_{600nm} of 1.5 ± 0.2. A triplicate of 54 mL TYIR media in serum bottles were each inoculated with 6 mL of the subculture. These bottles were incubated at 32°C in an anaerobic cabinet and sampled every 3–4 h for 72 h. pH and OD_{600nm} were measured at each sample point. Sugar, acids, and solvent concentrations were measured by HPLC.

Genome sequencing

At least 1 μg of purified gDNA with a concentration of at least 10 ng/μl was provided to GATC-Biotech for sequencing by Illumina HiSeq. Genome analysis was done in-house using bwa 0.5.9-r16 and samtools 0.1.17 to compare the sequence data to the reference genome sequence and to identify SNPs and InDels.

RESULTS

Identification of the CRISPR-Cas features in *C. saccharoperbutylacetonicum* N1-4(HMT)

Due to the relative ease of transforming plasmid DNA into *C. saccharoperbutylacetonicum* N1-4(HMT) (compared with the closely related *C. saccharoperbutylacetonicum* N1-504), we chose to develop a genome editing system for this strain. *Clostridium saccharoperbutylacetonicum* N1-4(HMT) has an easily identifiable CRISPR-Cas operon and with the presence of a *cas8a1* homologue, is classified as Type IA: gene order *cas6-cas8a-cas7-cas5-cas3-cas4-cas1-cas2* (Fig. 1A) (Makarova *et al.* 2011). The 'Ldr' sequence which drives transcription of the DR_spc cluster was presumed to be within the 181 bp between the end of the *cas2* gene and the beginning of the first DR sequence (Fig. 1A).

This strain has two potential spc clusters, the first is clear-cut and consists of 60 spcs, most of 36 or 37 bp (except for one significant exception, spc 34 is 102 bp), located immediately downstream of the cas array. For this cluster the DR sequence is consistent (Fig. 1B). The second putative cluster is not associated with the CRISPR-Cas array and has just three spcs, each of 34 or 35 bp, located between Cspa_c05630 and Cspa_c05640. In this case the DR sequences are all slightly different (Fig. 1C).

The PAM site which distinguishes 'invading' DNA (e.g. from a phage or plasmid) from the bacterial genomic DNA was identified through BLAST identification of each of the spcs located in the first *C. saccharoperbutylacetonicum* N1-4(HMT) DR_spc cluster and subsequent analysis of both up-and down-stream regions. Two spcs, 31 and 53, had hits to phage DNA and analysis of the sequences led us to identify CCN as a promising PAM candidate. The triplets CCC, CCA and CCT were cloned into pMTL83251 along with the spc 53 sequence copied from the *C. saccharoperbutylacetonicum* N1-4(HMT) array. Resulting plasmids were transformed into *C. saccharoperbutylacetonicum* N1-4(HMT) and all three PAM sequences caused significant reduction in transformation efficiencies compared with empty vector control and a plasmid carrying spc 53 with no functional PAM, 'GAC' (Fig. 1D). This shows the endogenous CRISPR-Cas machinery is able to recognise the plasmid as 'invading DNA' only when this CCN triplet is present. Subsequently we have shown through genome targeting that CCG also works as a PAM (not shown).

Genome editing

We used this information as a basis for developing a genome editing technique. For the examples described, we constructed targeting vectors that mimic the native CRISPR-Cas DR clusters, for example: Ldr_DR_targetingspacer_DR, where the targeting spc was directed to a protospacer sequence in the clostridial genome, adjacent to a CCN PAM site. This allowed us to specifically target sites within the WT clostridial genome. When this plasmid is transformed into the cell, the microbe is unable to efficiently repair the DNA cleavage and is therefore lost from the population. By combining this with a homologous recombination step, we could significantly improve the selection method for microbes containing the desired modifications: SNPs, deletions or integrations (Fig. 2A–E).

SNPs and deletions

To generate a SNP in the *C. saccharoperbutylacetonicum* N1-4(HMT) genome, the *spo0A* gene was targeted (Fig. 2A). Previous work has shown that a deletion of this gene results in a disruption not just to sporulation but also to solventogenesis as the two processes are tightly linked (Ravagnani *et al.* 2000; Harris, Welker and Papoutsakis 2002). However, modifications in the C-terminal helix-turn-helix and σA activator region have been shown to decouple these processes, generating asporogenic strains that are still capable of producing ABE at high titres, a phenotype that is desirable for the industrial production of solvents (Sandoval *et al.* 2015). Thus, as well as being an easy-to-screen for modification, it also has relevance to commercialisation of the clostridial ABE process.

We modified a single nucleotide at position 782 in the *C. saccharoperbutylacetonicum* N1-4(HMT) *spo0A* gene with the intention of changing a conserved isoleucine in the C-terminal protein domain to threonine (I261T) (Fig. 3A). Mutations within the last 15 bp of the *B. subtilis spo0A*, and specifically amino acids 257, 258 and 260, have been shown to play a role in transcription activation (Rowe-Magnus, Richer and Spiegelman 2000). A plasmid carrying homology arms either side of the desired SNP was generated with a silent mutation also being incorporated approx. 180 bp upstream of the SNP to 'knock out' the PAM site (Fig. 2A, C). Validation of the targeting spc (Ldr_DR_Spo0Aspc_DR) showed it was > 99.5% efficient at killing the WT cells (not shown). After transformation with the targeting plasmid, just five colonies were obtained and only one was confirmed as positive by HRM and Sanger sequencing (Fig. 3B). Genome sequencing of this colony showed no additional SNPs or InDels and the reason for the low efficiency of this target is currently unknown.

Deletions are much easier to generate and can be achieved using the same method (Fig. 2B). In this case the design of the HR vector resulted in a 248 amino acid in-frame deletion leaving behind just the first and last 13 amino acids of the Spo0A coding region to avoid inadvertently impacting on any up- or down-stream DNA regulatory sequences. The same targeting construct was used (Ldr_DR_Spo0Aspc_DR); construction of this mutant was more successful and of the 30 colonies tested, 100% were positive for the deletion (Fig. 3C).

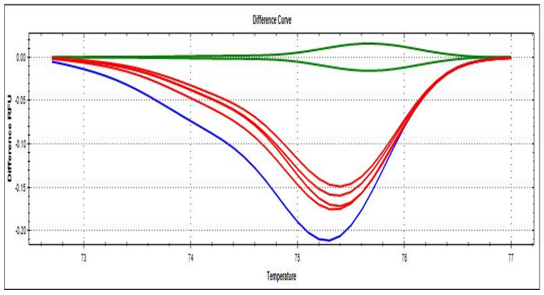
To compare the impact of the Spo0A SNP and the in-frame gene deletion on solventogenic and sporogenic phenotypes, small scale bottle screens and microscopy analyses were undertaken (Fig. 4): The WT and Spo0A SNP strains both produced ABE with low final acid titres as expected, whereas the Spo0A deletion strain was unable to enter solventogenesis and produced acids instead of solvents (Fig. 4A normalised data). Light

(A)

```
WT_Spo0A        MEDSKISVLIADDNKEFCSILNDYLLNQKDIVVTGIAKDGREALDLIVERKPDLVILDII
SNP_Spo0A       MEDSKISVLIADDNKEFCSILNDYLLNQKDIVVTGIAKDGREALDLIVERKPDLVILDII
                ************************************************************

WT_Spo0A        MPHLDGLGVLEKLNTMNLEKVPRIIILSAVGQDKITQQAITLGADYYTVKPFDMEVFTKR
SNP_Spo0A       MPHLDGLGVLEKLNTMNLEKVPRIIILSAVGQDKITQQAITLGADYYTVKPFDMEVFTKR
                ************************************************************

WT_Spo0A        IREMFNGAPAQESNVRASSYMQSPVMTSGEPKSKTPVDLETEITNIIHEVGVPAHIKGYM
SNP_Spo0A       IREMFNGAPAQESNVRASSYMQSPVMTSGEPKSKTPVDLETEITNIIHEVGVPAHIKGYM
                ************************************************************

WT_Spo0A        YLREAITMVVNDMELLSAVTKELYPSIAKKYNTTASRVERAIRHAIEVAWGRGQIDAINR
SNP_Spo0A       YLREAITMVVNDMELLSAVTKELYPSIAKKYNTTASRVERAIRHAIEVAWGRGQIDAINR
                ************************************************************

WT_Spo0A        LFGYTVHTEKGKPTNSEFIAIIADKLRLKNKVS
SNP_Spo0A       LFGYTVHTEKGKPTNSEFIATIADKLRLKNKVS
                ********************* ************
```

(B)

Well	Fluor	Content	Sample	Cluster	Percent Confidence
E01	SYBR	Unkn	col 4	Cluster 1	96.5
H01	SYBR	Pos Ctrl	SNP ctrl	Cluster 1	99.6
B01	SYBR	Unkn	col 1	Cluster 2	99.7
C01	SYBR	Unkn	col 2	Cluster 2	99.8
D01	SYBR	Unkn	col 3	Cluster 2	99.8
G01	SYBR	Neg Ctrl	WT ctrl	Cluster 2	99.1
F01	SYBR	Unkn	col 5	Cluster 3	95.1



(C)

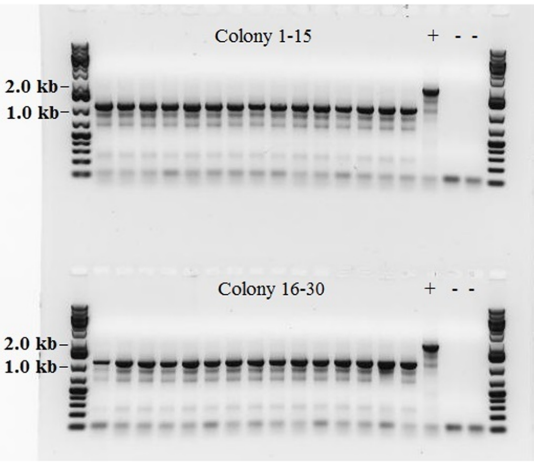


Figure 3. Analysis of *spo0A* SNP and deletion strains. (A) Spo0A amino acid alignment (Clustalx 2.1). Key C-terminal features are identified: DNA binding motif in bold, σA activator region highlighted (Sandoval *et al.* 2015), transcription activation region in italics (Rowe-Magnus, Richer and Spiegelman 2000). (B) HRM analysis of *spo0A* SNP colonies clustered using Precision Melt Analysis software. (C) Agarose gel showing colony PCR result. 100% of the tested colonies contain the deletion. Controls: WT gDNA (+ control), lysis reaction (-ve control) and dH_2O (-ve control). Ladder: GeneRuler 1 kb Plus DNA Ladder (ThermoFisher Scientific).

microscopy of the WT and Spo0A SNP strains confirmed the asporogenic phenotype of the SNP mutant (Fig. 4B).

Integrations

Clostridia are promising as host strains for the production of non-native chemicals, a key advantage being their ability to metabolise a wide range of feedstocks including both C5 and C6 sugars (Wayman and Yu 1985; Ezeji, Qureshi and Blaschek 2007; Sun *et al.* 2014). In order to use these microbes as biochemical factories in an industrial process it is preferable to be able to integrate new non-endogenous genes and pathways into the genome, rather than rely on plasmid-encoded pathways that require antibiotics for stability and may lack copy number control which can be vital for balancing pathway flux.

We attempted to integrate fragments of lambda phage DNA into a site 4.8 kb downstream of the *pyrE* gene in C. *saccharoperbutylacetonicum* N1-4(HMT). The specific site is flanked by transcriptional terminators (Fig. 2D), thereby reducing impact on neighbouring genes. Fragments of increasing size (1, 3 and 5 kb) were cloned into the recombination vector flanked by regions of homology. The targeting plasmid selectively targeted the integration region in the wild type strain (Fig. 2E). Validation of the Ldr_DR_IntSpc_DR targeting spc showed high, but not quite 100% efficiency for killing the WT strain (not shown). For the 1 and 3 kb lambda phage fragments integration was successful with 60% of tested colonies carrying the integrated 1 kb fragment and 53% of tested colonies carrying the integrated 3 kb fragment (Fig. 5). This has been confirmed by Sanger sequencing. The 5 kb fragment proved to be more difficult to integrate and although 35% of colonies tested indicated integration had occurred, all of them showed a smaller fragment of the lambda phage DNA than anticipated which may be due to instability of the relatively large (12.7 kb) plasmid (Gu *et al.* 2019). In addition a reduction in transformation efficiencies was observed compared with the empty vector as the insert size increased; empty vector = 63.7 cfu/pmol,

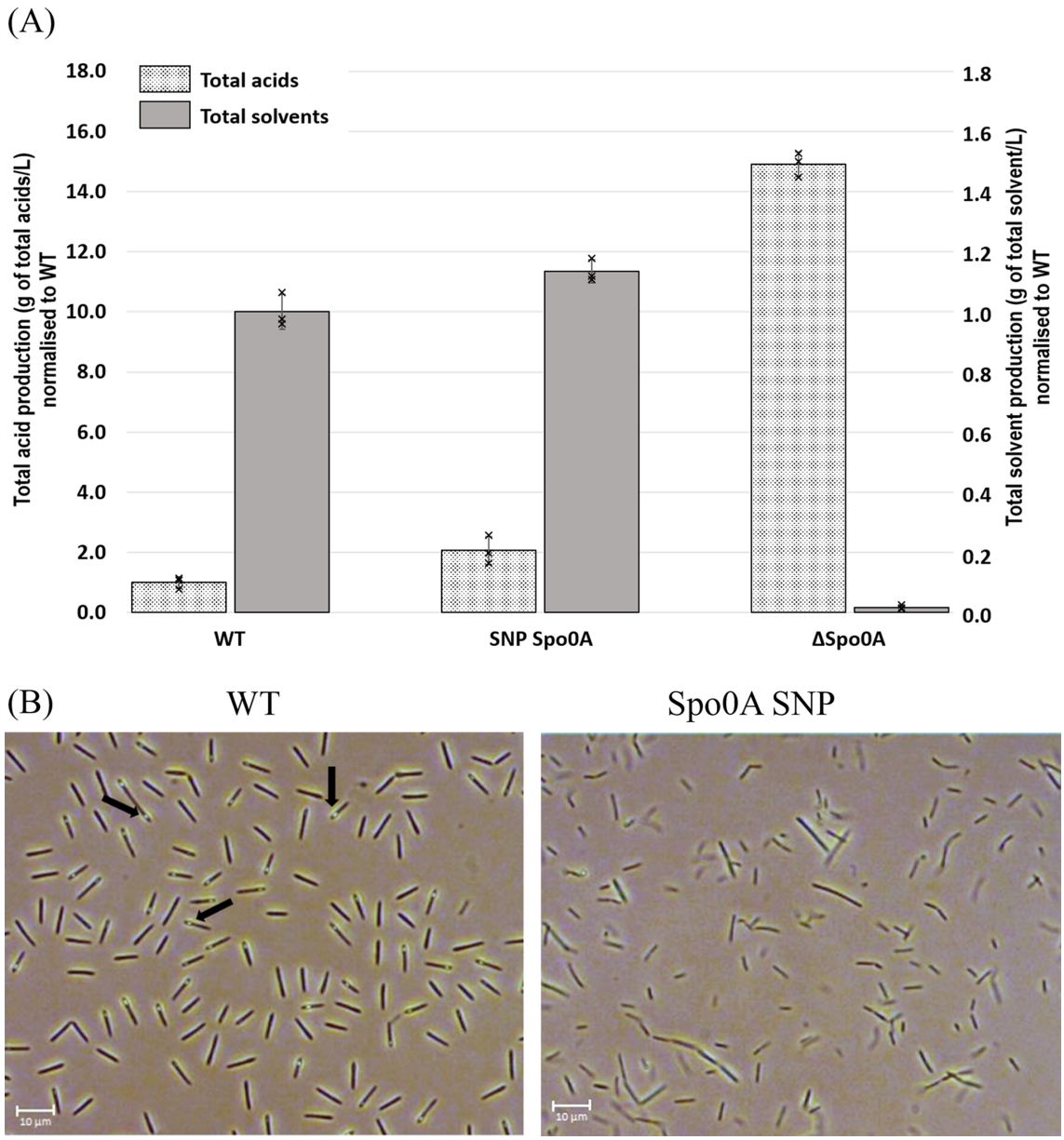


Figure 4. Phenotypic characterisation of *spo0A* mutants. (A) Bottle screen solvent and acid analysis for WT, *spo0A* SNP and *spo0A* deletion strains. Cultures were grown in TYIR medium and sampled at 72 h. Acetone, Butanol, Ethanol, Butyric Acid and Acetic Acid were measured by HPLC. (B) Light microscopy of WT and *spo0A* SNP cultures. After 20 h, the WT strain began to sporulate as indicated by the arrows. No spores were observed for the *spo0A* SNP strain after 22 h (image) or within the 72 h screen.

1 kb insert = 51.6 cfu/pmol, 3 kb insert = 19.2 cfu/pmol, 5 kb insert = 13.9 cfu/pmol (N.B. the amount of plasmid transformed was calculated in pmol to account for size differences).

DISCUSSION

We have shown for the first time that the endogenous CRISPR-Cas system of *C. saccharoperbutylacetonicum* N1-4(HMT) can be used to introduce any type of modification: SNPs, deletions and integrations, into the clostridial genome (Jenkinson and Krabben 2016). Through the use of a dual-vector process, in which the recombination and targeting stages are temporally separated, we have effectively circumvented the problem of low recombination frequency that has previously made engineering of clostridia difficult. By using the endogenous CRISPR-Cas to remove cells carrying the WT sequence from the population we can enrich for those that carry the desired modifications.

In general we have found this to be an efficient technology, however the CRISPR-Cas selection method is very powerful and there is a risk that the cell can overcome it through generating mutations either within the genomic *cas* gene cluster or within the plasmid carrying the 'Ldr_DR_spc_DR' cassette. To date, we have not seen mutations within the *cas* cluster, although as reported by Jiang *et al* we have seen that transformation of the targeting plasmid into WT cells can sometimes result in a small number of colonies being recovered (Jiang *et al*. 2013). In a number of these, sequencing the 'Ldr_DR_spc_DR' region of the plasmid revealed mutations within the DR or spc which prevents the selection from working. Spacer design continues to be an area for optimisation as some appear to be 100% efficient whereas

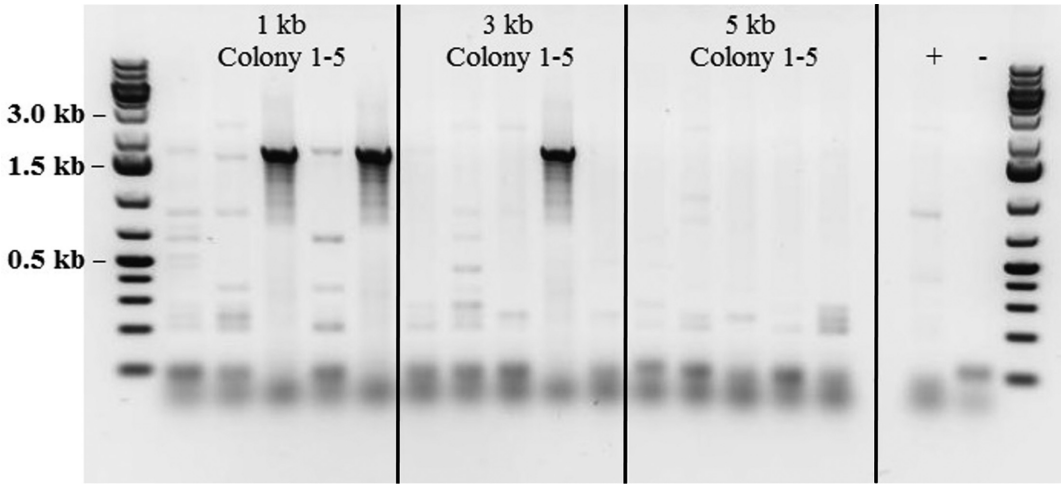


Figure 5. Identification of positive integration recombinants. Agarose gel showing colony PCR result for five colonies each of the 1, 3 and 5 kb integrants. The integration editing experiment was repeated three times (gel shown is from one experiment). Overall, 1 kb gave the highest number of positive hits (1.6 kb PCR product compared with no product for the WT). Controls: WT gDNA (+ control) and dH_2O (-ve control). Ladder: GeneRuler 1 kb Plus DNA Ladder (ThermoFisher Scientific).

others seem more likely to promote the 'escaper' phenotype (Jiang *et al.* 2013).

The primary application of the integration technology is in the development of industrial strains designed to carry new pathways. Given the challenges we faced in integrating fragments larger than 3 kb, further improvements are needed. Recently Gu *et al* used CRISPR-cas9 editing to cure *C. saccharoperbutylacetonicum* N1-4(HMT) of its endogenous megaplasmid (Gu *et al.* 2019). The megaplasmid-cured mutant was shown to have improved transformation efficiencies over the wild type, and, to some extent, the deletion mitigated instability of non-endogenous plasmids. Based on these observations, repeating the integration experiments in a megaplasmid-cured background may improve recovery of positive integrants with all insert sizes.

The site of integration also needs considerable thought as the structure of the bacterial chromosome can have a significant impact on transcription efficiency (Vora *et al.* 2009; Bryant *et al.* 2014). We have tested multiple regions of the genome with varying degrees of success. The site described within this paper is relatively reproducible in terms of integrating new DNA, however a site tested downstream of the macrolide export protein *macB1* (Cspa_c07400), consistently failed to yield integrants.

The solventogenic clostridia have the potential to be industrial workhorses, not just for the production of solvents and organic acids. Technologies such as this are a major step forwards, towards new industrial applications.

ACKNOWLEDGEMENTS

The authors are grateful to members of the molecular microbiology and analytical chemistry groups at Green Biologics, in particular Dr Anna Baker and Dr Chris Hills for support and useful discussions, and Dr Mohini Thite and Emily Rose for sample analysis. Thanks also to Dr Alan Goddard, Aston University, for reviewing this manuscript.

FUNDING

The work on *spo0A* was supported by the European Union Marie Sklodowska Curie Innovation Training Network (ITN)—CLOSPORE, contract number 642068, awarded to ANA.

Conflicts of interest. Authors ANA, VH, AJH, HKS and ERJ are employed by Green Biologics Ltd. The PAM identification work described in this article has been previously published as an example in a granted UK patent (GB2530831B). Several other patents are pending based on this work including **US** 20170037393 and **EP** 3132036.

REFERENCES

Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.

Barrick JE, Lenski RE. Genome dynamics during experimental evolution. *Nat Rev Genet* 2013;**14**:827–39.

Bryant JA, Sellars LE, Busby SJW *et al.* Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res* 2014;**42**:11383–92.

Cooksley CM., Zhang Y., Wang H *et al.* Targeted mutagenesis of the *Clostridium acetobutylicum* Acetone-Butanol-Ethanol fermentation pathway. *Metab Eng* 2012;**14**:630–41.

Croux C, Nguyen NPT, Lee J *et al.* Construction of a restriction-less, marker-less mutant useful for functional genomic and metabolic engineering of the biofuel producer *Clostridium acetobutylicum*. *Biotechnol Biofuels* 2016;**9**:23.

Ezeji TC, Qureshi N, Blaschek HP. Bioproduction of butanol from biomass: from genes to bioreactors. *Curr Opin Biotechnol* 2007;**18**:220–7.

Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 2007;**35**(Web Server issue):W52–7.

Gu Y, Feng J, Zhang Z-T *et al.* Curing the endogenous megaplasmid in *Clostridium saccharoperbutylacetonicum* N1-4(HMT) using CRISPR-Cas9 and preliminary investigation of the role of the plasmid for the strain metabolism, *Fuel* 2019;**236**:1559–66.

Gyulev IS, Willson BJ, Hennessy RC *et al.* Part by part: synthetic biology parts used in solventogenic Clostridia, *ACS Synth Biol* 2018;**7**:311–27.

Harris LM, Welker NE, Papoutsakis ET. Northern, morphological, and fermentation analysis of spo0A inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. *J Bacteriol* 2002;**184**:3586–97.

Heap JT, Ehsaan M, Cooksley CM *et al*. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Res* 2012;**40**:e59.

Heap JT, Kuehne SA, Ehsaan M *et al*. The ClosTron: Mutagenesis in Clostridium refined and streamlined. *J Microbiol Methods* 2010;**80**:49–55.

Heap JT, Pennington OJ, Cartman ST *et al*. Modular system for Clostridium shuttle plasmids. *J Microbiol Methods* 2009;**78**:79–85.

Jenkinson E, Krabben P, Targeted Mutations, UK granted patent GB2530831B (first published in 2015 as WO 2015159087A1), 2016.

Jiang W, Bikard D, Cox D *et al*. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 2013;**31**:233–9.

Jones DT, Woods DR. Acetone-butanol fermentation revisited. *Microbiol Rev* 1986;**50**:484–524.

Li Q, Chen J, Minton NP *et al*. CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnol J* 2016;**11**:961–72.

Lin YL, Blaschek HP. Butanol production by a butanol-tolerant strain of *Clostridium acetobutylicum* in extruded corn broth. *Appl Environ Microbiol* 1983;**45**:966–73.

Liu XB, Gu QY, Yu XB. Repetitive domestication to enhance butanol tolerance and production in *Clostridium acetobutylicum* through artificial simulation of bio-evolution. *Bioresour Technol* 2012;**130**:638–43.

Makarova KS, Haft DH, Barrangou R *et al*. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 2011;**9**:467–77.

Pyne ME, Bruder M, Moo-Young M *et al*. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Sci Rep* 2016;**6**:25666.

Ravagnani A, Jennert KC, Steiner E *et al*. Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. *Mol Microbiol* 2000;**37**:1172–85.

Rowe-Magnus D, Richer MJ, Spiegelman GB. Identification of a second region of the Spo0A response regulator of *Bacillus subtilis* required for transcription activation, *J Bacteriol* 2000;**182**:4352–5.

Sandoval NR, Venkataramanan KP, Groth TS *et al*. Whole-genome sequence of an evolved *Clostridium pasteurianum* strain reveals Spo0A deficiency responsible for increased butanol production and superior growth. *Biotechnol Biofuels* 2015;**8**:227.

Sun Z, Chen Y, Yang C *et al*. A novel three-component system-based regulatory model for D-xylose sensing and transport in *Clostridium beijerinckii*. *Mol Microbiol* 2014;**95**:576–89.

Vora T, Hottes AK, Tavazoie. Protein occupancy landscape of a bacterial genome. *Mol Cell* 2009;**35**:247–53.

Wang S, Dong S, Wang P *et al*. Genome editing in *Clostridium saccharoperbutylacetonicum* N1-4 using CRISPR-Cas9 system. *Appl Environ Microbiol* 2017;**83**:e00233–17.

Wang Y, Zhang ZT, Seo SO *et al*. Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *J Biotechnol* 2015;**200**:1–5.

Wang Y, Zhang ZT, Seo SO *et al*. Bacterial genome editing with CRISPR-Cas9: deletion, integration, single nucleotide modification, and desirable "Clean" mutant selection in *Clostridium beijerinckii* as an example. *ACS Synth Biol* 2016;**5**:721–32.

Wasels F, Jean-Marie J, Collas F *et al*. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *J Microbio Methods* 2017;**140**:5–11.

Wayman M, Yu S. Acetone-butanol fermentation of xylose and sugar mixtures. *Biotechnol Lett* 1985;**7**:255–60.

Woodley JM, Breuer M, Mink D. A future perspective on the role of industrial biotechnology for chemicals production. *Chem Eng Res Des* 2013;**91**:2029–36.

Zhang J, Zong W, Hong W *et al*. Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. *Metab Eng* 2018; doi:10.1016/j.ymben.2018.03.007.