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

$^1$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

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...
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 (Cookley 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 of this method to modify clostridial genome is gaining popularity.

Methods have evolved, a two-step (Wasels 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 II 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. saccharoperbutylicum (Jenkinson and Krabben 2016).

Here, we demonstrate for the first time the application of the endogenous CRISPR-Cas system from C. saccharoperbutylicum 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 saccharoperbutylicum 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 K₂HPO₄, 0.75 g L⁻¹ KH₂PO₄, 0.4 g L⁻¹ MgSO₄, 0.01 g L⁻¹ FeSO₄, 0.01 g L⁻¹ MnSO₄, 1 g L⁻¹ NaCl, 2 g L⁻¹ (NH₄)₂SO₄, 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 FeSO₄, 0.5 g L⁻¹ (NH₄)₂SO₄, 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. saccharoperbutylicum 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. saccharoperbutylicum 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. saccharoperbutylicum 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. saccharoperbutylicum N1-4(HMT).

**Electroporation**

Plasmids were transformed into C. saccharoperbutylicum 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.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMTL83251</td>
<td>Backbone for homologous recombination vectors</td>
<td>Heap et al. 2009</td>
</tr>
<tr>
<td>pMTL82154</td>
<td>Backbone for CRISPR-Cas targeting vectors</td>
<td>Heap et al. 2009</td>
</tr>
<tr>
<td>pMTL83251.Spc53.CCA</td>
<td>Confirmation of PAM sequence</td>
<td>This work</td>
</tr>
<tr>
<td>pMTL83251.Spc53.CCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pMTL83251.Spc53.CCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pMTL82154.Spo0A.SNP</td>
<td>Recombination vector for making Spo0A SNP</td>
<td>This work</td>
</tr>
<tr>
<td>pMTL82154.Spo0A_del</td>
<td>Recombination vector for making Spo0A deletion</td>
<td>This work</td>
</tr>
<tr>
<td>pMTL82154.int_1kb</td>
<td>Recombination vector for making 1 kb integration</td>
<td>This work</td>
</tr>
<tr>
<td>pMTL82154.int_3kb</td>
<td>Recombination vector for making 3 kb integration</td>
<td>This work</td>
</tr>
<tr>
<td>pMTL82154.int_5kb</td>
<td>Recombination vector for making 5 kb integration</td>
<td>This work</td>
</tr>
<tr>
<td>pMTL83251.Ldr.Spc</td>
<td>Targeting vector for selecting integrated mutants</td>
<td>This work</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ref</th>
<th>Primer name</th>
<th>Sequence</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Spc_53_PAM_T.F</td>
<td>CCTattactagaatgggaggaaaaagcaaattaagaaacA</td>
<td>Anneals with Spc_53_PAM_T.R to make spc region 53 with CCT PAM</td>
</tr>
<tr>
<td>2</td>
<td>Spc_53_PAM_T.R</td>
<td>AGCTTTgtttcttaatttgctttttcctcccattctagtaatAGG</td>
<td>Pair with Spc_53_PAM_T.F</td>
</tr>
<tr>
<td>3</td>
<td>Spc_53_PAM_A.F</td>
<td>CCAattactagaatgggaggaaaaagcaaattaagaaacA</td>
<td>Anneals with Spc_53_PAM_A.R to make spc region 53 with CCA PAM</td>
</tr>
<tr>
<td>4</td>
<td>Spc_53_PAM_A.R</td>
<td>AGCTTTgtttcttaatttgctttttcctcccattctagtaatTGG</td>
<td>Pair with Spc_53_PAM_A.F</td>
</tr>
<tr>
<td>5</td>
<td>Spc_53_PAM_C.F</td>
<td>attactagaatgggaggaaaaagcaaattaagaaacA</td>
<td>Anneals with Spc_53_PAM_C.R to make spc region 53. Clone into SmaI site to make CCC PAM</td>
</tr>
<tr>
<td>6</td>
<td>Spc_53_PAM_C.R</td>
<td>AGCTTTgtttcttaatttgctttttcctcccattctagtaat</td>
<td>Pair with Spc_53_PAM_C.F</td>
</tr>
<tr>
<td>7</td>
<td>spo0A_SNP_HR1.F</td>
<td>agggaaaattatggggtaag</td>
<td>Amplifies Spo0A 857 bp upstream of SNP. Pair with spo0A_SNP_HR2.R</td>
</tr>
<tr>
<td>8</td>
<td>spo0A_SNP_HR2.R</td>
<td>gttaaataaagttaccttacctgcctc</td>
<td>Amplifies Spo0A 680 bp downstream of SNP.</td>
</tr>
<tr>
<td>9</td>
<td>spoHR1.F</td>
<td>gatggacaaagagagaagaagaagatgg</td>
<td>Amplifies spo0A SNP region. Anneals in HR region.</td>
</tr>
<tr>
<td>10</td>
<td>spo0A_seq.R</td>
<td>agtggacttctggttaacactcacc</td>
<td>Pairs with spoHR1.F Will only anneal in the chromosome.</td>
</tr>
<tr>
<td>11</td>
<td>spo0A_SNP_HRM.F</td>
<td>ggtataactggttctgtaacagaaaaag</td>
<td>HRM primer for nested PCR</td>
</tr>
<tr>
<td>12</td>
<td>spo0A_SNP_HRM.R</td>
<td>ctaaataagtccgtctatgactcctag</td>
<td>Pair with spo0A_SNP_HRM.F</td>
</tr>
<tr>
<td>13</td>
<td>spo0A_del_HR1.F</td>
<td>atggacacgtcccagtaagttctgataag</td>
<td>Amplifies HR1 for Spo0A deletion</td>
</tr>
<tr>
<td>14</td>
<td>spo0A_del_HR1.R</td>
<td>cttactaggcatactctgcaaatgtaactag</td>
<td>Pair with spo0A_del_HR1.F</td>
</tr>
<tr>
<td>15</td>
<td>spo0A_del_HR2.F</td>
<td>ctattgcegatcaagctgtgtaagctgtgcttg</td>
<td>Amplifies HR2 for Spo0A deletion</td>
</tr>
<tr>
<td>16</td>
<td>spo0A_del_HR2.R</td>
<td>taaaacactctgctcttctcact</td>
<td>Pair with spo0A_del_HR2.F</td>
</tr>
<tr>
<td>17</td>
<td>intvectorHR1.F</td>
<td>GGGCCGctaaatcactttatatgcattactaggg</td>
<td>Amplifies HR1 for integration</td>
</tr>
<tr>
<td>18</td>
<td>intvectorHR1.R</td>
<td>ACGGCTtcgattacaacttacattacttag</td>
<td>Pair with intvectorHR1.F</td>
</tr>
<tr>
<td>19</td>
<td>intvectorHR2.F</td>
<td>CTCGAgttaaaccacttacatagaacct</td>
<td>Amplifies HR2 for integration</td>
</tr>
<tr>
<td>20</td>
<td>intvectorHR2.R</td>
<td>AGGCCCTctttaaatcactattacttttactgatc</td>
<td>Pair with intvectorHR2.F</td>
</tr>
<tr>
<td>21</td>
<td>LP_Skb_F1</td>
<td>ATTATAGacctccaaagcgcggagcgtggttgc</td>
<td>Pair with LP_Skb_F1 to amplify 1 kb fragment from λDNA</td>
</tr>
<tr>
<td>22</td>
<td>LP_intk_R1</td>
<td>cggccctattacccagaagccttcc</td>
<td>Pair with LP_Skb_F1 to amplify 3 kb fragment from λDNA</td>
</tr>
<tr>
<td>23</td>
<td>LP_Skb_R1</td>
<td>ccaacgcgtagcggcgtattgcttgg</td>
<td>Pair with LP_Skb_F1 to amplify 5 kb fragment from λDNA</td>
</tr>
<tr>
<td>24</td>
<td>Int_genSeq2</td>
<td>tctgggaacttcgctacaagacc</td>
<td>Binds upstream of integration site. Will only anneal in the chromosome</td>
</tr>
<tr>
<td>25</td>
<td>LP_seq_R1</td>
<td>gttcagggactgcctacc</td>
<td>Reverse primer out of lambda DNA fragment</td>
</tr>
</tbody>
</table>

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 Na2HPO4, 4.4 mM NaH2PO4, 10 mM MgCl2), re-centrifuged and resuspended in 2 ml of EPB_NS buffer (as for EPB_S but without MgCl2). 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 Smal 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 Smal 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 Ntot/Mul 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 Ncol/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 gDNA fragments were PCR amplified (DNA from NEB) using primers shown in Table 1, ref 21–24, and ligated into the SacI and Stul sites of the integration vector multiple cloning sites. All integration vectors were checked by PCR and Sanger sequencing before transforming into C. saccharoperbutylicum 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_SpoAspc_DR’ was also synthesised and cloned downstream of the Ldr using Stul/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. saccharoperbutylicum N1-4(HMT) and colony numbers compared to an empty pMTL83251 plasmid.

Genome editing

Recombination vectors were transformed into C. saccharoperbutylicum 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 CGM 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, dH₂O 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 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_K1′ (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_Spo0A_spc,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 recognisable CRISPR-Cas operon and with the presence of a cas cluster was ascribed 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.

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 α 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 (261T) (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_Spo0AAspc_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_Spo0AAspc_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
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 dH2O (-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. saccharoperbutylicum N1-4(HMT). The specific site is flanked by transpositional 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...
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. saccharoperbutylicum 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 Skłodowska 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


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


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.


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.