



OPEN ACCESS

EDITED BY
Stefan Pflügl,
Vienna University of Technology, Austria

REVIEWED BY
Byung-Kwan Cho,
Korea Advanced Institute of Science and
Technology, South Korea
Bastian Molitor,
University of Tübingen, Germany
Armin Ehrenreich,
Technical University of Munich,
Germany

*CORRESPONDENCE
Steven D. Brown,
Steve.Brown@lanzatech.com

[†]These authors have contributed equally
to this work

SPECIALTY SECTION
This article was submitted to Industrial
Biotechnology,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

RECEIVED 29 April 2022
ACCEPTED 29 June 2022
PUBLISHED 05 August 2022

CITATION
Nogle R, Nagaraju S, Utturkar SM,
Giannone RJ, Reynoso V, Leang C,
Hettich RL, Mitchell WP, Simpson SD,
Jewett MC, Köpke M and Brown SD
(2022), *Clostridium autoethanogenum*
isopropanol production via native
plasmid pCA replicon.
Front. Bioeng. Biotechnol. 10:932363.
doi: 10.3389/fbioe.2022.932363

COPYRIGHT
© 2022 Nogle, Nagaraju, Utturkar,
Giannone, Reynoso, Leang, Hettich,
Mitchell, Simpson, Jewett, Köpke and
Brown. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Clostridium autoethanogenum isopropanol production via native plasmid pCA replicon

Robert Nogle^{1†}, Shilpa Nagaraju^{1†}, Sagar M. Utturkar^{2†},
Richard J. Giannone^{3†}, Vinicio Reynoso¹, Ching Leang¹,
Robert L. Hettich³, Wayne P. Mitchell¹, Sean D. Simpson¹,
Michael C. Jewett^{4,5,6,7,8}, Michael Köpke¹ and Steven D. Brown^{1*}

¹LanzaTech Inc., Skokie, IL, United States, ²Center for Cancer Research, Purdue University, West Lafayette, IN, United States, ³Oak Ridge National Laboratory, Oak Ridge, TN, United States, ⁴Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, United States, ⁵Center for Synthetic Biology, Northwestern University, Evanston, IL, United States, ⁶Simpson Querrey Institute, Northwestern University, Evanston, IL, United States, ⁷Chemistry of Life Processes Institute, Northwestern University, Evanston, IL, United States, ⁸Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Evanston, IL, United States

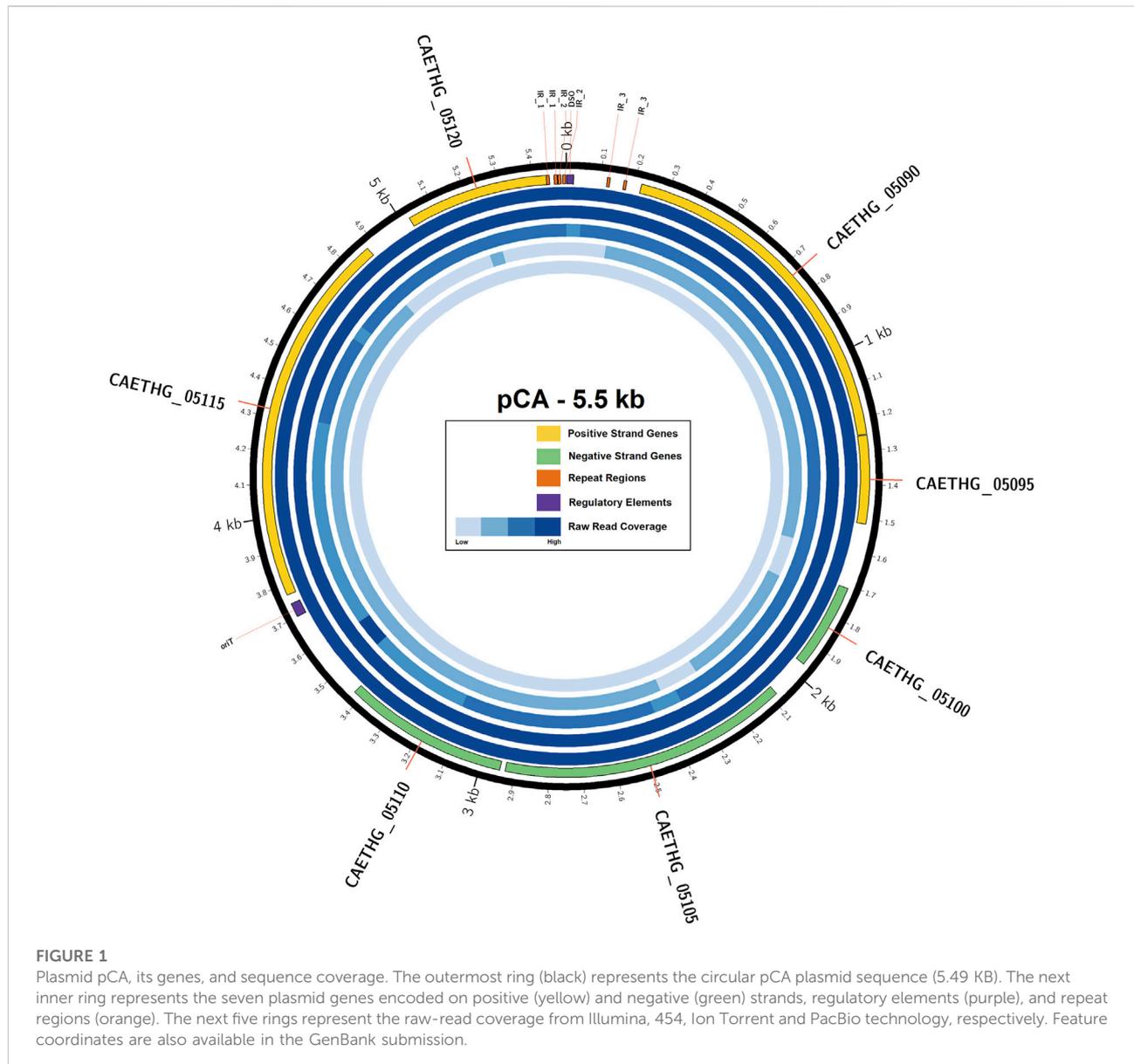
Clostridium autoethanogenum is a model gas-fermenting acetogen for commercial ethanol production. It is also a platform organism being developed for the carbon-negative production of acetone and isopropanol by gas fermentation. We have assembled a 5.5 kb pCA plasmid for type strain DSM10061 (JA1-1) using three genome sequence datasets. pCA is predicted to encode seven open-reading frames and estimated to be a low-copy number plasmid present at approximately 12 copies per chromosome. RNA-seq analyses indicate that pCA genes are transcribed at low levels and two proteins, CAETHG_05090 (putative replication protein) and CAETHG_05115 (hypothetical, a possible Mob protein), were detected at low levels during batch gas fermentations. Thiolase (*thlA*), CoA-transferase (*ctfAB*), and acetoacetate decarboxylase (*adc*) genes were introduced into a vector for isopropanol production in *C. autoethanogenum* using the native plasmid origin of replication. The availability of the pCA sequence will facilitate studies into its physiological role and could form the basis for genetic tool optimization.

KEYWORDS

acetogen, syngas, genome, clostridia, biofuel, ethanol, synthetic biology

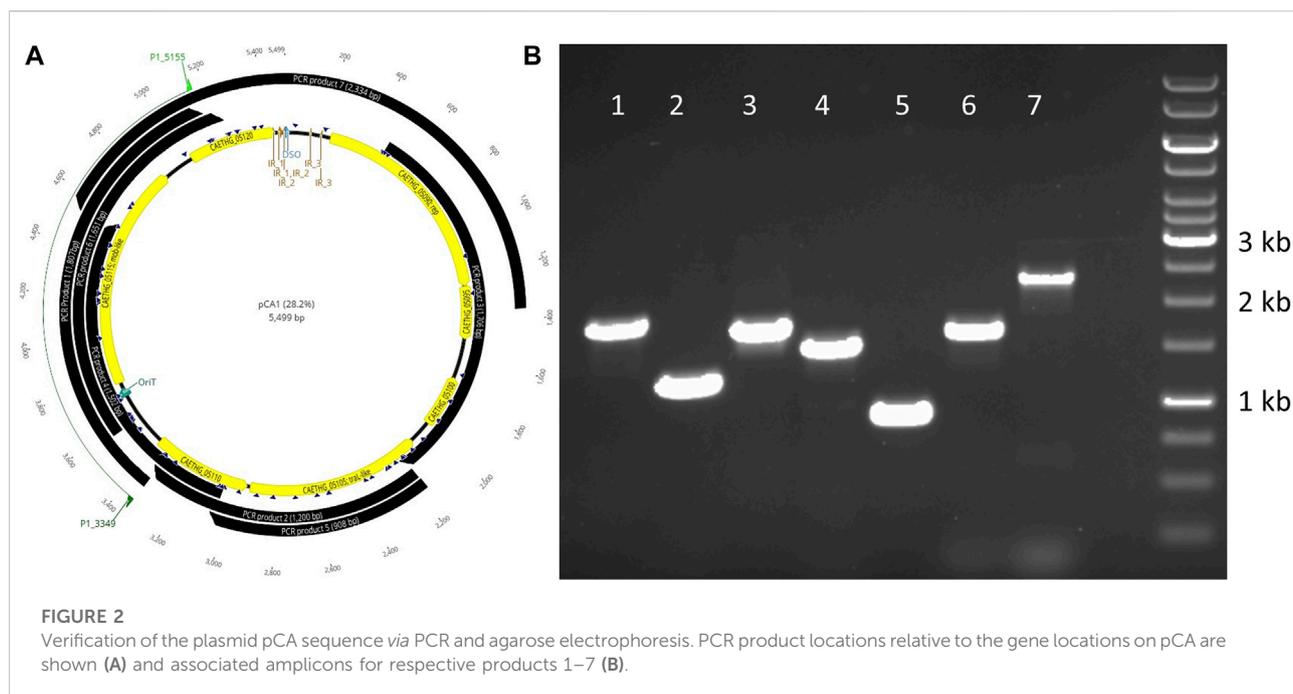
Introduction

Clostridium autoethanogenum is a model acetogen that uses a wide range of CO, CO₂, and H₂ gas mixes for carbon and energy sources *via* the Wood-Ljungdahl pathway (Marcellin et al., 2016). *C. autoethanogenum* has been used to investigate the fundamentals of electron bifurcation and acetogen bioenergetics (Mock et al., 2015). A genetic toolbox exists that includes heterologous expressions *via* plasmid and chromosomal integration, gene deletion, CRISPR systems, validated genetic parts, and codon adaptation algorithms (Joseph et al., 2018; Fackler et al., 2021). The *C.*



autoethanogenum DSM10061 restriction system has been characterized as part of a study to improve DNA transfer efficiencies into *Clostridium* spp. (Woods et al., 2019). Furthermore, plasmid DNA synthetic biology toolkit developments and the application of plasmids for clostridia genetics have been described recently (Joseph et al., 2018). *C. autoethanogenum* is used as a biocatalyst to produce ethanol in commercial scale gas fermentations (Fackler et al., 2021). Recently, the bacterium has been engineered through a series of gene deletions and *via* chromosomal heterologous expressions to generate strains for acetone or isopropanol production at commercially relevant production rates (Liew et al., 2022).

Because of its industrial relevance and status as an acetogen model organism, there has been significant efforts generating a highly polished and manually annotated genome sequence. The initial wild-type *C. autoethanogenum* strain JA1-1 (DSM10061) genome is a draft assembly that consisted of 100 contigs (Bruno-Barcena et al., 2013). The application of long-read sequencing technology facilitated the generation of a closed chromosome sequence for *C. autoethanogenum* DSM10061 (Brown et al., 2014; Utturkar et al., 2015). The DSM10061 genome sequence underwent further polishing *via* Illumina MiSeq and Sanger sequencing and manual annotation updates were applied (Humphreys et al., 2015). In an earlier survey, conducted before *C. autoethanogenum* was isolated, plasmid bands ranging from 3 to >100 kb were detected in 26 strains



representing 21 species across the genus *Clostridium* (Lee et al., 1987). In this study, we identify a native 5,499 bp plasmid present in the raw DNA sequence data from three prior genome-sequencing studies. We characterize the plasmid, expression of encoded proteins and demonstrate its utility for synthetic biology through the heterologous expression of genes encoding thiolase, CoA transferase subunits A and B, and acetoacetate decarboxylase for isopropanol production.

Results

Plasmid DNA assembly and estimated copy number

C. autoethanogenum pCA is a novel 5,499 bp circular plasmid identified in strain DSM10061. It is predicted to encode seven open reading frames (ORFs) and have a 28.2% GC content (Figure 1). Plasmid pCA was assembled from a previously published *C. autoethanogenum* DSM10061 Illumina dataset using plasmidSPAdes (Antipov et al., 2016), which was specifically developed to extract and assemble plasmid data from whole genome-sequencing projects. Previously reported Illumina data (Utturkar et al., 2015) were assembled into two plasmidSPAdes sequence outputs with different coverage values ($\times 132$ and $\times 672$), different GC contents (28 and 44%), and different lengths (5,598 bp and 5,626 bp). Both sequences appeared to be circular based on the initial dot-plot analysis. Data from an independent chromosome mapping study (Humphreys et al., 2015) underwent plasmidSPAdes assembly

as part of this study, which also produced two short similar sequence outputs (5,626 and 5,516 bp). After annotation, the higher coverage and G + C content sequences were determined to be phiX control DNA (G + C 45 %) and were thus discarded. The Circulator software (Hunt et al., 2015) did not assemble the pCA ends, so they were joined manually (see Methods). The number of reads mapped to the assembled plasmid sequence for each technology were 73,571 (Illumina), 7,945 (454), 5,705 (Ion-Torrent), and 72 (PacBio). The Illumina and PacBio datasets were sequenced to higher coverages ($>100\times$) relative to 454 and ion torrent datasets which had lower coverages ($<50\times$) (Utturkar et al., 2015). Illumina coverage for plasmid pCA is estimated to be approximately $\times 1,994$ (73,571 reads $\times 149$ bp or 10,962,079 bases/5.499 kb), or 15.8 plasmid copies per chromosome (based on reported $\times 126$ chromosome coverage). The estimated number of plasmid per chromosome sequence for the second study is ~ 10.4 copies ($\times 2,272.5/\times 219.4$). Genome data from a chemostat study (Martin et al., 2016) indicates ~ 11.7 copies ($\times 2,271/\times 193.6$). Taken together, the mean plasmid number is estimated to be 12.6 (± 2.8 S.D.) copies per cell, indicating that pCA is a low copy number plasmid. The presence of a circular pCA plasmid was verified through a series of overlapping PCR reactions (Figure 2).

Sequence similarity

A BLASTN analysis of the pCA sequence showed the highest similarity scores to the *Clostridium tyrobutyricum* strain Cirm BIA 2237 chromosome (GenBank: CP038158.1) matching

several regions [709/905 nt (78%) and 129/157 nt (82%)], which encode putative hypothetical proteins followed by plasmid sequences in other clostridia. Each of the seven predicted pCA proteins (CAETHG_05090-50120) have identical matches to previously described *C. autoethanogenum* DSM10061 proteins (NCBI accessions; OVY48499.1, OVY48498.1, OVY48497.1, OVY48503.1, OVY48502.1, OVY48501.1, and OVY48500.1, respectively) (Martin et al., 2016). CAETHG_05100, 05, 10, and 20, begin with ATG start sites, CAETHG_05090 has a TTG start site (Leu) and CAETHG_05095 and CAETHG_05115 begin with a GTG start site (Val).

Protein CAETHG_05090 is the 344 aa putative plasmid replication protein, which has a 48–61% identity to plasmid borne replication proteins from Clostridia species and 40–46% identity to other Gram-positive bacteria and particularly from *Bacillus* species (Figure 3). The CAETHG_05090 replication protein shares similarity with the pUB110 plasmid family replication proteins, including 100% identity across several conserved regions and an active site tyrosine residue. An 22 bp sequence in the 5' region of CAETHG_05090 is a possible double strand origin (DSO) of replication (*ori*) for the leading strand replication based on sequence similarity, with inverted repeats flanking it (Supplementary Figure S1). A possible minus single strand *ori* (SSO) for lagging strand replication is identified, although less sequence conservation is observed (Supplementary Figure S2), along with a region for a possible origin of transfer (*oriT*) (Figure 1). CAETHG_05095 and CAETHG_05100 are putative hypothetical proteins. ORF CAETHG_05105 is predicted to encode a 282 aa protein containing transmembrane domains and having 27% identity with the *Lactobacillus* spp. conjugal transfer pilus assembly protein TraL. CAETHG_05110 is predicted to encode a 160 aa protein with a signal peptide. Open reading frame CAETHG_05115 is predicted to encode a hypothetical protein that shares 35–40% identity with mobility (MOB) proteins from pUB110, pMV158, and *C. saccharoperbutylacetonicum* shuttle vector pNAK1. Sequences upstream of pCA_06 (3,700–3,741 bases) share >90% identity with the core *oriT*s from pUB110/pMV158 that consist of PalD inverted repeats and a recombination site (Supplementary Figure S3) and this may form a part of the pCA *oriT*. CAETHG_05120 encodes a putative 141 aa hypothetical protein.

Gene expression analysis

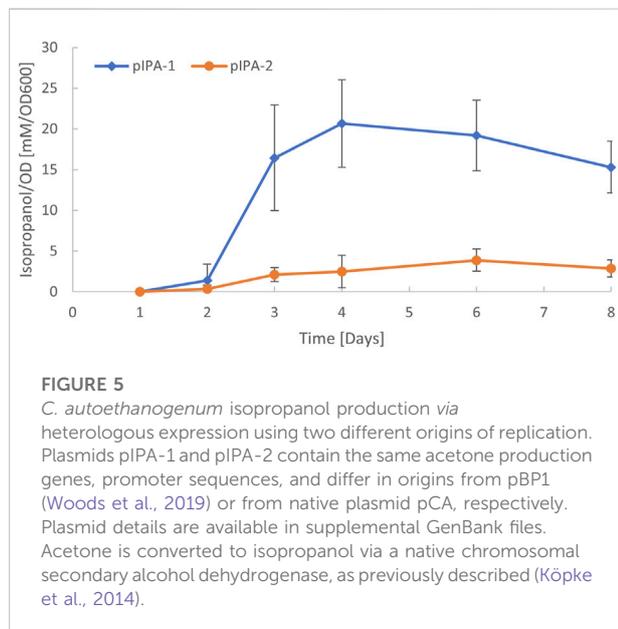
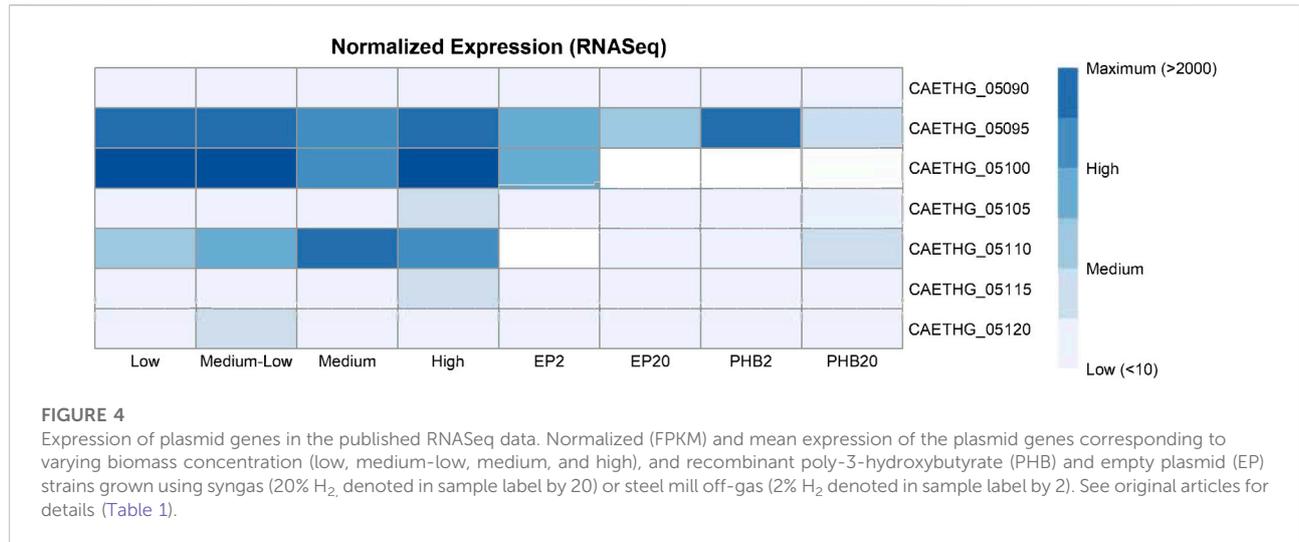
To examine the pCA gene expression, we re-analyzed the RNA-seq data from a study on recombinant poly-3-hydroxybutyrate (PHB)-producing *C. autoethanogenum* and control strains (de Souza Pinto Lemgruber et al., 2019) and another study with profiles for the steady-state chemostat cultures grown on syngas (Valgepea et al., 2017). DNA sequence read alignment rates to pCA genes were low, measuring between 0.5–0.7% and 0.9–1.3% across both datasets. Transcripts for the CAETHG_05090 putative

replication gene were detected at low levels for most samples (17/20) (Figure 4). CAETHG_05095 had the highest levels of expression for plasmid genes, but its detection was variable (detected in 11 of 20 samples), followed by CAETHG_05100. Normalized FPKM values indicate CAETHG_05105 (*traL*-like) was expressed in all *C. autoethanogenum* strain DSM 19630 samples and in five of eight samples for PHB-producing strains. CAETHG_05115 was detected at a low level in the PHB study (7 of 8), and infrequently (3 of 12) in the other study. CAETHG_05110 and CAETHG_05120 were at or below detection limit levels.

In an initial assessment to determine whether plasmid genes ultimately produced detectible protein products, batch-culture grown *C. autoethanogenum* was sampled at either lag- or early log-phases ($n = 3$) and processed for LC-MS/MS-based proteomic measurements. Across all samples and both growth states, a total of 13,693 peptide analytes (FDR <0.01) were identified. These high-confidence peptide identifications mapped to 1,892 proteins, with 1,647 at a protein-level FDR ≤ 0.05 , and 1492 at FDR ≤ 0.02 . Two plasmid proteins (CAETHG_05090 and CAETHG_05115) were detected at approximately the 28th and 22nd percentiles within the entire proteome (out of 1647), respectively, when sorted by decreasing the median abundance. This suggests that their presence is at a low level relative within the proteome. We present CAETHG_05090 and CAETHG_05115 proteomics data, along with the twenty most highly expressed proteins identified for the strain (Supplementary Table S1). As expected, Wood–Ljungdahl pathway proteins were among the most abundant. Across the two growth conditions, a total of 97 proteins exhibited statistically significant differences in abundance (Benjamini–Hochberg corrected p -value ≤ 0.05 ; $|\log_2$ difference| > 1), with 25 proteins higher during the lag-phase and 72 higher during the log-phase (Supplementary Table S2).

Utility of the plasmid pCA replicon for isopropanol production by gas fermentation

To test the functionality of the native plasmid origin of replication, it was amplified from *C. autoethanogenum* and used for the heterologous expressions of thiolase (*thlA*), CoA-transferase (*ctfAB*), and acetoacetate decarboxylase (*adc*) genes under control of the IPL-tet3n0 promoter (TCTATCATTGAT AGGTTATAATGAACATTGTAGAATTCCCATAATAAAGA AAGAATTTTAAATAAAGGAGGAACACA) (Nagaraju et al., 2016) for *C. autoethanogenum* isopropanol production [the final reduction step in the isopropanol pathway is catalyzed by a native secondary alcohol dehydrogenase (Köpke et al., 2014)]. We cloned the genes along with the identified pCA origin into the modular pMTL8X25X plasmid (Heap et al., 2009). As control, we used the same construct but with the pBP1 origin present in the



original pMTL8225X plasmid. Each plasmid was transformed into *C. autoethanogenum* and the resulting strains were tested for isopropanol production during gas fermentation. When the pathway was expressed from the original pMTL80000 plasmid with the pBP1 origin of replication, approximately 33 mM of isopropanol was observed (Figure 5). We also observed isopropanol production when the pathway was expressed from a plasmid using the native pCA plasmid origin of replication. However, production levels were lower using the pCA origin, with approximately 7 mM of isopropanol produced (Figure 5), indicating a lower copy number. This demonstrates the utility of the plasmid for gene expression in *C. autoethanogenum* and how it can be used to modulate gene

expression levels. It was not determined if the native plasmid was maintained or lost after the introduction of plasmid pIPA-2.

Discussion

In this study, we identify the novel 5.49 kb *C. autoethanogenum* plasmid pCA as being present, but previously unreported in three independent Illumina genome datasets. Plasmids are common among clostridia and pCA, similar to most (Davis et al., 2005), are also likely cryptic. Cell wall-associated DNase activity has been previously attributed as an impediment to the recovery of intact clostridial plasmid DNA (Blaschek and Klacik, 1984; Roberts et al., 1986; Hussain and Purnima, 1993) and this coupled with the low copy number of pCA may have contributed to it not being characterized previously. An earlier screen for plasmid DNA in the genus *Clostridium* identified a 5.4 kb plasmid in *Clostridium aceticum* and while the size is similar (5.5 kb) to the plasmid from this study, the restriction profiles are distinct (Lee et al., 1987).

The traditional methods for plasmid sequencing involved purification of plasmid DNA, end-fragment or shot-gun sequencing, followed by gap closure using primer-walking. *De novo* assembly of short-read next-generation sequence data often leads to fragmented genome assemblies that complicate the plasmid assembly and identification (Arredondo-Alonso et al., 2017). The *C. autoethanogenum* DSM10061 chromosome was assembled as one contig using single-molecule sequencing. The assembled chromosome featured repeats, putative prophage sequence, low G + C content, and nine copies of the rRNA gene operons, all of which present challenges to short-read sequencing technologies (Brown et al., 2014). Plasmid DNA was likely excluded during the PacBio library preparation, owing to a size selection for high molecular weight genomic DNA to produce long, single molecule sequence

reads, as observed previously (Margos et al., 2017). Furthermore, the hierarchical genome assembly process, when implemented with a seed read length cut-off greater than plasmid size, can result in small plasmids being lost from the main assembly (Forde et al., 2014; Koren and Phillippy, 2015). Different technologies, bioinformatics approaches, and challenges in plasmid reconstruction have been described (Hunt et al., 2015b). The *Escherichia coli* strain K-12 genome, arguably the most well-studied single organism (Riley et al., 2006), and others model organisms such as *Desulfovibrio alaskensis* G20 (Hauser et al., 2011), and *Zymomonas mobilis* ZM4, have similarly been improved over time (Yang et al., 2009, 2018). The description of plasmid pCA further improves the *C. autoethanogenum* DSM10061 genome sequence and the utility of its origin of replication was demonstrated for isopropanol production (Figure 5).

Bacillus species plasmids that show the most similarity to pCA are of the non-conjugative rolling-circle type that can be mobilized for conjugative transfer in the presence of conjugation machinery. They replicate by a rolling circle mechanism via a single strand DNA intermediate enabled by a replication protein, plus a double strand (DSO) origin of replication (*ori*) for leading strand replication and a minus single strand (SSO) *ori* for lagging strand replication. They are broadly classified into five families based on replication protein and DSO homology (Ilyina and Koonin, 1992). We have identified a putative replication protein that has conserved consensus motifs, including an active site tyrosine and conserved 18 bp DSO sequence flanked by inverted repeats. SSO sequences within a family share less homology and certain plasmids have more than one SSO (Ilyina and Koonin, 1992). pCA may replicate by a rolling circle mechanism and is potentially a non-conjugative plasmid, but this requires further study. An improved understanding of pCA replication could assist in determining whether it can coexist with other plasmids and transferability. The presence of pCA may prevent the efficient transformation of some heterologous plasmids that rely on the same replication mechanism, but replication co-existence (e.g., pCB102 of the pMTL80000 plasmids) could be useful for dual plasmid expressions similar to the pETDuet system for *E. coli* (Novy et al., 2002). The pCA copy number appears lower than with pMTL80000 plasmids (pBP1 origin), further expanding to the suite of tools to modulate the expression in clostridia.

Plasmid pCA appears to be a small, selfish DNA element with limited coding potential and expression, present at approximately 12 copies per cell, although this requires examining a greater range of conditions and the use of more quantitative methods. A proteomic analysis generated a comprehensive profile of *C. autoethanogenum* DSM10061 growing in a batch culture and under the conditions analyzed here (lag-vs. early log-phase); plasmid proteins were either detected at low levels or not at all. The physiological role of the pCA plasmid, if any, remains unclear. Any metabolic burden associated with plasmid maintenance could be relieved by generating a cured strain using strategies such as elevated growth temperature and screening procedures, targeting the replication origin through transformation of an incompatible plasmid, and potentially

through the application of CRISPR technologies. The application of Cell-Free Protein Synthesis (CFPS) technologies (Silverman et al., 2020) could enable a better understanding of its elements and future systems, and biology studies may shed light on the role of this native *C. autoethanogenum* plasmid. We have shown the utility of the native plasmid origin of replication for isopropanol production. The small size of pCA (5.5 kb), low G + C content, low-copy number, and native origin of replication may facilitate its further optimization as a genetic tool.

Methods

Data sources and sequence analysis

DNA, RNA-seq, and chromosome sequence data were downloaded from the National Center for Biotechnology Information databases and accession numbers are provided, along with the data generated in this study (Table 1). Plasmid pCA was assembled from Illumina data using plasmidSPAdes (Antipov et al., 2016), and overlapping contig ends were resolved manually using the Geneious software (version 8.1.6), as described previously (Utturkar et al., 2017). Gene predictions and annotations were conducted using the RAST pipeline (Overbeek et al., 2014), with the Prodigal gene caller option and default settings. Sequence similarity searches were performed using NCBI BLAST and PSI-BLAST tools. Plasmid sequences for pUB110/pMV158 and from this family were retrieved from NCBI and further alignments were conducted using CLUSTALW (<https://www.genome.jp/tools-bin/clustalw>).

For plasmid coverage estimates, DNA sequencing reads were trimmed using Trimmomatic (Bolger et al., 2014), (adapter removal and minimum quality score 30) and then mapped to the plasmid pCA sequence using the Bowtie2 (Langmead and Salzberg, 2012) (Illumina, Ion-Torrent) and blasr (Chaisson and Tesler, 2012) (PacBio, 454) algorithms with default parameters. Read coverages for every 50 bp intervals were determined using the deepTools-multiBamSummary (Ramírez et al., 2016). Genome overview and heatmap figures were generated using the Circos (Krzywinski et al., 2009) and R-heatmap package, respectively. RNA-seq reads were trimmed using Trimmomatic to remove sequencing adapters and bases below a quality score of 30. Paired reads were mapped to the pCA plasmid sequence using Bowtie2 with two alignments per read (k). The alignments were scored with maximum and minimum mismatch penalties (mp) of 5 and 1, respectively, and the natural log function (G). Mapped reads were filtered and sorted using SAMtools (Li et al., 2009), and FPKM values were estimated using Cufflinks utility in the Cufflinks package (Trapnell et al., 2010).

PCR

PCR primers were designed using Primer3 (version 0.4.0), synthesized by Integrated DNA Technologies (IDT) and the

TABLE 1 Sequences and chromosome data sources.

Data type	NCBI BioProject/GenBank Accession/Proteomics DB	References
RNA-seq	PRJNA355951	Valgepea et al. (2017)[15]
RNA-seq	PRJNA476240	de Souza Pinto Lemgruber et al. (2019)[14]
DNA sequence	PRJNA219420	Utturkar et al. (2017)
DNA sequence	PRJNA291963	Humphreys et al. (2015)[9]
DNA sequence	PRJNA280340	Martin et al. (2016)[12]
Chromosome sequence	CP012395	Humphreys et al. (2015)[9]
Plasmid sequence	CP097624.1	This study
Raw proteomics	MassIVE: MSV000089452	This study
Searched proteomics	ProteomeXchange: PXD033792	This study

TABLE 2 Oligonucleotides used to confirm plasmids via overlapping PCR reactions.

Primer	Gene	Product name	Product size (bp)	Sequence
P1_3349	CAETHG_05110	Product 1	1,807	TCGTAAAACCACTGCAGCC
P1_5155	CAETHG_05120	Product 1		ACTGGCCTGACTCTTCGAAA
P2_2169	CAETHG_05105	Product 2	1,200	TGAGCCAGTCGAATACCACA
P2_3368	CAETHG_05110	Product 2		GGCTGCAGTGGTTTAAACGA
P3_483	CAETHG_05090	Product 3	1,706	TGTGTGCATGGAGAAGGTCA
P3_2188	CAETHG_05105	Product 3		TGTGGTATTCGACTGGCTCA
P4_3050	CAETHG_05110	Product 4	1,502	TGTTAACTACCGCGTGCTCA
P4_4551	CAETHG_05115	Product 4		AGCAAGTCCACGTGAAGCTA
P5_2162	CAETHG_05105	Product 5	908	TTCTTCCTGAGCCAGTCGAA
P5_3069	CAETHG_05110	Product 5		TGACACGCCGGTAGTTAACA
P6_3580	intergenic	Product 6	1,651	GGAGCGAACCCCTTGACATTT
P6_5230	CAETHG_05120	Product 6		AGGCTCGGAAACAGGACAAT
P7_4533	CAETHG_05115	Product 7	2,334	AGCTTCACGTGGACTTGCTA
P7_1367	CAETHG_05095	Product 7		TGCGGTGCTAAACATAATGACA

sequences are provided (Table 2). PCR reactions were performed using the Q5 High-Fidelity DNA Polymerase under standard conditions (New England Biolabs).

Proteomics

C. autoethanogenum strain JA1-1 was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) culture collection (DSM10061). For proteomic samples, cultures were grown at 37°C using gaseous substrates (50% CO₂, 10% H₂, 30% CO₂, 10% N₂) and a previously described medium (Valgepea et al., 2017). After ~24 h, the “lag-phase” cultures had cell densities in the range of OD₆₀₀ 0.09–0.15 and after ~42 h, the “log-phase” cultures had cell densities in the range of OD₆₀₀ 0.2–0.27. Cell pellets (*n* = 3) from both time points were harvested for proteomics by centrifugation, followed by supernatant removal, rapid freezing in liquid nitrogen, and storage at –80°C until analysis. Cells were lysed by bead-beating in Tris-HCl (100 mM

at pH 8.0), adjusting the sample to 4% SDS, and heat-treatment (95°C for 10 min). Crude protein was obtained by centrifugation at ×21,000 g for 10 min followed by quantification with a Nanodrop OneC spectrophotometer (Thermo Scientific). Samples were adjusted to 10 mM dithiothreitol (10 min at 95°C) to reduce proteins, then 30 mM iodoacetamide (20 min at room temperature in darkness), and cleaned up *via* protein aggregation capture (Batth et al., 2019). Aggregated protein [on magnetic Sera-Mag (GE Healthcare) beads] was then digested with proteomic-grade trypsin (1:75 w/w; Promega) in 100 mM Tris-HCl, pH 8.0 overnight at 37°C, and again for 3 h at 37°C the following day. Tryptic peptides were then collected, filtered through a 10 kDa MWCO spin filter (Vivaspin 2; Sartorius), and quantified by Nanodrop OneC. Three micrograms of the peptides were then analyzed by 1D LC-MS/MS using a Vanquish uHPLC coupled directly to an Orbitrap Q Exactive mass spectrometer (Thermo Scientific) as previously described (Walker et al., 2020). Peptides were separated by a 180 min organic gradient across an in-house-pulled nanospray emitter packed with 15 cm of 1.7-micron Kinetex

reversed-phase resin (Phenomenex). Peptide fragmentation spectra were analyzed/sequenced by the Proteome Discoverer software (Thermo Scientific) and the peptides quantified by chromatographic area-under-the-curve. Peptide abundances were summed to their respective proteins, protein abundances \log_2 transformed, and normalized with InfernoRDN (Taverner et al., 2012). Statistical analyses were performed with Perseus (Tyanova et al., 2016).

Strains and growth experiments

Wild-type *C. autoethanogenum* DSM10061 is available from the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ; Braunschweig, Germany). Strain DSM19630 (derived from DSM10061) was used in these studies and plasmid pMTL80000 plasmid has been described previously (Heap et al., 2009). Methods for the heterologous product synthesis, along with anaerobic techniques and media have been described (Liew et al., 2022). Plasmids pIPA-1 and pIPA-2 were constructed using the described methods (Liew et al., 2022). Briefly, the primers NP_replicon_F (GTGTTCTTTCTTAACCTTGAATTGGCGCGCC) and NP_replicon_R (GTTTGAACCTTCTTCACATTCAATTGGCCGCGCC) were used to amplify the native plasmid origin, followed by DNA restriction using the AscI and FseI enzymes, gel extraction, and T4 ligation. The Supplementary Material provides Genbank files for the plasmids respectively. For the growth experiments, the strains were cultured in Schott bottles with a medium containing clarithromycin and using a gas blend (50% CO₂, 10% H₂, 30% CO₂, and 10% N₂; Airgas) at 37°C, shaking at 120 rpm, and isopropanol was measured by high performance liquid chromatography (HPLC), as described previously (Liew et al., 2022).

Data availability statement

C. autoethanogenum DSM10061 is available from DSMZ, the German Collection of Microorganisms and Cell Cultures GmbH. Sequence data used in this study have been previously described and are publicly available (Table 1). Raw peptide, raw protein and quantified proteomic data generated as part of this study are presented (Table 1, Supplementary Tables S1–S2). All other data generated or analyzed in this article are presented in this published article.

Author contributions

CL, SB, and MK designed the study. SN, SU, VR, WM, and SB assembled, annotated plasmid pCA, performed RNA-seq, and mapping analyses. SN conducted sequence motif and replication analysis. RN and CL engineered pCA for isopropanol production. SB cultured strain DSM10061 for

proteomics; RG and RH generated and analyzed the proteomics data. SN, SU, RN, RG, and SB drafted the initial manuscript with input from all authors. All authors read and approved the final manuscript.

Funding

This material by the *Clostridium* foundry for biosystem designs (cBioFAB) is based upon the work supported by the U.S. Department of Energy, Office of Biological and Environmental Research in the DOE Office of Science under Award Number DE-SC0018249. Oak Ridge National Laboratory is managed by UT-Battelle LLC for the DOE under contract DE-AC05-00OR22725. The funder had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Acknowledgments

We thank Joanna Cogan and Vicki Liu (LanzaTech) for the medium preparation.

Conflict of interest

LanzaTech has an interest in commercializing gas fermentation with *C. autoethanogenum*. SN, VR, RN, CL, WM, MK, SS and SB are employees of LanzaTech. MJ consults for and has joint funding with LanzaTech. MK is co-inventors on granted US patent 9,365,868 (assigned to LanzaTech) related to production of acetone and isopropanol by fermentation of a gaseous substrate.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2022.932363/full#supplementary-material>

References

- Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A., Pevzner, P. A., et al. (2016). PlasmidSPAdes: Assembling plasmids from whole genome sequencing data. *Bioinformatics* 32, 3380–3387. doi:10.1093/bioinformatics/btw493
- Arredondo-Alonso, S., Willems, R. J., van Schaik, W., and Schürch, A. C. (2017). On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb. Genom.* 3, e000128. doi:10.1099/mgen.0.000128
- Bath, T. S., Tollenare, M. A. X., Rütger, P., Gonzalez-Franquesa, A., Prabhakar, B. S., Bekker-Jensen, S., et al. (2019). Protein aggregation capture on microparticles enables multipurpose proteomics sample preparation. *Mol. Cell. Proteomics* 18, 1027–1035. doi:10.1074/mcp.TIR118.001270
- Blaschek, H. P., and Klacik, M. A. (1984). Role of DNase in recovery of plasmid DNA from *Clostridium perfringens*. *Appl. Environ. Microbiol.* 48, 178–181. doi:10.1128/aem.48.1.178-181.1984
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170
- Brown, S. D., Nagaraju, S., Utturkar, S., de Tissera, S., Segovia, S., Mitchell, W., et al. (2014). 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
- Bruno-Barcelona, J. M., Chinn, M. S., and Grunden, A. M. (2013). Genome sequence of the autotrophic acetogen *Clostridium autoethanogenum* JA1-1 strain DSM 10061, a producer of ethanol from carbon monoxide. *Genome Announc.* 1, e00628-13. doi:10.1128/genomeA.00628-13
- Chaisson, M. J., and Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. *BMC Bioinforma.* 13, 238. doi:10.1186/1471-2105-13-238
- Davis, I. J., Carter, G., Young, M., and Minton, N. P. (2005). “Gene cloning in clostridia,” in *Handbook on clostridia*. doi:10.1201/9780203489819-11
- de Souza Pinto Lemgruber, R., Valgepea, K., Tappel, R., Behrendorff, J. B., Palfreyman, R. W., Plan, M., et al. (2019). Systems-level engineering and characterisation of *Clostridium autoethanogenum* through heterologous production of poly-3-hydroxybutyrate (PHB). *Metab. Eng.* 53, 14–23. doi:10.1016/j.ymben.2019.01.003
- Fackler, N., Heijstra, B. D., Rasor, B. J., Brown, H., Martin, J., Ni, Z., et al. (2021). Annual review of chemical and biomolecular engineering stepping on the gas to a circular economy: Accelerating development of carbon-negative chemical production from gas fermentation. *Annu. Rev. Chem. Biomol. Eng.* 2021, 12.
- Forde, B. M., ben Zakour, N. L., Stanton-Cook, M., Phan, M. D., Totsika, M., Peters, K. M., et al. (2014). The complete genome sequence of *Escherichia coli* EC958: A high quality reference sequence for the globally disseminated multidrug resistant *E. coli* O25b:H4-ST131 clone. *PLoS ONE* 9, e104400. doi:10.1371/journal.pone.0104400
- Hauser, L. J., Land, M. L., Brown, S. D., Larimer, F., Keller, K. L., Rapp-Giles, B. J., et al. (2011). Complete genome sequence and updated annotation of *Desulfovibrio alaskensis* G20. *J. Bacteriol.* 193, 4268–4269. doi:10.1128/JB.05400-11
- Heap, J. T., Pennington, O. J., Cartman, S. T., and Minton, N. P. (2009). A modular system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods* 78, 79–85. doi:10.1016/j.mimet.2009.05.004
- Humphreys, C. M., McLean, S., Schatschneider, S., Millat, T., Henstra, A. M., Annan, F. J., et al. (2015). Whole genome sequence and manual annotation of *Clostridium autoethanogenum*, an industrially relevant bacterium. *BMC Genomics* 16, 1085. doi:10.1186/s12864-015-2287-5
- Hunt, M., Silva, N. D., Otto, T. D., Parkhill, J., Keane, J. A., Harris, S. R., et al. (2015). Circulator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* 16, 294. doi:10.1186/s13059-015-0849-0
- Hussain, A., and Purnima (1993). A comparison of methods for isolating plasmid DNA from *Clostridium perfringens*. *Vet. Res. Commun.* 17, 335–339. doi:10.1007/BF01839384
- Ilyina, T. V., and Koonin, E. V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria. *Nucleic Acids Res.* 20, 3279–3285. doi:10.1093/nar/20.13.3279
- Joseph, R. C., Kim, N. M., and Sandoval, N. R. (2018). Recent developments of the synthetic biology toolkit for *Clostridium*. *Front. Microbiol.* 9, 154. doi:10.3389/fmicb.2018.00154
- Köpke, M., Gerth, M. L., Maddock, D. J., Mueller, A. P., Liew, F. M., Simpson, S. D., et al. (2014). Reconstruction of an acetogenic 2,3-butanediol pathway involving a novel NADPH-dependent primary-secondary alcohol dehydrogenase. *Appl. Environ. Microbiol.* 80, 3394–3403. doi:10.1128/AEM.00301-14
- Koren, S., and Phillippy, A. M. (2015). One chromosome, one contig: Complete microbial genomes from long-read sequencing and assembly. *Curr. Opin. Microbiol.* 23, 110–120. doi:10.1016/j.mib.2014.11.014
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: An information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi:10.1101/gr.092759.109
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923
- Lee, C. K., Dürre, P., Hippe, H., and Gottschalk, G. (1987). Screening for plasmids in the genus *Clostridium*. *Arch. Microbiol.* 148, 107–114. doi:10.1007/BF00425357
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- Liew, F. E., Nogle, R., Abdalla, T., Rasor, B. J., Canter, C., Jensen, R. O., et al. (2022). Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. *Nat. Biotechnol.* 40, 335–344. doi:10.1038/s41587-021-01195-w
- Marcellin, E., Behrendorff, J. B., Nagaraju, S., Detissera, S., Segovia, S., Palfreyman, R. W., et al. (2016). Low carbon fuels and commodity chemicals from waste gases-systematic approach to understand energy metabolism in a model acetogen. *Green Chem.* 18, 3020–3028. doi:10.1039/c5gc02708j
- Margos, G., Hepner, S., Mang, C., Marosevic, D., Reynolds, S. E., Krebs, S., et al. (2017). Lost in plasmids: next generation sequencing and the complex genome of the tick-borne pathogen *Borrelia burgdorferi*. *BMC Genomics* 18, 422. doi:10.1186/s12864-017-3804-5
- Martin, M. E., Richter, H., Saha, S., and Angenent, L. T. (2016). Traits of selected *Clostridium* strains for syngas fermentation to ethanol. *Biotechnol. Bioeng.* 113, 531–539. doi:10.1002/bit.25827
- Mock, J., Zheng, Y., Mueller, A. P., Ly, S., Tran, L., Segovia, S., et al. (2015). Energy conservation associated with ethanol formation from H₂ and CO₂ in *Clostridium autoethanogenum* involving electron bifurcation. *J. Bacteriol.* 197, 2965–2980. doi:10.1128/jb.00399-15
- Nagaraju, S., Davies, N. K., Walker, D. J. F., Köpke, M., and Simpson, S. D. (2016). Genome editing of *Clostridium autoethanogenum* using CRISPR/Cas9. *Biotechnol. Biofuels* 9, 219. doi:10.1186/s13068-016-0638-3
- Novy, R., Yaeger, K., Held, D., and Mierendorf, R. (2002). Coexpression of multiple target proteins in *E. coli*. *Innovations.* 15, 2–6.
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi:10.1093/nar/gkt1226
- Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., et al. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165. doi:10.1093/nar/gkw257
- Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K. B., Blattner, F. R., Chaudhuri, R. R., et al. (2006). *Escherichia coli* K-12: A cooperatively developed annotation snapshot - 2005. *Nucleic Acids Res.* 34, 1–9. doi:10.1093/nar/gkj405
- Roberts, I., Holmes, W. M., and Hylemon, P. B. (1986). Modified plasmid isolation method for *Clostridium perfringens* and *Clostridium absconum*. *Appl. Environ. Microbiol.* 52, 197–199. doi:10.1128/aem.52.1.197-199.1986
- Silverman, A. D., Karim, A. S., and Jewett, M. C. (2020). Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet.* 21, 151–170. doi:10.1038/s41576-019-0186-3
- Taverner, T., Karpievitch, Y. v., Polpitiya, A. D., Brown, J. N., Dabney, A. R., Anderson, G. A., et al. (2012). DanteR: An extensible R-based tool for quantitative analysis of -omics data. *Bioinformatics* 28, 2404–2406. doi:10.1093/bioinformatics/bts449
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi:10.1038/nbt.1621
- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., et al. (2016). The Perseus computational platform for comprehensive analysis of (prote) omics data. *Nat. Methods* 13, 731–740. doi:10.1038/nmeth.3901
- Utturkar, S. M., Klingeman, D. M., Bruno-Barcelona, J. M., Chinn, M. S., Grunden, A. M., Köpke, M., et al. (2015). Sequence data for *Clostridium autoethanogenum* using three generations of sequencing technologies. *Sci. Data* 2, 150014. doi:10.1038/sdata.2015.14

Utturkar, S. M., Klingeman, D. M., Hurt, R. A., and Brown, S. D. (2017). A case study into microbial genome assembly gap sequences and finishing strategies. *Front. Microbiol.* 8, 1272. doi:10.3389/fmicb.2017.01272

Valgepea, K., de Souza Pinto Lemgruber, R., Meaghan, K., Palfreyman, R. W., Abdalla, T., Heijstra, B. D., et al. (2017). Maintenance of ATP homeostasis triggers metabolic shifts in gas-fermenting acetogens. *Cell Syst.* 4, 505–515.e5. doi:10.1016/j.cels.2017.04.008

Walker, C., Ryu, S., Giannone, R. J., Garcia, S., and Trinh, C. T. (2020). Understanding and eliminating the detrimental effect of thiamine deficiency on the oleaginous yeast *Yarrowia lipolytica*. *Appl. Environ. Microbiol.* 86, e02299-19. doi:10.1128/AEM.02299-19

Woods, C., Humphreys, C. M., Rodrigues, R. M., Ingle, P., Rowe, P., Henstra, A. M., et al. (2019). A novel conjugal donor strain for improved DNA transfer into *Clostridium* spp. *Anaerobe* 59, 184–191. doi:10.1016/j.anaerobe.2019.06.020

Yang, S., Pappas, K. M., Hauser, L. J., Land, M. L., Chen, G. L., Hurst, G. B., et al. (2009). Improved genome annotation for *Zymomonas mobilis*. *Nat. Biotechnol.* 27, 893–894. doi:10.1038/nbt1009-893

Yang, S., Vera, J. M., Grass, J., Savvakis, G., Moskvina, O. v., Yang, Y., et al. (2018). Complete genome sequence and the expression pattern of plasmids of the model ethanologen *Zymomonas mobilis* ZM4 and its xylose-utilizing derivatives 8b and 2032. *Biotechnol. Biofuels* 11, 125. doi:10.1186/s13068-018-1116-x