A highly efficient Cre-based Clostridial workflow for genomic integration and expression of large biosynthetic pathways

Patrick A. Sanford¹, Ian Blaby², Yasuo Yoshikuni², and Benjamin Woolston¹

¹Northeastern University
²DOE Joint Genome Institute

May 17, 2024

Abstract

Acetogenic Clostridia are obligate anaerobes that have emerged as promising microbes for the renewable production of biochemi-
cals owing to their ability to efficiently metabolize sustainable single-carbon feedstocks. Additionally, Clostridia are increasingly
recognized for their biosynthetic potential, with recent discoveries of diverse secondary metabolites ranging from antibiotics to
pigments to modulators of the human gut microbiota. Lack of efficient methods for genomic integration and expression of large
heterologous DNA constructs remains a major challenge in studying biosynthesis in Clostridia and using them for metabolic
engineering applications. To overcome this problem, we harnessed chassis-independent recombinase-assisted genome engineering
(CRAGE) to develop a workflow for facile integration of large gene clusters (>10 kB) into the human gut *Eubacterium limosum*. We then integrated a non-ribosomal peptide synthetase gene cluster from the gut anaerobe *Clostridium leptum*, which
previously produced no detectable product in traditional heterologous hosts. Chromosomal expression in *E. limosum* without
further optimization led to production of phevalin at 2.4 mg/L. These results further expand the molecular toolkit for a highly
tractable member of the Clostridia, paving the way for sophisticated pathway engineering efforts, and highlighting the potential
of *E. limosum* as a Clostridial chassis for exploration of anaerobic natural product biosynthesis.
gene cluster from the gut anaerobe *Clostridium leptum*, which previously produced no detectable product in traditional heterologous hosts. Chromosomal expression in *E. limosum* without further optimization led to production of phevalin at 2.4 mg/L. These results further expand the molecular toolkit for a highly tractable member of the Clostridia, paving the way for sophisticated pathway engineering efforts, and highlighting the potential of *E. limosum* as a Clostridial chassis for exploration of anaerobic natural product biosynthesis.

**Keywords**: *Eubacterium limosum*, gas fermentation, C1 metabolism, anaerobe natural products, chassis-independent recombinase-assisted genome engineering (CRAGE)

Acetogenic Clostridia are anaerobic bacteria of increasing academic and industrial interest thanks to their unique ability to metabolize single carbon substrates such as carbon dioxide at high energetic efficiency, giving them significant potential in the development of the circular bioeconomy and to combat climate change [1]. While basic plasmid-based genetic tools have enabled preliminary efforts to metabolically engineer acetogens to produce high-value compounds from these feedstocks, a major limitation is the inability to easily integrate large biosynthetic pathways into the genome, which allows for more stable expression and reduced metabolic burden. Among acetogens, *Eubacterium limosum* has grown to greater prominence recently thanks to its inherent metabolic capacity and contributions to metabolism in the human gut microbiota [2], [3]. *E. limosum* has seen significant recent advances in its domestication including development of engineering tools and analysis of its genome and transcriptome [4]–[7]. These developments have improved its engineerability, but it is still underdeveloped for applications in heterologous biosynthesis without an efficient method for genomic DNA integration.

Beyond metabolic engineering for biofuels and biochemicals, robust and efficient Clostridial genomic integration technology could be useful in the study of anaerobic natural product biosynthesis. While most characterized secondary metabolites come from aerobic microbes, recent analysis of biosynthetic gene clusters (BSGs) in anaerobe genomes suggests a large untapped biosynthetic potential, particularly among the Clostridia [8], [9]. While only a handful of clusters have been experimentally associated with a specific product [10], [11], these few characterized Clostridial products exemplify unique carbon skeletons, novel biosynthetic logic, and unprecedented biological activities [12]. In particular, increasing attention has come to the gut microbiome and its impact on human health, specifically in relation to secondary metabolites produced by the microbial community [13]. As gut bacteria are primarily either strict or facultative anaerobes of the phyla Bacteroidetes and Firmicutes, the bias of natural products research towards aerobic microbes has left a gap in our understanding of the production and impact of secondary metabolites produced by gut commensals [14]. Inability to culture and genetically manipulate many anaerobes, including those of the gut microbiota have further hampered efforts to understand these anaerobically produced metabolites [15]. While culturing is problematic, genomics approaches have enabled detection of BSGs from these strains, opening up the option for assessing them in heterologous production hosts [8], [16]. However, heterologous production in traditional engineering hosts such as *E. coli* have often failed, likely due to the phylogenetic dissimilarity between source and host [17]. While systems have been implemented for investigation of anaerobically derived BSGs, the community could benefit from a user-friendly Clostridial chassis for exploring biosynthesis from this class given its high representation among anaerobe BSGs [18]. *E. limosum* is ideally suited to take on this role. It is of the class Clostridia in the phylum Firmicutes, a classification encoding a high density of BSGs, and is more genetically tractable than other related species [19], [20]. To explore the utility of *E. limosum* as a platform strain for anaerobic biosynthesis and facilitate metabolic engineering efforts for sustainable biochemical production, here we implemented a Cre-recombinase landing pad for genomic integration of large DNA sequences, resulting in demonstration of biosynthesis of a natural product from an uncharacterized BSG from a gut microbe.

We previously developed homologous recombination-based strategies for integrating DNA fragments of up to 5 kB into the *E. limosum* genome using the native RecA, however these techniques are too cumbersome and lack the efficiency to integrate very large sequences, like BSGs [5]. To overcome this difficulty we adapted the ‘landing pad’ from the CRAGE system [21]. In this system a small DNA fragment containing a selection marker and Cre recombinase gene flanked by two mutually exclusive *lox* sites is integrated into
the genome (Figure 1a). Linear heterologous DNA of arbitrary size can then be integrated by placing it between lox sites and transforming it into Cre-expressing landing pad strain, replacing the landing pad (LP) between the same set of lox sites with the heterologous DNA in an irreversible manner. We modified this for *E. limosum* by designing a landing pad containing Tet(M) for selection on tetracycline and Cre recombinase driven by the p2TetO1 promoter for inducible expression by anhydrotetracycline (aTc), flanked by loxP and lox5171. In *E. limosum* we have the advantage of an efficient electroporation protocol and RecA homologous recombination. Therefore, we integrated this landing pad into the genome by cloning it with homology arms and transforming it in as a linear fragment and relying on the native RecA DNA repair machinery for targeted genome integration [5].

Next, we tested the efficiency of the LP strain with two different integration constructs flanked by loxP and lox5171, one encoding ermB for clarithromycin resistance, and the other encoding catP for thiamphenicol resistance. Integration constructs were encoded on plasmids including the lox sites and were simply PCR amplified intact from the plasmid in a system which allows easy re-cloning of the vector for any desired cargo (Figure S1). We induced expression of Cre in the LP strain and transformed these linear PCR amplified fragments into distinct strains before selecting on the relevant antibiotics. We observed between 6000 and 8000 colony forming units (CFUs) per transformation for both fragments, which is over an order of magnitude more efficient than our previous RecA-mediated genomic integration in *E. limosum* (Figure 1b) [5], [22]. Additionally, these transformations resulted in equivalent numbers of CFUs on non-selective plates between experimental conditions and control transformations in wild type cells, indicating that expression of Cre recombinase is not toxic and imparts no significant metabolic burden (Figure S2). Finally, we observed no colonies on selective agar plates for the LP strain transformed with the linear DNA without Cre induction, indicating that Cre expression is very tightly regulated and genomic integration is dependent on Cre. This integration strategy has several advantages over other similar systems like serine integrase as it does not rely on any heterologous enzymes for integration of the recognition sites, is entirely plasmid-free in the Clostridial host, and operates at a higher integration efficiency [23].

Given that the purpose of the LP strain is expression of heterologous DNA, we sought to understand expression level of genes integrated at this locus. We utilized the catP integration strain with the gene driven by the very strong constitutive nifJp promoter and performed a chloramphenicol acetyltransferase assay to quantify enzyme activity as compared to CatP driven by nifJp and p2tetO1 in a plasmid system [4]. When adjusted for copy number against the medium-high copy number plasmid the activity observed in the LP strain indicates that at worst this strain exhibits equivalent expression to a plasmid-borne system on a per-copy basis (Figure 1c) [4].

Figure 1: A Cre/Landing Pad Strategy Enables Efficient Targeted Genomic Integration and Strong Expression in *E. limosum*. a) Outline of landing pad strain function: landing pad is integrated into genome containing inducible Cre flanked by mutually exclusive lox sites. Then, heterologous DNA may be integrated into the genome by inducing Cre to replace the LP with the foreign DNA bounded by the same lox sites. b) Integration efficiency of heterologous cargo into landing pad strain as determined by integration of either catP or ermB and plating on their respective antibiotics. c) Enzyme activity from chloramphenicol acetyltransferase assay as a proxy for gene expression in different expression systems. U is $\Delta A_{412}/\text{min}$. Error bars are standard deviation of biological triplicates.
To test our new LP platform strain, we selected a non-ribosomal peptide synthetase (NRPS) BSG found in the genome of *Clostridium leptum* isolated from the human gut microbiota [8], [19]. We chose this BSG as it was previously part of a microbiome-derived BSG panel tested in model organisms *E. coli* and *B. subtilis*, and in those experiments 7 out of 14 BSGs failed to produce identifiable products, including the one from *C. leptum* [8], [19]. We thus reasoned that expression in a phylogenetically similar host may help to identify the native product. We obtained DNA encoding the NRPS BSG codon harmonized for expression in *E. limosum* with putative regulatory genes removed and integrated it into the LP strain under the nifJp promoter ([Figure 2a](#)). PCR and Sanger sequencing confirmed correct insertion of the 11 kB construct at the desired locus ([Figure S3](#)). After growth in liquid culture, we assessed the strain for heterologous product biosynthesis via LC-HRMS by comparison against the LP strain and the same cluster expressed in *E. coli* ([Figure 2a](#)). XCMS analysis revealed a unique feature at 229.1341 m/z in the NRPS-expressing strain which was not observed in either control strain ([Figure S4](#)) [24]. HPLC-UV analysis at 300 nm also revealed a single novel peak ([Figure S5](#)). Previous heterologous expression of clusters with similar NRPS domain architecture led to production of a class of dipeptide aldehydes that cyclized and oxidized to the corresponding pyrazinone with a characteristic 300 nm absorption peak [19]. To test whether the peak identified in our analysis shared a similar structure, we cross-referenced the identified mass against a database we generated for every possible pyrazinone dipeptide product derived from the 20 canonical amino acids, and found a mass match for a phenylalanine-valine derived pyrazinone. This structure is identical to that of phevalin, a secondary metabolite previously isolated from a different BSG in *Staphylococcus aureus* [25]. MS1 comparison between the unknown product from our NRPS strain matched a commercial phevalin standard ([Figure 2b](#)). To further confirm the heterologous production of phevalin in *E. limosum* we performed MS/MS fragmentation analysis and compared against both the phevalin standard and previously published phevalin fragmentation pattern [19], [26]. This analysis resulted in a match between our product, the standard, and the previously published pattern ([Figure 2c](#)). HPLC-UV quantification of phevalin from *E. limosum* cultures revealed that our biosynthesis strain produced 2.4 ± 1.1 mg/L phevalin.

The ability to rapidly integrate large DNA fragments and produce and identify a product from a previously expression-recalcitrant heterologous BSG highlights the potential of the *E. limosum* LP strain for discovery of novel Clostridial natural products. Additionally, the product titer attained here from genomic expression with no further optimization is similar to those previously obtained from high-level plasmid-borne expression, suggesting *E. limosum* could be a useful host for large-scale production efforts. Interestingly, the product obtained from this BSG in *E. limosum* did not match predictions from NRPS domain predictors, despite being consistent with previous characterization of similar clusters. Specifically, domain analysis from three different predictors suggests the NRPS contains two adenylation domains, consistent with a dipeptide product ([Figure 2a](#)) [27]–[29]. However, it also likely encodes a condensation starter domain, suggestive of N-acylation by endogenous acyl-CoA’s as was found by *in vitro* characterization of a similar cluster from *Ruminococcus bromii* [30]. We found no masses consistent with the acylated dipeptide aldehyde here. Another intriguing possibility is that, natively, the starter condensation domain could interact with the standalone acyl-carrier protein to load an atypical carboxylic acid precursor that is absent from *E. limosum* metabolism, analogous to Obafluorin or Clostrylpyrone biosynthesis [31], [32]. Resolving these discrepancies is the subject of future work. Overall, the addition of facile genome integration to the already advanced synthetic biology toolkit for *E. limosum* establishes a user-friendly platform for Clostridial heterologous biosynthesis, which will facilitate detailed mechanistic investigations, rapid screening of BSGs from anaerobic genomes, and metabolic engineering efforts targeting renewable production of biofuels and chemicals from single-carbon feedstocks.
Figure 2: Integration of an uncharacterized NRPS BSG from *C. leptum* in *E. limosum* results in production of phevalin. a) Gene and module layout of NRPS from biosynthetic gene cluster along with amino acid module specificity as predicted by three different predictors (C, condensation domain; A, adenylation domain; T, termination domain; R, reductase domain). The original gene cluster was refactored to remove putative regulatory genes as described in [19], and codon harmonized for expression in *E. limosum*. b) LC/MS extracted ion chromatograms at the m/z for phevalin (229.13410), for both the *E. limosum* BSG strain and the phevalin standard compared against the LP platform strain. c) MS/MS fragmentation analysis compared between *E. limosum* BSG strain and phevalin standard for m/z 229.13410 at retention time 3.9 minutes reveals an identical fragmentation pattern for both, verifying phevalin production from the BSG in *E. limosum*.

Materials and Methods

**Reagents, Bacterial Strains, and Growth Conditions:** All enzymes for cloning were obtained from New England Biolabs (NEB), all oligos from Genewiz/Azenta. Synthesized DNA was obtained from Teist Bioscience. PCR cleanup and miniprep kits were obtained from Zymo Research, phevalin standard from Cayman Chemical, and all other reagents from Fisher Scientific. All plasmid propagation was conducted in *E. coli* CopyCutter EPI400 or DH-10β, and all work in *E. limosum* was conducted in strain ATCC 8486. *E. coli* was grown at 37°C on LB media, or LB agar containing 50 μg/mL chloramphenicol, or 30°C for the BSG-carrying strain. *E. limosum* was grown strictly anaerobically at 37°C without agitation on DSMZ 135 medium without sulfide in liquid culture and reinforced clostridial medium (RCM) agar in plates. Thiamphenicol was used at a concentration of 5 μg/mL and clarithromycin at 0.5 μg/mL, and aTc at 30 ng/mL. All work involving *E. limosum* was undertaken in a Coy Lab Products flexible anaerobic chamber with a 10% CO₂, 5% H₂, 85% N₂ gas mix. Electroporation was conducted using a Bio-Rad GenePulser XCell electroporator following a protocol and electrocompetency preparation reported previously [4].

**Cre-recombinase Landing Pad Strain Development:** Plasmid pEL72 encoding the landing pad was assembled from synthetic DNA fragments into pEL1, linearized by SfoI BamHI digestion, and assembled using NEB-builde Hi-Fi Assembly. Candidate clones were sequence validated using the Pacific BioSciences Sequel IIe platform and analyzed with custom pipelines (Table S1). Landing pad DNA was amplified from pEL72 using oligos oPAS484 and 485 and integrated into the *E. limosum* genome via electroporation and RecA recombinating and confirmed by sequencing (Table S2, Figure S1). Cultures were grown both with and without aTc induction and competent cells prepared. DNA encoding ermB with *lox* sites was amplified from pEL84 and *catP* with *lox* sites from pLP-Cat both using oligos oPAS484 and 485. Linear DNA was transformed into induced and uninduced LP strain as well as WT in biological triplicate and plated onto plates containing thiamphenicol, clarithromycin, or no selection in a ten-fold dilution series spanning 1:10 to 1:100000. Plates were allowed to grow for six days before colonies were counted to assess LP function (Tables S3, S4). Proper replacement of the LP with the desired DNA was confirmed by Sanger sequencing. The strain encoding *catP* was then grown in liquid culture and compared against strains carrying plasmids pEL2.1 and pEL3.1 (nifJp and p2TetO1 respectively) for gene expression by chloramphenicol acetyltransferase assay.
in biological triplicate as described previously [4].

**Heterologous NRPS Product Biosynthesis in E. limosum:** Plasmid pBGC37 encoding the NRPS BSG was assembled from synthesized DNA identically to pEL72 (Table S5). Linear DNA encoding the NRPS BSG from *C. leptum* with flanking *lox* sites was amplified from pBGC37 using oPAS101 and 102, transformed into the induced LP strain, and confirmed by sequencing. Biological triplicate 50 mL liquid cultures were grown in DSMZ 135 medium at 37°C for 48 hours before extraction with equal volumes of ethyl acetate. The organic layer was separated by centrifugation at 3500g for 2 minutes, collected, and filtered through a 0.22 μm filter into a fresh vessel before drying under nitrogen and resuspension in 200 μL 80/20 v/v methanol/dimethyl sulfoxide. Samples were diluted 1:20 in 90/10 v/v acetonitrile/water before running on an Agilent 1290 Infinity II HPLC system connected to an Agilent 6545 quadrupole time of flight (Q-TOF) mass spectrometer with a mass scan range of 50 – 750 m/z in positive TOF mode utilizing a Zorbax Eclipse Plus C18 column 2.1 x 50 mm, 1.8μm with mobile phase A of 0.1% formic acid in water and B of 0.1% formic acid in acetonitrile according to the gradient in Table S6. Unique features were screened using XCMS [24]. Samples were run on UPLC-MS/MS with the same chromatography and TOF conditions with a collision energy of 20 eV in targeted MS/MS mode for parent m/z 229.1341. LP strain and *E. coli* controls were extracted and analyzed in the same way with the exception that *E. coli* was grown in LB medium at 30°C for 48 hours before extraction. Phevalin production was quantified by HPLC utilizing an Agilent 1260 Infinity II system with a diode array detector (DAD) scanning at 300nm with a 10nm bandwidth. Analysis was done with a Poroshell 120 C18 2.7μm 3.0x100mm column held at 40°C, and a 10μL sample injection following a gradient method with mobile phase A of 90% water, 10% methanol, and 0.25% acetic acid and B of 90% methanol, 10% water, and 0.25% acetic acid described in Table S7.

**Conflict of Interest**
None declared.

**Data Availability Statement**
The data that support the findings of this study are available in the supporting material of this article and from the corresponding author upon reasonable request.

**Acknowledgements**
This work was financially supported by Northeastern University startup funds and Department of Energy, Advanced Research Projects Agency-Energy ECOSynBIO under award number #DE-AR000151. The work (proposal: https://doi.org/10.46936/10.25585/60008422) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231.

**Author Contributions**
PAS: Conceptualization, conducted experiments, manuscript preparation, figure generation. IB: Construct cloning, DNA synthesis, manuscript revision. YY: Conceptualization, manuscript revision. BW: Conceptualization, manuscript revision.

**References**


