frontiers | Research Topics

CRISPR-aided bioengineering for value-added product development

Edited by
Anindya Bandyopadhyay, Michael Köpke and Tae Seok Moon

Published in
Frontiers in Bioengineering and Biotechnology

ISSN 1664-8714
ISBN 978-2-8325-4211-8
DOI 10.3389/978-2-8325-4211-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

CRISPR-aided bioengineering for value-added product development

Topic editors

Anindya Bandyopadhyay — International Maize and Wheat Improvement Center, Mexico

Michael Köpke — LanzaTech, United States

Tae Seok Moon — Engineering Biology Research Consortium, EBRC, United States

Citation

Bandyopadhyay, A., Köpke, M., Moon, T. S., eds. (2024). *CRISPR-aided bioengineering for value-added product development*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-4211-8

Table of contents

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Editorial
PUBLISHED 13 December 2023
DOI 10.3389/fbioe.2023.1340377

OPEN ACCESS

EDITED AND REVIEWED BY
Jean Marie François,
Institut Biotechnologique de Toulouse (INSA), France

*CORRESPONDENCE
Anindya Bandyopadhyay,
anindya.biotech@yahoo.com
Michael Köpke,
michael.koepke@lanzatech.com
Tae Seok Moon,
tsmoon7@gmail.com

RECEIVED 17 November 2023
ACCEPTED 04 December 2023
PUBLISHED 13 December 2023

CITATION
Bandyopadhyay A, Köpke M and Moon TS (2023), Editorial: CRISPR-aided bioengineering for value-added product development.
Front. Bioeng. Biotechnol. 11:1340377.
doi: 10.3389/fbioe.2023.1340377

Editorial: CRISPR-aided bioengineering for value-added product development

Anindya Bandyopadhyay[1]*, Michael Köpke[2]* and Tae Seok Moon[3]*

[1]International Maize and Wheat Improvement Center, Texcoco, Mexico, [2]LanzaTech Inc., Skokie, IL, United States, [3]Moonshot Bio Inc., North Andover, MA, United States

KEYWORDS

CRISPR, gene editing, biotechnology, bioprospecting, computational design, CRISPRi, base editor, CRISPR methylation

Editorial on the Research Topic
CRISPR-aided bioengineering for value-added product development

Since its discovery and development, CRISPR technology has revolutionized biology with the possibility of editing the genome of organisms from any kingdom of life (Wang and Doudna, 2023). The technology allows for editing multiple targets in one shot, base editing at single nucleotide resolution, selectively activating or repressing gene expression, and modifying the epigenetic landscape, considered impossible or difficult to achieve previously. This wave of CRISPR technology has reached bioengineering and synthetic biology with the promise of rapid and multiplexed cellular engineering, generation of gene variants, screening for certain functions or traits, selective modulation of metabolism, and easy and fast removal of competitive pathways. A myriad of fields of biology, from the medical sector to agriculture and industrial biotechnology, are embracing this tool, leading to rapid development and discovery, and a vast CRISPR-based market is supposed to surpass 15.84 billion USD by 2028 (Bloomberg, 2022).Clinical applications, such as treating diseases or finding novel drug targets, gained momentum with the first government approval of CRISPR-based treatment for sickle-cell disease and β-thalassaemia in the UK (Wong, 2023), as did the development of new diagnostics tools. Industrial biotechnology and agriculture immensely benefit from this technology for the development and optimization of highly productive bacteria, yeasts, algae, filamentous fungi, and crops. With the advent of CRISPR tools and a decade-long development (Wang and Doudna, 2023), it's a timely topic to discuss recent advancements. The nine articles in this Research Topic cover the latest progress in the development and application of CRISPR tools in sustainability and biomanufacturing, diagnostics and drug development, and fundamental tools.

The first category is the development and applications of CRISPR tools in metabolic engineering for sustainable chemical and biofuel production. For example, gene expression control in metabolic engineering is a critical challenge. Kim and Lee address this issue by developing and implementing CRISPR interference (CRISPRi) tools to repress target genes in a multiplexed way to enhance the biomanufacturing of isoprenol, a fragrance molecule and biofuel precursor. Specifically, the authors target 32 essential and nonessential genes in *E. coli*, enabling simultaneous gene repression and achieving up to a 4.5-fold increase in an isoprenol titer. Additionally, a 2L fed-batch fermentation leads to a record titer of 12.4 g/L, demonstrating the tool's utility. Another challenge is that not all host platforms are as

advanced as model organisms such as *E. coli*. CRISPR tools open up an opportunity to access such non-traditional hosts with unique properties such as the fixation of CO_2 for bioengineering. Dhokane et al. review CRISPR-based bioengineering strategies and approaches for microalgae, which are attractive candidates for biomanufacturing due to their high photosynthetic efficiency. Promising recent progress has been made in the adaptation and use of CRISPR tools to several microalgae species, including *Chlamydomonas reinhardtii*, *Nannochloropsis* spp., *Chlorella* spp., and *Phaeodactylum tricornutum* for the production of chemicals, cosmetics, nutrition, and biofuels, but further optimization is still needed.

Acetogenic bacteria are another emerging platform for the conversion of CO_2 and industrial waste streams into fuels, chemicals, proteins, and materials. Three articles describe the growing applications of different CRISPR tools for various industrially relevant acetogens. Nwaokorie et al. describe the use of CRISPR-Cas9 in *Clostridium autoethanogenum* to study genes involved in carbon fixation. Even though this organism is used at the commercial scale today, there are still gaps in our understanding of carbon fixation that CRISPR tools can help to unravel. By deleting two genes of unknown function associated with a known C1-fixation gene cluster, the authors were able to show a significant impact on growth, product, and proteome profiles, contributing to a better understanding of genotype-phenotype relationships. Seys et al. expanded on the CRISPR-Cas9 system in *C. autoethanogenum* by adding a base editor function. Base editors Target-AID and Target-AID-NG derived from the Cas9 nuclease were used to introduce nonsense mutations into four different coding sequences. The authors also demonstrated that the system can be multiplexed, but they also highlighted some important limitations of current base editors that need further research, including off-target mutations or mixed genotypes, and they suggested best practices that should be considered. In some acetogens such as *Acetobacterium woodii*, however, attempts to install the CRISPR-Cas9 system were unsuccessful. Poulalier-Delavelle et al. explored endogenous CRISPR-Cas systems in acetogens. They developed an algorithm to automate the identification of PAM candidate sequences. With the help of the algorithm, they were able to successfully hijack the endogenous Type I-B CRISPR/Cas system of *A. woodie* and create in-frame deletions. The devised workflow was also successfully applied to the Type I-B CRISPR/Cas system of *C. autoethanogenum*.

Beyond applications in metabolic engineering and sustainability, CRISPR-Cas9 systems also have numerous applications in the medical field and can accelerate drug discovery and diagnostics. Li et al. summarize advances in CRISPR technology for rapid diagnosis of pathogenic infections via Recombinase-aided amplification (RAA). RAA is an isothermal nucleic acid amplification technology that offers advantages such as simplicity, speed, precision, energy efficiency, and convenient operation and can be integrated with CRISPR technology, enabling more convenient and intuitive determination of detection results. This integration has significantly expanded the application of RAA in pathogen detection. The discovery of new pharmaceuticals, antibiotics, therapeutic proteins, vaccine adjuvants, and bioactive natural products is often hampered by the inability to genetically engineer natural organisms, and prototyping pathways in other systems can be challenging. Hu et al. developed a highly efficient CRISPR/Cas9-mediated genome editing system in a vancomycin-producing strain of *Amycolatopsis keratiniphila*. The system enables the deletion of large fragments (up to 87.5 kb), which was not possible previously, and the generation of mutants with up to a 40% increase in the vancomycin yield. Plant cell-culture-based biomanufacturing is rapidly becoming an effective approach for the production of high-value plant-derived products, but their biomanufacturing often suffers from limited metabolic flux if they are not directly derived from the core metabolism. Brzycki Newton et al. developed a strategy using CRISPR-guided DNA methylation and chemical inhibitors to control flux to target pathways. They demonstrated the effectiveness of this approach by optimizing the biosynthesis of the potent anticancer drug paclitaxel (Taxol) via the phenylpropanoid pathway in *Taxus chinensis*. To do this, they knocked down the expression of specific enzymes in metabolism using a CRISPR-guided plant DNA methyltransferase (NtDRM) and by chemical inhibition, leading to a 25-fold increase in paclitaxel accumulation.

Since CRISPR-Cas9 as a genome-editing technology was first described (Jinek et al., 2012), it has been applied in many innovative ways including the ones described above–CRISPR interference Kim and Lee, CRISPR multiplexing Kim and Lee; Seys et al., CRISPR base editing Seys et al., and CRISPR-guided methylation Brzycki Newton et al. Another interesting opportunity is the use of pooled CRISPR gRNA libraries to do genome-wide fitness screening. Fundamental to support all these tools are CRISPR design algorithms. Simirenko et al. describes a new open-access design tool for targeted and genome-scale gRNA design ("gRNA-SeqRET"). This *in silico* tool aids the automatic extraction of target regions and the construction of pooled gRNA assemblies, and it is universally applicable for any prokaryote or eukaryote.

This Research Topic of these nine articles is just a snapshot of the current development and myriad of applications of CRISPR tools. With more real-world applications of powerful CRISPR technologies, careful consideration of their safety and ethical consequences should be made (Moon, 2023). The biological research fields will continue to be disrupted by innovative applications of CRISPR.

Author contributions

AB: Writing–original draft, Writing–review and editing. MK: Writing–original draft, Writing–review and editing. TM: Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

Author MK was employed by LanzaTech Inc. Author TM was co-founder of Moonshot Bio Inc.

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Bloomberg, (2022). Statistics on global CRISPR genome editing market size and share to surpass USD 15.84 billion by 2028, exhibit a CAGR of 29.50%. https://www.bloomberg.com/press-releases/2022-06-22/statistics-on-global-crispr-genome-editing-market-size-share-to-surpass-usd-15-84-billion-by-2028-exhibit-a-cagr-of-29-50 (Available at: [Accessed November 14, 2023].

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi:10.1126/science.1225829

Moon, T. S. (2023). EBRC: enhancing bioeconomy through research and communication. *N. Biotechnol.* 78, 150–152. doi:10.1016/J.NBT.2023.10.012

Wang, J. Y., and Doudna, J. A. (2023). CRISPR technology: a decade of genome editing is only the beginning. *Science* 379, eadd8643. doi:10.1126/science.add8643

Wong, C. (2023). UK first to approve CRISPR treatment for diseases: what you need to know. *Nature* 623, 676–677. doi:10.1038/D41586-023-03590-6

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 03 March 2023
DOI 10.3389/fbioe.2023.1141176

OPEN ACCESS

EDITED BY
Tae Seok Moon,
Engineering Biology Research Consortium (EBRC), United States

REVIEWED BY
Eric Young,
Worcester Polytechnic Institute, United States
Jeongwook Lee,
Pohang University of Science and Technology, Republic of Korea

*CORRESPONDENCE
Xiuping Qian,
qianxp@sjtu.edu.cn

†These authors contributed equally to this work and share first authorship

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

RECEIVED 10 January 2023
ACCEPTED 20 February 2023
PUBLISHED 03 March 2023

CITATION
Hu M, Chen S, Ni Y, Wei W, Mao W, Ge M and Qian X (2023), CRISPR/Cas9-mediated genome editing in vancomycin-producing strain *Amycolatopsis keratiniphila*. *Front. Bioeng. Biotechnol.* 11:1141176. doi: 10.3389/fbioe.2023.1141176

CRISPR/Cas9-mediated genome editing in vancomycin-producing strain *Amycolatopsis keratiniphila*

Mengyi Hu[1†], Shuo Chen[1†], Yao Ni[1], Wei Wei[2], Wenwei Mao[1], Mei Ge[2] and Xiuping Qian[1*]

[1]School of Pharmacy, Shanghai Jiao Tong University, Shanghai, China, [2]Shanghai Laiyi Center for Biopharmaceutical R&D, Shanghai, China

Amycolatopsis is an important source of diverse valuable bioactive natural products. The CRISPR/Cas-mediated gene editing tool has been established in some *Amycolatopsis* species and has accomplished the deletion of single gene or two genes. The goal of this study was to develop a high-efficient CRISPR/Cas9-mediated genome editing system in vancomycin-producing strain *A. keratiniphila* HCCB10007 and enhance the production of vancomycin by deleting the large fragments of ECO-0501 BGC. By adopting the promoters of *gapdh*p and *ermE**p which drove the expressions of *scocas9* and sgRNA, respectively, the all-in-one editing plasmid by homology-directed repair (HDR) precisely deleted the single gene *gtfD* and inserted the gene *eGFP* with the efficiency of 100%. Furthermore, The CRISPR/Cas9-mediated editing system successfully deleted the large fragments of *cds13-17* (7.7 kb), *cds23* (12.7 kb) and *cds22-23* (21.2 kb) in ECO-0501 biosynthetic gene cluster (BGC) with high efficiencies of 81%–97% by selecting the sgRNAs with a suitable PAM sequence. Finally, a larger fragment of *cds4-27* (87.5 kb) in ECO-0501 BGC was deleted by a dual-sgRNA strategy. The deletion of the ECO-0501 BGCs revealed a noticeable improvement of vancomycin production, and the mutants, which were deleted the ECO-0501 BGCs of *cds13-17*, *cds22-23* and *cds4-27*, all achieved a 30%–40% increase in vancomycin yield. Therefore, the successful construction of the CRISPR/Cas9-mediated genome editing system and its application in large fragment deletion in *A. keratiniphila* HCCB10007 might provide a powerful tool for other *Amycolatopsis* species.

KEYWORDS

amycolatopsis, CRISPR/Cas9, genome editing, large fragment deletion, Eco-0501, vancomycin

1 Introduction

Amycolatopsis, a crucial genus of actinomycetes, was firstly defined as a new genus in 1986 with the feature of type IV cell wall composition and lacking mycolic acids (Lechevalier et al., 1986), and 86 species with validly published names have been described as of 2022 (https://lpsn.dsmz.de/genus/amycolatopsis). The genus *Amycolatopsis* is regarded as an important source for the generation of diverse valuable bioactive secondary metabolites (Song et al., 2021), such as antibiotics [vancomycin (Barna and Williams, 1984), rifamycin (Saxena et al., 2014), chloroeremomycin (Lu et al., 2004), balhimycin (Frasch et al., 2015), and ECO-0501 (Banskota et al., 2006)]. Several *Amycolatopsis* species were applied in

bioremediation (heavy metal immobilization, herbicide and polymer biodegradation) and bioconversion (wuxistatin and vanillin production) (Kisil et al., 2021).

The genome sequence analysis of *Amycolatopsis* spp. (http://wwws.ncbi.nlm.nih.gov/genomes/lproks.cgi) has revealed that the strains of *Amycolatopsis*, which have comparatively large genomes (from 5 to 10 Mb) in the form of a circular chromosomes, contain over 20 BGCs of natural products and the majority of BGCs are rarely or even not expressed under typical laboratory culture conditions (Kumari et al., 2016). The efficient genome editing tools not only can discover new valuable compounds by activating silent BGCs (Choi et al., 2015; Katz and Baltz, 2016; Nguyen et al., 2020), but also can improve the yield and purity of target metabolites and enhance the strain stability by amplifying BGC copy numbers, deleting genes for competing pathways, manipulating positive and negative regulatory genes, expressing BGCs in heterologous hosts, or refactoring the transcription and translation process, and so on (Baltz, 2016; Horbal et al., 2018). Among actinomycetes, CRISPR/Cas-based genetic engineering has been the most extensively investigated and widely applicated system in *Streptomyces* species, and it accelerated the natural product discovery, strain improvement, and functional genome research by single or multiplex gene/genome editing with higher efficiencies (Cobb et al., 2015; Tong et al., 2015; Wang et al., 2016; Tao et al., 2018; Alberti and Corre, 2019). The genetic manipulation in *Amycolatopsis* has progressed slowly due to the scarcity of sophisticated genetic tools and methods, such as strain-compatible tools, cloning methods for high GC-content DNA sequence, and transfer methods (Malhotra and Lal, 2007; Meyer et al., 2017; Mitousis et al., 2020). At present, the highly efficient CRISPR/Cas12a-based genome editing systems were developed in *A. mediterranei* U32 and *A. orientalis* AO-1, and deleted *rifZ*, *glnR*, and *gtfDE* genes successfully (Zhou et al., 2020; Qian et al., 2021). Furthermore, a CRISPR/Cas9 system could delete *vdh* gene with the efficiency of 10% in *Amycolatopsis* sp. (Zheng et al., 2021). However, these studies only conducted the knock-out of small fragments of about 1–3 kb. It is still crucial and challenging to develop a highly efficient CRISPR/Cas system for manipulating large DNA fragments in *Amycolatopsis*. The industrial strain of *A. keratiniphila* HCCB10007 has been used for large-scale production of the vital antibiotic vancomycin, and 26 gene clusters related to secondary metabolism were identified in the genome (Xu et al., 2014). The glycosidic polyketide antibiotic of ECO-0501, which discovered from the vancomycin-producer strain by genome scanning, was another important bioactive secondary metabolite (Banskota et al., 2006; Shen et al., 2014). The aim of this study was to develop a highly efficient CRISPR/Cas9-mediated editing system for deleting large fragments of ECO-0501 BGC in *A. keratiniphila* HCCB10007 and improve the production of vancomycin.

The all-in-one single plasmid pKCcas9dO consists of a target-specific guide RNA (sgRNA), a codon-optimized *cas9* (*scocas9*), two HDR templates and the temperature-sensitive replicon pSG5, and the system was employed by Huang et al., 2015 to create single/double gene deletions, single/double large-size gene cluster deletions, and point mutations in *S. coelicolor* with high efficiencies. The homologous regions flanking the editing sites in the system were used as a template for DNA double-strand breaks (DSBs) recombination repair (Huang et al., 2015), and provided more efficient and accurate target gene editing (Tong et al., 2015; Wang et al., 2016; Zhang et al., 2020; Zhou et al., 2020).

Here, the CRISPR/Cas9-mediated genome editing system derived from the plasmid pKCcas9dO was established in *A. keratiniphila* HCCB10007. The CRISPR/Cas9-mediated editing system could delete and insert single gene precisely, accomplish a high-efficient deletion of large-size DNA fragments of 21 kb by choosing proper sgRNA, and achieve a larger DNA fragment of 87.5 kb deletion by dual-sgRNA-guided cleavage strategy. It drastically improved the genome editing efficiency in *A. keratiniphila* and increased the production of vancomycin by deleting the competing biosynthetic pathway of ECO-0501.

2 Materials and methods

2.1 Strains, plasmids, and cultivation conditions

The strains and plasmids used in this study are listed in Supplementary Table S1. The cells of *E. coli* DH5α and *E. coli* JM110 were grown in Luria–Bertani medium at 37°C for 12–16 h. All *A. keratiniphila* strains were grown at 28°C. For sporulation, the cells were grown on Gauze's synthetic agar medium (GM). For preparing the competent cells, *A. keratiniphila* HCCB10007 was cultured in complete pre-cultivation medium (CRM) for 48 h. For the selection of transformants, the cells of *A. keratiniphila* were grown on Bennet's medium for 4 days, and then were cultivated in tryptic soy broth (TSB) liquid medium for 2 days. All of the above strains were cultured with agitation at 220 rpm in liquid medium. When necessary, 50 μg/mL or 100 μg/mL of apramycin (Apr) was added in the liquid medium or solid medium. For vancomycin and ECO-0501 production, *A. keratiniphila* HCCB10007 and the mutants were cultivated in seed medium for about 60 h with shaking at 250 rpm, and were then incubated at 200 rpm for 4 days (Shen et al., 2014; Xu et al., 2014).

2.2 Primers and reagents

The primers used in this study are listed in Supplementary Table S2. Restriction enzymes, Taq enzymes and ligases, and other common molecular biology reagents were purchased from TaKaRa. The ClonExpress MultiS One Step Cloning Kit purchased from Vazyme Biotech Co., Ltd. was used for ligation of fragments and vectors by homologous recombination. High fidelity polymerase KOD FX and PrimeSTAR (Toyobo) were used to amplify target gene for cloning purposes and to perform PCR screening of mutant strains according to the manufacturer's protocol. PCR reactions were carried out in a PCR instrument (Eppendorf). The sequencing of DNA and the synthesis of all primers were conducted by GENEWIZ. DNA recovery kits and plasmid extraction kits were purchased from Toyobo, and a DNA Marker (GenerulerTM 1 kb DNA ladder) was purchased from Fermentas. All chemicals used were analytical grade and commercially available.

2.3 DNA manipulations

Isolation of genomic DNA from *Amycolatopsis* strains and plasmid DNA from *E. coli* were carried out using standard protocols (Kieser et al., 2000). Restriction enzymes and molecular biology reagents were used according to recommendation of suppliers (Takara, Vazyme, Toyobo, Eppendorf).

2.4 sgRNA design

The CRISPR/Cas9 target online predictor CCTop (http://crispr.cos.uni-heidelberg.de), which present the rapid selection of high quality target sites for NHEJ as well as HDR, was selected for the guide sequences and protospacer adjacent motif (PAM) of sgRNA design (Stemmer et al., 2015; Labuhn et al., 2018).

2.5 Construction of CRISPR/Cas9 editing plasmids for *gtfD* deletion and *eGFP* insertion

On the basis of pKCcas9dO (Huang et al., 2015), the plasmids containing different promoters to drive the expression of *cas9* and sgRNA were constructed as follows. 1) Using the plasmid pKCcas9dO as the template, the sgRNA fragment containing the promoter *J23119*, the crRNA scaffold and 19-nt direct repeat was amplified with the primers gRNADNrecom/gtfDgRNAspc2. The resulting sgRNA fragment was cloned into the *Spe*I/*Hin*dIII-digested pKCcas9dO by the Solution-I ligation and the plasmid pKCcas9dgtfD-NA was generated. 2) The sgRNA fragment containing the crRNA scaffold and 19-nt direct repeat was amplified from pKCcas9dO with the primers gRNADNrecom/gtfDgRNArecom. The *ermE** promoter was amplified from pLYZWG using the primers ermE-F/ermE-R (Xu et al., 2015). With the Gibson method, the resulting sgRNA fragment and the *ermE** promoter was recombined into the plasmid pKCcas9dO which was doubly digested by *Xba*I/*Hin*dIII, and the plasmid pKCcas9EgdgtfD-NA was generated. 3) The endogenous *gapdh* promoter was amplified from the genome of *A. keratiniphila* HCCB10007 using the primers gapdh-F/gapdh-R, and it was cloned into the *Xba*I/*Nde*I-digested pKCcas9EgdgtfD-NA by Gibson ligation, thus generating the plasmid pKCpGcas9EgdgtfD-NA.

Using the genomic DNA of *A. keratiniphila* HCCB10007 as the template, the upstream and downstream homologous arm fragments of the gene *gtfD* were obtained by PCR amplification using the primers Vcm-8F/Vcm-8R and Vcm-10F/Vcm-10R. The two fragments were recovered with a DNA recovery kit and then fused into the pMD19-Tsimple vector by overlapping extension PCR. The resulting plasmid was doubly digested by *KpnI*/*PstI* and ligated with *eGFP* fragment, which was originated from pLYZWG by double digestion with *Kpn*I/*Pst*I (Xu et al., 2015). The fragment of *eGFP* was inserted between the upstream and downstream homologous arms of *gtfD* by the Solution-I ligation. The homologous arms inserted by *eGFP* were integrated into pKCcas9dgtfD-NA, pKCcas9EgdgtfD-NA, and pKCpGCas9EgdgtfD-NA by using a homologous recombination kit to generate the CRISPR/Cas9 editing plasmids pLYNY02, pLYNY03, and pLYNY04. The *cas9* was driven by *tipA* or *gadph* promoter and the target-specific sgRNA was driven by *J23119* or *ermE** promoter, respectively. Figure 1A showed the detailed CRISPR/Cas9 editing plasmids pLYNY02, pLYNY03, and pLYNY04.

2.6 Construction of CRISPR/Cas9 editing plasmids for deleting the fragments of ECO-0501

Based on pLYNY04, a series of CRISPR/Cas9-mediated editing plasmids containing different target-specific sgRNAs and homologous DNA arms for deleting the fragments of *cds13-17* (7.7 kb), *cds23* (12.7 kb), *cds22-23* (21.2 kb), and *cds4-27* (87.5 kb) in ECO-0501 BGC were constructed, respectively. Taking the editing plasmid for knocking out *cds13-17* as an example, the upstream and downstream homologous regions of *cds13-17* were amplified from the genome of *A. keratiniphila* HCCB10007 using the primers 13-17-arm-AF/13-17-arm-AR and 13-17-arm-ZF/13-17-arm-ZR, respectively. These two fragments were recombined with the linear pLYNY04 vector generated by *Hind*III digestion. This resulting plasmid was linearized by *Spe*I digestion and ligated with annealed sgRNA oligonucleotide by overlapping recombination to obtain the final *cds13-17*-specific editing plasmid.

The dual sgRNA-guided plasmid pLYHMY87-5-I was constructed as follows. The plasmid pLYHMY7-5 was amplified with the primers of sgRNA5-F/sgRNA5-R, and the fragment containing sgRNA 5 and *ermE** promoter was purled and ligated into pLYHMY21-I digested by *Xba*I.

2.7 Construction of the *A. keratiniphila* HCCB10007 mutants

The plasmids were transferred into *E. coli* DH5α for cloning and *E. coli* JM110 for demethylation by heat shock following the manufacturer's suggested protocol. The competent cells of *A. keratiniphila* HCCB10007 for electroporation were prepared as previously described (Xu et al., 2015). The CRISPR/Cas9 editing plasmids were transformed into the competent cells of *A. keratiniphila* HCCB10007 by a Gene Pulser Xcell (Bio-Rad, Inc.) and electroporation was performed at a field strength of 7.5 kV/cm (25 μF, 600 Ω) with a pulse of about 13 ms. After electroporation, 1 mL of liquid TSB was added to the cell suspension, followed by 7 h incubation. The transformants were grown on Bennet's medium containing Apr for 5 days, and the incubation was continued in 3 mL of liquid TSB medium containing Apr at 200 rpm for 2 days. The resulting colonies were subsequently checked by PCR with a set of primers outside or inside the region of recombination and subsequent Sanger sequencing. These verified colonies were cultivated on Bennet's medium at 37°C overnight for one or two rounds to eliminate the plasmid, and subsequently incubated on Bennet's medium with or without Apr for 3 days at 28°C to select the correct mutants.

2.8 Analyses of vancomycin and ECO-0501

The analyses of vancomycin and ECO-0501 produced by the strains of *A. keratiniphila* HCCB10007 and the mutants were conducted as previously described (Shen et al., 2014; Xu et al., 2014).

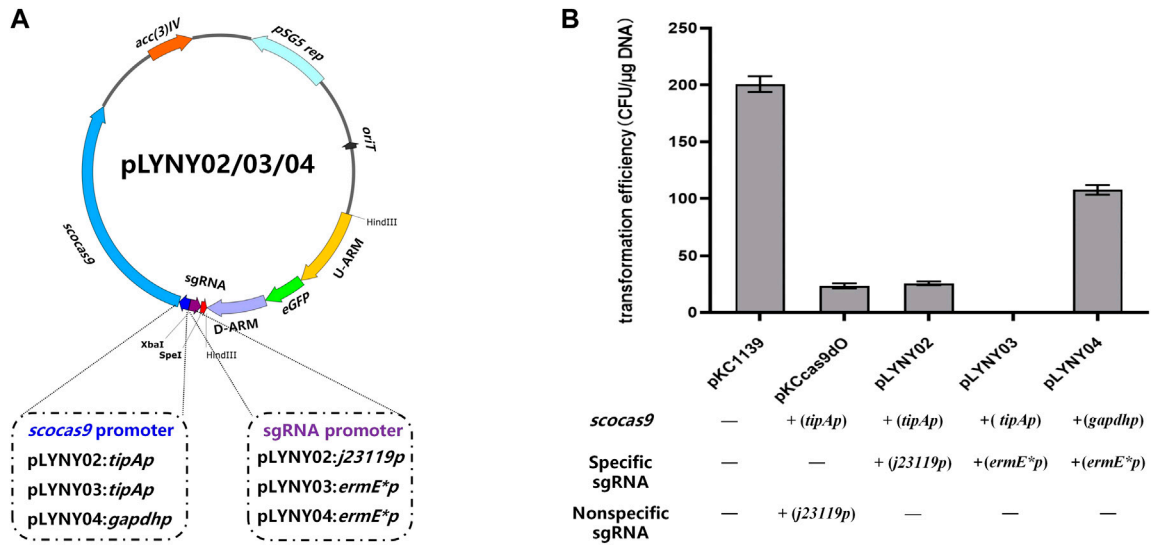


FIGURE 1
The CRISPR/Cas9-mediated editing plasmids for targeted gene editing in *A. keratiniphila*. **(A)** Schematic diagram of the editing plasmids pLYNY02, pLYNY03, and pLYNY04. Notable features included a codon-optimized *scocas9* driven by *tipA* promoter or *gapdh* promoter, target-specific sgRNA expression cassette driven by *j23119* promoter or *ermE** promoter, the homologous arms flanking the target gene, a temperature-sensitive pSG5 origin, and selection marker *aac(3)IV*. **(B)** Effects of different CRISPR/Cas9-mediated editing plasmids on the transformation efficiency. The high expression of *scocas9* under control of the inducible promoter *tipA* had toxicity and prevented the cell growth. The number of transformants was greatly decreased with pLYNY02, where no transformants were obtained with pLYNY03. When *scocas9* was driven by the endogenous promoter *gapdh* from *A. keratiniphila* HCCB10007, the transformation efficiency was significantly improved. The plasmids pKC1139 and pKCcas9dO were used as the controls. Error bars represented the standard deviations from three independent biological replicates.

2.9 Fluorescence microscopy

The strains of *A. keratiniphila* HCCB10007 and *A. keratiniphila* HCCB10007 *eGFPΔgtfD* were cultured in TSB medium for 5 days. The mycelia were observed under a Leica TCS-SP8 confocal laser-scanning microscope with a ×40 objective lens using excitation wavelength (485 nm) and emitting wavelength (535 nm) (Santos-Beneit and Errington, 2017).

3 Results

3.1 CRISPR/Cas9-mediated deletion and insertion of single gene

The CRISPR/Cas9-mediated editing plasmid in this study was derived from the high-efficiency plasmid pKCcas9dO, which originated from pKC1139 and successfully applied in *Streptomyces* (Huang et al., 2015). However, the transformation efficiency was much lower than that of the control plasmid pKC1139 in *A. keratiniphila* HCCB10007 (Figure 1B). To drive the expression of the *cas9* gene and the target-specific sgRNA in *A. keratiniphila*, the three editing plasmids, which *cas9* and sgRNA were expressed under control of different promoters, were transformed into *A. keratiniphila* HCCB10007 by electroporation. Compared with the empty vector pKC1139, the introduction of pKCcas9dO and pLYNY02 caused about 88% decrease in the transformation efficiency, and no transformants were observed in three independent experiments with pLYNY03. It was indicated that Cas9 toxicity was apparent for *A. keratiniphila*. When *cas9* was expressed under control of endogenous *gapdh* promoter from *A. keratiniphila* HCCB10007, the transformation efficiency was improved. About 46.5% decrease of transformation efficiency was caused when pLYNY04 was introduced into *A. keratiniphila* HCCB10007. The number of transformants was acceptable for the genetic manipulation. The deletion of glycosyltransferase gene *gtfD* and insertion of *eGFP* with the editing plasmid pLYNY04 was illustrated in Figure 2A.

After eliminating the plasmid of pLYNY04 by high-temperature culturing, the colonies were confirmed by PCR with three pairs of primers (Figure 2B). Because of the replacement of *gtfD* by *eGFP*, PCR amplification from the genomic DNA of the original strain produced no specific products, and PCR of the mutants had the amplicons of 1795, 3340, and 3357 bp, respectively. It was confirmed that the introduction of pLYNY04 into *A. keratiniphila* HCCB10007 caused a 100% editing efficiency of *gtfD* gene deletion and *eGFP* gene insertion from three independent replicates. The mutant strain was named *A. keratiniphila* HCCB1007 *eGFP* Δ*gtfD*.

As expected, the HPLC assay for the fermentation products showed that unlike the *A. keratiniphila* HCCB10007, no vancomycin peak was observed in the mutant of *A. keratiniphila* HCCB1007 *eGFP* Δ*gtfD* (Figure 2C). The absence of vancomycin in the metabolites of the mutant strain proved that the pLYNY04 plasmid successfully deleted the *gtfD* and further led to the block of vancomycin biosynthesis. The green fluorescence was detected clearly in the mycelia of *A. keratiniphila* HCCB1007 *eGFP* Δ*gtfD* by confocal laser microscopy, which further confirmed the completion of expected insertion of *eGFP* and the successful expression of green fluorescent protein (Figure 2D). It

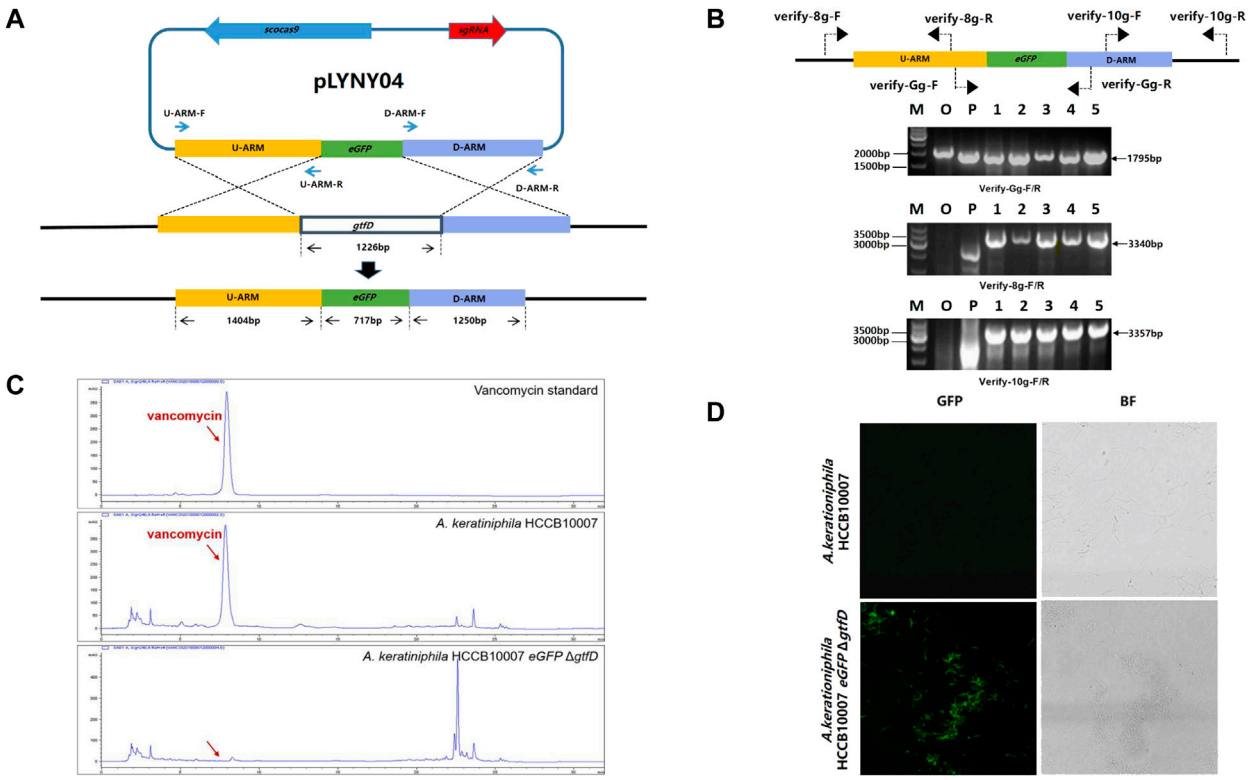


FIGURE 2
CRISPR/Cas9-mediated the deletion of *gtfD* and the insertion of *eGFP* in *A. keratiniphila*. **(A)** Strategy for replacement of *gtfD* by *eGFP* with CRISPR/Cas9-mediated editing plasmid pLYNY04. Scocas9 introduced DSB within *gtfD* gene, which could then be precisely repaired by a donor DNA fragment containing *eGFP* and the flanking upstream and downstream homologous arms. **(B)** PCR amplification analysis of randomly selected transformants verified the simultaneous *gtfD* deletion and *eGFP* insertion. The 1795 bp amplicons represented the replacement of *gtfD* by *eGFP*, and the 3340bp and 3357bp PCR amplicons verified the insertion of *eGFP*. M, GeneRuler 1 kb DNA Ladder (Thermo Scientific). O, original strain *A. keratiniphila* HCCB10007. P, pLYNY04. Lanes 1–5, colonies transformed with pLYNY04. **(C)** HPLC analysis of vancomycin from the original strain *A. keratiniphila* HCCB10007 and the mutant *A. keratiniphila* HCCB10007 *eGFPΔgtfD*. The *gtfD* deleted mutant did not produce vancomycin. The position indicated by the red arrow was vancomycin peak. **(D)** Mycelia observation of the strain *A. keratiniphila* HCCB10007 and the mutant *A. keratiniphila* HCCB10007 *eGFPΔgtfD* by confocal laser-scanning microscopy. The fluorescence of the mutant mycelia confirmed the expression of *eGFP*.

could be therefore concluded that the CRISPR/Cas9-mediated editing plasmid pLYNY04, harboring *cas9* driven by *gadph* promoter and the target-specific sgRNA driven by *ermE** promoter, was capable of HDR-dependent gene editing in *A. keratiniphila* HCCB1007 and the system could achieve high efficiency for the deletion and insertion of single gene.

3.2 CRISPR/Cas9-mediated deletion of large fragments in ECO-0501 BGC with single sgRNA

The editing ability of CRISPR/Cas9 is mainly determined by the capacity of sgRNA to recognize and cleave at specific sites, and a great quantity of online sgRNA design websites are available to predict and improve the efficiency of gene editing (Cui et al., 2018; Liu et al., 2020). The sgRNA score of CCTop is based on the off-target quality and the distribution of mismatches, and guides the user towards selecting the optimal target site (Stemmer et al., 2015). To achieve the deletion of large chromosomal fragments in *A. keratiniphila* HCCB1007, *cds13-17* (7.7 kb) and *cds22-23* (21.2 kb) in the ECO-0501 BGC were selected as targeting clusters, and nine sgRNAs targeting specific sites of *cds13-17* and *cds22-23*, which scores were over 0.75 and there were no predicted off-target effects were designed. The detailed information of location, sequence composition, PAM sequence, scores in CCTop and GC contents were shown in Figure 3A. On the basis of pLYNY04, the plasmids harboring the corresponding sgRNAs and the homologous arms of *cds13-17* or *cds22-23* were constructed, respectively. No successful deletion of *cds13-17* was observed using the plasmids of pLYHMY7-2 and pLYHMY7-4. With the plasmids of pLYHMY7-1, pLYHMY7-3, pLYHMY21-II, and pLYHMY21-III, the editing efficiencies of 5% to 29% were relatively low. The plasmids of pLYHMY7-5, pLYHMY7-6, and pLYHMY21-I exhibited the more efficient editing ability, and the editing efficiencies of 87% ± 12%, 74% ± 20%, and 97% ± 3% were accomplished, respectively for the deletion of *cds13-17* and *cds22-23* clusters (Figure 3A). The correct deletions of the clusters of *cds13-17* or *cds22-23* were verified by PCR with two pairs of primers located inside and outside of the targeting deletion clusters (Figure 3B). PCR of the *cds13-17* deleted colonies could not produce a 1491bp DNA fragment with the primers of Verify-7in-F/Verify-7in-R, while amplified a 6092bp DNA fragment with the

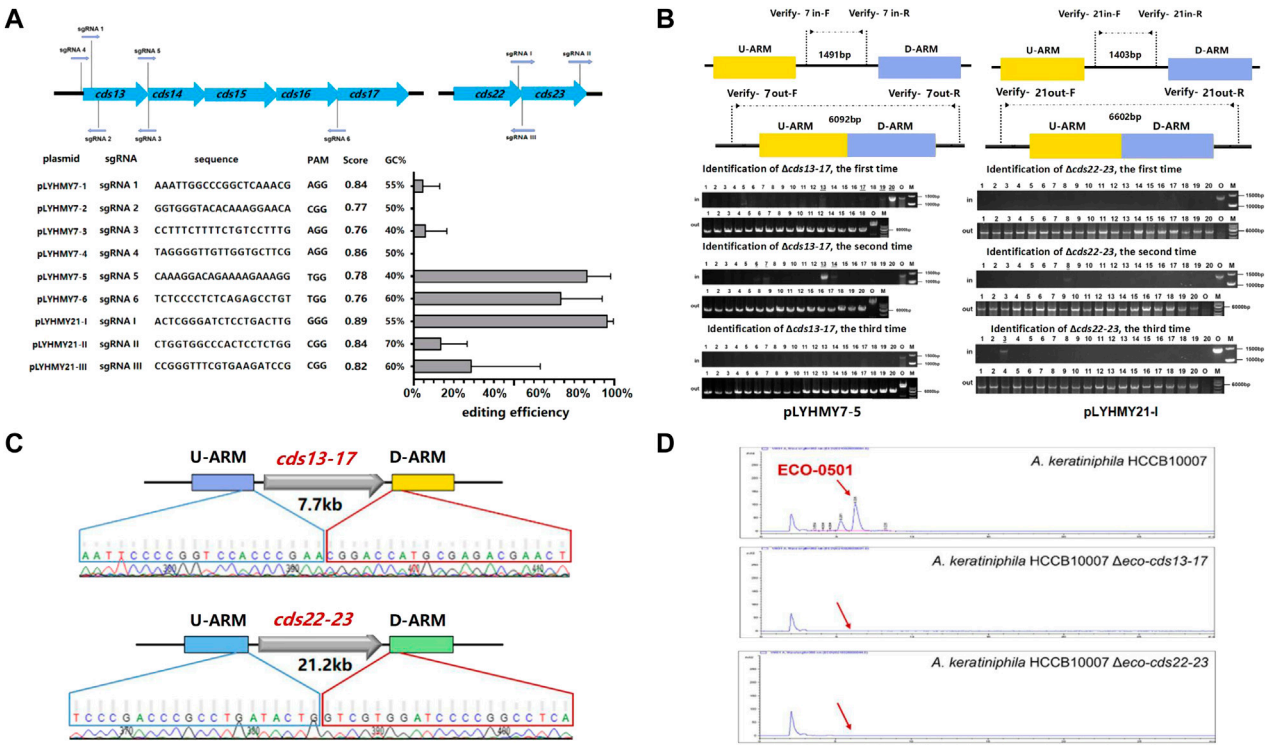


FIGURE 3
CRISPR/Cas9-mediated deletions of large fragments in ECO-0501 BGC. **(A)** The deletion efficiencies of *cds13-17* and *cds22-23* obtained by the corresponding editing plasmids. The editing plasmids contained different sgRNAs expression cassettes and up/downstream homologous arms of target clusters. Error bars represented the standard deviations from three independent biological replicates. The information of sgRNAs included the binding position, sequence composition, PAM sequence, score and GC contents. **(B)** PCR analysis of randomly selected colonies verified the deletions of *cds13-17* and *cds22-23*. The amplicons of 1491 bp and 1403 bp were the products of the original strain, respectively, and 6092 bp and 6602 bp amplicons represented that clusters *cds13-17* and *cds22-23* of colonies were deleted, respectively. The three repeated experiments of pLYHMY7-5 and pLYHMY21-I listed as representative. The underline below the number indicated that the DNA fragments of the colonies had not been deleted after the verification of Verify-7in-F/Verify-7in-R and Verify-21in-F/Verify-21in-R. M, GeneRuler 1 kb DNA Ladder (Thermo Scientific). O, original strain *A. keratiniphila* HCCB10007. Lanes 1–20, 20 colonies transformed by pLYHMY7-5 with *cds13-17*-specific sgRNA5 and pLYHMY21-I with *cds22-23*-specific sgRNAI. **(C)** Sanger sequencing results confirmed the correct deletion of the *cds13-17* and *cds22-23* clusters. **(D)** HPLC detection of ECO-0501 of the secondary metabolites in the mutants indicated that there were no ECO-0501 production owing to the successful deletion of the *cds13-17* and *cds22-23* clusters. The position indicated by the red arrow was ECO-0501 peak.

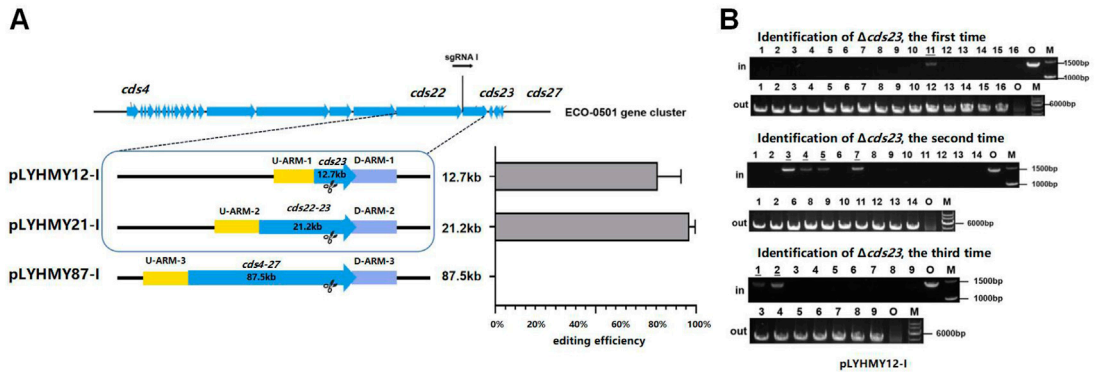


FIGURE 4
The deleting efficiencies of the same sgRNA expression cassette on different size of fragments. **(A)** The deleting efficiencies of *cds23*, *cds22-23*, and *cds4-27* obtained by the editing plasmids containing the same sgRNA1 expression cassette and different up/downstream homologous arms of target clusters. Error bars represented the standard deviations from three independent biological replicates. **(B)** PCR analysis of randomly selected colonies transformed with pLYHMY12-I verified the deletions of *cds23*. The amplicons of 1481 bp were the product of the original strain and 5169 bp amplicons represented that cluster *cds23* of colonies were deleted. The underline below the number indicated that the DNA fragments of the colonies had not been knocked out after the verification of Verify-12in-F/Verify-12in-R. M, GeneRuler 1 kb DNA Ladder (Thermo Scientific). O, original strain *A. keratiniphila* HCCB10007. Lanes 1–16, colonies transformed by pLYHMY12-I with *cds23*-specific sgRNA1.

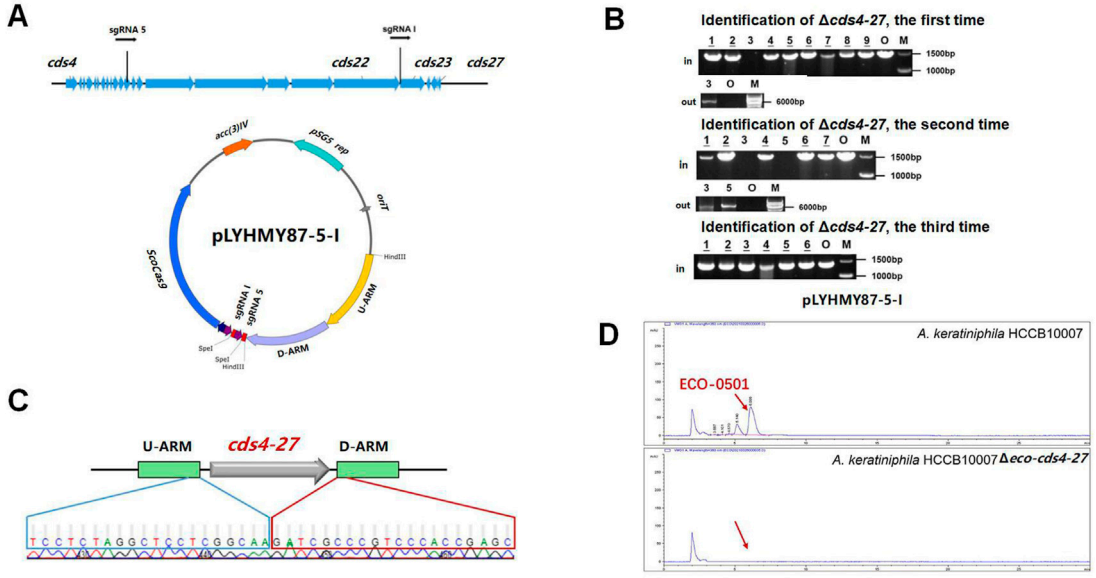


FIGURE 5
CRISPR/Cas9-mediated deletion of large fragment in ECO-0501 BGC with dual sgRNAs expression cassettes. **(A)** The location of sgRNA5 and sgRNAI on the ECO-0501 cluster and the diagram of editing plasmid PLYHMY87-5-I with dual sgRNAs expression cassette. **(B)** PCR analysis of randomly selected colonies transformed with pLYHMY87-5-I verified the deletions of *cds4-27*. The amplicons of 1403 bp were the products of the original strain and 6777 bp amplicons represented *cds4-27* deleted mutants. M, GeneRuler 1 kb DNA Ladder (Fermentas). O, original strain *A. keratiniphila* HCCB10007. Lanes 1–9, colonies transformed with pLYHMY87-5-I targeting CDS4-27. The underline below the number indicated that the gene of the colony had not been knocked out after the verification of Verify-87in-F/R. **(C)** Sanger sequencing results confirmed the correct deletion of the *cds4-27* cluster. **(D)** HPLC detection of ECO-0501 of the secondary metabolites in the mutants indicated that there were no ECO-0501 production owing to the successful deletion of the *cds4-27* cluster. The position indicated by the red arrow was ECO-0501 peak.

primers of Verify-7out-F/Verify-7out-R. Similarly, the results showed that the *cds22-23* deleted colonies gave a specific 6602bp amplicon with the primers Verify-21out-F/Verify-21out-R, and did not have 1403bp amplicon with the primers Verify-21in-F/Verify-21in-R. The sequencing results of the DNA fragments deleted colonies also confirmed that the deletion of the DNA fragments of *cds13-17* or *cds22-23* and homologous recombination repair were achieved (Figure 3C). The HPLC analyses of the metabolites fermented by the two clusters deleted mutants *A. keratiniphila* HCCB10007 Δ*eco-cds13-17* and *A. keratiniphila* HCCB10007 Δ*eco-cds22-23* demonstrated that the mutant strains were unable to biosynthesize ECO-0501 and the genomic targets had been deleted correctly (Figure 3D).

To detect the editing efficiency of the same sgRNA expression cassette with the corresponding homologous arms, the plasmids of pLYHMY12-I targeting to *cds23* (12.7 kb) and pLYHMY87-I targeting to *cds4-27* (87.5 kb) were transformed into *A. keratiniphila* HCCB10007. The deletion efficiency of *cds23* was 81% ± 12%, which was comparable to the editing efficiency of 97% ± 3% which achieved by the plasmid of pLYHMY21-I for deleting *cds22-23* (21.2 kb) (Figure 4A). The correct deletion of *cds23* was validated by PCR with the corresponding verification primers. The *cds23* deleted colonies could amplify a 5169 bp fragment with the primer of Verify-12out-F/Verify-12out-R and did not obtain the amplicon of 1481 bp with primer Verify-12in-F/Verify-12in-R (Figure 4B). No mutant with deletion of *cds4-27* was obtained (Figure 4A). The plasmid of pLYHMY87-I failed to delete the fragment of 87.5 kb, which was likely due to the inefficient of single sgRNA targeting to much larger gene cluster. So, the plasmid pLYNY04 derivatives with single sgRNA and two homologous arms flanking the targeted clusters could delete large DNA fragment, which covers a size of 21.2 kb region in ECO-0501 BGC with a higher efficiency.

3.3 CRISPR/Cas9-mediated deletion of large fragment in ECO-0501 BGC with dual sgRNAs

The dual-sgRNA strategy which introduced DSBs at both ends and bridged the gap with homologous arms clearly demonstrated that it could significantly improve the deletion efficiency of CRISPR/Cas9 system in editing larger fragments (Cobb et al., 2015; Huang et al., 2015). ECO-0501 BGC spans approximately 100 kb of DNA in *A. keratiniphila* HCCB10007 (Xu et al., 2014), The deletion of larger DNA fragments of ECO-0501 BGC is particularly valuable for increasing the availability of the precursors beneficial for vancomycin biosynthesis. To further delete the cluster of *cds4-27*, which covers 87.5 kb of DNA fragments in ECO-0501 BGC, the tandem *cds4-27*-specific sgRNA expression cassettes driven by two copies of the promoters *ermE** to increase the cleavage was considered. The plasmid pLYHMY87-5-I, containing the efficient sgRNAs of both sgRNA 5 and sgRNA I with corresponding homologous arms for *cds4-27* (Figure 5A), was constructed and transformed into the strain of *A. keratiniphila* HCCB10007. The average deletion

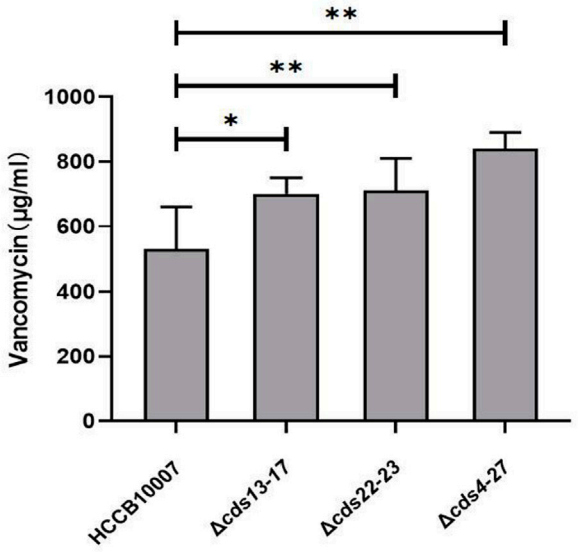


FIGURE 6
Improvement of vancomycin production. HPLC detection of vancomycin in the mutants indicated that the yield of vancomycin was increased owing to the successful deletions of gene clusters *cds13-17*, *cds22-23*, and *cds4-27* in ECO-0501 BGC. The position indicated by the red arrow was ECO-0501 peak. Original, original strain *A. keratiniphila* HCCB10007; Δ*eco-cds13-17*, the mutant *A. keratiniphila* HCCB10007 Δ*eco-cds13-17*; Δ*eco-cds22-23*, the mutant *A. keratiniphila* HCCB10007 Δ*eco-cds22-23*; Δ*eco-cds4-27*, the mutant *A. keratiniphila* HCCB10007 Δ*eco-cds4-27*. Error bars represented the standard deviations from three independent biological replicates of each strain. *$p < 0.05$, **$p < 0.01$.

efficiency of *cds4-27* cluster was 13% ± 11%, and the cluster deleted colonies were validated by PCR with the corresponding verification primers. There was no amplicons of 1403 bp with the primer of Verify-21in-F/Verify-21in-R, and the amplicon of 6777 bp was amplified with the primer of Verify-87out-F/Verify-87out-R (Figure 5B). The DNA sequencing result of the strain *A. keratiniphila* HCCB10007 Δ*eco-cds4-27* demonstrated that the mutant deleted the cluster of *cds4-27* completely (Figure 5C), and could not biosynthesize the metabolite of ECO-0501 (Figure 5D). So, the pLYHMY87-5-I with dual sgRNAs and two homologous arms flanking the targeted cluster could improve the deletion efficiency in deleting larger fragment, which covers a size of 87.5 kb region in ECO-0501 BGC.

3.4 Improvement of vancomycin production through the deletion of ECO-0501 BGC

A. keratiniphila HCCB10007 is well known for producing vancomycin and ECO-0501, and the both biosynthesis pathways share some common precursor (Shen et al., 2014; Xu et al., 2014). Because of the deletions of gene clusters *cds13-17*, *cds22-23* and *cds4-27* in the BGC of ECO-0501, the ability of vancomycin biosynthesis in the three mutants were all significantly improved, and the production of vancomycin increased by 30.54% ± 6.2%, 33.99% ± 8.6%, and 40.58% ± 7.5%, respectively (Figure 6).

4 Discussion

The CRISPR/Cas-mediated gene/genome editing system exhibits powerful efficiency and has been successfully applied in the actinomycetes *Streptomyces* species (Cobb et al., 2015; Wang et al., 2016). Both CRISPR/Cas9 and CRISPR/Cas12a-mediated gene editing systems have been reported deleting single gene (*rifZ*, *glnR* and *vdh*) and two genes *gtfDE* in the rare actinomycetes of *A. mediterranei* U32, *A. orientalis* AO-1 and *Amycolatopsis* sp. (Zhou et al., 2020; Qian et al., 2021; Zheng et al., 2021), however it is more desirable to develop a highly efficient CRISPR/Cas system to knock out large chromosomal fragments in *Amycolatopsis*.

All-in-one CRISPR/Cas9-mediated genome editing plasmid pKCcas9dO harbored the temperature-sensitive characteristic of replicon pSG5, apramycin-resistance gene (*aac3IV*), the thiostrepton-inducible promoter *tipA* and a codon optimized *cas9* from *S. pyogenes*, the promoter *J23119* and a target specific sgRNA, along with a pair of homologous recombination repair templates for HDR after DSB. It created the single gene deletions as well as whole antibiotic BGC deletions of up to 82.8 kb with an efficiency of 60%–100% in *S. coelicolor* M145 (Huang et al., 2015). It was also successfully applied in *S. pristinaespiralis* and *S. cinnamonensis* for deleting a 25.5 kb-long gene cluster and *dasR* with higher efficiency (Huang et al., 2015; Zhang et al., 2016). However, the gene editing plasmid pLYZYP01 derived from pKCcas9dO deleted the *vdh* gene with a lower efficiency of 10% in *Amycolatopsis* sp. (Zheng et al., 2021). Here, an efficient CRISPR/Cas9-mediated genome editing system, derived from pKCcas9dO, was reported in vancomycin-producing strain *A. keratiniphila* HCCB10007. Employing the established method, the deletions of large-size gene clusters were successfully achieved, and the production of vancomycin was increased 30%–40% by deleting BGCs of ECO-0501.

The critical issues to CRISPR/Cas9 systems are the toxicity of Cas9 in the specific strain and the poor expression of the *cas9* gene or that of sgRNA(s) (Alberti and Corre, 2019; Ding et al., 2020). Cas9/dCas9 has been demonstrated to be toxic in *Amycolatopsis* species, and no transformants were obtained when Cas9/dCas9 was expressed in *A. orientalis* AO-1 and *A. mediterranei* U32 (Zhou et al., 2020; Qian et al., 2021). In this study, the transformation efficiency of the plasmid pKCcas9dO was found much lower than that of the control plasmid. Therefore, the expressions of *scocas9* and sgRNA were needed to be tuned. It was required to confirm whether the *tipA* and J23119 promoters would be appropriate to drive the expressions of the *scocas9* and sgRNA of *A. keratiniphila*. Considering the susceptibility of the synthetic promoter J23119 to the surrounding sequence context (Jin et al., 2019), the strong constitutive promoter *ermE**, which was proven to drive the overexpression of a type II thioesterase gene (*ECO-orf27*) to enhance the yield of ECO-0501 of *A. orientalis* dA9 (Shen et al., 2014), was used to transcribe target-specific sgRNA. However, the high-level expression of *scocas9* under control of *tipA* promoter still led to negative influence on the growth of the strain *A. keratiniphila* HCCB10007. In order to improve genome editing efficiency, the Cas9 protein was usually highly expressed by a strong promoter (Ding et al., 2020). However, the expression of *cas9* with strong promoter showed toxic effect to the host cells (Wang et al., 2016; Ding et al., 2020; Mitousis et al., 2020). To address this issue, one of strategies was to modulate *cas9* expression at the transcriptional levels (Ye et al., 2020; Zhao et al., 2020). Given that the stronger promoter

gapdh drove the transcription of heterologous genes at high level in *Streptomyces* (Shao et al., 2013), and it was selected to drive the sgRNA expression of CRISPR/Cas9 plasmid pCRISPomyces-2 (Cobb et al., 2015; Wang et al., 2016), the endogenous *gapdh* promoter was attempted to lower the toxicity of Scocas9 to the cells of *A. keratiniphila* HCCB10007. As seen in the experiment, the transformation efficiency was improved. This result suggested that the Cas9 toxicity to the strain *A. keratiniphila* HCCB10007 somehow could be addressed by using endogenous *gapdh* promoter, as long as the expression level of Scocas9 was kept on the range of tolerance of the cells.

Furthermore, one dominant challenge is active, reliable and sufficient expression of Cas9 protein and sgRNA when CRISPR/Cas system is applied in non-model microorganisms (Ding et al., 2020). Generally, the strong promoters should be considered to guarantee sufficient Cas9 abundance for efficient CRISPR editing rate, and the strong expression of gRNA was also recommended for an efficient DNA target binding and CRISPR complex activation (Cobb et al., 2015; Huang et al., 2015; Ding et al., 2020). The CRISPR/Cas9 toolkits of pCRISPR-Cas9 and pCRISPomyces-2 were applied in *Streptomyces* successfully (Cobb et al., 2015; Tong et al., 2015; Wang et al., 2016), and their promoter combinations to drive the expression of *cas9* and sgRNA were *tipA*p/*ermE**p, *rpsL*p (XC)/*gapdh*p (EL), respectively. The promoter *tipA* showed higher activity compared to *ermE**p in *S. lividans* (Liu et al., 2016), while the strong promoters of *rpsL*p (XC) and *gapdh*p (EL) were confirmed similar activities in *S. lividans* (Shao et al., 2013). However, the promoter combination of the plasmid pCM4.4, which performed gene editing with high efficiency in *S. coelicolor,* was *ermE**p/*gapdh*p (EL), and the activity of *ermE**p for *cas9* expression was weaker than that of *gapdh*p (EL) for sgRNA expression (Ye et al., 2020). Thus, medium strength or weak promoters of *cas9* also showed high editing efficiency (Ding et al., 2020). The promoters which were selected to drive the expressions of CRISPR/Cas9 elements correlated with the strains. In *Amycolatopsis* sp. ATCC 39116, the Gene expression activated by *Km*^{r}p led to a six-fold lower glucuronidase activity in comparison to *ermE**p (Fleige and Steinbüchel, 2014), and the deleting efficiency of *vdh* gene was only 10% with the promoter combination of *Km*^{r}p/*ermE**p for driving the expressions of the *scocas9* and sgRNA (Zheng et al., 2021). In this study, with the help of the homologous arms and selected sgRNA, the promoter combination of *gapdh*p/*ermE**p achieved a high-efficient deletion of large-size DNA fragments, but it is worth further studying the optimal promoters for driving CRISPR/Cas9 elements in *A. keratiniphila*.

Besides, the design of the sgRNA appeared however to influence considerably the efficiency of deletions (Tong et al., 2015; Alberti and Corre, 2019; Ding et al., 2020), and the GC content, binding site and sequence composition of sgRNA might influence on the editing efficiency (Tong et al., 2015; Ding et al., 2020; Liu et al., 2020; Zhang et al., 2020). But, it was not clearly observed in this study. So, more sgRNAs should be tested for high-efficient deletions of large fragments in *A. keratiniphila* HCCB10007.

It was proposed that the editing efficiency was somehow unrelated to the edited DNA size when knocking out several specifications of fragments ranging from 1.0–82.8 kb in *Streptomyces* (Huang et al., 2015). It seems possible to find an efficient sgRNA to achieve the purpose of knocking out a fragment of whatever size, but there was an upper limit of the edited DNA size with the same sgRNA in this study. When targeting large fragments of 12.7 kb and 21.2 kb, the plasmids could perform well and reach the high editing efficiency of more than 81%. However, when the targeting fragment became larger, such as 87.5 kb tried in this experiment, the editing plasmid did not work. In *S. lividans*, it was possible for excision of larger chromosomal segments of 31 kb *red* cluster by introducing a DSB at both ends and bridging the gap with a plasmid-borne editing template, and all four exconjugants displayed the edited genotype by using a dual-targeting pCRISPomyces-2 plasmid (Cobb et al., 2015). Similar results were observed that the deletion rates of the 31.6 kb *red* cluster and 52.9 kb *act/red* clusters rose up to 67% and 45% in *S. coelicolor* by dual-sgRNA strategy (Huang et al., 2015). In this study, the dual-sgRNA strategy could improve the deletion efficiency of CRISPR/Cas9 system in editing the larger DNA fragment of 87.5 kb. The simultaneous action of two sgRNAs located at both ends of 87.5 kb fragment might be the possible reason for the increase in efficiency. Although the dual sgRNA strategy was effective in larger gene fragment deletion, only modest editing efficiency was observed and the transformants greatly reduced. To improve the editing efficiency in *A. keratiniphila*, the optimization of the *cas9* codon and the extension of homologous arms lengths should be carefully considered in the case of larger DNA fragment deletion.

It is not unusual for actinomycetes that the two or more unrelated senondary metabolic pathways compete for common precursors, cofactors, energy sources, reducing power, etc., thus limiting the potential yield of the desired product (Baltz, 2016). The biosynthesis pathways of vancomycin and ECO-0501, belonging to the non-ribosomal peptide synthetase system and type I polyketide synthase system, respectively, share some common precursors, including malonyl-CoA and D-glucose (Banskota et al., 2006; Shen et al., 2014; Xu et al., 2014). malonyl-CoA is the initial precursor not only bond with the acyl carrier protein for the biosynthesis of the polyketide backbone of ECO-0501 (Banskota et al., 2006), but also catalyze to form 3,5-dihydroxyphenylglycine (L-Dpg), which involved in the heptapeptide backbone of vancomycin (Xu et al., 2014). Both vancomycin and ECO-0501 have the glycosyl groups which transformed from D-glucose (Banskota et al., 2006; Xu et al., 2014). Because of the relationship between the biosynthetic pathways of vancomycin and ECO-0501, the disrupted pathway of ECO-0501 would redirect the precursor flow into vancomycin biosynthetic pathway and lead to the higher level of vancomycin.

In conclusion, a highly efficient CRISPR/Cas9-mediated genome editing for large DNA fragment of ~21 kb deletion was demonstrated in *A. keratiniphila* HCCB10007, and it accomplished larger DNA fragment of 87.5 kb deletion by dual sgRNA strategy. The improvement of vancomycin was realized by disrupting the competing secondary metabolic pathway of ECO-0501. This system would facilitate a wide variety of future studies in rare actinomycetes *Amycolatopsis* species, such as analysis of metabolic pathways, enhancement of secondary metabolites production, activation of silent BGCs, and industrial strain improvement.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

MH: Investigation, Validation, Data curation, Writing—original draft. SC: Investigation, Validation, Data curation. YN: Investigation, Visualization, Writing—original draft. WW: Methodology, Formal analysis, Writing—review and editing. WM: Visualization, Methodology, Formal analysis. MG: Supervision, Writing—review and editing. XQ: Project administration, Funding acquisition, Writing—review and editing.

Funding

This work was financially supported by National Key R&D Program of China (2019YFA0905400).

Conflict of interest

The 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.2023.1141176/full#supplementary-material

References

Alberti, F., and Corre, C. (2019). Editing streptomycete genomes in the CRISPR/Cas9 age. *Nat. Prod. Rep.* 36, 1237–1248. doi:10.1039/c8np00081f

Baltz, R. H. (2016). Genetic manipulation of secondary metabolite biosynthesis for improved production in *Streptomyces* and other actinomycetes. *J. Ind. Microbiol. Biotechnol.* 43, 343–370. doi:10.1007/s10295-015-1682-x

Banskota, A. H., McAlpine, J. B., Sorensen, D., Ibrahim, A., Aouidate, M., Piraee, M., et al. (2006). Genomic analyses lead to novel secondary metabolites Part 3. ECO-0501, a novel antibacterial of a new class. *J. Antibiot.* 59, 533–542. doi:10.1038/ja.2006.74

Barna, J. C. J., and Williams, D. H. (1984). The structure and mode of action of glycopeptide antibiotics of the vancomycin group. *Ann. Rev. Microbiol.* 38, 339–357. doi:10.1146/annurev.mi.38.100184.002011

Choi, S. S., Kim, H. J., Lee, H. S., Kim, P., and Kim, E. S. (2015). Genome mining of rare actinomycetes and cryptic pathway awakening. *Process Biochem.* 50, 1184–1193. doi:10.1016/j.procbio.2015.04.008

Cobb, R. E., Wang, Y., and Zhao, H. (2015). High-efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* 4, 723–728. doi:10.1021/sb500351f

Cui, Y., Xu, J., Cheng, M., Liao, X., and Peng, S. (2018). Review of CRISPR/Cas9 sgRNA design tools. *Interdiscip. Sci.* 10, 455–465. doi:10.1007/s12539-018-0298-z

Ding, W., Zhang, Y., and Shi, S. (2020). Development and application of CRISPR/Cas in microbial biotechnology. *Front. Bioeng. Biotechnol.* 8, 711. doi:10.3389/fbioe.2020.00711

Fleige, C., and Steinbüchel, A. (2014). Construction of expression vectors for metabolic engineering of the vanillin-producing actinomycete *Amycolatopsis sp* ATCC 39116. *App Microbiol. Biotechnol.* 98, 6387–6395. doi:10.1007/s00253-014-5724-5

Frasch, H. J., Kalan, L., Kilian, R., Martin, T., Wright, G. D., and Stegmann, E. (2015). Alternative pathway to a glycopeptide-resistant cell wall in the balhimycin producer *Amycolatopsis balhimycina*. *ACS Infect. Dis.* 1, 243–252. doi:10.1021/acsinfecdis.5b00011

Horbal, L., Marques, F., Nadmid, S., Mendes, M. V., and Luzhetskyy, A. (2018). Secondary metabolites overproduction through transcriptional gene cluster refactoring. *Metab. Eng.* 49, 299–315. doi:10.1016/j.ymben.2018.09.010

Huang, H., Zheng, G., Jiang, W., Hu, H., and Lu, Y. (2015). One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochim. Biophys. Sin.* 47, 231–243. doi:10.1093/abbs/gmv007

Jin, L., Nawab, S., Xia, M., Ma, X., and Huo, Y. (2019). Context-dependency of synthetic minimal promoters in driving gene expression: A case study. *Microb. Biotechnol.* 12, 1476–1486. doi:10.1111/1751-7915.13489

Katz, L., and Baltz, R. H. (2016). Natural product discovery: Past, present, and future. *J. Ind. Microbiol. Biotechnol.* 43, 155–176. doi:10.1007/s10295-015-1723-5

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000). Practical Streptomyces genetics. *Norwich John Innes Found*, 161–210.

Kisil, O. V., Efimenko, T. A., and Efremenkova, O. V. (2021). Looking back to *Amycolatopsis*: History of the antibiotic discovery and future prospects. *Antibiotics* 10, 1254. doi:10.3390/antibiotics10101254

Kumari, R., Singh, P., and Lal, R. (2016). Genetics and genomics of the genus *Amycolatopsis*. *Indian J. Microbiol.* 56, 233–246. doi:10.1007/s12088-016-0590-8

Labuhn, M., Adams, F. F., Ng, M., Knoess, S., Schambach, A., Charpentier, E. M., et al. (2018). Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. *Nucleic Acids Res.* 46, 1375–1385. doi:10.1093/nar/gkx1268

Lechevalier, M. P., Prauser, H., Labeda, D. P., and Ruan, J. S. (1986). Two new genera of nocardioform actinomycetes: *Amycolata* gen. nov. and *Amycolatopsis* gen. nov. *Int. J. Syst. Bacteriol.* 36, 29–37. doi:10.1099/00207713-36-1-29

Liu, G., Zhang, Y., and Zhang, T. (2020). Computational approaches for effective CRISPR guide RNA design and evaluation. *Comput. Struct. Biotechnol. J.* 18, 35–44. doi:10.1016/j.csbj.2019.11.006

Liu, S., Wang, M., Du, G., and Chen, J. (2016). Improving the active expression of transglutaminase in *Streptomyces lividans* by promoter engineering and codon optimization. *BMC Biotechnol.* 16, 75. doi:10.1186/s12896-016-0304-7

Lu, W., Oberthür, M., Leimkuhler, C., Tao, J., Kahne, D., and Walsh, C. T. (2004). Characterization of a regiospecific epivancosaminyl transferase *GtfA* and enzymatic reconstitution of the antibiotic chloroeremomycin. *Proc. Natl. Acad. Sci.* 101, 4390–4395. doi:10.1073/pnas.0400277101

Malhotra, S., and Lal, R. (2007). The genus *Amycolatopsis*: Indigenous plasmids, cloning vectors and gene transfer systems. *Indian J. Microbiol.* 47, 3–14. doi:10.1007/s12088-007-0003-0

Meyer, F., Pupkes, H., and Steinbüchel, A. (2017). Development of an improved system for the generation of knockout mutants of *Amycolatopsis* sp. strain ATCC 39116. *Appl. Environ. Microbiol.* 83, 026600–e2716. doi:10.1128/aem.02660-16

Mitousis, L., Thoma, Y., and Musiol-Kroll, E. M. (2020). An update on molecular tools for genetic engineering of actinomycetes—The source of important antibiotics and other valuable compounds. *Antibiotics* 9, 494. doi:10.3390/antibiotics9080494

Nguyen, C. T., Dhakal, D., Pham, V. T. T., Nguyen, H. T., and Sohng, J. K. (2020). Recent advances in strategies for activation and discovery/characterization of cryptic biosynthetic gene clusters in *Streptomyces*. *Microorganisms* 8, 616. doi:10.3390/microorganisms8040616

Qian, H., Wei, W., Chen, X. A., Mo, X. T., Ge, M., Zhao, Q. W., et al. (2021). Strategy for producing the high-quality glycopeptide antibiotic A82846B in *Amycolatopsis orientalis* based on the CRISPR-Cas12a system. *ACS Synth. Biol.* 10, 3009–3016. doi:10.1021/acssynbio.1c00317

Santos-Beneit, F., and Errington, J. (2017). Green fluorescent protein as a reporter for the spatial and temporal expression of *actIII* in *Streptomyces coelicolor*. *Arch. Microbiol.* 199, 875–880. doi:10.1007/s00203-017-1358-1

Saxena, A., Kumari, R., Mukherjee, U., Singh, P., and Lal, R. (2014). Draft genome sequence of the rifamycin producer *Amycolatopsis rifamycinica* DSM 46095. *Genome announc.* 2, 006622–e714. doi:10.1128/genomea.00662-14

Shao, Z., Rao, G., Li, C., Abil, Z., Luo, Y., and Zhao, H. (2013). Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Synth. Biol.* 2, 662–669. doi:10.1021/sb400058n

Shen, Y., Huang, H., Zhu, L., Luo, M., and Chen, D. (2014). A regulatory gene (Eco-Orf4) required for ECO-0501 biosynthesis in Amycolatopsis orientalis. *J. Basic Microbiol.* 54, 104–110. doi:10.1002/jobm.201200314

Song, Z., Xu, T., Wang, J., Hou, Y., Liu, C., Liu, S., et al. (2021). Secondary metabolites of the genus *Amycolatopsis*:structures,bioactivities and biosynthesis. *Molecules* 7, 1884–1918. doi:10.3390/molecules26071884

Stemmer, M., Thumberger, T., Keyer, M. D., Wittbrodt, J., and Mateo, J. L. (2015). CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS ONE* 10, e0124633. doi:10.1371/journal.pone.0124633

Tao, W., Yang, A., Deng, Z., and Sun, Y. (2018). CRISPR/Cas9-based editing of *Streptomyces* for discovery, characterization, and production of natural products. *Front. Microbiol.* 9, 1660. doi:10.3389/fmicb.2018.01660

Tong, Y., Charusanti, P., Zhang, L., Weber, T., and Lee, S. Y. (2015). CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth. Biol.* 4, 1020–1029. doi:10.1021/acssynbio.5b00038

Wang, Y., Cobb, R. E., and Zhao, H. (2016). High-efficiency genome editing of *Streptomyces* species by an engineered CRISPR/Cas system. *Method Enzymol.* 575, 271–284. doi:10.1016/bs.mie.2016.03.014

Xu, L., Huang, H., Wei, W., Zhong, Y., Tang, B., Yuan, H., et al. (2014). Complete genome sequence and comparative genomic analyses of the vancomycin-producing *Amycolatopsis orientalis*. *BMC Genomics* 15, 363. doi:10.1186/1471-2164-15-363

Xu, L., Li, Y., Zhu, L., Zhao, W., Chen, D., Huang, W., et al. (2015). Characterization of plasmid pXL100 from *Amycolatopsis orientalis* HCCB10007 and construction of a shuttle vector. *J. Basic Microbiol.* 55, 247–254. doi:10.1002/jobm.201400210

Ye, S., Enghiad, B., Zhao, H., and Takano, E. (2020). Fine-tuning the regulation of Cas9 expression levels for efficient CRISPR-Cas9 mediated recombination in. *Streptomyces. J. Ind. Microbiol. Biot.* 47, 413–423. doi:10.1007/s10295-020-02277-5

Zhang, S., Guo, F., Yan, W., Dai, Z., Dong, W., Zhou, J., et al. (2020). Recent advances of CRISPR/Cas9-based genetic engineering and transcriptional regulation in industrial biology. *Front. Bioeng. Biotechnol.* 7, 459. doi:10.3389/fbioe.2019.00459

Zhang, Y., Lin, C. Y., Li, X. M., Tang, Z. K., Qiao, J. J., and Zhao, G. R. (2016). DasR positively controls monensin production at two-level regulation in *Streptomyces cinnamonensis. J. Ind. Microbiol. Biotechnol.* 43, 1681–1692. doi:10.1007/s10295-016-1845-4

Zhao, Y., Li, G., Chen, Y., and Lu, Y. (2020). Challenges and advances in genome editing technologies in *Streptomyces*. *Biomolecules* 10, 734. doi:10.3390/biom10050734

Zheng, Y., Wu, D., and Zheng, P. (2021). A vanillin producing *Amycolatopsis* sp. with the *vdh* gene knockout by CRISPR-Cas9 technology. *Acta Microbiol. Sin.* 61, 3583–3593. doi:10.13343/j.cnki.wsxb.20210082

Zhou, Y., Liu, X., Wu, J., Zhao, G., and Wang, J. (2020). CRISPR-Cas12a-Assisted genome editing in *Amycolatopsis mediterranei*. *Front. Bioeng. Biotechnol.* 8, 698. doi:10.3389/fbioe.2020.00698

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 15 May 2023
DOI 10.3389/fbioe.2023.1167892

OPEN ACCESS

EDITED BY
Isabelle Meynial-Salles,
Institut National des Sciences Appliquées de Toulouse (INSA), France

REVIEWED BY
Byung-Kwan Cho,
Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea
Armin Ehrenreich,
Technical University of Munich, Germany

*CORRESPONDENCE
Kaspar Valgepea,
kaspar.valgepea@ut.ee

†These authors have contributed equally to this work

RECEIVED 16 February 2023
ACCEPTED 03 May 2023
PUBLISHED 15 May 2023

CITATION
Nwaokorie UJ, Reinmets K, de Lima LA, Pawar PR, Shaikh KM, Harris A, Köpke M and Valgepea K (2023), Deletion of genes linked to the C_1-fixing gene cluster affects growth, by-products, and proteome of *Clostridium autoethanogenum*. *Front. Bioeng. Biotechnol.* 11:1167892. doi: 10.3389/fbioe.2023.1167892

Deletion of genes linked to the C_1-fixing gene cluster affects growth, by-products, and proteome of *Clostridium autoethanogenum*

Ugochi Jennifer Nwaokorie[1†], Kristina Reinmets[1†], Lorena Azevedo de Lima[1], Pratik Rajendra Pawar[1], Kurshedaktar Majibullah Shaikh[1], Audrey Harris[2], Michael Köpke[2] and Kaspar Valgepea[1*]

[1]ERA Chair in Gas Fermentation Technologies, Institute of Technology, University of Tartu, Tartu, Estonia, [2]LanzaTech Inc., Skokie, IL, United States

Gas fermentation has emerged as a sustainable route to produce fuels and chemicals by recycling inexpensive one-carbon (C_1) feedstocks from gaseous and solid waste using gas-fermenting microbes. Currently, acetogens that utilise the Wood-Ljungdahl pathway to convert carbon oxides (CO and CO_2) into valuable products are the most advanced biocatalysts for gas fermentation. However, our understanding of the functionalities of the genes involved in the C_1-fixing gene cluster and its closely-linked genes is incomplete. Here, we investigate the role of two genes with unclear functions—hypothetical protein (*hp;* LABRINI_07945) and CooT nickel binding protein (*nbp;* LABRINI_07950)—directly adjacent and expressed at similar levels to the C_1-fixing gene cluster in the gas-fermenting model-acetogen *Clostridium autoethanogenum.* Targeted deletion of either the *hp* or *nbp* gene using CRISPR/nCas9, and phenotypic characterisation in heterotrophic and autotrophic batch and autotrophic bioreactor continuous cultures revealed significant growth defects and altered by-product profiles for both Δ*hp* and Δ*nbp* strains. Variable effects of gene deletion on autotrophic batch growth on rich or minimal media suggest that both genes affect the utilisation of complex nutrients. Autotrophic chemostat cultures showed lower acetate and ethanol production rates and higher carbon flux to CO_2 and biomass for both deletion strains. Additionally, proteome analysis revealed that disruption of either gene affects the expression of proteins of the C_1-fixing gene cluster and ethanol synthesis pathways. Our work contributes to a better understanding of genotype-phenotype relationships in acetogens and offers engineering targets to improve carbon fixation efficiency in gas fermentation.

KEYWORDS

gas fermentation, acetogen, CRISPR/Cas9, genetic engineering, metabolomics, proteomics, chemostat, syngas

1 Introduction

The high concentration of greenhouse gases (GHG) in the atmosphere is leading to potentially irreversible climate change threatening Earth's sustainability. Fossil fuel combustion for energy, heat, and transportation is widely recognised as the primary source of GHG (mainly CO_2) from human activities (Sixth Assessment Report IPCC, 2021). Likewise, increasing population leads to higher volume of generated waste [e.g., plastic waste and municipal solid waste (MSW)] that is negatively impacting ecosystems. Although decarbonising energy generation is important to lowering GHG emissions, we also need to adopt sustainable technologies for the production of fuels and chemicals as these will mostly stay carbon-based for the foreseeable future. The advancement of such technologies will be vital in achieving the United Nations Sustainable Development Goal (SDG) 7—Affordable and Clean Energy and SDG 13—Climate Action by the target year 2050.

Gas fermentation has emerged as an attractive route for sustainable production of low-carbon fuels and chemicals from gaseous one-carbon (C_1) waste feedstocks (e.g., industrial waste gases (CO_2, CO, and CH_4) and syngas [from gasified biomass or MSW (CO, H_2, and CO_2)] using gas-fermenting microbes (Liew et al., 2016; Redl et al., 2017; Fackler et al., 2021; Pavan et al., 2022). Anaerobic acetogens are the most advanced biocatalysts for gas fermentation as they can naturally use gas as their sole energy and carbon source (Drake et al., 2006). Acetogens use the most energy-efficient CO_2 fixation pathway, the Wood-Ljungdahl pathway (WLP), to fix carbon oxides (CO_2 and/or CO) into many metabolic products via acetyl-CoA. Notably, the model-acetogen *Clostridium autoethanogenum* is being used as a commercial-scale cell factory for ethanol production (Köpke and Simpson, 2020) and has been demonstrated at pilot scale to produce two valuable industrial solvents: acetone and isopropanol (Liew et al., 2022).

The WLP is considered to be the first biochemical pathway on Earth (Russell & Martin, 2004) and it plays a vital role in global carbon cycles by converting ~20% of CO_2 on Earth each year to billions of tons of acetate (Drake et al., 2006; Ljungdahl, 2009). The genes encoding the enzymes involved in the WLP belong to a large C_1-fixing gene cluster (Figure 1) that has been annotated as CAETHG_1606–1621 (Brown et al., 2014). The authors proposed that the cluster contains 16 genes, including the WLP central enzyme—the bifunctional CO dehydrogenase/acetyl-CoA synthase (CODH/ACS; CAETHG_1620-21/1608) complex—that catalyses CO oxidation/CO_2 reduction and acetyl-CoA synthesis (Ragsdale and Pierce, 2008) and is essential for autotrophic growth (Liew et al., 2016). The cluster also encodes the pathway's most abundant but slowest enzyme in *C. autoethanogenum* (Valgepea et al., 2022): the formyl-THF ligase (Fhs; CAETHG_1618). Though the biochemistry of the WLP is well-described (Ragsdale, 2008; Ragsdale and Pierce, 2008; Can et al., 2014), our understanding of the functionalities of genes within the proposed C_1-fixing gene cluster is incomplete. For instance, functions of genes CAETHG_1607 and 1613 are unclear in *C. autoethanogenum*. In fact, the C), our understanding of the functionalities of genes within the proposed C_1-fixing gene cluster is incomplete. For instance, functions of genes CAETHG_1607 and 1613 are unclear in *C. autoethanogenum*. In fact, the C), our understanding of the functionalities of genes within the proposed C_1-fixing gene cluster is incomplete. For instance, functions of genes CAETHG_1607 and 1613 are unclear in *C. autoethanogenum*. In fact, the C_1-fixing gene cluster might be larger as several adjacent genes with unknown functions show similarly high gene and protein expression levels (Valgepea et al., 2017a-fixing gene cluster might be larger as several adjacent genes with unknown functions show similarly high gene and protein expression levels (Valgepea et al., 2017a-fixing gene cluster might be larger as several adjacent genes with unknown functions show similarly high gene and protein expression levels (Valgepea et al., 2017a; 2022). Two of such genes, found right downstream of the currently annotated cluster, are hypothetical protein (*hp*; CAETHG_1604) and CooT family nickel binding protein (*nbp*; CAETHG_1605). Both *nbp* and *hp* genes are

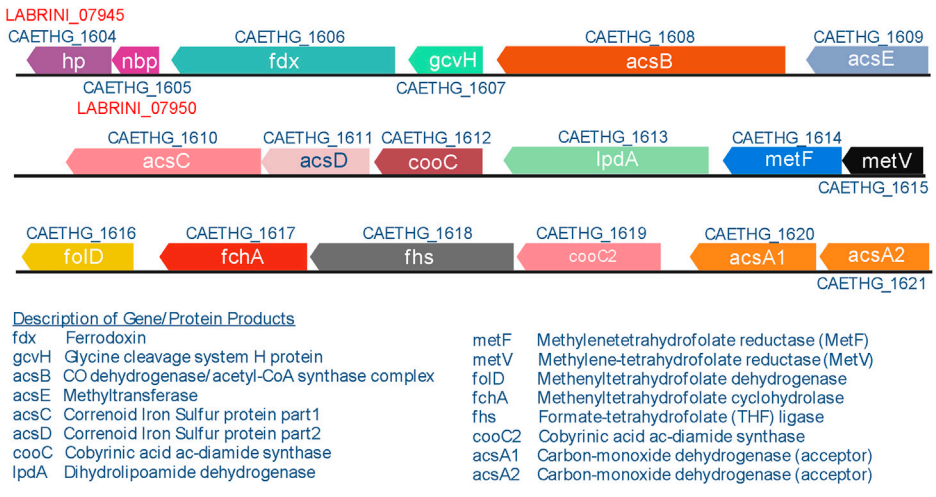


FIGURE 1
C_1-fixing gene cluster in *Clostridium autoethanogenum*. The gene cluster contains 16 genes, including five genes with unconfirmed biochemical functions (CAETHG_1606; 1607; 1612; 1613; 1619). Gene names and annotations based on Valgepea et al., 2022. Figure created using BioRender.com.

expressed highly in *C. autoethanogenum* and their homologs accompany the WLP genes of other commonly studied acetogens like *Clostridium ljungdahlii*, *Clostridium drakei*, *Clostridium carboxidivorans*, *Eubacterium limosum*, as well as thermophilic acetogens *Acetobacterium woodii* and *Thermoanaerobacter kivui*. Despite the clear genomic conservation of the *hp* and *nbp* genes among acetogens, their function has not yet been described. While nickel binding is predicted for Nbp based on gene sequence annotation, and histidine-rich amino acid sequence of the Hp protein suggests a metal-binding activity, the validity of these functional predictions remains to be demonstrated.

Determination of gene functionalities in acetogens through genotype-phenotype studies has been limited mostly due to slow, labour-intensive, and challenging workflows for constructing genetically modified strains (Pavan et al., 2022). However, gene-editing tools for acetogens have rapidly improved recently (Jin et al., 2020; Bourgade et al., 2021; Pavan et al., 2022). We thus aimed to take advantage of these developments to shed light on the functionalities of the *hp* (CAETHG_1604) and the CooT family nickel binding gene *nbp* (CAETHG_1605), referred to as LABRINI_07945 for *hp* and LABRINI_07950 for *nbp* based on the *C. autoethanogenum* base strain "LAbrini" used in this work (Ingelman et al., 2023).

Here, we used CRISPR/nCas9 in *C. autoethanogenum* for the targeted deletion of two genes adjacent to the C_1-fixing gene cluster with unclear functions: named *hp* (LABRINI_07945) and *nbp* (LABRINI_07950). Phenotypic characterisation in heterotrophic and autotrophic batch and autotrophic bioreactor continuous cultures revealed significant growth defects and altered by-product profiles for both deletion strains compared to the base strain LAbrini. Additionally, proteomics showed that disruption of either gene affects the expression of proteins of the C_1-fixing gene cluster and ethanol synthesis pathways. Our findings contribute to a better understanding of genotype-phenotype relationships which is needed to advance the metabolic engineering of acetogen factories for gas fermentation.

2 Materials and methods

2.1 Bacterial strains, growth media and cultivation conditions

A derivate of *C. autoethanogenum* DSM 10061 strain deposited in the German Collection of Microorganisms and Cell Cultures (DSMZ) named "LAbrini" (Ingelman et al., 2023) was stored as a glycerol stock at −80°C and used as the base strain for genetic engineering. *Escherichia coli* strains NEB Turbo and NEB Express (New England BioLabs) were used for cloning, plasmid assembly, and propagation. For transformant selection, the cultivation of *E. coli* was done in Lysogeny Broth (LB) medium supplemented with ampicillin (100 µg/mL).

For *C. autoethanogenum* cultivation on solid medium, YTF agar containing 10 g/L yeast extract (YE), 16 g/L tryptone, and 5 g/L fructose (Liew et al., 2017) was used and supplemented with 4 µg/mL clarithromycin if needed (see below). For liquid batch cultures, cells were grown in chemically defined PETC-MES medium (Valgepea et al., 2017b) with or without YE (1.5 g/L) and supplemented with 0.4 g/L of cysteine-HCl·H_2O as the reducing agent while the carbon source was either 5 g/L of fructose or syngas (50% CO, 20% H_2, 20% CO_2, and 10% Ar; AS Eesti AGA). The PETC-MES medium contained per litre: 1 g NH_4Cl, 0.1 g KCl, 0.2 g $MgSO_4·7H_2O$, 0.2 g KH_2PO_4, 0.02 g $CaCl_2$, 20 g 2-(N-morpholino) ethanesulfonic acid (MES), 0.25 g sodium acetate, 0.05 g $Fe(SO_4)·7H_2O$, 0.05 g nitriolotriacetic acid, 0.5 mL of 2 g/L resazurin, 10 mL trace metal solution (TMS) and 10 mL of Wolfe's vitamin solution. The TMS solution contained per litre: 2 g nitrilotriacetic acid, 1 g $MnSO_4·H_2O$, 0.8 g $Fe(SO_4)·7H_2O$, 0.2 g $CoCl_2·6H_2O$, 0.2 mg $ZnSO_4·7H_2O$, 0.02 g $CuCl_2·2H_2O$, 0.02 g $NaMoO_4·2H_2O$, 0.03 g $Na_2SeO_3·5H_2O$, 0.02 g $NiCl_2·6H_2O$ and 0.02 g $Na_2WO_4·2H_2O$. The Wolfe's vitamin solution contained per litre: 2 mg biotin, 2 mg folic acid, 10 mg pyridoxine hydrochloride, 5 mg thiamine-HCl, 5 mg riboflavin, 5 mg nicotinic acid, 5 mg calcium pantothenate, 0.1 mg vitamin B_{12}, 5 mg 4-aminobenzoic acid and 5 mg thioctic acid. The medium was prepared anaerobically, and the pH was adjusted to 5.7. Batch fermentations were performed in 250 mL Schott bottles with 50 mL liquid volume in biological triplicates incubated horizontally at 37°C with orbital shaking at 120 RPM under strictly anaerobic conditions unless stated otherwise. For autotrophic batch cultures, bottle headspace was pressurised to 140 kPa with syngas. Cultures were sampled during the exponential growth phase for growth characterisation and extracellular metabolome analysis to determine the maximum specific growth rate (μ_{max}) and production yields of acetate, ethanol, and 2,3-butanediol. At least five time points were used for μ_{max} calculation resulting in correlation coefficients $R^2 \geq 0.98$ between cultivation time and natural logarithm (ln) of the culture's optical density (OD_{600}). Product yields (mmol/gram of dry cell weight [DCW]) were estimated using at least three time points by linear regression between the respective product concentration (mmol/L) and biomass concentration (gDCW/L) with $R^2 \geq 0.92$. Product yields were estimated across time-points that showed no consumption of any of the three quantified products (i.e., products did not serve as substrates).

Continuous chemostat fermentations were carried out as described in detail before (Valgepea et al., 2017a). Shortly, all three *C. autoethanogenum* strains were grown in bioreactors on syngas under strictly anaerobic conditions at 37°C and pH 5 in chemically defined medium (without YE) at a dilution rate of 1 day^{-1}. Chemostat continuous cultures were performed in 1.4 L Multifors bioreactors (Infors AG) at a working volume of 750 mL connected to a Hiden HPR-20-QIC mass spectrometer (Hiden Analytical) for online high-resolution off-gas analysis. Steady-state results were collected after OD, gas uptake and production rates had been stable for at least 3–5 working volumes.

2.2 Genetic engineering

2.2.1 Genomic and plasmid DNA manipulations

Routine *C. autoethanogenum* genomic DNA extraction for PCR diagnostics was done with PureLink Genomic DNA extraction kit (Invitrogen™, Thermo Fischer Scientific), while plasmid DNA was extracted from *E. coli* using FavorGen FavorPrep™ Plasmid DNA Extraction Mini Kit (Favorgen Biotech Corp). All purified genomic and plasmid DNA fragments were quantified using the NanoDrop™

1000 Spectrophotometer (Thermo Fischer Scientific). PCR for amplification of DNA fragments for sequencing and cloning, routine screening, and analytical procedures was performed using the Phusion™ High-Fidelity DNA Polymerase (Thermo Fischer Scientific). All primers used in this study (see Supplementary Table S1) were designed using the NetPrimer software (PREMIER Biosoft International) and synthesised by Integrated DNA Technologies. PCR products and DNA fragments were purified from agarose gels with FavorPrep Gel/PCR Purification kit (Favorgen Biotech Corp).

2.2.2 Assembly of CRISPR/nCas9 plasmids

A CRISPR/nCas9 gene editing plasmid for *C. autoethanogenum* was constructed based on the pNickClos2.0 plasmid (Addgene plasmid #73228) (Li et al., 2016) with the following modifications to the backbone sequence: origin of transfer and the required traJ coding sequence from pFX01 (Xia et al., 2020) was added to allow potential transfer of the resulting plasmids into recipient cells *via* conjugation. Additionally, one of the two NotI restriction sites was removed to enable convenient cloning of the homology arms (HAs) using NotI and XhoI sites.

To specify the target site for the nCas9 enzyme, single guide RNAs (sgRNAs) were designed using CRISPR RGEN Tools (Park et al., 2015). We used the Basic Local Alignment Search Tool (Altschul et al., 1990) to cross-check that the selected spacer sequences would lead to unique cuts in the genome, thus minimising off-target cleavages. Inverse PCR (iPCR) was employed to clone the unique sgRNA for each target gene into the backbone plasmid to create intermediate plasmids: pGFT021 for LABRINI_07945 (*hp*) and pGFT022 for LABRINI_07950 (*nbp*). The two HAs were obtained by PCR amplification of the 1 kb regions upstream and downstream of each target gene from the *C. autoethanogenum* genome using various primer pairs (Supplementary Table S1). The derived 5′ and 3′ HA fragments were fused using overlap extension PCR as described by Bryksin and Matsumura (2013) and moved into the appropriate sgRNA-containing vectors using the same technique or via restriction cloning. The resulting plasmids containing all necessary parts for CRISPR/nCas9-mediated gene deletion were named pGFT036 for *hp* and pGFT048 for *nbp* KO. See Supplementary File S1 (pGFT036) and Supplementary File S2 (pGFT048) for plasmid maps. Plasmids constructed in this study are listed in Supplementary Table S2.

2.2.3 Plasmid transformation into *C. autoethanogenum* using electroporation

Before electroporation, plasmids were propagated and isolated from the *E. coli* NEB Express strain to lose the Dcm methylation pattern, which has been shown to be targeted by the native type IV restriction system of *C. autoethanogenum* (Woods et al., 2019). Preparation of electrocompetent *C. autoethanogenum* cells and transformation using electroporation was performed as previously described (Leang et al., 2013). Electroporated cells were recovered in 10 mL YTF medium for 18–24 h at 37°C with 120 RPM of agitation. Recovery cultures were mixed with 1.5% molten agar YTF supplemented with clarithromycin (4 μg/mL). Transformation plates were incubated inside Oxoid™ AnaeroJars™ (Thermo Fischer Scientific) at 37°C until colonies became visible (in less than 10 days).

2.2.4 Transformant colony screening and plasmid curing

Colonies were first inoculated into liquid YTF medium containing clarithromycin and then plated on selective YTF agar to isolate individual transformant colonies. Next, transformant colonies were randomly chosen for colony PCR screening to verify *C. autoethanogenum* (primers *acs*B-F and -R), plasmid presence (primers oriT-F and -R), and successful gene deletion. Primers *hp*-F, *hp*-R and, *nbp*-F, *nbp*-R were used to confirm deletions of *hp* and *nbp* genes, respectively. PCR products obtained with the gene deletion control primers were purified and verified by Sanger sequencing.

Plasmid-carrying *C. autoethanogenum* colonies with target gene deletions were then subcultured three or more times in 10 mL YTF medium without the antibiotic before plating on non-selective YTF agar to isolate individual colonies. Next, colonies were randomly chosen and patch-plated on both YTF agar without and with clarithromycin (4 μg/mL). Plasmid loss in colonies that failed to grow in the presence of clarithromycin was confirmed by colony PCR using primers oriT-F and oriT-R as above. In total, 10 out of 10 colonies for *hp* gene deletion and 1 out of 30 colonies for *nbp* gene deletion had lost their plasmid. The plasmid-cured colonies were then inoculated into liquid YTF medium for outgrowth and preparation of glycerol stocks, stored as Δ*hp* and Δ*nbp*.

2.3 Analytical methods

2.3.1 Biomass concentration analysis

Biomass concentration (gDCW/L) was estimated by measuring culture OD at 600 nm and using the correlation coefficient of 0.23 between OD and DCW using the methodology established before (Peebo et al., 2014).

2.3.2 Extracellular metabolome analysis

Extracellular metabolome analysis from filtered broth samples was performed as described before (Valgepea et al., 2017b). We note that cells produced 2R,3R-butanediol.

2.3.3 Bioreactor off-gas analysis

Bioreactor off-gas analysis for the determination of specific gas uptake (CO and H_2) and production rates (CO_2, ethanol) (mmol/gDCW/h) have been described before (Valgepea et al., 2017a). The Faraday Cup detector monitored the intensities of H_2, CO, ethanol, H_2S, Ar, and CO_2 at 2, 14, 31, 34, 40, and 44 amu, respectively.

2.3.4 Carbon balance analysis

Carbon recoveries and balances were determined as described before (Valgepea et al., 2017a). Briefly, carbon balancing of substrate carbon (CO, cysteine) between growth products (acetate, ethanol, 2,3-BDO, CO_2, biomass) was calculated as C-mol fraction of each product from total C-mol carbon uptake. Carbon recoveries were calculated as the fraction of summed C-mol products from total C-mol substrates. Ethanol stripping and the total soluble CO_2 fraction in culture broth were also considered to achieve more accurate carbon balancing.

2.4 Proteome analysis

2.4.1 Sample preparation

Proteome analysis was carried out for four biological replicate cultures for each of the three strains: LAbrini, Δ*hp*, and Δ*nbp*. *C. autoethanogenum* chemostat cultures were sampled for proteomics by immediately pelleting 2 mL of the culture by centrifugation (20,000 × g for 1 min at 4°C) and stored at −80°C until analysis. Cell pellets were suspended with 10 volumes (relative to pellet volume) of chaotrope-based lysis buffer (6 M guanidine-HCl, 100 mM Tris-HCl pH 8.0, 50 mM dithiothreitol), heated at 95°C for 10 min and sonicated with the Bioruptor sonicator (Diagenode) for 15 cycles (30 s ON, 60 s OFF; "High" setting) at 4°C. Next, 0.5 volumes (relative to lysate volume) of Silibeads TypZY-s 0.4–0.6 mm (Sigmund Lindner) beads were added to the lysate and bead beating was carried out with the FastPrep24 device (MP Biomedicals) by 2 × 40 s at 6 m/s. Next, samples were centrifuged at 17,000 × g for 5 min, and the supernatant was separated from the beads and transferred to a new tube. A small aliquot of the lysate was then precipitated with trichloroacetic acid-deoxycholate (TCA-DOC) precipitation and the protein concentration was determined with the Micro BCA assay (Thermo Fisher Scientific). Next, 20 µg of lysate protein was alkylated with 100 mM chloroacetamide by incubating for 1 h in the dark at room temperature. Samples were then processed, and proteins were digested to peptides by the SP3 protocol as described elsewhere (Hughes et al., 2018).

2.4.2 LC-MS/MS analysis

Sample order for LC-MS/MS analysis was randomised to avoid batch effects and blanks were flanked to minimise carry-over. 1 µg of peptides for each sample was injected to an Ultimate 3500 RSLCnano system (Dionex) using a 0.3 × 5 mm trap-column (5 µm C18 particles, Dionex) and an in-house packed (3 µm C18 particles, Dr Maisch) analytical 50 cm × 75 µm emitter-column (New Objective). Columns were operated at 45°C. Peptides were eluted at 300 nL/min with an 8%–45% B 90 min gradient (buffer B: 80% acetonitrile + 0.1% formic acid, buffer A: 0.1% formic acid) to a Q Exactive HF (Thermo Fisher Scientific) mass spectrometer (MS) using a nano-electrospray source (spray voltage of 2.5 kV in positive mode). The MS was operated using a data-independent acquisition (DIA) approach (Gillet et al., 2012) with variable isolation windows over a range of 400–1,100 m/z. Briefly, one full range 400–1,100 m/z MS1 scan was collected at a resolution setting of 60,000 (max ion injection time of 60 ms, max of 3e6 ions), followed by 25 overlappings (overlap of 1 m/z) DIA isolation windows as follows (start–end m/z, isolation width): 399–472, 25; 471–510, 20; 509–608, 15; 607–703, 20; 702–790, 30; 789–868, 40; 867–966, 50; and 965–1101, 136. Each DIA scan was collected at a resolution setting of 30,000 (max ion injection time of 41 ms, max of 3e6 ions), the default charge state was set to +3 and normalised collision energy to 27. The overall method cycle time was approximately 2.3 s.

2.4.3 DIA MS data analysis and differential protein expression analysis

DIA MS data was analysed using version 1.8 of the software suite DIA-NN (Demichev et al., 2019) with default settings. The spectral library was generated *in silico* using smart profiling from the protein sequence database of *C. autoethanogenum* LAbrini (NCBI Genbank CP110420) (Ingelman et al., 2023) with a precursor of 1% false discovery rate (FDR) filter. The following settings were used for analysis: full trypsin specificity with 1 missed cleavage allowed for peptides with a length of 7–30 AAs; fixed modifications of cysteine carbamidomethylation and methionine N-terminal excision; precursor charge range 1–4 and m/z range 400–1,100; fragment m/z range 200–1,800; double-pass mode for neural network classifier; Robust LC (high precision) for quantification strategy; and RT-dependent cross-run normalization. Label-free protein quantification and normalisation were performed within the DIA-NN workflow using the MaxLFQ method (Cox et al., 2014). We confidently quantitated 28,752 peptides and 2,127 proteins across all samples, and 27,358 peptides and 2,107 proteins on average within each sample after the removal of shared peptides from the analysis.

Protein expression fold-changes (FC) with q-values between Δ*hp* and LAbrini, and Δ*nbp* and LAbrini were determined using the software Perseus (Tyanova et al., 2016) with Student's T-test. Only proteins with at least two peptides in all 12 samples (1,835) analysed were used to ensure higher quantification accuracy. Proteins were considered to be differentially expressed by an FC > 1.5 and a q-value < 0.05 after FDR correction (Benjamini and Hochberg, 1995). Differentially expressed proteins are presented in Supplementary Table S3. Proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Perez-Riverol et al., 2022) with the data set identifier PXD040151.

3 Results and discussion

3.1 Single-gene deletion of *hp* and *nbp* in *C. autoethanogenum* using CRISPR/nCas9 knock-out plasmids

To study the role of two genes adjacent to the C_1-fixing gene cluster (Figure 1) with unclear functions, LABRINI_07945 (*hp*) and LABRINI_07950 (*nbp*), each gene was targeted for deletion using CRISPR/nCas9. Plasmids pGFT036 and pGFT048 carrying required components (Figure 2A) for *hp* and *nbp* deletion, respectively, were transferred into *C. autoethanogenum* by electroporation that yields transformation efficiencies up to 200 CFU/µg DNA in our laboratory. We obtained 40 and 2 colonies for *hp* and *nbp* deletion, respectively, and colony PCR screening of 6 colonies for *hp* and 2 for *nbp* deletion confirmed *C. autoethanogenum* genotype and the presence of our relatively large plasmid (~12.5 kb) in all colonies. The gene deletion control PCR showed that all the *hp* colonies screened had a mixed base strain/knock-out (KO) genotype evidenced by the presence of both the 2444-bp fragment of the base strain allele and the predicted 2103-bp fragment of the KO allele (Figure 2B). Six rounds of liquid selective sub-culturing of the mixed colonies was needed to obtain 1 colony (from 20 screened) with a clean *hp* KO genotype (Figure 2C). In contrast, the 2 *nbp* colonies (nbp-C_1 and nbp-C2) obtained showed a 1444-bp fragment of the desired KO allele, i.e., a clean KO genotype (Figure 2B). Sanger

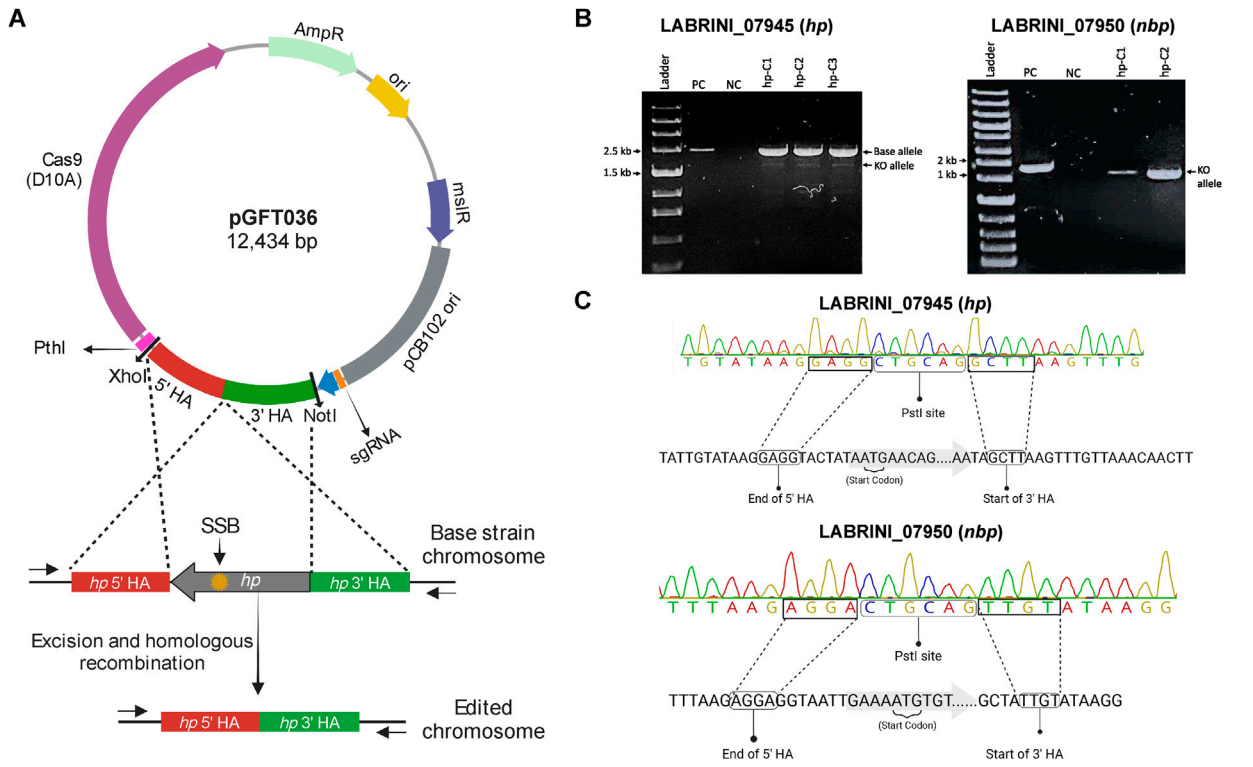


FIGURE 2
Successful deletion of target genes using CRISPR/nCas9 KO plasmids in *C. autoethanogenum*. **(A)** Representative plasmid map and an overview of CRISPR/nCas9 mediated gene deletion mechanism. The yellow star indicates the genomic recognition site of the Cas9-sgRNA complex. **(B)** PCR Screening of transformant colonies using control primers to amplify gene regions. **(C)** Confirmation of *hp* (LABRINI_07945) and *nbp* (LABRINI_07950) deletion via Sanger Sequencing. PC, positive control (non-transformed *C. autoethanogenum*); NC, negative control (no template); LAbrini, *C. autoethanogenum* base strain allele; KO, knock-out/deletion allele; SSB, single-strand break. Figure partly created using BioRender.com.

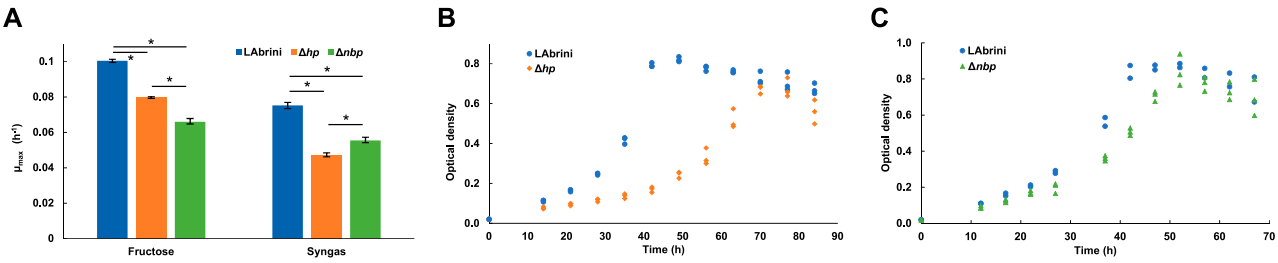


FIGURE 3
Batch growth of Δ*hp*, Δ*nbp*, and LAbrini strains on PETC-MES with YE. **(A)** Maximum specific growth rate (μ_{max}) in heterotrophic (fructose) and autotrophic (syngas) growth conditions. Data are average ± standard deviation between three biological replicates. Asterisk denotes values statistically different according to *t*-test (*p*-value ≤ 0.01). **(B, C)** Autotrophic growth profiles on syngas. LAbrini, *C. autoethanogenum* base strain; YE, yeast extract.

sequencing confirmed both *hp* and *nbp* gene deletions in selected colonies and the presence of a PstI site, proving that the provided plasmid-based DNA repair template, in the form of the fused 1 kb HA interspaced with the PstI site, was used during homologous recombination to repair the nick from nCas9 (Figure 2C). Successful deletion of both genes was also confirmed by proteome analysis (see below). Finally, the obtained Δ*hp* and Δ*nbp* cells were cured of the plasmid before proceeding with phenotypic characterisation.

3.2 Deletion of *hp* and *nbp* genes confer growth defects in rich-medium batch cultures

Next, we grew Δ*hp* and Δ*nbp* strains in batch cultures on PETC-MES medium with YE (i.e., rich medium) in heterotrophic (fructose) and autotrophic (syngas: $CO+H_2+CO_2$) conditions to investigate potential phenotypic effects of *hp* and *nbp* deletions. Interestingly, the maximum specific growth rate (μ_{max}) of both mutant strains was significantly

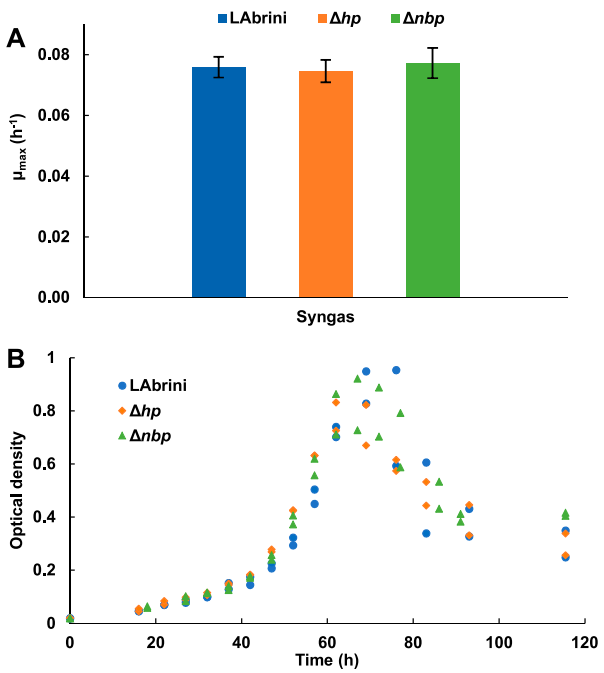


FIGURE 4
Autotrophic batch growth of Δ*hp*, Δ*nbp*, and LAbrini strains in PETC-MES without YE. **(A)** Maximum specific growth rates (μ_{max}) on syngas. Data are average ± standard deviation between two biological replicates. **(B)** Growth profiles on syngas. LAbrini, *C. autoethanogenum* base strain; YE, yeast extract.

impaired compared to the *C. autoethanogenum* base strain "LAbrini" on both fructose and syngas (Figure 3A). The μ_{max} for the Δ*hp* strain was 0.08 ± 0.001 (average ± standard deviation) and 0.05 ± 0.001 on fructose and syngas, respectively, while for the Δ*nbp* strain, the μ_{max} was 0.07 ± 0.002 on fructose and 0.06 ± 0.001 on syngas (Figure 3A). The LAbrini base strain showed μ_{max} values of 0.10 ± 0.001 on fructose and 0.08 ± 0.001 on syngas. Thus, μ_{max} for deletion strains were substantially lower compared to base strain: for Δ*hp*, 21% and 37% lower on fructose and syngas and for Δ*nbp*, 34% and 26% lower on fructose and syngas, respectively (Figure 3A). Interestingly, Δ*hp* had a more pronounced growth defect on syngas, while Δ*nbp* was affected more on fructose. The μ_{max} values for both growth conditions were also statistically different between deletion strains. In addition to the lower μ_{max}, Δ*hp* strain exhibited a longer lag phase and slightly lower peak OD on syngas compared to the base strain LAbrini (Figure 3B). The growth profile and peak OD of the Δ*nbp* strain, however, were similar to LAbrini despite the lower μ_{max} values (Figure 3C). Altogether, our results demonstrate that the deletion of either the *hp* or the *nbp* gene confers significant growth defects for hetero- and autotrophic rich medium batch cultures, suggesting a notable role for both genes in the physiology of *C. autoethanogenum*.

3.3 Deletion of *hp* and *nbp* show no effects on growth in minimal medium autotrophic batch cultures

We also investigated the growth of all three strains in PETC-MES medium without YE to assess the effects of *hp* and *nbp* deletions on autotrophic growth on syngas in minimal medium batch cultures. Intriguingly, there was no difference in the μ_{max} values between the three strains: 0.08 ± 0.004, 0.08 ± 0.005, and 0.08 ± 0.003 for the Δ*hp*, Δ*nbp*, and LAbrini strains, respectively (Figure 4). Although one would expect slower growth in minimal medium, Δ*hp* and Δ*nbp* strains grew faster in minimal medium compared to rich medium, whereas LAbrini showed similarly high μ_{max} values for autotrophic growth of acetogens on both media. Like the μ_{max} data, the three strains did not show significant variations in peak OD and growth curves (i.e., lag phase and entry into stationary phase) (Figure 4B). Thus, both genes seem to play a role in *C. autoethanogenum* for utilisation of complex nutrients due to significant growth defects in rich medium with no phenotypic effects in minimal medium for Δ*hp* and Δ*nbp* strains.

3.4 Deletion of *hp* and *nbp* alter growth by-product yields in autotrophic batch cultures

We determined the production yields (mmol/gDCW) of *C. autoethanogenum* growth by-products — ethanol, acetate, and 2,3-butanediol (2,3-BDO) — during the above-described autotrophic batch cultures to check if these are affected by the deletion of either the *hp* or the *nbp* gene. Remarkably, the Δ*hp* strain showed a >5-fold higher acetate yield (550 ± 45 mmol/gDCW) relative to LAbrini (92 ± 7 mmol/gDCW) during growth in rich medium (Figure 5A). The production of 2,3-BDO was also slightly higher for Δ*hp* compared to LAbrini (10 ± 1 vs. 7 ± 1 mmol/gDCW).

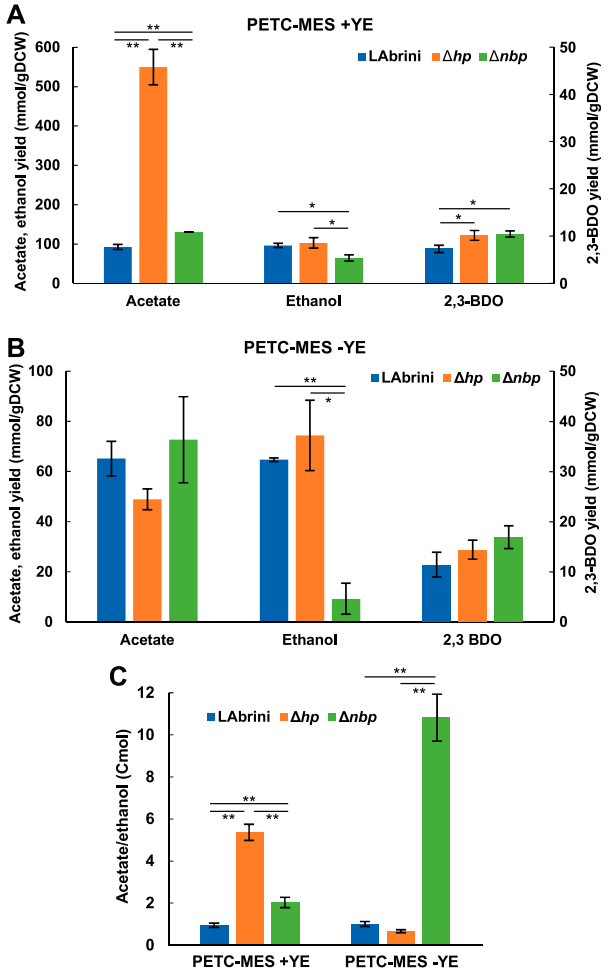


FIGURE 5
Growth by-product yields and acetate-to-ethanol ratios of Δ*hp*, Δ*nbp*, and LAbrini strains in autotrophic batch cultures with and without YE. **(A, B)** Growth by-product yields. **(C)** Acetate/ethanol (Cmol). Data are average ± standard deviation between three biological replicates on PETC-MES + YE or two biological replicates on PETC-MES -YE. Asterisk denotes values statistically different according to *t*-test: *, *p*-value < 0.05; **, *p*-value < 0.01. LAbrini, *C. autoethanogenum* base strain; YE, yeast extract; gDCW, gram of dry cell weight; 2,3-BDO, 2,3-butanediol.

While the ethanol yield was not different for Δ*hp* when compared to LAbrini (96 ± 5 vs. 102 ± 13 mmol/gDCW), ethanol production did not coincide with acetate production in the Δ*hp* strain (data not shown). Namely, Δ*hp* was producing acetate in the early exponential growth phase and started producing ethanol only in the late exponential phase.

Also, the Δ*nbp* strain showed significantly different product yields compared to LAbrini during growth in rich medium: 41% higher acetate (130 ± 0.2 vs. 92 ± 7 mmol/gDCW), 44% higher 2,3-BDO (10 ± 1 vs. 7 ± 1 mmol/gDCW), and 33% lower ethanol compared to LAbrini (65 ± 8 vs. 96 ± 6 to mmol/gDCW) (Figure 5A). In addition, both acetate and ethanol production yields were different between the two deletion strains. The substantially higher acetate production yield of the Δ*hp* strain might indicate that deletion of the *hp* gene improves the uptake of YE components and their catabolism to acetate. However, this did not lead to faster growth, thus, acetate production was potentially not linked to ATP production.

We observed no difference in the product yields between the Δ*hp* strain and LAbrini on minimal medium (Figure 5B), which is not surprising as growth was also comparable (Figure 4). However, despite similar growth for Δ*nbp* and LAbrini strains on minimal medium (Figure 4), the Δ*nbp* strain showed an 86% lower ethanol yield than LAbrini (9 ± 6 vs. 65 ± 1 to mmol/gDCW) while acetate and 2,3-BDO yields were similar (Figure 5B). Only the ethanol production yield was different between the deletion strains on minimal medium. Interestingly, while LAbrini maintained the acetate-to-ethanol ratio at ~1 on both media, the ratio was higher for the Δ*hp* strain (~5 vs. ~1) and lower for the Δ*nbp* strain (~2 vs. ~11) on rich medium compared to minimal medium during autotrophic growth (Figure 5C). These ratios are in the range of those previously seen in autotrophic batch and continuous bioreactor cultures of *C. autoethanogenum* (Valgepea et al., 2017a; Valgepea et al., 2018; Ingelman et al., 2023). Our results imply that both *hp* and *nbp* genes play a notable role in carbon distribution between the two major by-products — acetate and ethanol — during autotrophic batch growth of *C. autoethanogenum*.

3.5 Deletion of *nbp* does not affect autotrophic growth in nickel-supplemented minimal medium batch cultures

Since Nbp is annotated as a nickel binding protein and deletion of *nbp* strongly affected growth in rich medium cultures (Figures 3, 5), we hypothesised that the effects could arise from a potentially lower capacity of the Δ*nbp* strain to catabolise the higher nickel levels in rich medium, i.e., YE contains nickel. We thus tested the effect of nickel supplementation on autotrophic batch growth of Δ*nbp* and LAbrini strains in PETC-MES without YE. We added 0.6 mg/L $NiCl_2{\cdot}6H_2O$ to the 0.2 mg/L in PETC-MES (0.8 mg/L final concentration) to facilitate potential effects on growth as a previous study showed variable influence of nickel in the range of 0.1–0.4 mg/L ($NiCl_2{\cdot}6H_2O$) on heterotrophic growth of *C. ljungdahlii* (Mann et al., 2022).

We observed no difference between μ_{max} values of Δ*nbp* and LAbrini strains growing on syngas and in PETC-MES medium without YE with supplemented nickel (Figure 6A). These values were not statistically significantly (*p*-value > 0.05) different from respective values during growth without additional nickel (Figure 4A). Likewise, by-product yields between the two strains with nickel supplementation did not differ (Figure 6B). Both acetate and ethanol production yields with nickel addition did, however, differ from yields observed during growth without extra nickel (Figure 5B). Effect of nickel levels on acetate and ethanol production is also seen during heterotrophic growth of *C. ljungdahlii* (Mann et al., 2022). Our results showed that deletion of *nbp* did not affect autotrophic growth in nickel-supplemented minimal medium batch cultures. As several enzymes in the WLP are dependent on nickel, the potential role of *nbp* in *C. autoethanogenum* nickel metabolism remains to be further elucidated.

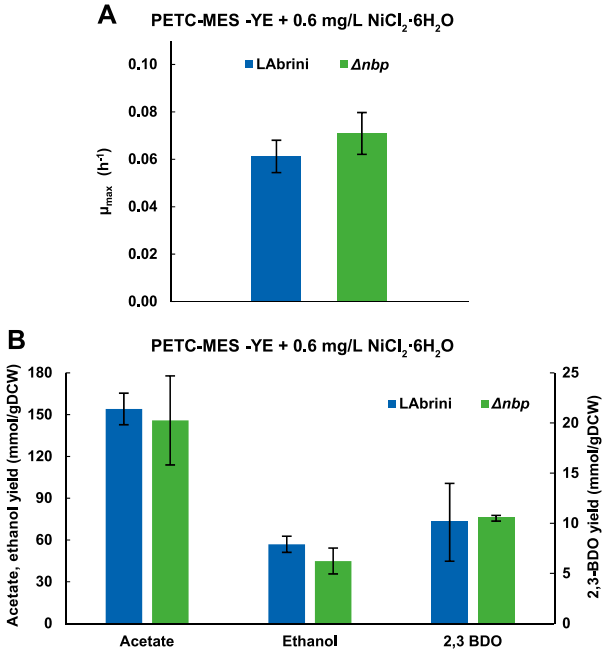


FIGURE 6
Effect of nickel supplementation on autotrophic (syngas) batch growth and growth by-product yields of Δ*nbp* and LAbrini strains in PETC-MES without YE. **(A)** Maximum specific growth rates (μ_{max}). **(B)** Growth by-product yields. Data are average ± standard deviation between three biological replicates. LAbrini, *C. autoethanogenum* base strain; YE, yeast extract; gDCW, gram of dry cell weight; 2,3-BDO, 2,3-butanediol.

3.6 Deletion of *hp* and *nbp* affect growth characteristics of steady-state chemostat cultures

We also compared the performance of the mutant strains to the LAbrini base strain in steady-state bioreactor continuous cultures as these yield high-quality physiological data (Adamberg et al., 2015; Valgepea et al., 2018) and are also the industrially relevant fermentation mode for acetogens (Köpke and Simpson, 2020). Cells were grown in minimal medium in syngas chemostats at a dilution rate of 1 day-1 and steady-state data at biomass concentrations ~1.5 gDCW/L were collected for by-product production (ethanol, acetate, 2,3-BDO) and gas uptake and production (CO, H_2, CO_2) analysis.

While acetate production in minimal medium batch cultures was similar across all three strains, both Δ*hp* and Δ*nbp* strains showed lower specific acetate production rates (q_{ace}; mmol/gDCW/day) in chemostats (Figure 7A). Specific ethanol production rates (q_{EtOH}) were also lower for both deletion strains compared to LAbrini (Figure 7A), whereas ethanol production differed only for the Δ*nbp* strain in batch cultures (Figure 5). Production of 2,3-BDO ($q_{2,3\text{-BDO}}$) was similar across the three strains in chemostats (Figure 7A). These steady-state data are consistent with above-described batch data by showing that both *hp* and *nbp* genes affect carbon distribution during autotrophic growth of the model-acetogen *C. autoethanogenum*. Still, differences in bioreactor chemostat and bottle batch data demonstrate the importance of testing strains in the industrially relevant fermentation mode for acetogens (Köpke and Simpson, 2020).

Bioreactor off-gas analysis revealed that the specific CO uptake rates (q_{CO}; mmol/gDCW/day) were slightly lower for both deletion strains compared to LAbrini (873 ± 19): by 8% for the Δ*hp* strain (805 ± 19) and by 5% for the Δ*nbp* strain (830 ± 15) (Figure 7B). At the same time, we detected no differences in the specific H_2 uptake rates (q_{H2}; mmol/gDCW/day) and CO_2 production rates (q_{CO2}; mmol/gDCW/day) among the three strains (Figure 7B). The slightly higher q_{CO2}/q_{CO} and q_{H2}/q_{CO} for both deletion strains indicates that cells needed to increase the generation of reducing power in the form of reduced ferredoxin as a response to gene deletion since both CO and H_2 oxidation supply reduced ferredoxin (Bertsch and Müller, 2015; Pavan et al., 2022). Quantification of carbon flows from substrates to products showed that both Δ*hp* and Δ*nbp* strains diverted more carbon to biomass and CO_2 with a concomitant lower carbon flux to ethanol and acetate (Figure 7C). The slightly lower acetate-to-ethanol ratios for Δ*hp* (1.2) and Δ*nbp* (1.1) compared to LAbrini (1.3) are consistent with higher q_{CO2}/q_{CO} for deletion strains since ethanol production generally results in a greater loss of carbon from CO as CO_2 (Bertsch and Müller, 2015).

3.7 Proteome analysis reveals intriguing effects of deletion of *hp* and *nbp* genes

We used proteomics to determine the effects of *hp* and *nbp* gene deletions on protein expression in the above-described chemostat cultures of *C. autoethanogenum*. Our analysis quantified 93 differentially expressed proteins (fold-change > 1.5 and a q-value < 0.05) between Δ*hp* and LAbrini, and 77 between Δ*nbp*

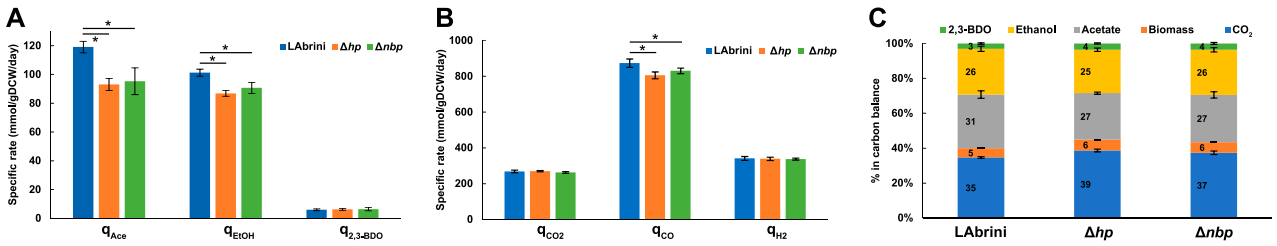


FIGURE 7
Steady-state growth characteristics of Δ*hp*, Δ*nbp*, and LAbrini strains in syngas-grown chemostats. **(A)** Specific by-product production rates (mmol/gDCW/day). **(B)** Specific gas uptake (CO, H_2) and production (CO_2) rates. Asterisk denotes values statistically different according to *t*-test (*p*-value < 0.05). **(C)** Carbon balances. Carbon recoveries were normalised to 100% to have a fair comparison of carbon distributions between different strains. Data are average ± standard deviation between four biological replicates. LAbrini, *C. autoethanogenum* base strain; EtOH, ethanol; 2,3-BDO, 2,3-butanediol; Ace, acetate; gDCW, gram of dry cell weight; q, specific rate.

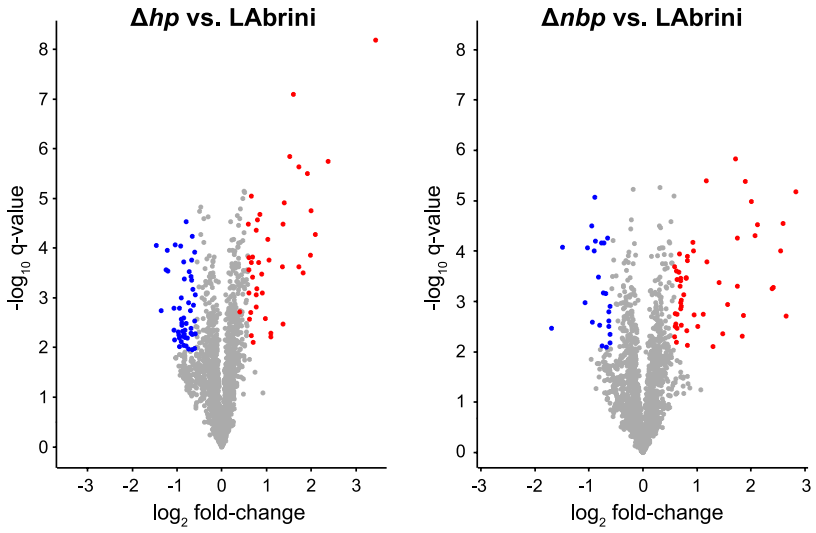


FIGURE 8
Volcano plots showing differentially expressed proteins in Δ*hp* and Δ*nbp* strains compared to LAbrini. Differentially expressed proteins are with fold-change > 1.5 and a q-value < 0.05. Blue markers, downregulated proteins; red markers, upregulated proteins; grey, proteins not expressed differentially. See Supplementary Table S3 for data.

and LAbrini (Figure 8; Supplementary Table S3). Proteomics also confirmed the deletion of each gene in respective strains as no peptides for either protein in corresponding strains were detected. Additionally, the data showed no effect of deletion of either gene on expression of the protein encoded by the other gene, i.e., *hp* deletion on Nbp expression or *nbp* deletion on Hp expression. Intriguingly, deletion of either gene increased the expression of the methyltransferase (AcsE; 07970) and a component of the correnoid iron sulphur protein (AcsD; 07980) (Table 1) that are linked to a critical activity of the WLP: synthesis of acetyl-CoA. AcsE and AcsD supply the methyl group to the carbon monoxide dehydrogenase (AcsA; 08025) and acetyl-CoA synthase (AcsB; 07965) CODH/ACS enzyme complex (Lemaire and Wagner, 2021), that is essential for autotrophic growth of *C. autoethanogenum* (Liew et al., 2016). At the same time, the expression of AcsB was repressed in both deletion strains (Table 1). These are all among the top 10 most abundant proteins for autotrophic growth of *C. autoethanogenum*, except AcsA (Valgepea et al., 2022). These results show that the two genes adjacent to the C_1-fixing cluster studied here—*hp* and *nbp*—affect the expression of key proteins for C_1-fixation. Notably, the effect of *hp* and *nbp* on C_1-fixation extends beyond the WLP as the expression of the most abundant CODH in *C. autoethanogenum* (Valgepea et al., 2022)—CooS1; 15015—was repressed in both mutant strains (Table 1). Further work is needed to understand the interplay between the *hp*, *nbp,* and the affected enzymes with a central role in the autotrophic growth of *C. autoethanogenum*.

Deletion of our target genes also affected the expression of proteins involved in the synthesis of the most attractive native product of *C. autoethanogenum*: ethanol. In contrast to the lowered ethanol production in the deletion strains, proteome analysis revealed that

TABLE 1 Key protein expression changes in Δ*hp* and Δ*nbp* strains compared to LAbrini.

Protein ID	Description of protein product	Δ*hp* vs. LAbrini			Δ*nbp* vs. LAbrini		
		Log_2 FC	FC	q-value	Log_2 FC	FC	q-value
C_1-fixation							
LABRINI_07970	Methyltransferase for CODH	0.61	1.5	0.01	0.63	1.6	0.03
LABRINI_07980	Correnoid iron sulphur protein part2	0.58	1.5	0.01	0.63	1.5	0.02
LABRINI_07965	Acetyl-CoA synthase (ACS)	−0.60	−1.5	< 0.01	−0.88	−1.8	< 0.01
LABRINI_15015	Carbon-monoxide dehydrogenase CooS1	−0.55	−1.5	0.02	−0.83	−1.8	0.01
Ethanol synthesis							
LABRINI_09125	Iron-containing alcohol dehydrogenase	1.71	3.3	0.01	1.75	3.4	0.01
LABRINI_19700	Iron-containing alcohol dehydrogenase	0.82	1.8	0.01	0.93	1.9	0.01
LABRINI_18005	Iron-containing alcohol dehydrogenase	0.61	1.5	0.01	0.82	1.8	0.01
LABRINI_00495	Aldehyde ferredoxin oxidoreductase	1.10	2.1	0.03	1.47	2.8	0.03
LABRINI_00445	Aldehyde ferredoxin oxidoreductase	0.61	1.5	0.02	0.02	1.6	0.02
LABRINI_18700	Bifunctional AdhE1	0.77	1.7	0.01	2.07	4.2	< 0.01

ID, identifier; FC, fold-change. Positive FC means upregulation, negative downregulation. See Supplementary Table S3 for all differentially expressed proteins.

expression of alcohol dehydrogenases (19700 and 18005) was elevated in both mutant strains, including the most abundant alcohol dehydrogenase in *C. autoethanogenum* (Valgepea et al., 2022): Adh4, 09125 (Table 1). Similarly, increased expression of both aldehyde ferredoxin oxidoreductases (AOR1, 00445 and AOR2, 00495), that couple ethanol synthesis to ATP production and the bifunctional acetaldehyde-CoA/alcohol dehydrogenase AdhE1 (18700) were measured in deletion strains. Such seeming discrepancies between either the transcripts or proteins linked to ethanol synthesis and the measured ethanol production have been detected in *C. autoethanogenum* before (Valgepea et al., 2017a; Valgepea et al., 2018). Thus, more systemic and in-depth investigations are required to decipher the genotype-phenotype relationships for ethanol synthesis in acetogens.

Among other differentially expressed proteins, it was unexpected to see repression of 13 ribosomal proteins in the Δ*hp* strain (1.7- to 2.3-fold, $q < 0.05$; Supplementary Table S3). In addition, several oxidoreductases (05235, 14800, and 18000; ~2-fold, $q < 0.03$) and sigma-54-related transcriptional regulators (14025 and 00450; ~3.2- and 1.6-fold, $q < 0.05$) were upregulated in the deletion strains. Also noteworthy is the ~1.8-fold ($q < 0.01$) upregulation of a PocR ligand-binding domain-containing protein (16380) in both Δ*hp* and Δ*nbp* strains, as we recently identified a missense mutation in the PocR gene of a CO-evolved *C. autoethanogenum* strain obtained via adaptive laboratory evolution (Ingelman et al., 2023). Moreover, PocR has an alternative annotation in *C. autoethanogenum* as a histidine kinase (Humphreys et al., 2015) that regulates sporulation and metabolism in Clostridia (Xin et al., 2020).

4 Conclusion

In summary, our study shows that CRISPR/nCas9 can effectively be used for targeted gene deletion in the industrially relevant acetogen *C. autoethanogenum* and provides valuable insights to accelerate the rational genetic engineering of acetogen cell factories. Additionally, the characterisation of the constructed deletion strains offers valuable clues to the *in vivo* functionalities of the *hp* and *nbp* genes that seem to influence both growth and product patterns of *C. autoethanogenum* during autotrophic growth. Furthermore, our proteomics results shed light on enzymes that potentially play a key role in ethanol production. Future work will need to confirm the biochemical function of closely linked genes in the WLP and identify possible targets for improving carbon fixation efficiencies.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

Conceptualization: UJN, KR, and KV; Methodology: UJN, KR, KS, AH, MK, and KV; Formal analysis: UJN, KR, LD, PP, and KV; Investigation: UJN, KR, LD, PP, and KV; Resources: KV; Writing—Original Draft: UJN, KR, and KV; Writing—Review and Editing: UJN, KR, LD, PP, KS, AH, MK, and KV; Supervision: KV; Project Administration: KV; Funding Acquisition: KV. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This work was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement N810755 and the Estonian Research Council's grant agreement PSG289.

Acknowledgments

We thank Bastian Molitor and Peng-Fei Xia for providing us the pFX01 plasmid.

Conflict of interest

LanzaTech has interest in commercial gas fermentation with *C. autoethanogenum*. AH and MK are employees of LanzaTech.

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.2023.1167892/full#supplementary-material

References

Adamberg, K., Valgepea, K., and Vilu, R. (2015). Advanced continuous cultivation methods for systems microbiology. *Microbiol* 161, 1707–1719. doi:10.1099/mic.0.000146

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bertsch, J., and Müller, V. (2015). Bioenergetic constraints for conversion of syngas to biofuels in acetogenic bacteria. *Biotechnol. Biofuels* 8, 210. doi:10.1186/s13068-015-0393-x

Bourgade, B., Minton, N. P., and Islam, M. A. (2021). Genetic and metabolic engineering challenges of C_1-gas fermenting acetogenic chassis organisms. *FEMS Microbiol. Rev.* 45, fuab008. fuab008. doi:10.1093/femsre/fuab008

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

Bryksin, A., and Matsumura, I. (2013). "Overlap extension PCR cloning," in *Methods in molecular Biology*. Editors K. Polizzi and C. Kontoravdi (Totowa, NJ: Humana Press), 31–42. doi:10.1007/978-1-62703-625-2_4/COVER

Can, M., Armstrong, F. A., and Ragsdale, S. W. (2014). Structure, function, and mechanism of the nickel metalloenzymes. *Chem. Rev.* 8, 4149. doi:10.1021/cr400461p

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* 13, 2513–2526. doi:10.1074/MCP.M113.031591

Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S., and Ralser, M. (2019). DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* 171, 41–44. doi:10.1038/s41592-019-0638-x

Drake, H. L., Küsel, K., and Matthies, C. (2006). "Acetogenic Prokaryotes," in *Prokaryotes* (Berlin, Germany: Springer), 354–420.

Fackler, N., Heijstra, B. D., Rasor, B. J., Brown, H., Martin, J., Ni, Z., et al. (2021). Stepping on the gas to a circular economy: Accelerating development of carbon-negative chemical production from gas fermentation. *Annu. Rev. Chem. Biomol. Eng.* 12, 439–470. doi:10.1146/annurev-chembioeng-120120-021122

Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., et al. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* 11, O111.016717. doi:10.1074/mcp.O111.016717

Hughes, C. S., Moggridge, S., Müller, T., Sorensen, P. H., Morin, G. B., and Krijgsveld, J. (2018). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* 141, 68–85. doi:10.1038/s41596-018-0082-x

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

Ingelman, H., Heffernan, J. K., Harris, A., Brown, S. D., Yar Saqib, A., Pinheiro, M. J., et al. (2023). Autotrophic adaptive laboratory evolution of the acetogen Clostridium autoethanogenum delivers the gas-fermenting strain LAbrini with superior growth, products, and robustness. bioRxiv. doi:10.1101/2023.01.28.526018

IPCC (2021). Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment Report of the intergovernmental panel on climate change. Available at: https://report.ipcc.ch/ar6/wg1/IPCC_AR6_WGI_FullReport.pdf.

Jin, S., Bae, J., Song, Y., Pearcy, N., Shin, J., Kang, S., et al. (2020). Synthetic Biology on acetogenic bacteria for highly efficient conversion of C_1 gases to biochemicals. *Int. J. Mol. Sci.* 21, 7639. doi:10.3390/ijms21207639

Köpke, M., and Simpson, S. D. (2020). Pollution to products: Recycling of 'above ground' carbon by gas fermentation. *Curr. Opin. Biotechnol.* 65, 180–189. doi:10.1016/j.copbio.2020.02.017

Leang, C., Ueki, T., Nevin, K. P., and Lovley, D. R. (2013). A genetic system for Clostridium ljungdahlii: A chassis for autotrophic production of biocommodities and a model homoacetogen. *Appl. Environ. Microbiol.* 79, 1102–1109. doi:10.1128/AEM.02891-12

Lemaire, O. N., and Wagner, T. (2021). Gas channel rerouting in a primordial enzyme: Structural insights of the carbon-monoxide dehydrogenase/acetyl-CoA synthase complex from the acetogen Clostridium autoethanogenum. *Biochim. Biophys. Acta - Bioenerg.* 1862, 148330. doi:10.1016/j.bbabio.2020.148330

Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., et al. (2016). CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnol. J.* 11, 961–972. doi:10.1002/BIOT.201600053

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

Liew, F., Henstra, A. M., Köpke, M., Winzer, K., Simpson, S. D., and Minton, N. P. (2017). Metabolic engineering of Clostridium autoethanogenum for selective alcohol production. *Metab. Eng.* 40, 104–114. doi:10.1016/j.ymben.2017.01.007

Liew, F., Henstra, A. M., Winzer, K., Köpke, M., Simpson, S. D., and Minton, N. P. (2016). Insights into CO2 fixation pathway of Clostridium autoethanogenum by targeted mutagenesis. *MBio* 7, 004277–e516. doi:10.1128/mBio.00427-16

Ljungdahl, L. G. (2009). A life with acetogens, thermophiles, and cellulolytic anaerobes. *Annu. Rev. Microbiol.* 63, 1–25. doi:10.1146/annurev.micro.091208.073617

Mann, M., Effert, D., Kottenhahn, P., Hüser, A., Philipps, G., Jennewein, S., et al. (2022). Impact of different trace elements on metabolic routes during heterotrophic growth of C. ljungdahlii investigated through online measurement of the carbon dioxide transfer rate. *Biotechnol. Prog.* 38, e3263. doi:10.1002/BTPR.3263

Park, J., Bae, S., and Kim, J. S. (2015). Cas-designer: A web-based tool for choice of CRISPR-cas9 target sites. *Bioinformatics* 31, 4014–4016. doi:10.1093/BIOINFORMATICS/BTV537

Pavan, M., Reinmets, K., Garg, S., Mueller, A. P., Marcellin, E., Köpke, M., et al. (2022). Advances in systems metabolic engineering of autotrophic carbon oxide-fixing biocatalysts towards a circular economy. *Metab. Eng.* 71, 117–141. doi:10.1016/j.ymben.2022.01.015

Peebo, K., Valgepea, K., Nahku, R., Riis, G., Õun, M., Adamberg, K., et al. (2014). Coordinated activation of PTA-ACS and TCA cycles strongly reduces overflow metabolism of acetate in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 98, 5131–5143. doi:10.1007/s00253-014-5613-y

Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., et al. (2022). The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* 50, D543–D552. doi:10.1093/NAR/GKAB1038

Ragsdale, S. W. (2008). Enzymology of the Wood-Ljungdahl pathway of acetogenesis. *Ann. N. Y. Acad. Sci.* 1125, 129–136. doi:10.1196/annals.1419.015

Ragsdale, S. W., and Pierce, E. (2008). Acetogenesis and the wood–ljungdahl pathway of CO2 fixation. *Biochim. Biophys. Acta* 1784, 1873–1898. doi:10.1016/j.bbapap.2008.08.012

Redl, S., Sukumara, S., Ploeger, T., Wu, L., Jensen, T. Ø., Nielsen, A. T., et al. (2017). Thermodynamics and economic feasibility of acetone production from syngas using the thermophilic production host Moorella thermoacetica. *Biotechnol. Biofuels* 10, 150. doi:10.1186/s13068-017-0827-8

Russell, M. J., and Martin, W. (2004). The rocky roots of the acetyl-CoA pathway. *Trends biochem. Sci.* 29, 358–363. doi:10.1016/j.tibs.2004.05.007

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

Valgepea, K., de Souza Pinto Lemgruber, R., Abdalla, T., Binos, S., Takemori, N., Takemori, A., et al. (2018). H2 drives metabolic rearrangements in gas-fermenting Clostridium autoethanogenum. *Biotechnol. Biofuels* 11, 55. doi:10.1186/s13068-018-1052-9

Valgepea, K., de Souza Pinto Lemgruber, R., Meaghan, K., Palfreyman, R. W., Abdalla, T., Heijstra, B. D., et al. (2017a). 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

Valgepea, K., Loi, K. Q., Behrendorff, J. B., Lemgruber, R. de S. P., Plan, M., Hodson, M. P., et al. (2017b). Arginine deiminase pathway provides ATP and boosts growth of the gas-fermenting acetogen Clostridium autoethanogenum. *Metab. Eng.* 41, 202–211. doi:10.1016/j.ymben.2017.04.007

Valgepea, K., Talbo, G., Takemori, N., Takemori, A., Ludwig, C., Mahamkali, V., et al. (2022). Absolute proteome quantification in the gas-fermenting acetogen Clostridium autoethanogenum. *mSystems* 7, e0002622–22. doi:10.1128/msystems.00026-22

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

Xia, P.-F., Casini, I., Schulz, S., Klask, C.-M., Angenent, L. T., and Molitor, B. (2020). Reprogramming acetogenic bacteria with CRISPR-targeted base editing via deamination. *ACS Synth. Biol.* 9, 2162–2171. doi:10.1021/acssynbio.0c00226

Xin, X., Cheng, C., Du, G., Chen, L., and Xue, C. (2020). Metabolic engineering of histidine kinases in Clostridium beijerinckii for enhanced butanol production. *Front. Bioeng. Biotechnol.* 8, 214. doi:10.3389/fbioe.2020.00214

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 23 June 2023
DOI 10.3389/fbioe.2023.1213236

OPEN ACCESS

EDITED BY
Michael Köpke,
LanzaTech, United States

REVIEWED BY
Byung-Kwan Cho,
Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea
Bastian Molitor,
University of Tübingen, Germany

*CORRESPONDENCE
Margaux Poulalier-Delavelle,
mbxmp6@exmail.nottingham.ac.uk
Nigel P. Minton,
nigel.minton@nottingham.ac.uk

†PRESENT ADDRESS
Jonathan P. Baker, CHAIN Biotechnology Ltd. MediCity, Nottingham, United Kingdom

RECEIVED 27 April 2023
ACCEPTED 06 June 2023
PUBLISHED 23 June 2023

CITATION
Poulalier-Delavelle M, Baker JP, Millard J, Winzer K and Minton NP (2023), Endogenous CRISPR/Cas systems for genome engineering in the acetogens *Acetobacterium woodii* and *Clostridium autoethanogenum*. *Front. Bioeng. Biotechnol.* 11:1213236. doi: 10.3389/fbioe.2023.1213236

Endogenous CRISPR/Cas systems for genome engineering in the acetogens *Acetobacterium woodii* and *Clostridium autoethanogenum*

Margaux Poulalier-Delavelle*, Jonathan P. Baker†, James Millard, Klaus Winzer and Nigel P. Minton*

Clostridia Research Group, BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Biodiscovery Institute, University of Nottingham, Nottingham, United Kingdom

Acetogenic bacteria can play a major role in achieving Net Zero through their ability to convert CO_2 into industrially relevant chemicals and fuels. Full exploitation of this potential will be reliant on effective metabolic engineering tools, such as those based on the *Streptococcus pyogenes* CRISPR/Cas9 system. However, attempts to introduce *cas9*-containing vectors into *Acetobacterium woodii* were unsuccessful, most likely as a consequence of Cas9 nuclease toxicity and the presence of a recognition site for an endogenous *A. woodii* restriction–modification (R-M) system in the *cas9* gene. As an alternative, this study aims to facilitate the exploitation of CRISPR/Cas endogenous systems as genome engineering tools. Accordingly, a Python script was developed to automate the prediction of protospacer adjacent motif (PAM) sequences and used to identify PAM candidates of the *A. woodii* Type I-B CRISPR/Cas system. The identified PAMs and the native leader sequence were characterized *in vivo* by interference assay and RT-qPCR, respectively. Expression of synthetic CRISPR arrays, consisting of the native leader sequence, direct repeats, and adequate spacer, along with an editing template for homologous recombination, successfully led to the creation of 300 bp and 354 bp in-frame deletions of *pyrE* and *pheA*, respectively. To further validate the method, a 3.2 kb deletion of *hsdR1* was also generated, as well as the knock-in of the fluorescence-activating and absorption-shifting tag (FAST) reporter gene at the *pheA* locus. Homology arm length, cell density, and the amount of DNA used for transformation were found to significantly impact editing efficiencies. The devised workflow was subsequently applied to the Type I-B CRISPR/Cas system of *Clostridium autoethanogenum*, enabling the generation of a 561 bp in-frame deletion of *pyrE* with 100% editing efficiency. This is the first report of genome engineering of both *A. woodii* and *C. autoethanogenum* using their endogenous CRISPR/Cas systems.

KEYWORDS

acetogen, *Acetobacterium woodii*, *Clostridium autoethanogenum*, endogenous CRISPR/Cas, protospacer adjacent motif (PAM), in-frame deletion

1 Introduction

In view of the current climate crisis, Net Zero defines the global commitment to cut greenhouse gas (GHG) emissions to as close to zero as possible by 2050. It is recognized that GHG emissions associated with the continued exploitation of fossil resources are the main driver of climate change. Therefore, innovative solutions for production of fuels and chemicals that participate in atmospheric CO_2 reduction are required.

The Wood–Ljungdahl Pathway (WLP) allows acetogenic bacteria to grow on CO_2 as their sole carbon source. This ancient pathway supports the conversion of two molecules of CO_2 to one molecule of acetyl-CoA, which may then be converted to different C_2-products (acetate or ethanol) or elongated to C_4 (butyrate) or C_5-products (caproate), depending on the species (Schuchmann and Muller, 2016). Acetogens, therefore, represent a particularly attractive option for achieving Net Zero. *Clostridium autoethanogenum*, for instance, produces ethanol in addition to acetate and now forms the basis of a commercial process for its production from steel mill off-gas (Liew et al., 2016). However, for yield improvements and extension of the range of chemicals that can be produced by acetogens, effective genome editing tools are required.

In recent years, genome editing systems based on the Type II CRISPR/Cas9 system of *Streptococcus pyogenes* (SpCas9) (Nishimasu et al., 2014) have played a pivotal role in genetic engineering advances. They have simplified the rapid generation of markerless knock-out (KO), knock-in (KI), and point mutations in multiple bacterial species (Wang et al., 2015; Li et al., 2016; McAllister and Sorg, 2019), including acetogens like *Clostridium ljungdahlii* and *C. autoethanogenum* (Huang et al., 2016; Seys et al., 2020). Despite their benefits, these systems have many drawbacks, most notably poor transfer rates due to the large size of the encoding *cas9* gene and its toxicity when expressed in cells.

Among acetogens, *A. woodii* is a homoacetogen, in that it primarily produces acetate under autotrophic and heterotrophic conditions (Schuchmann and Muller, 2016). Attempts to use the SpCas9-based RiboCas system (Canadas et al., 2019) in *A. woodii* yielded low transformation efficiencies, despite the *cas9* gene being under tight regulatory control. Preliminary experiments indicated that nuclease toxicity and a Type I restriction-modification system, a N^6-adenine DNA methyltransferase (encoded by AWO_c08800, AWO_c08810, and AWO_c08870) with target sequence 5′-$TAAGN_5TCC$-3′, were hindering strain transformation (data not shown). Accordingly, an alternative method, not reliant on the *S. pyogenes* Cas9, was required for gene editing in *A. woodii*. CRISPR systems are present in approximately 45% of sequenced bacterial genomes (Grissa et al., 2007a; Rousseau et al., 2009), which has led to the creation and improvement of databases of predicted CRISPR loci and tools to allow prediction of CRISPR loci in newly sequenced genomes (Grissa et al., 2007a; Grissa et al., 2007b; Rousseau et al., 2009; Couvin et al., 2018). As a result, endogenous CRISPR/Cas systems are becoming better characterized and represent an attractive basis for the development of gene editing systems in many bacteria. In *Clostridium* spp., 74% of strains have an endogenous CRISPR system (Pyne et al., 2016), and endogenous Type I-B systems have successfully been used to genetically modify *Clostridium pasteurianum* (Pyne et al., 2016), *Clostridium butyricum* (Zhou et al., 2021), *Clostridioides difficile* (Maikova et al., 2019), and *Clostridium thermocellum* (Walker et al., 2020).

In Type I systems, the CRISPR locus is composed of a Cas cluster and one or more CRISPR arrays. The Cas cluster contains all the Cas subunits, including the subunits of the Cascade effector complex necessary for interference; the array(s) contain multiple spacers separated by identical direct repeats (DRs) (Nishimasu and Nureki, 2017). A leader sequence controls the expression of CRISPR arrays and is usually located between the cluster and the array (Zhang et al., 2018; Moon et al., 2019). The spacers are derived from previous invader sequences and allow sequence-specific targeting of new invaders. The array is transcribed into pre-crRNA and further processed by the Cas machinery to form mature crRNA. Foreign DNA can be targeted when it contains a sequence matching one of the spacers and an appropriate protospacer adjacent motif (PAM) to distinguish it from endogenous DNA (Nishimasu and Nureki, 2017). In the presence of a protospacer and a PAM, the Cascade–crRNA complex binds to the target DNA, and the Cas3 nuclease is recruited (Mulepati and Bailey, 2013). Cas3 nicks both strands of DNA upstream of the protospacer (Mulepati and Bailey, 2013). It was found that the interference process is directed by a seed sequence adjacent to the PAM rather than the full spacer (Semenova et al., 2011; Wiedenheft et al., 2011), possibly enabling faster screening of DNA entering the cells. It was also shown that in the genome, new spacers are inserted at the beginning of the CRISPR array (Alkhnbashi et al., 2016) by a Cas1–Cas2 complex, making the new spacers likely to have a relatively higher level of expression.

This study describes the steps undertaken to facilitate the exploitation of endogenous CRISPR/Cas systems as genome engineering tools in two acetogens, *A. woodii* and *C. autoethanogenum*. A Python script was developed for PAM prediction and validated against previously published functional PAMs. *A. woodii* Type I-B endogenous CRISPR/Cas system was analyzed, an interference assay confirmed potential PAMs, and RT-qPCR confirmed the functionality of the leader sequence. For genome engineering, the requisite knock-out vectors were built with two components, a synthetic array and an editing template. The synthetic array mimics native arrays; it contains one spacer flanked by direct repeats and is under the control of the native leader sequence. The editing template consists of two homology arms (HAs) to allow homologous recombination. Knock-outs and a reporter gene knock-in were achieved in *A. woodii* while validating different parameters to improve editing efficiency. Subsequently, the workflow was successfully applied to *C. autoethanogenum*, demonstrating the transferability of the techniques presented.

2 Materials and methods

2.1 Python script development

The Python script described in Supplementary Figures S1, S2 was created in the PyCharm IDE (Integrated Development Environment) with Biopython modules (the Python script will be made available upon request). An object called Spacer was created to easily compile all the information collected on each hit for each spacer (See Supplementary Table S1 for a full list of attributes collected) and to facilitate submission, retrieval, and filtering of information from databases. INPUT—the user needs to enter an identification email and the GenBank accession number of the studied organism, as well as a set of information about the CRISPR system which can be retrieved from the CRISPRFinder database (Grissa et al., 2007a; Grissa et al., 2007b; Couvin et al., 2018). ARRAY—the array direction is set either with the CRISPRFinder predictions or an internal function submitting the

flanks to BPROM (Solovyev and Salamov, 2011). If no orientation is found, the user is prompted to enter an orientation. BLAST—the protospacer list is submitted to BLASTn with a filtering function: global mismatch lower than 20% and no more than one mismatch in the first seven nucleotides. PHASTER—information on the hits is retrieved to update the phage attribute, whether directly from the position of the hits or after submission to PHASTER (Zhou et al., 2011; Arndt et al., 2019). PAM—if at least two of the hits are in phage regions, a consensus sequence and WebLogo graphical representations are then created to represent the potential PAMs (Crooks et al., 2004). OUTPUT—confirmation of intermediate steps is displayed in the console, and the final output of the script is an Excel summary document.

2.2 Chemicals

All the chemicals were purchased from Sigma-Aldrich, unless otherwise stated.

2.3 Bacterial strains and plasmids

The following strains were used: *Escherichia coli* TOP10 (Invitrogen), wild-type (WT) *A. woodii* strain DSM1030 (DSMZ), and WT *C. autoethanogenum* strain DSM10061 (DSMZ). Strains are listed in Supplementary Table, S2.

2.4 Growth media and conditions

E. coli was grown aerobically at 37°C in LB medium supplemented with the appropriate antibiotic: chloramphenicol 25 μg/mL in plates or 12.5 μg/mL in liquid cultures and spectinomycin 250 μg/mL. *A. woodii* was grown anaerobically at 30°C in either liquid *A. woodii* medium (AWM) or modified ATCC medium 1019 plates, supplemented when required with thiamphenicol 15 μg/mL. For phenotypic work, ATCC medium lacking yeast extract was used supplemented, when required, with uracil or phenylalanine at 20 μg/mL. The trace element solution SL9, the selenite–tungstate solution, and the vitamin solution were prepared separately and stored at 4°C. *C. autoethanogenum* was grown anaerobically at 37°C in YTF media supplemented with the appropriate antibiotic: 15 μg/mL thiamphenicol in plates or 7.5 μg/mL in liquid cultures and 250 μg/mL D-cycloserine in plates. For phenotypic work, PETC MES agar plates were used, supplemented when required with 20 μg/mL uracil. The list of ingredients for the media and the different solutions is given in Supplementary Tables S3–S5.

2.5 DNA manipulations

All the primers used for the constructs are listed in Supplementary Table S6. Some of the plasmid backbones were taken from the culture collection of the laboratory. All the plasmids built were built according to the pMTL80000 shuttle plasmid standards (Heap et al., 2009). All plasmids used are listed in Supplementary Table S7. The Monarch Plasmid Miniprep Kit, Monarch DNA Gel Extraction Kit, Q5 High-Fidelity Master Mix, restriction enzymes, and T4 DNA Ligase were purchased from New England Biolabs and used according to the manufacturer's protocols. The plasmids required for this study and the colony PCRs to verify their assembly or their presence are described hereafter. Colony PCRs to verify the corresponding genome locus in *A. woodii* or *C. autoethanogenum* are also described.

2.5.1 Interference assay plasmids

The following primer pairs were annealed with Q5: FW_proto8/RV_proto8, FW_proto8-5′/RV_proto8, FW_proto20/RV_proto20, FW_proto20-5′/RV_proto20, FW_proto22/RV_proto22, and FW_proto22-5′/RV_proto22 to yield the three protospacers with and without their PAM candidate. They were then inserted by HiFi in a SacII/SalI-digested pMTL82151 backbone to create pMTL-MPD1 to 6. Colony PCRs and Sanger sequencing were performed with primers ColE1+tra-F2 and pBP1-R1.

2.5.2 Endogenous leader characterization

The MCS from pMTL82151 was inserted after the leader sequence in pMTL-MPD21 by digestion with SacI/NheI and ligation with T4 DNA Ligase, creating pMTL-MPD15 (the full sequence of the vector pMTL-MPD15 is available at www.plasmidvectors.com (RRID: SCR_023475), where it may be sourced). The FAST gene and the *catP* gene were inserted in the MCS after the leader sequence for characterization of the leader sequence by RT-qPCR and to function as a negative control, respectively. The *catP* and FAST genes were amplified with primer pairs FW_catP_SacI/RV_catP_NheI and FW_FAST_SacI/RV_FAST_NheI, respectively. The pMTL-MPD15 backbone and both genes were digested by SacI and NheI. The *catP* and FAST genes were ligated in the backbone with T4 DNA ligase (NEB), yielding pMTL-MPD30 and pMTL-MPD29, respectively. All plasmids were confirmed by Sanger sequencing with primers ColE1+tra-F2 and pCD6-R1.

2.5.3 *pyrE* KO in *A. woodii*

The editing template consisted of two homology arms of 471 bp (LHA) and 491 bp (RHA). The spacer was designed to be 36 bp and at a distance of 29 bp from the RHA. Two Gram-positive replicons were selected, pBP1 and pCD6.

The *cas9* gene, together with its P_{fdxE} promoter, was removed from pMTL8215_pRECas1_MCS (Canadas et al., 2019) by XbaI/NotI digestion, blunting, and re-ligation, with T4 DNA polymerase and T4 DNA Ligase, respectively, yielding the vector pMTL8215_P1339_MCS. The editing cassette was assembled in this plasmid. HAs were amplified with primer pairs FW_LHA_pyrE/RV_LHA_pyrE+BM4 and FW_BM4+1+RHA_pyrE/RV_RHA_pyrE from *A. woodii* genomic DNA, assembled by SOEing PCR (Splicing by Overlap-Extension PCR), and ligated into the vector after AscI/AatII digestion, creating pMTL-MPD7.

A. woodii leader sequence was amplified with primers FW_leader/RV_leader; the synthetic array was created by annealing primers FW_pyrE_endo2 and RV_pyrE_endo2. Both the leader sequence and synthetic array were assembled by HiFi into pMTL-MPD7 digested by SacII/AatII, yielding pMTL-MPD8.

The editing cassette from pMTL-MPD8, amplified with primers FW_leader and RV_RHA_pyrE+MCS, pMTL82151, and pMTL84151, were digested by AatII/NheI, SacII/AatII, or SacII/NheI and ligated together to obtain constructs with HAs alone, expressed spacer alone, or the whole editing cassette, respectively, with either Gram-positive replicon, yielding plasmids pMTL-MPD9 to 14.

Colony PCRs were performed with primers ColE1+tra-F2/pBP1-R1 or ColE1+tra-F2/pCD6-R1 and plasmids confirmed by Sanger sequencing. The same primers were used to check for plasmid presence in *A. woodii,* and the *pyrE* locus was analyzed with primers AW_CRISPR_pyrE_armF and AW_pyrEcomp_RHAR.

2.5.4 *pheA* KO in *A. woodii* with 500 bp, 1.0 kb, and 1.5 kb HA

The spacer was designed to be 35 bp and at a distance of 26 bp from the RHA. Homology arms of 0.5, 1.0, and 1.5 kb were designed. The resulting editing cassettes were inserted into a backbone with the pCD6 Gram-positive replicon.

The spacer sequence was created by annealing primers FW_pheA_spacer and RV_pheA_spacer by PCR and digestion by SacI and AatII; the leader sequence was created by amplification with FW_leader/RV_leader and SacII and SacI digestion; both fragments were assembled in pMTL84151 digested by SacII/AatII. Homology arms of 0.5 kb, 1.0 kb, and 1.5 kb were designed. Left homology arms were amplified with primers FW_L0.5/L1.0/L1.5 and RV_LHA_pheA; right homology arms were amplified with primers FW_RHA_pheA and RV_R0.5/R1.0/R1.5. The LHA and RHA were assembled by SOEing PCR with primer pairs FW_L0.5/RV_R0.5, FW_L1.0/RV_R1.0, and FW_L1.5/RV_R1.5. Assembled HAs were ligated in both pMTL84151 and pMTL84151 containing the assembled leader and spacer after AatII/NheI digestion, yielding three plasmids containing the HAs alone (pMTL-MPD16 to 18) and three editing plasmids containing the expressed spacer and the HAs (pMTL-MPD19 to 21).

Colony PCRs were performed with primers ColE1+tra-F2/pCD6-R1 and plasmids confirmed by Sanger sequencing. *A. woodii* colonies were analyzed by colony PCR both for plasmid presence with primers ColE1+tra-F2 and pCD6-R1 and for the *pheA* locus with primers FW_pheAHA1.5_out and RV_pheAHA1.5_out.

2.5.5 *hsdR1* KO in *A. woodii*

Homology arms of 1,455 bp (LHA) and 1,489 bp (RHA) and a 36 bp spacer starting 40 bp from the LHA were designed.

Primers FW_hsdR1_spacer and RV_hsdR1_spacer were annealed by PCR and inserted in the pMTL84151 vector with the assembled leader and spacer used for *pheA* KO by SacI and AatII digestion and ligation with T4 DNA Ligase, resulting in the vector pMTL8415_leader_hsdR1_spacer. The left and right homology arms were amplified with the primers FW_LHA_hsdR1/RV_LHA_hsdR1 and FW_RHA_hsdR1/RV_RHA_hsdR1, respectively, and assembled by SOEing PCR with the FW_LHA_hsdR1/RV_RHA_hsdR1 primer pair. The vector containing the spacer and the assembled HAs were digested by AatII and NheI and ligated together with T4 DNA Ligase, resulting in the final editing vector for *hsdR1* KO, pMTL-MPD22.

Colony PCRs were performed with primers ColE1+tra-F2/pCD6-R1 and plasmids confirmed by Sanger sequencing. Cell densities and the amount of DNA were investigated as potential factors to increase editing efficiency as follows: 100 µL of cells and 1 µg of DNA (T1); 100 µL of high-density cells and 1 µg of DNA (T2); 200 µL of high-density cells and 4 µg of DNA (T3). *A. woodii* colonies were analyzed by PCR for plasmid presence with primers ColE1+tra-F2/pCD6-R1 and the *hsdR1* locus with primers FW_hsdR1_screen_1.5/RV_hsdR1_screen_1.5 and FW_hsdR1_del_screen/RV_hsdR1_del_screen.

2.5.6 FAST gene KI at the *pheA* locus in *A. woodii*

The spacer is 36 bp, and homology arms are 1,496 bp and 1,465 bp and were designed for insertion of the cargo in place of the PAM. The cargo consists of the FAST reporter gene flanked by the thiolase (*thl*) promoter (P_{thl}) and *pyrE-hydA* (orotate phosphoribosyltransferase and hydrogenase I) terminator from *C. acetobutylicum* and was inserted between the homology arms.

Primers FW_AWO_pheA_CCG_KI_spacer and RV_AWO_pheA_CCG_KI_spacer were annealed by PCR and inserted in the pMTL84151 vector with the assembled leader and spacer used for *pheA* KO by SacI and AatII digestion and ligation with T4 DNA Ligase. The left and right homology arms were amplified with primers FW_AWO_LHA_pheA_cargo/RV_AWO_LHA_pheA_CCG_cargo_MCS and FW_AWO_RHA_pheA_CCG_cargo_MCS/RV_AWO_RHA_pheA_cargo, respectively, and assembled by SOEing-PCR with the FW_AWO_LHA_pheA_cargo/RV_AWO_RHA_pheA_cargo primer pair. The assembled HAs were inserted in the vector containing the spacer by AatII and NheI digestion and ligation by T4 DNA Ligase, resulting in a KI vector (pMTL-MPD23) containing an MCS with BamHI, NotI, NdeI, and XbaI restrictions sites between the two homology arms.

A P_{thl}_FAST_term cassette was amplified with the primer pair FW_BamHI_FAST_cassette/RV_FAST_NotI_AscI from plasmid pMTL8415X_P_{thl}_FAST with the FAST gene flanked by the *thl* promoter and *pyrE-hydA* terminator. The cassette was then inserted into the MCS of the pMTL-MPD23 vector by XbaI and NotI digestion and ligation with T4 DNA Ligase, yielding pMTL-MPD24.

Plasmids were confirmed by Sanger sequencing with ColE1+tra-F2/pCD6-R1/FW_pheA_cargo_seq/RV_pheA_cargo_seq.

WT *A. woodii* was transformed with the final editing vector, pMTL-MPD24, following the protocol named T3 in the previous section, with 200 µL of high-density cells and 4 µg of DNA. *A. woodii* colonies were analyzed by colony PCR for plasmid presence with primers ColE1+tra-F2/pCD6-R1 and the *pheA* locus with primers FW_pheA_KI_seq and RV_pheA_KI_seq.

2.5.7 *pyrE* KO in *C. autoethanogenum*

Homology arms of 868 bp and 797 bp were used, and the synthetic array was designed to mimic *C. autoethanogenum* native array 4, with the native leader sequence controlling the expression of a 35 bp spacer flanked by direct repeats. The editing cassette was inserted into a pMTL83151 backbone.

pMTL83151 and CRISPR array leader amplified with FW_leader4_CLAU/RV_leader4_CLAU were digested by SacII/SacI and ligated together to yield pMTL-MPD25 (the full sequence of the vector pMTL-MPD25 is available at www.plasmidvectors.com (RRID: SCR_023475), where it may be sourced). Primers FW_

CLAUpyrE_spacer1 and RV_CLAUpyrE_spacer1 were annealed by PCR and inserted into pMTL-MPD25 by SacI and BamHI digestion and ligation with T4 DNA Ligase, yielding pMTL-MPD27. This plasmid and the homology arms amplified with FW_CLAUpyrE_HA/RV_CLAUpyrE_HA from plasmid vFS67 (provided by Dr. Francois Seys) were digested by BamHI/NheI and ligated together to yield the final editing vector, pMTL-MPD28. An intermediate vector (pMTL-MPD26) containing the HAs alone was obtained by digesting pMTL83151 and the HAs by BamHI/NheI and ligation.

Colony PCRs were performed with primers ColE1+tra-F2 and pCB102-R1 and plasmids confirmed by Sanger sequencing. *C. autoethanogenum* colonies were analyzed by PCR for plasmid presence with ColE1+tra-F2 and pCB102-R1 and the *pyrE* locus analyzed with FW_pyrECLAU_screening and RV_pyrECLAU_screening.

2.6 Strain engineering

2.6.1 *Escherichia coli*

2.6.1.1 Chemically competent cells

An overnight culture of *E. coli* TOP10 was used to inoculate 100 mL of LB medium. Once the OD_{600nm} reached 0.4–0.5, the culture was incubated at 4°C for 30 min. The cells were then collected by centrifugation at 5,000 rpm for 5 min at 4°C. The cells were washed and incubated for 15 min with ice-cold 0.1 M $MgCl_2$. A second wash and incubation was performed with ice-cold 0.1 M $CaCl_2$. Finally, the cells were resuspended in 1 mL of ice-cold 0.1 M $CaCl_2$ with 20% glycerol. The cells were then aliquoted for storage and kept at −80°C until further use.

2.6.1.2 Transformation and colony PCR

A volume of 50 µL of chemically competent *E. coli* TOP10 cells was incubated with DNA for 30 min on ice, then 30 s at 42°C, 5 min on ice, and a final 1 h recovery in SOC medium before plating on selective LB agar plates. Colonies were checked by colony PCRs to verify the presence of the plasmids using Taq with standard buffer (New England Biolabs). The plasmids were then extracted using the Monarch Plasmid Miniprep Kit (New England Biolabs) and confirmed by Sanger sequencing.

2.6.2 *Acetobacterium woodii*

2.6.2.1 Competent cells: standard and high density

A 48 h liquid culture of the appropriate strain of *A. woodii* was used to inoculate a 300 mL culture to reach an OD_{600nm} of between 0.3 and 0.4 overnight. The cells were collected and then washed twice with filter-sterilized SMP5.8 buffer (1 mM sodium phosphate pH 5.8, 1 mM $MgCl_2$, and 270 mM sucrose). In the final step, they were resuspended in 2 mL of SMP5.8 buffer for standard competent cells or in 666 µL for high-density competent cells, and 100% DMSO was added to reach 10% concentration. The cells were aliquoted for storage and kept at −80°C until further use.

2.6.2.2 Transformation and colony PCR

A. woodii was transformed by electroporation using either 100 µL of competent cells and 1 µg of plasmid in 20 µL of water or 200 µL of high-density competent cells and 4 µg of plasmid in 20 µL of water. Electroporation was performed using a Gene Pulser Xcell™ electroporator (Bio-Rad). After 5 h (unless stated otherwise) of recovery in AWM, transformations were plated on modified AWM containing the appropriate selection or supplementation. After 10 days, the transformation efficiencies were recorded. The colonies were then cultured in liquid AWM, and after 48 h, colony PCRs were performed for analyzing plasmid presence and genome locus, when appropriate, using Taq DNA polymerase (NEB). The PCRs were subsequently confirmed by Sanger sequencing.

2.6.2.3 Plasmid loss

After mutant strain confirmation, a series of inoculations into liquid broth and agar plates were performed for plasmid loss. Single colonies were streaked on plates with and without selection to check for plasmid loss.

2.6.3 *Clostridium autoethanogenum*

2.6.3.1 Conjugation and colony PCR

YTF was inoculated to an OD_{600nm} of 0.02 from a late exponential (48–72 h) preculture of the *C. autoethanogenum* recipient strain and incubated overnight. An *E. coli* NEB sExpress conjugation donor (Woods et al., 2019) harboring the shuttle vector was inoculated and incubated overnight in LB supplemented with spectinomycin and additional antibiotics for the shuttle vector.

The next day, LB broth containing the appropriate antibiotic selection was inoculated with the *E. coli* donor strain and incubated until OD_{600nm} 0.2–0.4 was reached. A 1 mL aliquot of the *E. coli* donor strain culture was collected at 3,000 g for 3 min, washed in 500 µL of PBS through gentle flicking of the tube, and centrifuged again at 3,000 g for 3 min. The pellet was then transferred into an anaerobic cabinet and resuspended in 200 µL of the overnight culture of *C. autoethanogenum*. The mixture was spotted onto non-selective YTF agar plates and incubated overnight (16–24 h on a mating plate).

The next day, the growth on the non-selective mating plate was collected using 500 µL of anaerobic PBS and a wedge-shaped spreader. The slurry was then spread onto agar plates with appropriate selection for the shuttle vector and D-cycloserine as an *E. coli* counter-selection. The plates were incubated under anaerobic conditions for 3–4 days. Once the colonies were visible, they were inoculated onto fresh agar plates with selection, and single colonies were inoculated in liquid cultures with appropriate antibiotic. Glycerol stocks and colony PCRs were performed after 48 h using Taq DNA polymerase (NEB). The PCRs were subsequently confirmed by Sanger sequencing.

2.6.3.2 Plasmid loss

After mutant strain confirmation, a series of inoculations into liquid broth and agar plates were performed for plasmid loss. Colonies were inoculated onto non-selective plates, either directly from the liquid cultures with selection used for colony PCRs, or from a subculture of it without selection. From those non-selective plates, single colonies were streaked onto non-selective plates. Single colonies were streaked on plates with and without selection to check for plasmid loss.

2.7 Flow cytometry

A. woodii cells were grown until the early- to mid-exponential phase and then pelleted at 7,000 g for 5 min. The cells were washed twice in PBS before being resuspended in PBS. A volume of 100 µL of 25 µM ^{TF}Amber (The Twinkle Factory) or PBS was added to 400 µL of resuspended cell samples immediately before fluorescence analysis. Cell fluorescence was analyzed using an Astrios EQ cell sorter (Beckman Coulter) with a 488 nm laser and a 576/21 nm filter. Data were analyzed using the Kaluza 2.1 analysis software (Beckmann Coulter). The population was gated to eliminate outlier cells and obtain consistent results for each sample, a first gate to select cells and a second gate to select single cells when represented on a side scatter/forward scatter plot. The percentage of cells gated was represented as a function of the FAST Amber fluorescence intensity. The number of events for each graph is as follows: WT without/with ligand: 549,749/441,371; plasmid without/with ligand: 2,568,561/6,791,597; KI without/with ligand: 1,285,872/5,630,864.

2.8 Endogenous leader characterization

WT *A. woodii* was separately transformed with the plasmids described in Section 2.5.2, colonies were grown in liquid cultures, and samples were taken at different stages of cell growth. RNA was extracted, and RT-qPCR was performed as described in the following sections.

2.8.1 RNA extraction

A. woodii cells were collected by centrifugation at 7,000 g for 5 min at 4°C, and the pellet was resuspended in 2 mL of RNAprotect (QIAGEN). After a 5 min incubation at room temperature, the cells were collected again, and the pellet was stored at −80°C.

RNA extraction was carried out using a FastRNA Pro Blue kit (MP Bio). Cells were homogenized at 6,400 rpm for 45 s. All the centrifugation steps were carried out for 15 min at 4°C. The RNA was precipitated for 2 h at −80°C, and the RNA pellet was dried for 30 min at room temperature and resuspended in 50 µL of DEPC-H_2O.

To reduce genomic DNA contamination, samples were digested by TURBO DNase (Ambion). The resulting RNA was purified using the RNeasy Mini kit (QIAGEN). 350 µL of RLT buffer containing β-mercaptoethanol and 200 µL of absolute ethanol were added to the digested samples before purification according to the manufacturer's instructions.

To check for genomic DNA contamination, a PCR (35 cycles, 50°C, and 2:00 min extension) with primers gDNA_FW and gDNA_RV (Supplementary Table S8), amplifying a 1,997 bp fragment in the *A. woodii pyrE* region, was carried out with Taq DNA polymerase (NEB). When genomic DNA was present, the samples were treated with TURBO DNase and purified with the RNeasy Mini kit a second time. The RNA samples were stored at −80°C.

The RNA concentration was assessed using a Nanodrop Lite spectrophotometer (Thermo Fisher). The quality was evaluated using an Agilent RNA 6000 Nano kit on an Agilent 2100 Bioanalyzer, and the RNA integrity number (RIN) was recorded for all the samples.

2.8.2 Quantitative reverse transcription real-time PCR (RT-qPCR)

Complementary DNA was synthesized from 1 µg of RNA using the Omniscript RT Kit (QIAGEN), random hexamer (Thermo Fisher), and RNase inhibitor (NEB) following the manufacturer's instructions.

Power SYBR Green (Thermo Fisher) was used for the qPCR, and primer pairs targeting the *gyrA* gyrase gene in the *A. woodii* genome, *orfB* of the plasmid backbones, and the FAST gene employed are listed in Supplementary Table S8. Standard curves allowed for efficiencies of primer pairs to be calculated. Gene expression was calculated for each sample using the Pfaffl method without a calibrator gene:

$$Expression = \frac{E_{target}{}^{-Cp_{target}}}{E_{reference}{}^{-Cp_{reference}}}.$$

3 Results

3.1 Automated pipeline for *in silico* analysis of endogenous CRISPR systems

3.1.1 Python script development

A Python script was developed to collect the host organism genome identifier, along with the information found on the CRISPRFinder database (Grissa et al., 2007b; Couvin et al., 2018), and compile all the information collected on each match for each spacer. A schematic overview is presented in Supplementary Figure S1, and a complete diagram is available in Supplementary Figure S2. Briefly, each spacer is run through BLAST and then filtered to reduce the number of false-positives and retain biologically relevant hits.

False-positives were eliminated by removing hits with more than 20% mismatch between spacers and hits, and only one mismatch was allowed in the seven nucleotides seed region adjacent to the PAM. Biologically relevant hits are then used to compile the PAM consensus sequence, based on filtering capabilities for mobile genetic element (MGE) sequences. Hits are kept based on three elements: the title of the hit containing the keywords "phage" or "plasmid;" their presence in a phage sequence predicted by PHASTER; and the annotation of the hit containing the keywords "transposase," "phage," "integrase," or "terminase."

3.1.2 Evaluation of Python script performance against published literature

The performance of the script was assessed using data available in the literature on Type I-B systems from *Clostridium* spp. that have been successfully used for genome editing. The results are presented in Supplementary Table S9. The script successfully identified the majority of the functional PAMs. For *C. tyrobutyricum,* it failed to identify the TCG alternative (Zhang et al., 2018), and for *C. thermocellum,* it identified only TTA/TTG/TCA instead of TTN and TNA, which have been experimentally validated as functional (Walker et al., 2020). For a majority of those organisms, the script predicted PAMs that had not been identified by manual analysis in the literature (marked with asterisks in Supplementary Table S9). *In vivo* tests are necessary to determine if those PAMs allow for interference and to conclude on the performance of the script.

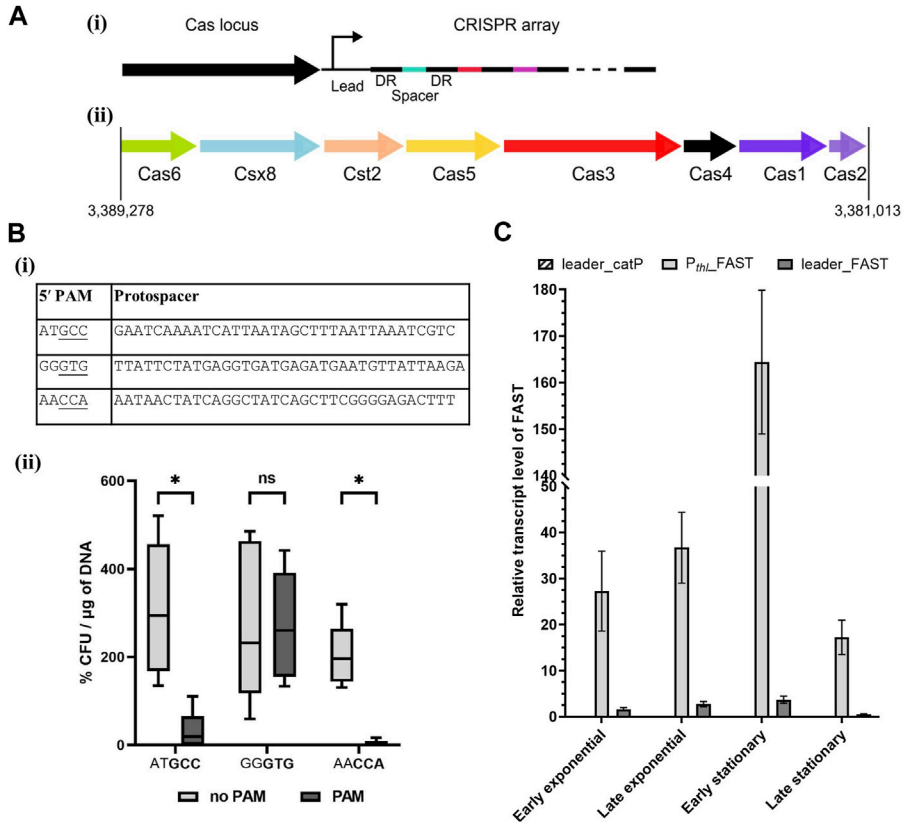


FIGURE 1
Acetobacterium woodii DSM1030 Type I-B CRISPR system. **(A)** Organization of the Type I-B CRISPR system of *A. woodii* DSM1030. (i) Cas locus and CRISPR array relative positions. Both the Cas locus and CRISPR array are in reverse direction in the genome, and the CRISPR array is directly downstream of the Cas locus. (ii) Gene organization of the Cas locus. **(B)** Interference assay to determine functional PAMs in *A. woodii*. (i) Table of the protospacers and PAM candidates used in the interference assay in *A. woodii*. Three of the protospacers identified in Supplementary Table S11 were inserted in a backbone with, in the 5′ position, either the native five nucleotides from the direct repeat "CAAAT" or candidate PAMs. The native 3′ sequence from the direct repeat "ATTTA" was used. Potential functional trinucleotide PAMs are underlined. (ii) Transformation efficiency of WT *A. woodii* obtained with the plasmids containing the different combinations of the protospacer, with (dark gray) or without (light gray) their corresponding candidate PAMs (bold). Data normalized and expressed in percentages with the following parameters: transformation efficiency obtained with the empty backbone control pMTL82151 set to 100% and smallest value of the replicate as 0%. $n = 5$; the 5th and 95th percentile are represented; two-way ANOVA test performed to test for significance: ns: non-significant, *: $p < 0.033$. **(C)** FAST transcript level normalized with *orfB* as the reference gene measured by RT-qPCR at different stages of cell growth. Samples were collected in early and late exponential stages and early and late stationary stages for the pMTL-MPD30 plasmid (dark gray), the negative control pMTL-MPD29 (dashed), and positive control pMTL8415_P_{thl}_FAST (light gray). The transcript levels were calculated with the Pfaffl method without a calibrator sample. Error bars represent technical triplicates.

3.1.3 Analyses of other *Clostridium* spp.

Clostridium spp. of interest containing a Type I-B system, *Clostridium butyricum*, *Clostridium limosum*, and *Clostridium novyi* were analyzed using the script. The results are presented in Supplementary Table S10. The results obtained indicate that PAMs containing the nucleotide pairs TC or TT are common in Type I-B CRISPR/Cas systems. The number of candidate PAMs for each organism is quite high, indicating that those CRISPR/Cas systems are likely to recognize multiple PAMs (Fischer et al., 2012).

3.1.4 Proof-of-concept: analysis of the *A. woodii* Type I-B CRISPR system

The CRISPRFinder database (Grissa et al., 2007b; Couvin et al., 2018) identified a Type I-B system in *A. woodii,* as presented in Figure 1A. This system contains a Csx8, a Cas8a homolog, confirmed by protein BLAST with homology to multiple Cas8a proteins from *Clostridium* spp. Cas8a is specific to Type I-A systems, but the gene organization of the *A. woodii* Cas locus is characteristic to Type I-B systems (Makarova and Koonin, 2015). Only one CRISPR array, named array 2, with an evidence level of 4 is identified in the database. It is located directly downstream of the Cas locus; it contains 47 spacers and its direct repeat is "ATTTAC ATTCCAATATGGATCTACTCAAAT." This CRISPR array was analyzed using the Python script, and the results are compiled in Supplementary Table S11.

The script identified multiple matches with invading elements but none with less than 17% mismatch. Although the script computed no clear consensus sequence for the 5′ PAM, three out of the eight matches contain two consecutive cytosines. Interestingly, a match with a 5′ CCA PAM in a phage gene was found, and this PAM was shown to be functional for the *C. difficile* Type I-B system (Boudry et al., 2015; Maikova et al., 2019). Thus, it is the main candidate for harnessing the endogenous CRISPR system in *A. woodii*.

3.2 Harnessing of the *A. woodii* endogenous CRISPR system

3.2.1 PAM validation by interference assay

Interference assays are used to assess PAM functionality by transformation of the WT strain with plasmids containing a protospacer and different candidate PAMs. Functional PAMs trigger interference and result in lower transformation efficiencies (Pyne et al., 2016; Walker et al., 2020). Three candidate 5′ PAMs identified by the Python script were tested experimentally in *A. woodii*, as shown in Figure 1Bi. Mismatches adjacent to the PAM reduce the likelihood of effective interference. Therefore, protospacers with no mismatch in the first position and with only one mismatch in the first 10 nucleotides were selected for testing. Each protospacer was assembled into a plasmid with the 3′ sequence from the direct repeat "ATTTA" and either the 5′ sequence of the direct repeat "CAAAT" or its corresponding PAM candidate. WT *A. woodii* was separately transformed with the six plasmids, pMTL-MPD1 to 6; transformation efficiencies were expressed as a percentage of the transformation efficiency obtained with the pMTL82151 positive control and normalized to the highest and lowest values.

Although high variations were observed between replicates (Figure 1Bii), the reduction in transformation efficiency in the presence of the GCC and CCA PAMs was shown to be statistically significant (p-value <0.033) compared to the presence of the protospacers alone. Sanger sequencing of colony PCR amplicons showed a high prevalence of mutations in the CCA PAM, with one of the cytosines missing. Those colonies with mutated PAMs are thus escape mutants of the endogenous CRISPR/Cas system. Taken together, these results strongly suggest that the double cytosine is essential to the interference phase of the studied Type I-B CRISPR/Cas system. The CCA 5′ PAM was selected for subsequent genome editing experiments in *A. woodii*.

3.2.2 Leader sequence characterization

For efficient targeting of invading DNA, the CRISPR array needs to be transcribed into pre-crRNA before maturation and Cascade assembly. In order to harness the *A. woodii* CRISPR system, a better characterization of the leader sequence responsible for the transcription of the array is required. For that purpose, the plasmid pMTL-MPD30 was built with the FAST reporter gene placed downstream of the leader of the aforementioned array 2. Control plasmids were built with either the *catP* reporter gene replacing the FAST reporter gene downstream of the leader sequence or with the thiolase promoter replacing the leader sequence for expression of the FAST-encoding reporter gene. These controls confirm primer specificity and efficiency, respectively. The transcript levels of the FAST gene were expressed (Figure 1C) with *orfB* as the reference gene to take into account the plasmid copy number. Transcript levels of the controls confirmed adequate annealing and specificity of the primers. The leader sequence allows for a low level of transcription in all cell growth stages with a clear reduction in the RNA level in the late stationary phase. This experiment confirms the functionality of the leader sequence for expression of CRISPR arrays for targeting DNA.

3.2.3 Genome engineering capability validation: *pyrE* knock-out

To establish the potential of the *A. woodii* Type I-B CRISPR/Cas system for genome editing, the *pyrE* gene was targeted for inactivation through the creation of a 300 bp deletion known to lead to uracil auxotrophy (Baker et al., 2022). Six plasmids were constructed (Figure 2A) with HAs alone, spacer alone, or the whole editing cassette and either the pCD6 or pBP1 Gram-positive replicon. The vectors with HAs and spacer alone function as controls to assess the impact of homologous recombination and self-targeting by Cas3, respectively. Only the whole editing cassette with HAs and expressed spacer allows for homologous recombination and selection against WT by self-targeting. WT *A. woodii* was separately transformed with the aforementioned plasmids in biological duplicates and plated after 5 h and 7.5 h of recovery.

The presence of the expressed spacer targeting the *pyrE* gene, with either replicon pCD6 or pBP1, with or without HAs, significantly reduced transformation efficiencies to between 0.06% and 2.17% of the transformation efficiency obtained with the corresponding plasmid containing HAs alone (data not shown). This shows the efficient self-targeting by the endogenous CRISPR/Cas system of the *pyrE* gene, leading to cell death. Colony numbers obtained varied significantly between the different replicates; colony PCR results are presented in Figure 2B. Successful gene knock-out was obtained with the full editing cassette, with either Gram-positive replicon pBP1 or pCD6. The latter was found to lead to 100% plasmid loss after one subculture without selection (Baker et al., 2022); it was, therefore, chosen for subsequent editing vectors.

Analysis of the escape mutants revealed single crossover events and frequent deletion of the spacer. The deletion of the spacer can be explained by the presence of the identical direct repeats flanking the spacer, facilitating recombination under Cas3 selective pressure.

3.2.4 Impact of homology arm length on editing efficiency: *pheA* knock-out

The impact of homology arm length on editing efficiency was studied, tested with 0.5 kb, 1.0 kb, and 1.5 kb, for a 354 bp in-frame deletion in the *pheA* gene, previously shown to lead to phenylalanine auxotrophy (Baker et al., 2022).

WT *A. woodii* was separately transformed with six vectors (Figure 3A), three containing the HAs alone and three editing plasmids also containing the expressed spacer. The transformation efficiency obtained with each editing vector was normalized to the transformation efficiency obtained with the corresponding vector containing only the HAs. While low transformation efficiencies and high variations in colony numbers between each replicate were observed, the presence of the spacer reduced the transformation efficiency to between 0.053% and 0.32% of the efficiency obtained with the vectors carrying HAs alone (data not shown), showing efficient targeting of the *A. woodii* genome by the selected spacer. Characteristics of the colonies analyzed for the three replicates of this experiment are shown in Figures 3B, C, as well as the results of the colony PCRs of the *pheA* locus for the first replicate of this experiment.

At both 5 h and 7.5 h recovery, the 1.5 kb HAs resulted in more consistent knock-out generation with between 33.3% and 100% gene editing efficiency. After the plasmid loss protocol, 91.7% of colonies tested (72 in total) had successfully lost the plasmid. The phenotype

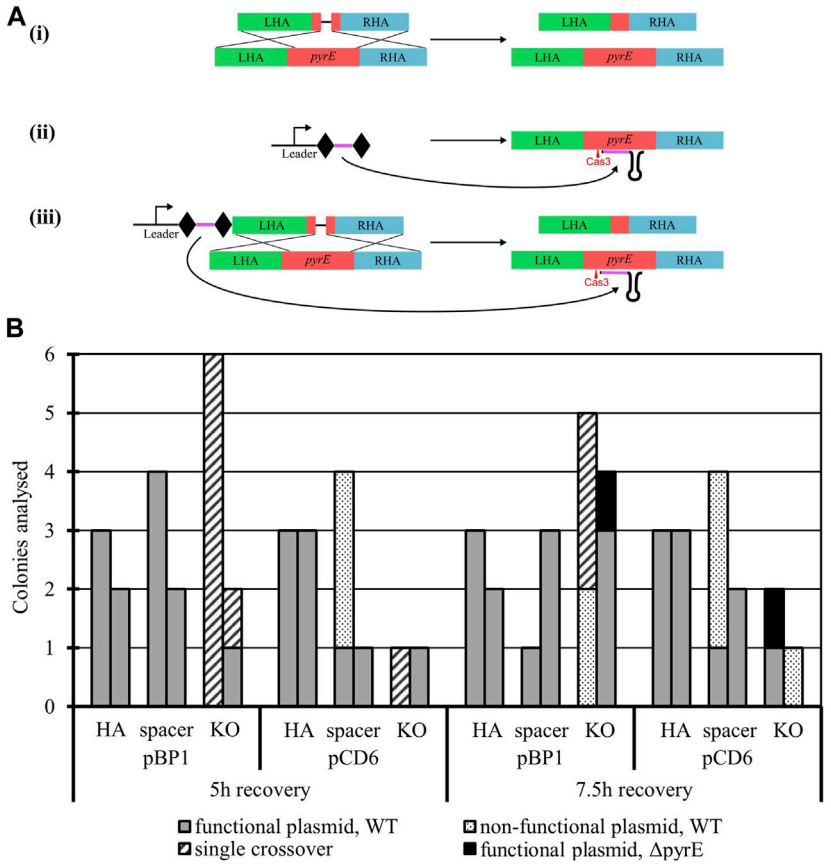


FIGURE 2
Endogenous CRISPR-based deletion of the *A. woodii pyrE* gene. **(A)** *pyrE* deletion strategy in *A. woodii* using its native Type I-B CRISPR system. The constructs represented correspond to the application-specific module of the following plasmids, each of them built with either pBP1 or pCD6 as the Gram-positive replicon: (i) pMTL-MPD9 and 12 HAs alone (471 bp LHA in green; 491 bp RHA in blue) allow for homologous recombination in some of the cells of the population; (ii) pMTL-MPD10 and 13, the expression of the 36 bp spacer (purple) flanked by DRs (diamonds) alone leads to self-cleavage by Cas3 (red) at the *pyrE* gene (red) and cell death; (iii) pMTL-MPD11 and 14, the full editing cassette allows for homologous recombination, deletion of the 300 bp region, and selection of the successful KO. **(B)** Summary of colonies analyzed for the 300 bp deletion in the *pyrE* gene by colony PCRs with primers AW_CRISPR_pyrE_armF and AW_pyrEcomp_RHAR. Data show two replicates of transformation assays with pBP1 and pCD6 Gram-positive replicon for recovery times of 5 h and 7.5 h. Where possible, up to six colonies were analyzed. Gray: WT *pyrE* locus with functional corresponding plasmid; dots: WT *pyrE* locus with non-functional corresponding plasmid; dashed: single crossover of the plasmid in the *pyrE* locus; black: Δ*pyrE* with the functional corresponding plasmid.

of the strains was confirmed by plating on minimal media with and without phenylalanine supplementation (Figure 3D).

These results are the first report of successful gene editing using the *A. woodii* endogenous Type I-B CRISPR/Cas system and represent a promising basis for further engineering of this chassis. The knock-out of *pheA* followed by plasmid loss and confirmation of the phenotype highlights how straightforward and rapid the technique is, while illustrating the importance of the length of the editing template.

3.2.5 Validation of the system: 3.2 kb *hsdR1* knock-out

To test the robustness of the system for gene KO in *A. woodii*, the restriction subunit gene *hsdR1* responsible for cleavage of the *cas9*-containing plasmids was targeted. Deletion of this gene has never previously been generated and at 3.2 kb represents a 10-fold increase in the size of the DNA deleted. WT *A. woodii* was transformed with the editing vector pMTL-MPD22 (Figure 4A), following protocols with increasing cell densities and DNA amounts referred to as T1, T2, and T3. Plating after 7.5 h of recovery yielded only WT colonies (data not shown). The generation of knock-out mutants was demonstrated (Figure 4B) with all three methods of transformation plated at 5 h of recovery, with 16.7%, 50%, and 75% efficiency for T1, T2, and T3, respectively. Sanger sequencing of the PCR-amplified plasmid sequences for T1 colonies 1 and 2; all T2 colonies; and T3 colonies 1, 2, and 4 revealed the absence of the spacer in the plasmid of colonies T1 colony 2 and T3 colony 4. As previously observed, recombination between the direct repeats flanking the spacer facilitates the excision of the spacer from the plasmid, rendering it non-functional. All Δ*hsdR1* colony PCR amplicons were confirmed by Sanger sequencing. Plasmid loss was undertaken for the six confirmed Δ*hsdR1* colonies, with 80.6% of those tested (36 in total) being shown to have successfully lost the plasmid.

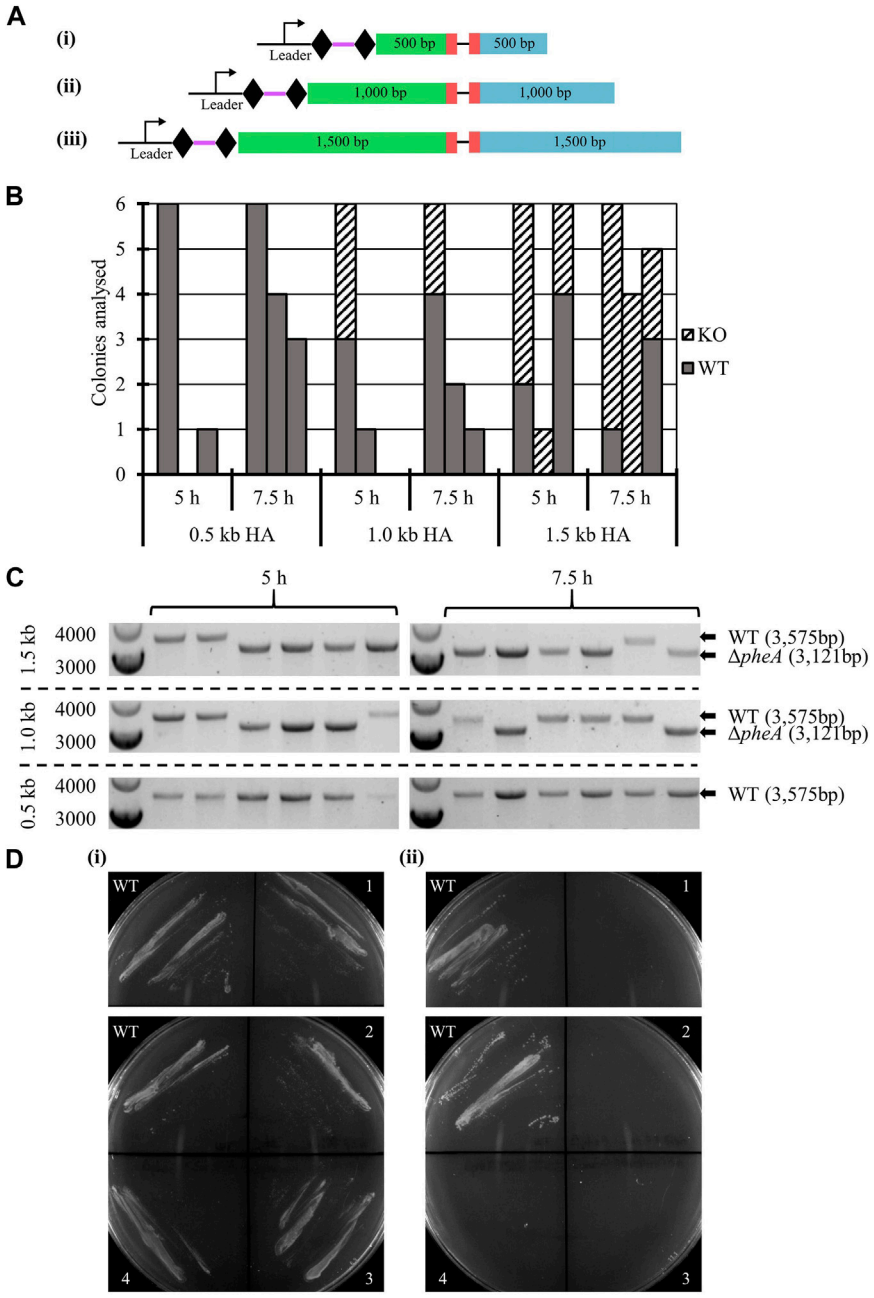


FIGURE 3
Endogenous CRISPR-based deletion of the *A. woodii pheA* gene. **(A)** Representation of the application-specific modules of the three different editing plasmids containing different size HAs for *pheA* (red) KO in *A. woodii*. The 35 bp spacer (purple) is flanked by the native DR (diamond) and expressed under the control of the native leader sequence. HAs (LHA in green; RHA in blue) of different lengths allow for homologous recombination and deletion of the 354 bp region: (i) pMTL-MPD19, (ii) pMTL-MPD20, and (iii) pMTL-MPD21. **(B)** Summary of colonies analyzed for the 354 bp deletion in the *pheA* gene by colony PCRs with primers FW_pheAHA1.5_out and RV_pheAHA1.5_out. Data show three replicates of transformation assays with homology arms of 0.5 kb, 1.0 kb, and 1.5 kb for recovery times of 5 h and 7.5 h. Where possible, up to six colonies were analyzed; missing bars show conditions where no colonies were obtained. Gray: WT *pheA* locus; dashed: Δ*pheA*. **(C)** Colony PCR results for the first replicate of the *pheA* KO experiment with homology arms of 0.5 kb, 1.0 kb, or 1.5 kb by amplification with primers FW_pheAHA1.5_out and RV_pheAHA1.5_out. The WT amplicon is expected at 3,575 bp and the KO amplicon at 3,121 bp. **(D)** Phenotypic analysis of the *A. woodii* Δ*pheA* strains obtained. 1 and 2 are *A. woodii* Δ*pheA* strains obtained with 1.0 kb HAs with 5 h or 7.5 h recovery time, respectively; 3 and 4 are *A. woodii* Δ*pheA* strains obtained with 1.5 kb HAs with 5 h or 7.5 h recovery time, respectively; WT is the WT *A. woodii* strain. (i) Minimal media supplemented with 20 μg/mL of phenylalanine; (ii) minimal media without supplementation.

To confirm the impact of the R-M site on transformation efficiency and the phenotype of the *hsdR1* deletion, the R-M site 5′-TAAGN_5TCC-3′ was inserted in one, two, or three copies in the backbone pMTL82151. WT and Δ*hsdR1 A. woodii* strains were separately transformed with the four plasmids. For two of the replicates, the transformation efficiency of the Δ*hsdR1* strain

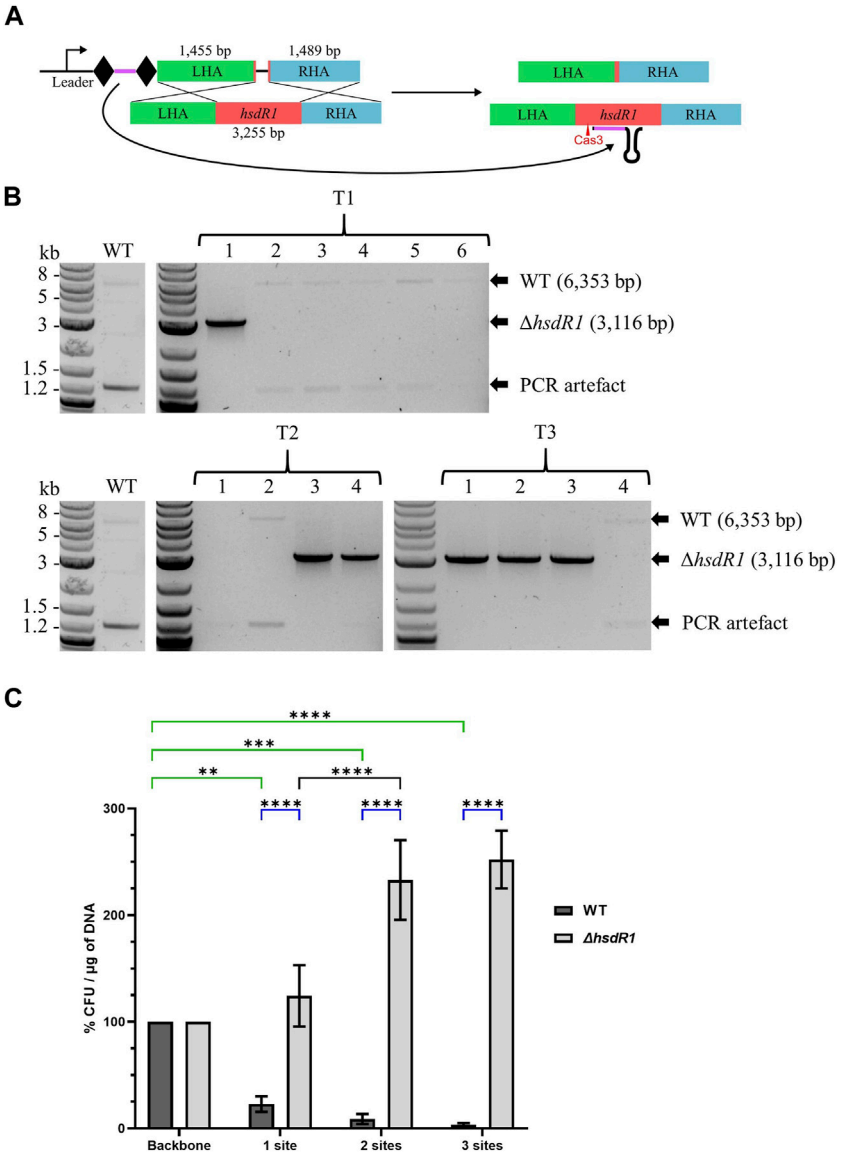


FIGURE 4
Endogenous CRISPR-based deletion of the *A. woodii hsdR1* gene. **(A)** Representation of the application-specific module of the editing plasmid for *hsdR1* (red) KO in *A. woodii*, pMTL-MPD22. The 36 bp spacer (purple) is flanked by the native DR (diamond) and expressed under the control of the native leader sequence. The editing template with a 1,455 bp LHA (green) and a 1,489 bp RHA (blue) allows for homologous recombination and deletion of a 3,237 bp region. **(B)** Colony PCR results with primers FW_hsdR1_screen_1.5 and RV_hsdR1_screen_1.5 of the *hsdR1* locus of colonies transformed with the editing vector for *hsdR1* KO. The WT amplicon is expected at 6,353 bp, and KO amplicon is expected at 3,116 bp. T1: standard *A. woodii* electroporation; T2: high-efficiency protocol with 100 µL cells and 1 µg of DNA; T3: high-efficiency protocol with 200 µL cells and 4 µg of DNA. **(C)** Transformation efficiencies of *A. woodii* WT and Δ*hsdR1* obtained with plasmids containing up to three R-M sites normalized to the transformation efficiency obtained with the empty backbone pMTL82151. Data represent technical triplicates for WT and biological triplicates for Δ*hsdR1*. Two-way ANOVA tests were performed to test for significance, in green, of the impact of the presence of 1–3 R-M sites compared to the empty backbone on transformation efficiency of WT *A. woodii*; in blue, for confirmation of the phenotype of *A. woodii* Δ*hsdR1* compared to WT for each plasmid. **: $p = 0.0010$; ***: $p = 0.0002$; ****: $p < 0.0001$.

with the empty backbone was lower than the transformation efficiency of the WT, representing 20.59% and 52.94% of the WT transformation efficiency. More experiments are needed to determine if this is due to the *hsdR1* deletion or to competent cell preparation variations.

As shown by the two-way ANOVA test presented in green in Figure 4C, the presence of the R-M site on plasmids has a significant impact on the transformation efficiency of WT *A. woodii*. Increasing numbers of recognition sites negatively impact the transformation efficiency. Deletion of the HsdR1 subunit of the R-M system significantly increased relative transformation efficiencies obtained with plasmids containing the R-M recognition site compared to WT, as shown by the two-way ANOVA test shown in blue. This

10.3389/fbioe.2023.1213236

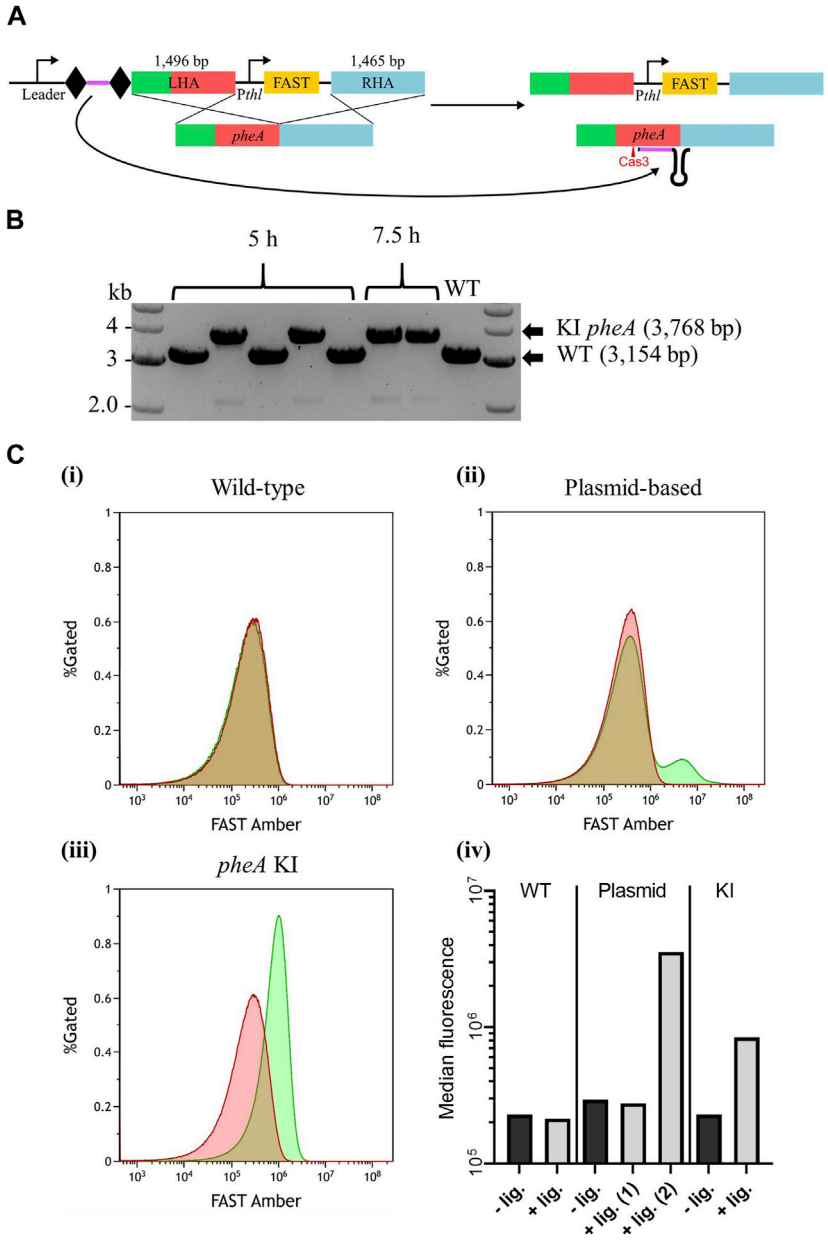


FIGURE 5
Endogenous CRISPR-based insertion of P_{thl}_FAST at the *pheA* locus in the *A. woodii* genome. **(A)** Representation of the application-specific module of the editing plasmid for P_{thl}_FAST insertion at the *pheA* (red) locus in *A. woodii*, pMTL-MPD24. The 36 bp spacer (purple) is flanked by the native DR (diamond) and expressed under the control of the native leader sequence. The editing template with a 1,496 bp LHA (containing *pheA*) and a 1,465 bp RHA (blue) allows for homologous recombination and insertion of the cargo. **(B)** Colony PCR results with primers FW_pheA_KI_seq and RV_pheA_KI_seq of the *pheA* locus of colonies transformed with the editing vector. The WT amplicon is expected at 3,154 bp, and the KI amplicon is expected at 3,768 bp. **(C)** Flow cytometry results for (i) *A. woodii* WT, (ii) *A. woodii* WT with pMTL8415_P_{thl}_FAST, and (iii) *A. woodii* P_{thl}_FAST *pheA* KI. Cell populations with PBS only (red) or with the ^{TF}Amber ligand (green), and the percentage of cells gated is represented as a function of the FAST Amber fluorescence intensity. The number of events represented is >440,000 for WT and >1,200,000 for all other strains; and exact numbers can be found in Materials and Methods. (iv) Median fluorescence of the populations with PBS only (dark gray) or with the ^{TF}Amber ligand (light gray). For the plasmid-based expression strain, the median fluorescence was calculated separately for the two peaks observed and annotated (1) and (2).

confirms recognition of the 5′-TAAGN_5TCC-3′ motif by the HsdR1 subunit. Surprisingly, the presence of two or three R-M sites on the plasmids resulted in an increased relative transformation efficiency of *A. woodii* Δ*hsdR1* compared to plasmids with none or one site.

3.2.6 Validation of the system: reporter gene knock-in at the *pheA* locus

As a final test to the robustness of *A. woodii* Type I-B CRISPR/Cas system as a genetic engineering tool, its capacity to knock-in a reporter gene was tested. As no 5′ CCA trinucleotide could be found in the target

intergenic region downstream of *pheA*, an alternative 5′ CCG PAM was chosen for spacer design. Its selection was based on the aforementioned interference assay, suggesting that the double cytosine alone is essential for efficient targeting. WT *A. woodii* was transformed with the editing vector pMTL-MPD24 (Figure 5A), which was designed to insert a 614 bp DNA fragment expressing the FAST reporter gene downstream of *pheA* in place of the targeted PAM.

Based on the PCR analysis of six colonies from plating at 5 h and 7.5 h, successful knock-ins were obtained with both recovery times (Figure 5B), with an overall efficiency of approximately 57%. Following plasmid loss, plating of the clones onto minimal media with and without supplementary phenylalanine confirmed that the insertion had not disrupted *pheA* (Supplementary Figure S3). Expression of FAST was assayed using flow cytometry (Figure 5C), with WT *A. woodii* and WT *A. woodii* with P_{thl}_FAST in a pMTL84151 plasmid being used as negative and positive controls, respectively. As expected, WT *A. woodii* displayed the same level of fluorescence with and without the ligand. In the presence of the ligand, the plasmid-based population fluorescence was heterogeneous with two clear cell populations; one population displayed a level of fluorescence similar to the level of fluorescence without a ligand and a smaller one displayed a higher level of fluorescence. The KI strain, on the contrary, displayed a homogeneous population with a level of fluorescence higher than that without the ligand. It is important to note that the fluorescence of the KI population (KI + ligand in Figure 5Civ) is lower than the fluorescence of the plasmid-based fluorescent population (plasmid + ligand (2) in Figure 5Civ). The auto-fluorescence of the plasmid-based population is also slightly higher than that of both the WT and KI strains.

3.3 Proof-of-concept: harnessing the endogenous CRISPR system of *Clostridium autoethanogenum*

3.3.1 Description of the CRISPR locus in *Clostridium autoethanogenum* and PAM prediction with Python script

In *C. autoethanogenum*, the CRISPRFinder database identified a Type I-B system between positions 1,491,001 and 1,499,580 in reverse orientation. The database did not predict a Cas8 gene, which is essential for a functional CRISPR system. A protein BLAST of the gene situated between the Cas7 and the Cas6 revealed homologies to proteins labeled Cas8b in other *Clostridium* spp. For instance, an alignment with the Cas8b protein of the Type I-B system of *Clostridium tetani* has an E-value of 1e-148. Overall, the gene organization is the same as the one described for *A. woodii*.

Four CRISPR arrays were identified, including three with an evidence level of 4; array 2 is directly downstream of the Cas locus with 21 spacers, while array 3 (42 spacers) and array 4 (33 spacers) are immediately upstream of the Cas locus. All these arrays have a similar direct repeat, with only one nucleotide varying: "ATTTAAATACATCT(C/T)ATGTTGAGGTTCAAC." This suggests that they are related and should yield coherent results when run in the Python script. Hits with less than 15% mismatch are presented in Supplementary Table S12 (hits with mismatch between 15% and 20% are presented in Supplementary Table S13 for further information). Array 4 had the most hits with 15 in phage regions or phage-related genes.

Overall, the Python script predicted a TCH/TTR putative 5′ PAM. Most notably, the spacer in position 27 of this array is a perfect match to a phage-related gene in *Clostridium ljungdahlii* with, in position 5′, the sequence "CTTCA." The 5′ TCA PAM was selected for harnessing the endogenous system as a genetic engineering tool in *C. autoethanogenum*.

3.3.2 Genome engineering capacity validation by *pyrE* knock-out in *C. autoethanogenum*

Design of the *pyrE* KO editing vector was based both on a 561 bp in-frame *pyrE* deletion already performed in *C. autoethanogenum* in our laboratory with CRISPR/Cas9 and on its native CRISPR array 4. WT *C. autoethanogenum* was conjugated with three plasmids containing HAs alone, expressed spacer alone, and both HAs and expressed spacer, as described for *A. woodii* in Figure 2A. Colony PCR analysis of transformants (Figure 6A) demonstrated that the editing vector containing both the HAs and the expressed spacer resulted in an editing efficiency of 100%. In contrast, vectors harboring only one part of the editing cassette did not result in any Δ*pyrE* mutants. Plasmid loss was achieved; the additional subculture without selection as described in Materials and Methods increased the plasmid loss rate from 33.3% to 58.3% (out of 12 colonies tested in both conditions). The phenotype of the strains obtained after plasmid loss was confirmed by plating on minimal media with and without uracil supplementation (Figure 6B). This is the first report of the successful use of the *C. autoethanogenum* Type I-B endogenous system for genome editing and demonstrates the robustness of the developed workflow—from the automated PAM identification pipeline to gene knock-out.

4 Discussion

4.1 *In silico* analyses

Endogenous CRISPR/Cas systems offer alternative genome editing tools for microorganisms that contain them. They have, in some cases, proven to be superior to using the highly toxic *S. pyogenes* Cas9 (Pyne et al., 2016; Zhou et al., 2021) and have enabled genome editing of otherwise genetically recalcitrant bacteria (Hidalgo-Cantabrana et al., 2019). PAM identification is the first necessary step. *In silico* strategies are widely employed (Zhou et al., 2011; Arndt et al., 2019), requiring the analysis of considerable volumes of data to derive an appropriate catalog of potential PAM candidates. A variant of that strategy is the use of CRISPRTarget, a tool created to predict the targets of CRISPR RNAs (Biswas et al., 2013; Biswas et al., 2015) and to characterize the exposure history of the strains (Zheng et al., 2019; Xu et al., 2020). *In vivo* techniques are also employed to identify PAMs such as plasmid depletion assays (Walker et al., 2020) or PAM-SCANR (Leenay et al., 2016). These require extensive library construction or cloning and rely on large-

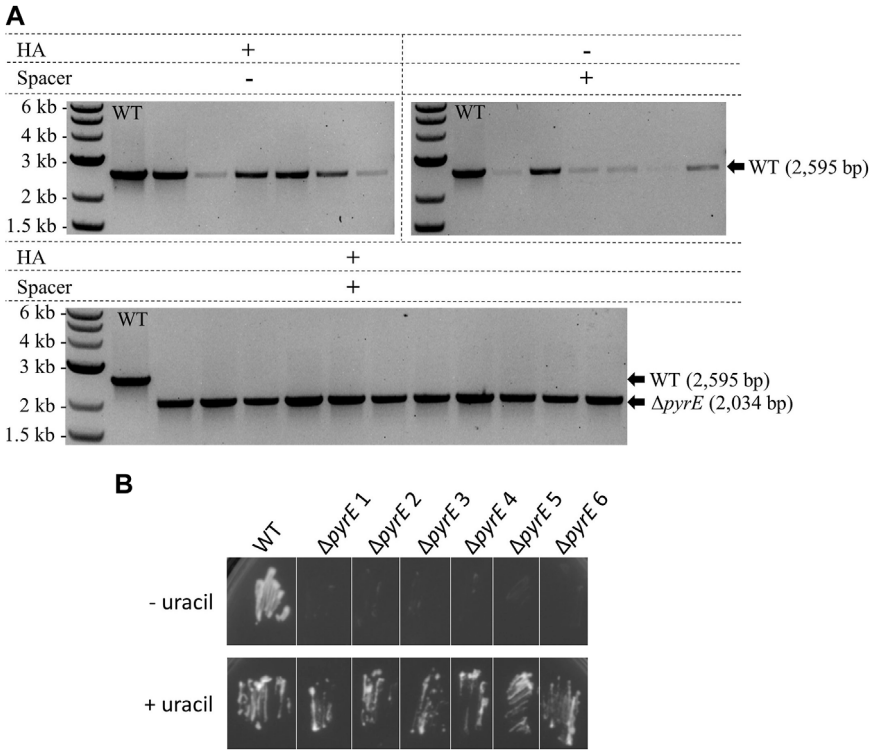


FIGURE 6
Endogenous CRISPR-based deletion of the *C. autoethanogenum pyrE* gene. **(A)** Colony PCR results for the *pyrE* locus of *C. autoethanogenum* with primers FW_pyrECLAU_screening and RV_pyrECLAU_screening; the WT amplicon is expected at 2,595 bp and the KO amplicon at 2,034 bp. + and —indicate the presence or absence of the corresponding element, respectively. **(B)** Phenotypic analysis of the *C. autoethanogenum* Δ*pyrE* strains obtained. WT and six colonies of Δ*pyrE C. autoethanogenum* strains obtained after plasmid loss were plated on minimal media supplemented (+ uracil) or not (- uracil) with 20 μg/mL of uracil.

scale sequencing. Accordingly, PAM identification is recognized as one of the challenges of exploiting endogenous CRISPR systems for genetic engineering (Zhou et al., 2021).

In this study, a dedicated bioinformatics tool was successfully developed to allow fully automated predictions based on data from the CRISPRFinder database. In the meantime, two other tools to identify PAMs have been published (Vink et al., 2021; Rybnicky et al., 2022). In all three tools, spacers are run through BLAST, and the flanking sequences of the matches are retrieved to compile a prediction of the PAM. The main advancement made over preceding tools is filtering of hits. False-positives are filtered based either on global mismatch (Vink et al., 2021) or a combination of parameters, including gap number, e-value, length of the hit, and starting position of the hit (Rybnicky et al., 2022). Our Python script filters on global mismatch but is more stringent in the seed region adjacent to the PAM, as interference relies on the recognition of a seed region (Semenova et al., 2011; Xue et al., 2015). As CRISPR/Cas systems are prokaryotic defense mechanisms against invading elements (Makarova and Koonin, 2015), filtering of biologically relevant hits relies on their presence in mobile genetic elements (MGEs), including plasmids, phages, prophages, and any related genes. The similar performances between Spacer2PAM (Rybnicky et al., 2022) and our Python script, despite a more comprehensive filtering parameter selection for false-positives in Spacer2PAM, could be explained by our refined selection for biologically relevant hit sequences.

As *C. autoethanogenum* is an important industrial strain, its CRISPR/Cas system has been of interest in recent years (Pyne et al., 2016; Rybnicky et al., 2022) and can serve as a case study to compare the different methods of PAM elucidation. The output of our Python script was two putative 5′ PAMs, TCH and TTR, and TCA was confirmed *in vivo* by genome editing. In previously performed manual PAM prediction (Pyne et al., 2016), the array direction was incorrect, leading to the 3′ PAM being wrongly selected as the potential PAM. Spacer2PAM only predicted a TTNN PAM consensus, while a NYCN consensus was experimentally confirmed (Rybnicky et al., 2022). Lastly, Vink et al. (2021). showed a direct link between the direct repeat sequence and the identified PAMs. Spacers from different microorganisms with identical direct repeats were clustered and PAM prediction achieved for each cluster. The PAM prediction corresponding to *C. autoethanogenum* direct repeats was TCNX, corroborated by *in vivo* experiments performed in this study and by Rybnicky et al. (2022).

For the *A. woodii* Type I-B CRISPR system, the Python script did not manage to compute a clear consensus, but allowed the rapid identification of matches with MGEs and their corresponding PAM candidates. *In vivo* experiments subsequently confirmed the functionality of one of these PAMs.

CRISPR systems have been shown to recognize a variety of PAMs. Mismatched PAMs can lead to interference or primed adaptation depending on the spacer analyzed (Xue et al., 2015),

and it is hypothesized that a wider variety of PAMs can be recognized for interference compared to adaptation. This would result in a better defense system for the host microorganism, which would be less likely to suffer from evasion from mutated MGEs or closely related MGEs (Xue et al., 2015; Rybnicky et al., 2022). This diversity of functional PAMs is proving to be challenging for both *in silico* prediction and *in vivo* validation and could explain the complex results obtained from the Python script for some microorganisms. Conveniently, Spacer2PAM (Rybnicky et al., 2022) provides a PAM score to help the user judge the quality of each PAM within one CRISPR/Cas system. Future tools could use that score but also optimize the output to reflect the functional PAM diversity without compromising user-friendliness and clarity.

A better understanding of the adaptation and the interference mechanisms is necessary to adapt the parameters for PAM identification, but all of those recently developed tools and their outputs pave the way for a PAM prediction pipeline combining (1) spacer BLAST to find MGE matches and (2) direct repeat homology to find candidate PAMs based on both *de novo* identification and mining of published PAMs.

4.2 Proof-of-concept for the use of the endogenous CRISPR/Cas system for genome engineering in *A. woodii* and *C. autoethanogenum*

In this study, proof-of-concept for the complete workflow of genome editing with the endogenous CRISPR/Cas system was established in *A. woodii*. After *in silico* analyses, the PAM was confirmed *in vivo* and the leader sequence was characterized. Similar to the interference assay completed by Rybnicky et al. (2022), the transformation efficiencies observed were variable between the five replicates. However, the assay showed that the double cytosine is essential for interference in *A. woodii,* and this result was confirmed by successful genome engineering experiments with both CCA and CCG PAMs.

When using the endogenous CRISPR system, providing homology arms for homologous recombination allows evasion from self-targeting and cell death. Other pathways exist for CRISPR-based targeting evasion, including disruption of targeting by mutation of either the target site or the targeting mechanisms (Wimmer and Beisel, 2019). Analysis of the escape mutants obtained in our study predominantly revealed inactivation of the CRISPR-based targeting by loss of the plasmid-encoded spacer, likely through recombination of the direct repeats (Gomaa et al., 2014; Canez et al., 2019). Consequently, optimizing DNA transfer and homologous recombination is essential; different parameters are reported to have an impact on editing efficiency, including spacer design and expression, Gram-positive replicon, editing template length, and cell density; they were each investigated with successive deletion targets and are discussed below.

In *A. woodii* and *C. autoethanogenum*, the spacers were chosen to have lengths corresponding to the average length of the protospacers in the native arrays, and a GC content of between 40% and 60% (Javaid and Choi, 2021; Zhou et al., 2021). Other parameters for spacer design appear to have an impact on interference efficiency but remain elusive. In bacteria, the spacer sequence was shown to have an impact on interference (Xue et al., 2015; Bourgeois et al., 2020). Studies performed in mammalian cells with CRISPR/Cas9 have demonstrated that the sgRNA sequence impacts interference, likely through sgRNA transcription and stability (Doench et al., 2014; Wu et al., 2014). Target sequence accessibility also modulates interference efficiency (Wu et al., 2014; Chari et al., 2015).

Spacer expression replicating the native arrays was effective in *A. woodii* and *C. autoethanogenum,* as was also found in previous studies (Pyne et al., 2016; Hidalgo-Cantabrana et al., 2019; Zheng et al., 2019). Other studies have found that expression with the leader sequence can be inadequate (high or low), leading either to cell death, potentially because spacer expression happens before successful double crossover occurs (Maikova et al., 2019; Zhou et al., 2021), or to inefficient genome editing (Walker et al., 2020; Xu et al., 2020). In these cases, genome editing was successfully achieved by expression of the array with an inducible promoter or a strong constitutive promoter, respectively.

The initial generation of *pyrE* KO in *A. woodii* was successful using both the pCD6 and pBP1 replicons. Subsequent editing experiments were confined to plasmids based on the former replicon in view of its comparative instability (Baker et al., 2022), a desirable property when plasmid loss is needed. Targeting of *pheA* confirmed that the use of editing templates comprising larger HAs (1.5 kb compared to 0.5 kb and 1.0 kb) has a positive impact on genome editing efficiency (Zheng et al., 2019). Not unexpectedly, editing efficiencies were also enhanced through the use of more effective transformation procedures. In *C. autoethanogenum*, Gram-positive replicon, homology arms, and transformation protocol previously used for CRISPR/Cas9 genome editing led to successful editing with its endogenous CRISPR system.

As expected, deletion of the *hsdR1* gene in *A. woodii* resulted in an increased relative frequency of transfer of plasmids carrying its corresponding restriction recognition site(s). Interestingly, plasmids containing two or three R-M sites resulted in slightly higher transformation efficiencies of the *A. woodii* Δ*hsdR1* strain compared to the backbone alone or with just one R-M site. This suggests that the methylation and specificity subunits of the system confer an advantage for transformation of *A. woodii* with plasmids containing the R-M recognition site in the absence of the restriction subunit. In *E. coli* K-12, a point mutation in the restriction subunit resulted in an r^-m^* strain with enhanced modification (Kelleher et al., 1991). Adenine-specific methylation systems have complex and diverse functions in bacterial cells, including, but not limited to, regulation of conjugation, DNA replication, and various transcriptional and posttranscriptional gene regulations (Wion and Casadesus, 2006; Low and Casadesus, 2008; Vasu and Nagaraja, 2013; Adhikari and Curtis, 2016). Further study is necessary for a full understanding of the impact of the *hsdR1* deletion in *A. woodii.*

It is interesting to note that when the FAST reporter was present on an autonomous plasmid in *A. woodii,* FAST expression showed a heterogeneous population. Similar studies have observed the same phenomenon in *A. woodii* (Mook et al., 2022), in *C. acetobutylicum* (Streett et al., 2019), and *Eubacterium limosum* (Flaiz et al., 2021). The reason for this heterogeneity is not entirely clear, but it was not observed when the FAST gene was integrated into the genome, indicating that it is a consequence of the reporter gene being localized on an autonomous plasmid. Accordingly, if population heterogeneity is to be avoided in the future, desirable metabolic pathways should be integrated into the genome.

5 Conclusion

This study reports the development of a Python script for fast and efficient PAM elucidation for the use of endogenous CRISPR/Cas systems as genome engineering tools. After validation against published literature, it allowed for characterization of Type I-B endogenous systems of two industrially relevant microorganisms for their use as genome engineering tools in only a few steps. Editing cassettes were designed to be synthetic CRISPR arrays, replicating the native arrays of each host microorganism. They consisted of the leader sequence expressing an adequate spacer, flanked by direct repeats. The editing efficiencies varied between targets and between microorganisms, but this work is the first report of the successful use of the endogenous CRISPR/Cas systems of *A. woodii* and *C. autoethanogenum* for genetic engineering.

Compared to most CRISPR/Cas9-based techniques, the workflow presented here represents a major improvement for genome editing tools for recalcitrant microorganisms like *A. woodii* and will allow the leveraging of a widespread range of bacteria containing endogenous CRISPR systems. It removes the need for sophisticated inducible promoter systems necessary to mitigate *cas9* toxicity (Canadas et al., 2019) and reduces plasmid size; with no induction step required, it also reduces the number of subcultures performed and the risk of SNP introduction. Overall, the method is fast and results in high editing efficiencies.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization: NM, JB, and MP-D; investigation, methodology, bioinformatics, and data analysis: MP-D; supervision: JB, KW, and NM; writing—original draft preparation: MP-D; writing—review and editing: JM and NM; project administration and funding acquisition: NM. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Biotechnology and Biological Sciences Research Council (Grant Numbers BB/L013940/1 and BB/W01453X/1) and by the European Commission HORIZON 2020 Grant 760994 ENGICOIN.

Acknowledgments

MP-D acknowledges the receipt of a studentship from the University of Nottingham, Synthetic Biology Research Center (SBRC) Doctoral Training Program. The authors thank Dr. David Onion and Maria-Isabel Haig for the flow cytometry results used in Figure 5C.

Conflict of interest

The 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.

The handling editor MK declared a past collaboration with the author NPM.

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.2023.1213236/full#supplementary-material

References

Adhikari, S., and Curtis, P. D. (2016). DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol. Rev.* 40, 575–591. doi:10.1093/femsre/fuw023

Alkhnbashi, O. S., Shah, S. A., Garrett, R. A., Saunders, S. J., Costa, F., and Backofen, R. (2016). Characterizing leader sequences of CRISPR loci. *Bioinformatics* 32, 576–i585. doi:10.1093/bioinformatics/btw454

Arndt, D., Marcu, A., Liang, Y., and Wishart, D. S. (2019). PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. *Brief. Bioinform* 20, 1560–1567. doi:10.1093/bib/bbx121

Baker, J. P., Saez-Saez, J., Jensen, S. I., Nielsen, A. T., and Minton, N. P. (2022). A clean in-frame knockout system for gene deletion in *Acetobacterium woodii*. *J. Biotechnol.* 353, 9–18. doi:10.1016/j.jbiotec.2022.05.013

Biswas, A., Fineran, P. C., and Brown, C. M. (2015). Computational detection of CRISPR/crRNA targets. *Methods Mol. Biol.* 1311, 77–89. doi:10.1007/978-1-4939-2687-9_5

Biswas, A., Gagnon, J. N., Brouns, S. J., Fineran, P. C., and Brown, C. M. (2013). CRISPRTarget: Bioinformatic prediction and analysis of crRNA targets. *RNA Biol.* 10, 817–827. doi:10.4161/rna.24046

Boudry, P., Semenova, E., Monot, M., Datsenko, K. A., Lopatina, A., Sekulovic, O., et al. (2015). Function of the CRISPR-cas system of the human pathogen *Clostridium difficile*. *mBio* 6, 01112–01115. doi:10.1128/mbio.01112-15

Bourgeois, J., Lazinski, D. W., and Camilli, A. (2020). Identification of spacer and protospacer sequence requirements in the *Vibrio cholerae* type I-E CRISPR/cas system. *mSphere* 5, 00813–00820. doi:10.1128/msphere.00813-20

Canadas, I. C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., and Minton, N. P. (2019). RiboCas: A universal CRISPR-based editing tool for *Clostridium*. *ACS Synth. Biol.* 8, 1379–1390. doi:10.1021/acssynbio.9b00075

Canez, C., Selle, K., Goh, Y. J., and Barrangou, R. (2019). Outcomes and characterization of chromosomal self-targeting by native CRISPR-Cas systems in *Streptococcus thermophilus*. *FEMS Microbiol. Lett.* 366, 105. doi:10.1093/femsle/fnz105

Chari, R., Mali, P., Moosburner, M., and Church, G. M. (2015). Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat. Methods* 12, 823–826. doi:10.1038/nmeth.3473

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Neron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version,

enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251. doi:10.1093/nar/gky425

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: A sequence logo generator: Figure 1. *Genome Res.* 14, 1188–1190. doi:10.1101/gr.849004

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., et al. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* 32, 1262–1267. doi:10.1038/nbt.3026

Fischer, S., Maier, L. K., Stoll, B., Brendel, J., Fischer, E., Pfeiffer, F., et al. (2012). An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. *J. Biol. Chem.* 287, 33351–33363. doi:10.1074/jbc.m112.377002

Flaiz, M., Ludwig, G., Bengelsdorf, F. R., and Durre, P. (2021). Production of the biocommodities butanol and acetone from methanol with fluorescent FAST-tagged proteins using metabolically engineered strains of *Eubacterium limosum*. *Biotechnol. Biofuels* 14, 117. doi:10.1186/s13068-021-01966-2

Gomaa, A. A., Klumpe, H. E., Luo, M. L., Selle, K., Barrangou, R., and Beisel, C. L. (2014). Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *mBio* 5, 00928–1013. doi:10.1128/mbio.00928-13

Grissa, I., Vergnaud, G., and Pourcel, C. (2007b). CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi:10.1093/nar/gkm360

Grissa, I., Vergnaud, G., and Pourcel, C. (2007a). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinforma.* 8, 172. doi:10.1186/1471-2105-8-172

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

Hidalgo-Cantabrana, C., Goh, Y. J., Pan, M., Sanozky-Dawes, R., and Barrangou, R. (2019). Genome editing using the endogenous type I CRISPR-Cas system in *Lactobacillus crispatus*. *Proc. Natl. Acad. Sci. U. S. A.* 116, 15774–15783. doi:10.1073/pnas.1905421116

Huang, H., Chai, C., Li, N., Rowe, P., Minton, N. P., Yang, S., et al. (2016). CRISPR/Cas9-Based efficient genome editing in *Clostridium ljungdahlii*, an autotrophic gas-fermenting bacterium. *ACS Synth. Biol.* 5, 1355–1361. doi:10.1021/acssynbio.6b00044

Javaid, N., and Choi, S. (2021). CRISPR/Cas system and factors affecting its Precision and efficiency. *Front. Cell. Dev. Biol.* 9, 761709. doi:10.3389/fcell.2021.761709

Kelleher, J. E., Daniel, A. S., and Murray, N. E. (1991). Mutations the confer *de Novo* activity upon a maintenance methyltransferase. *J. Mol. Bio* 221, 431–440. doi:10.1016/0022-2836(91)80064-2

Leenay, R. T., Maksimchuk, K. R., Slotkowski, R. A., Agrawal, R. N., Gomaa, A. A., Briner, A. E., et al. (2016). Identifying and visualizing functional PAM diversity across CRISPR-cas systems. *Mol. Cell.* 62, 137–147. doi:10.1016/j.molcel.2016.02.031

Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., et al. (2016). CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnol. J.* 11, 961–972. doi:10.1002/biot.201600053

Liew, F., Martin, M. E., Tappel, R. C., Heijstra, B. D., Mihalcea, C., and Kopke, M. (2016). Gas fermentation-A flexible Platform for commercial scale production of low-carbon-fuels and chemicals from waste and renewable feedstocks. *Front. Microbiol.* 7, 694. doi:10.3389/fmicb.2016.00694

Low, D. A., and Casadesus, J. (2008). Clocks and switches: Bacterial gene regulation by DNA adenine methylation. *Curr. Opin. Microbiol.* 11, 106–112. doi:10.1016/j.mib.2008.02.012

Maikova, A., Kreis, V., Boutserin, A., Severinov, K., and Soutourina, O. (2019). Using an endogenous CRISPR-cas system for genome editing in the human Pathogen *Clostridium difficile*. *Appl. Environ. Microbiol.* 85, 01416–01419. doi:10.1128/aem.01416-19

Makarova, K. S., and Koonin, E. V. (2015). Annotation and classification of CRISPR-cas systems. *Methods Mol. Biol.* 1311, 47–75. doi:10.1007/978-1-4939-2687-9_4

McAllister, K. N., and Sorg, J. A. (2019). CRISPR genome editing systems in the genus *Clostridium*: A timely advancement. *J. Bacteriol.* 201, 00219–19. doi:10.1128/jb.00219-19

Mook, A., Beck, M. H., Baker, J. P., Minton, N. P., Dürre, P., and Bengelsdorf, F. R. (2022). Autotrophic lactate production from H(2) + CO(2) using recombinant and fluorescent FAST-tagged *Acetobacterium woodii* strains. *Appl. Microbiol. Biotechnol.* 106, 1447–1458. doi:10.1007/s00253-022-11770-z

Moon, S. B., Kim, D. Y., Ko, J. H., and Kim, Y. S. (2019). Recent advances in the CRISPR genome editing tool set. *Exp. Mol. Med.* 51, 1–11. doi:10.1038/s12276-019-0339-7

Mulepati, S., and Bailey, S. (2013). *In vitro* reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J. Biol. Chem.* 288, 22184–22192. doi:10.1074/jbc.m113.472233

Nishimasu, H., and Nureki, O. (2017). Structures and mechanisms of CRISPR RNA-guided effector nucleases. *Curr. Opin. Struct. Biol.* 43, 68–78. doi:10.1016/j.sbi.2016.11.013

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., et al. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell.* 156, 935–949. doi:10.1016/j.cell.2014.02.001

Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A., and Chou, C. P. (2016). Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. *Sci. Rep.* 6, 25666. doi:10.1038/srep25666

Rousseau, C., Gonnet, M., Le Romancer, M., and Nicolas, J. (2009). Crispi: A CRISPR interactive database. *Bioinformatics* 25, 3317–3318. doi:10.1093/bioinformatics/btp586

Rybnicky, G. A., Fackler, N. A., Karim, A. S., Kopke, M., and Jewett, M. C. (2022). Spacer2PAM: A computational framework to guide experimental determination of functional CRISPR-cas system PAM sequences. *Nucleic Acids Res.* 50, 3523–3534. doi:10.1093/nar/gkac142

Schuchmann, K., and Muller, V. (2016). Energetics and application of heterotrophy in acetogenic bacteria. *Appl. Environ. Microbiol.* 82, 4056–4069. doi:10.1128/aem.00882-16

Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., et al. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10098–10103. doi:10.1073/pnas.1104144108

Seys, F. M., Rowe, P., Bolt, E. L., Humphreys, C. M., and Minton, N. P. (2020). A gold standard, CRISPR/Cas9-Based complementation strategy reliant on 24 nucleotide bookmark sequences. *Genes. (Basel)* 11, 458. doi:10.3390/genes11040458

Solovyev, V., and Salamov, A. (2011). "Automatic annotation of microbial genomes and metagnomic sequences," in *Metagenomics and its applications in agriculture, biomedicine and environmental studies*. Editors W. Robert and P. Li (New York: Nova Science Publishers, Inc), 61–78.

Streett, H. E., Kalis, K. M., and Papoutsakis, E. T. (2019). A strongly fluorescing anaerobic reporter and protein-tagging system for *Clostridium* organisms based on the Fluorescence-Activating and Absorption-Shifting Tag (FAST) protein. *Appl. Environ. Microbiol.* 85, 00622–719. doi:10.1128/aem.00622-19

Vasu, K., and Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.* 77, 53–72. doi:10.1128/mmbr.00044-12

Vink, J. N. A., Baijens, J. H. L., and Brouns, S. J. J. (2021). PAM-repeat associations and spacer selection preferences in single and co-occurring CRISPR-Cas systems. *Genome Biol.* 22, 281. doi:10.1186/s13059-021-02495-9

Walker, J. E., Lanahan, A. A., Zheng, T., Toruno, C., Lynd, L. R., Cameron, J. C., et al. (2020). Development of both type I-B and type II CRISPR/Cas genome editing systems in the cellulolytic bacterium *Clostridium thermocellum*. *Metab. Eng. Commun.* 10, 00116. doi:10.1016/j.mec.2019.e00116

Wang, Y., Zhang, Z. T., Seo, S. O., Choi, K., Lu, T., Jin, Y. S., et al. (2015). Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *J. Biotechnol.* 200, 1–5. doi:10.1016/j.jbiotec.2015.02.005

Wiedenheft, B., Van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., et al. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10092–10097. doi:10.1073/pnas.1102716108

Wimmer, F., and Beisel, C. L. (2019). CRISPR-cas systems and the paradox of self-targeting spacers. *Front. Microbiol.* 10, 3078. doi:10.3389/fmicb.2019.03078

Wion, D., and Casadesus, J. (2006). N6-methyl-adenine: An epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.* 4, 183–192. doi:10.1038/nrmicro1350

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

Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C., Hsu, P. D., Dadon, D. B., et al. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* 32, 670–676. doi:10.1038/nbt.2889

Xu, Z., Li, Y., and Yan, A. (2020). Repurposing the native type I-F CRISPR-cas system in *Pseudomonas aeruginosa* for genome editing. *Star. Protoc.* 1, 100039. doi:10.1016/j.xpro.2020.100039

Xue, C., Seetharam, A. S., Musharova, O., Severinov, K., Brouns, S. J., Severin, A. J., et al. (2015). CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.* 43, 10831–10847. doi:10.1093/nar/gkv1259

Zhang, J., Zong, W., Hong, W., Zhang, Z. T., and Wang, Y. (2018). Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. *Metab. Eng.* 47, 49–59. doi:10.1016/j.ymben.2018.03.007

Zheng, Y., Han, J., Wang, B., Hu, X., Li, R., Shen, W., et al. (2019). Characterization and repurposing of the endogenous Type I-F CRISPR-Cas system of *Zymomonas mobilis* for genome engineering. *Nucleic Acids Res.* 47, 11461–11475. doi:10.1093/nar/gkz940

Zhou, X., Wang, X., Luo, H., Wang, Y., Wang, Y., Tu, T., et al. (2021). Exploiting heterologous and endogenous CRISPR-Cas systems for genome editing in the probiotic *Clostridium butyricum*. *Biotechnol. Bioeng.* 118, 2448–2459. doi:10.1002/bit.27753

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). Phast: A fast phage search tool. *Nucleic Acids Res.* 39, W347–W352. doi:10.1093/nar/gkr485

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 10 July 2023
DOI 10.3389/fbioe.2023.1211197

OPEN ACCESS

EDITED BY
Michael Köpke,
LanzaTech, United States

REVIEWED BY
Marie Schoelmerich,
University of California, Berkeley,
United States
Benjamin Michael Woolston,
Northeastern University, United States

*CORRESPONDENCE
Nigel P. Minton,
nigel.minton@nottingham.ac.uk

RECEIVED 24 April 2023
ACCEPTED 06 June 2023
PUBLISHED 10 July 2023

CITATION
Seys FM, Humphreys CM, Tomi-Andrino C, Li Q, Millat T, Yang S and Minton NP (2023), Base editing enables duplex point mutagenesis in *Clostridium autoethanogenum* at the price of numerous off-target mutations. *Front. Bioeng. Biotechnol.* 11:1211197. doi: 10.3389/fbioe.2023.1211197

Base editing enables duplex point mutagenesis in *Clostridium autoethanogenum* at the price of numerous off-target mutations

François M. Seys[1], Christopher M. Humphreys[1], Claudio Tomi-Andrino[1,2,3], Qi Li[4], Thomas Millat[1], Sheng Yang[5] and Nigel P. Minton[1]*

[1]Clostridia Research Group, BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Biodiscovery Institute, University of Nottingham, Nottingham, United Kingdom, [2]Centre for Analytical Bioscience, Advanced Materials and Healthcare Technologies Division, School of Pharmacy, University of Nottingham, Nottingham, United Kingdom, [3]Nottingham BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Mathematical Sciences, University of Nottingham, Nottingham, United Kingdom, [4]College of Life Sciences, Sichuan Normal University, Chengdu, China, [5]Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai, China

Base editors are recent multiplex gene editing tools derived from the Cas9 nuclease of *Streptomyces pyogenes*. They can target and modify a single nucleotide in the genome without inducing double-strand breaks (DSB) of the DNA helix. As such, they hold great potential for the engineering of microbes that lack effective DSB repair pathways such as homologous recombination (HR) or non-homologous end-joining (NHEJ). However, few applications of base editors have been reported in prokaryotes to date, and their advantages and drawbacks have not been systematically reported. Here, we used the base editors Target-AID and Target-AID-NG to introduce nonsense mutations into four different coding sequences of the industrially relevant Gram-positive bacterium *Clostridium autoethanogenum*. While up to two loci could be edited simultaneously using a variety of multiplexing strategies, most colonies exhibited mixed genotypes and most available protospacers led to undesired mutations within the targeted editing window. Additionally, fifteen off-target mutations were detected by sequencing the genome of the resulting strain, among them seven single-nucleotide polymorphisms (SNP) in or near loci bearing some similarity with the targeted protospacers, one 15 nt duplication, and one 12 kb deletion which removed uracil DNA glycosylase (UDG), a key DNA repair enzyme thought to be an obstacle to base editing mutagenesis. A strategy to

Abbreviations: ABE adenine base editors; AID activation-induced cytidine deaminase; BER Base excision repair; Cas CRISPR-associated protein; CBE cytosine base editors; CDS coding sequences; CRISPR clustered regularly interspersed repeats; CDA cytosine deaminase; Cm chloramphenicol; D-cyc D-cycloserine; DR direct repeat; DSB double-strand breaks; 5-FOA 5-Fluoroorotic acid; HR homologous recombination; Kan kanamycin; LB Luria-Bertani broth; nCas9 Cas9 nickase; NHEJ non-homologous end-joining; kb kilobase; nt nucleotide; MMR mismatch repair; mCRISPR multiplexing with a synthetic *S. pyogenes* CRISPR array; msgRNA multiplexing with multiple sgRNA expression cassettes; mtRNA multiplexing with sgRNA-tRNA fusions mtRNA; NEB New England Biolabs Ltd.; PAM protospacer adjacent motifs; PCR polymerase chain reactions; SNP single-nucleotide polymorphism; Th theophylline; Tm thiamphenicol; tracrRNA trans-activating CRISPR RNA; UDG uracil DNA glycosylase; UGI uracil glycosylase inhibitor; YTF yeast tryptone fructose.

process prokaryotic single-guide RNA arrays by exploiting tRNA maturation mechanisms is also illustrated.

KEYWORDS

base editor, target-AID, *Clostridium*, multiplex mutagenesis, Cas9, off-target mutagenesis, alternative PAM, tRNA

1 Introduction

Base editors are recent gene editing tools derived from the Cas9 nuclease of *Streptomyces* pyogenes (Komor et al., 2016; Nishida et al., 2016). They can target and modify a single nucleotide in the genome without inducing double-strand breaks (DSB) of the DNA helix. This can be exploited to change a single amino acid for *in vivo* protein engineering, or, more commonly, to disrupt protein expression with nonsense mutations. Consequently, base editors are promising tools for the engineering of microbes that lack effective DSB repair pathways such as homologous recombination (HR) or non-homologous end-joining (NHEJ) (Wang et al., 2018; Luo et al., 2020). Additionally, base editors are simple to customize to a new target gene, requiring only to swap their 20 nt protospacer. This makes them readily compatible with high-throughput automated workflows (Wang et al., 2018). These combined features make them ideal candidates for multiplex genome editing tools, allowing the engineering of several loci in a single step and thereby significantly reducing the duration, cost and effort involved in mutagenesis procedures (Wang et al., 2020). In view of their advantages, in the current study the potential of base editors as multiplex genome editing tools was tested in *Clostridium* autoethanogenum. This Gram-positive acetogen is the process chassis used in the large-scale, commercial manufacture of ethanol from industrial off-gas (Fackler et al., 2021; Liew et al., 2022), characterised as slow-growing and challenging to engineer (Bourgade et al., 2021).

Several base editors have been developed to date, enabling C-to-T (cytosine base editors, CBE) or A-to-G (adenine base editors, ABE) (Gaudelli et al., 2017) targeted point mutations within various editing windows around the protospacer-adjacent motifs (PAMs) of their respective CRISPR-associated (Cas) nucleases (Banno et al., 2018; Eid et al., 2018; Jiang et al., 2018; Li et al., 2018; Marx, 2018; Molla & Yang, 2019; Chatterjee et al., 2020). Target-AID, a CBE, exploits the targeting capabilities of single guide RNAs (sgRNAs) and the activity of a cytidine deaminase (CDA), to deaminate a cytosine on a single strand of the DNA helix (Figure 1) in between the positions −20 and −16 from a PAM with the sequence NGG (Nishida et al., 2016). Then, the mismatch repair pathway (MMR) is hijacked by nicking the unedited strand with Cas9^{D10A} nickase (nCas9 ^{D10A}), which is thought to trick the MMR pathway into using the edited strand as template to repair the unedited strand (Modrich, 2006; Li, 2008; Spampinato et al., 2009; Fukui, 2010; Williams & Kunkel, 2014). However, the base excision repair pathway (BER), initiated by the specialized enzyme uracil DNA glycosylase (UDG), can sometimes remove the deaminated cytosine before the MMR pathway has had a chance mutate the unedited strand. For this reason, many CBEs are fused to another enzyme called uracil glycosylase inhibitor (UGI) to protect the targeted locus from UDG and the BER pathway during the early stages of

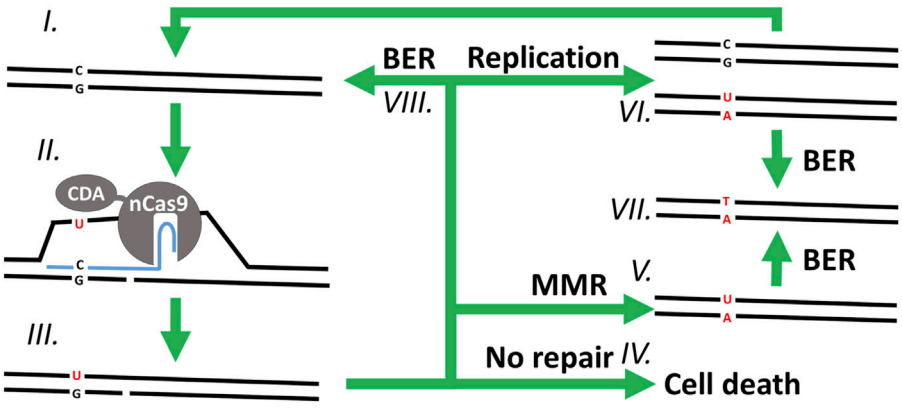


FIGURE 1
Overview of the proposed mutagenesis mechanism of CBEs. In a clockwise order starting from (I.) with the WT chromosomic DNA. (II.) The CBE-sgRNA duplex unwinds the DNA double helix around the protospacer, creating a R-loop, nicking the non-edited strand, and exposing a cytosine (C) to the deaminase activity of CDA on the edited strand—which (III.) changes it into uracil (U). *(IV.)* Without repair of the nick on the non-edited strand, the cell is unable to replicate its DNA and dies. *(V.)* If the DNA helix undergoes MMR, the nicked non-edited strand is repaired to match the edited strand, replacing the guanine (G) of the non-edited strand with an adenine (A). (VI.) Alternatively, the nick can be ligated, and the DNA replicated semi-conservatively into two double helixes: one mutated and the other one WT. (VII.) In the last step, the uracil of the edited strand is finally removed and replaced by a thymine (T) through DNA replication (not shown) or BER. Finally, immediate ligation of the nicked strand and subsequent repair of the edited strand through (VIII.) BER (or (VI.) DNA replication) would result in a stable WT chromosome which would be exposed to another cycle of mutagenesis for as long as the base editor and its sgRNA cassette are being expressed.

mutagenesis (Zhigang et al., 1991; Komor et al., 2016; Nishida et al., 2016; Wang et al., 2017).

Unfortunately, UGI has also been associated with extra toxicity and off-target mutagenesis in prokaryotes, even after the addition of a GLVA protein degradation tag (Banno et al., 2018). Consequently, we opted to avoid using a UGI fusion in the initial stages of our study. It has also been reported that altering the size of the spacer region of the sgRNA could change the base editing profile of Target-AID in *Escherichia coli*: instead of preferentially mutating the base in position −18 from the PAM with a standard 20 nt spacer, a 18 nt spacer would preferentially edit the base in position −17, and a 22 nt spacer would preferably edit the base in position −19 (Banno et al., 2018). We opted to test if this flexibility could also be exploited in *Clostridium autoethanogenum*.

To take full advantage of Target-AID's simplicity, several multiplexing strategies (Adiego-Pérez et al., 2019) that could target several protospacers at once with a single plasmid were also tested. Among more traditional strategies such as using multiple sgRNA transcriptional units (msgRNA) and reproducing the native SpCas9 CRISPR array (mCRISPR), the exploitation of a polycistronic array of sgRNAs and prokaryotic tRNAs fusions (mtRNA) was also explored with the expectation that individual sgRNAs would be released as a consequence of tRNA maturation. This is a common strategy in eukaryotic systems (Xie et al., 2015; Dong et al., 2017; Zhang et al., 2019), but, to the best of our knowledge, it has not yet been used in prokaryotes except for a single recent report in a nonmainstream journal (Lu et al., 2022).

Finally, and perhaps most importantly, the mutant strains were validated by whole-genome sequencing to look for potential off-target mutations.

2 Materials and methods

2.1 Strains and media

Vector assembly and cloning was conducted in *E. coli* strain DH5α, cultivated in Luria-Bertani (LB) broth. *E. coli* strain sExpress (Woods et al., 2019) was used as the conjugal DNA donor strain to transfer plasmids into *C. autoethanogenum* strain DSM10061. *C. autoethanogenum* was recovered from cryostocks and cultivated in pre-reduced yeast tryptone fructose (YTF) medium in an anaerobic cabinet (Don Whitley Scientific Ltd., Bingley, United Kingdom) at 37°C (Humphreys et al., 2015). Antibiotics and other additives to LB and YTF are summarized in Table 1.

The mutant *C. autoethanogenum* strains generated over the course of this study are summarised in Table 2.

2.2 Cloning and assembly

In silico design of constructs was achieved with A plasmid Editor (ApE) (RRID:SCR_014266) (Davis & Jorgensen, 2022). Unless otherwise specified, all kits, enzymes and buffers were purchased from New England Biolabs Ltd. (NEB, Hitchin, United Kingdom) and used following the manufacturer's instructions. DNA oligos were synthesised by Integrated DNA Technologies, Inc. (Coralville, United States) and were designed to have an annealing region with a melting temperature of 65°C using NEB Tm calculator (RRID:SCR_017969, tmcalculator.neb.com) and no secondary structure with a melting temperature higher than 57°C (modelled with Mfold, RRID:SCR_008543 (Zuker, 2003)). All oligos and vectors used in this study

TABLE 1 Antibiotics and additives used in this study.

Additive	Abbreviation	Working concentration (μg/mL)	Purpose
Chloramphenicol	Cm	12.5	Maintains pMTL83151-derived plasmids in *E. coli*
Kanamycin	Kan	50	Maintains the conjugative R702 plasmid in *E. coli*
D-cycloserine	D-cyc	250	Counter-selects *E. coli* sExpress during transconjugants selection
Thiamphenicol	Tm	7.5	Maintains pMTL83151-derived plasmids in *C. autoethanogenum*
Theophylline	Th	900	Induces Target-AID in *C. autoethanogenum*
5-Fluoroorotic acid	5-FOA	1,000	Selects for Δ*pyrE* genotype

TABLE 2 Mutant strains generated in this study.

Strain	Background	Genotype	Purpose	Engineered with plasmid
cFS02	*C. autoethanogenum*	*pyrE* (Q130X)	Proof of concept of initial Target-AID construct and knock-out strategy	vFS36
cFS04	*C. autoethanogenum*	*CLAU_532*(Q243X), *CLAU_534*(A242V)	Proof of concept of multiplex Target-AID	vFS50
cFS05	*C. autoethanogenum*	*CLAU_532*(Q243X), *CLAU_534*(A242V), *CLAU_1794*(S251L, W278X)	Proof of concept of multiplex Target-AID-NG	vFS50, vFS74
cFS15	*C. autoethanogenum*	*CLAU_532*(Q243X), *CLAU_534*(A242X)	Proof of concept of multiplex Target-AID	vFS50

are summarized in Supplementary Tables S1, S2. Amplification of all parts was achieved by polymerase chain reactions (PCR) with Q5® high-fidelity polymerase 2x master mix and an annealing temperature of 60°C. All parts were assembled with NEBuilder® HiFi assembly, designed to share between 25 nt and 40 nt of homologous overlap, and purified using the NEB Monarch® DNA gel purification kit. All vectors were transformed into chemically competent *E. coli* DH5α following NEB's protocol. *E. coli* colony PCR was done with DreamTaq Green PCR Master Mix (2X) from Thermofisher Scientific (Waltham, United States) after resuspending each colony into 40 µL of sterile ddH2O and using 1 µL as DNA template. After gel electrophoresis of the PCR products to screen for amplicons of the right size, these same colony aliquots were used to inoculate overnight cultures before proceeding to cryopreservation in Microbank® from Pro-lab diagnostics (Richmond Hill, Canada) and plasmid extraction with NEB Monarch® plasmid miniprep kit. Plasmids were then subjected to Sanger sequencing by Eurofins Genomics (Ebersberg, Germany). A detailed description of all cloning steps for each vector is available in the Supplementary Material, alongside key sequences highlighted in the text in Supplementary Tables S3, S4, S5. The full sequence of each vector is available upon request and at www.plasmidvectors.com (RRID: SCR_023475) where plasmids may be sourced.

2.2.1 Codon optimization

The sequence coding for UGI-GLVA and the activation-induced cytidine deaminase 1 from *Petromyzon marinus* (PmCDA1, or AID) with its long protein linker (Nishida et al., 2016; Banno et al., 2018) were codon optimized by Genscript (Piscataway, United States) to match the codon usage of *C. autoethanogenum*. The codon utilization table of *C. autoethanogenum* was obtained by extracting all the coding sequences (CDS) from its published genome (GenBank: CP012395.1) (Humphreys et al., 2015) and processing them with the CUSP algorithm (Rice et al., 2000).

2.2.2 Design of a *Clostridium* Target-AID plasmid

The Target-AID CDS was generated by replacing the STOP codon of the $Cas9^{D10A}$ nickase (nCas9) CDS with the CDS of the codon-optimized AID and its long protein linker (Nishida et al., 2016). Target-AID was then cloned inside the pMTL83151 backbone of our RiboCas system (Cañadas et al., 2019). Consequently, expression of Target-AID was subordinated to the tight control of a theophylline-inducible riboswitch, and expression of the custom sgRNA was driven by the strong P*araE* constitutive promoter (Huang et al., 2016).

2.2.3 Protospacer design and modelling of total protospacers and genomic coverage

Potential Target-AID protospacer targets for each gene were initially identified with Benchling (RRID:SCR_013955) CRISPR design tool, which lists all the 20 nt sequences directly upstream of an NGG PAM in a given DNA sequence. This list was then pasted onto a custom Microsoft Office Excel™ spreadsheet, which flagged the protospacers with the appropriate cytosines in positions −19 to −16 from the PAM that could lead to a TAA, TAG or TGA codon if turned into a thymine. Target-AID-NG protospacers were pulled from our whole-genome analysis of potential Target-AID targets for various PAMs in *C. autoethanogenum* using MATLAB™ (RRID:SCR_001622). For future work on individual genes, we recommend the use of BE-designer (RRID:SCR_023389, rgenome.net/be-designer/), an online tool with a user-friendly interface.

2.2.4 Design of multiplex sgRNA-tRNA array

Monocystronic tRNA sequences from *Clostridium pasteurianum* were complemented with 20 nt of their pre-tRNA sequence at their 3′end using the GtRNAdb—Genomic tRNA Database (RRID: SCR_006939) (Chan & Lowe, 2009; 2016). Individual tRNA structures were modelled with the online tool RNAfold (Gruber et al., 2008), and manually examined using a variety of parameters. The pre-tRNAs Thr-TGT-1-1 and fMet-CAT-1-1 were selected for (in decreasing order of importance) having a weak A-U rich stem-loop approximately 16 nt from the CCA-3′-end of the mature tRNA (Sekiya et al., 1979), for lacking a 3′ poly-U tail which could induce the termination of transcription, for not forming a strong DNA secondary structure with the binding sequence of the oligonucleotide primers used to synthesize the tRNA array during PCR amplification (Zuker, 2003), and for being associated with a codon in relatively high usage within *C. autoethanogenum* to avoid upsetting excessively the balance of its tRNA pool.

2.3 Conjugation in *Clostridium autoethanogenum*

Two days before mating, one 0.2 mL *C autoethanogenum* cryostock of 10% DMSO was thawed and inoculated into 2 mL of pre-reduced liquid YTF. Early the next day, each vector was transformed into 20 µL of *E. coli* sExpress and the *C. autoethanogenum* inoculum was subcultured into 4 mL of pre-reduced liquid YTF at starting OD 0.05. On the day of the mating, a single sExpress transformant colony of each construct was inoculated into 5 mL of room-temperature liquid LB + Cm + Kan until OD reached 0.2. 1 mL of culture was then centrifuged at 3,000 rcf for 3 min and washed in 0.5 mL PBS before being centrifuged again. The resulting pellet was finally resuspended into 0.2 mL of *C. autoethanogenum* culture and gently spread onto a YTF plate without antibiotics. 20 h later, 0.6 mL of PBS was vigorously spread over the mating plate to resuspend the cells and constitute the mating slurry. The mating slurry was normalized to 0.6 mL with PBS in a microcentrifuge tube, and 0.2 mL of it was finally transferred onto YTF + D-cyc + Tm + Th plates for selection of transconjugants and induction of Target-AID constructs. For the preliminary characterisation of Target-AID, the transconjugant colonies were subsequently transferred onto a YTF + D-cyc + Tm + Th+5-FOA in order to select for the *ΔpyrE* genotype.

2.4 Plasmid loss

After mutagenesis and colony PCR, plasmids were lost by re-streaking mutants on YTF plates without antibiotic, then patching 10 single colonies on YTF plates with and without Tm. Patches which failed to grow on Tm were considered to have lost their plasmid. They were then inoculated in liquid YTF with and without Tm to prepare cryopreservation in 10% pre-reduced DMSO once they reached OD600~0.4 without antibiotics, if there was still no growth in liquid cultures complemented with thiamphenicol.

2.5 Estimation of mutagenesis efficiency

The efficiency of Target-AID mutagenesis was roughly estimated by calculating the proportion of colonies which survived exposure to 5 mM 5-FOA after induction on 5 mM theophylline. Subsequent mutagenesis efficiencies were only estimated from the Sanger sequencing results of five transconjugants colonies per construct.

2.6 Sanger sequencing

The targeted genes of fifteen *C. autoethanogenum* vFS36_TA_pyrE colonies and five *C. autoethanogenum* colonies conjugated with each of the other plasmids were screened by Sanger sequencing after a colony PCR with Q5® DNA polymerase and purification with QIAquick PCR Purification Kit from Qiagen (Venlo, Netherlands). Each colony aliquot was first boiled in 40 µL sterile ddH2O for 10 min at 98°C and centrifuged at 2,500 RCF for 1 min before using 1 µL of supernatant as DNA template for the PCR. Sanger Sequencing was performed by Eurofins Genomics (Ebersberg, Germany) and sequencing results were aligned with their respective WT sequences using Benchling (RRID:SCR_013955). Sequencing primers are summarized in Supplementary Table S1 and raw reads are available in the Supplementary Material.

2.7 Determination of non-essentiality

Prior to plasmid design the three genes CLAU_532, CLAU_534 and CLAU_1794 were cross-checked in the published list of essential genes determined by transposon insertion sequencing (Woods et al., 2022) and confirmed to be non-essential under heterotrophic or autotrophic conditions.

2.8 Whole genome sequencing

Whole genome sequencing was performed at the Deepseq Next-Generation Sequencing Facility of the University of Nottingham.

2.8.1 Library preparation, library QC and sequencing protocol

DNA concentration was measured using the Qubit Fluorometer and the Qubit dsDNA BR Assay Kit (ThermoFisher; Q32853). 250 ng input DNA was used in the sequencing library preparation. Indexed sequencing libraries were prepared using the Nextera DNA Flex Library Prep Kit (Illumina; 20018705) and IDT® for Illumina Nextera™ DNA CD Indexes (Illumina; 20018708). 5 cycles of PCR were used. Libraries were quantified using the Qubit Fluorometer and the Qubit dsDNA HS Kit (ThermoFisher Scientific; Q32854). Library fragment-length distribution was assessed using the Agilent TapeStation 4,200 and the Agilent High Sensitivity D1000 ScreenTape Assay (Agilent; 5067-5584 and 5067-5585). Final library quantification was performed using the KAPA Library Quantification Kit for Illumina (Roche; KK4824) and the library was sequenced on an Illumina MiSeq using the MiSeq Reagent Kit v3 (600 cycle) (Illumina; MS-102-3,003) to generate 300-bp paired-end reads.

2.8.2 Data analysis for resequencing

2.8.2.1 Quality check and trimming of reads

Sequencing produced 1,919,451 raw reads that were trimmed of Illumina adapters and low quality (Q < 30) nucleotides using TrimGalore (v 0.6.6) (Andrews, 2010; Martin, 2011; Kreuger, 2012) nucleotide clip was performed at the 3′end of reads. Reads shorter than 20 bp were discarded. After quality filtering, 98.1% of reads remained.

2.8.2.2 Alignment of reads to reference genome

Trimmed reads were aligned to the *Clostridium autoethanogenum* reference genome (NCBI accession: NZ_CP012395.1) using "bwa mem" (v0.7.17) (Li & Durbin, 2009; 2010; Li, 2013). Putative SNPs were filtered using "bcftools mpileup", with minimum base and mapping quality of 20 applied and duplicate reads removed. SNPs were typed using 'bcftools call' with low quality genotypes filtered out (SNP quality <30 OR Mapping quality <30) (van der Auwera and O'Connor, 2020; Danecek et al., 2021). Mapping to the reference genome achieved coverage of 99.7% of reference bases at an average depth of 213x.

2.8.2.3 Annotation of variants

Genome annotation associated with NZ_CP012395.1 was downloaded from the NCBI database and appropriate annotations added to variants.

2.8.3 Identification of putative off-target mutations

All protospacers sequences targeted in cFS05 were submitted to Cas-Offinder (RRID: SCR_023390) to generate a list of putative off-target sites with up to 9 mismatches, 2 nt DNA gaps and 2 nt RNA gaps. This generated 14,376 putative off-target sites. Sites within 50 nt of the undesired SNP detected by whole genome sequencing were considered as potential off-targets of their associated protospacer. Additionally, the sequences of all protospacers targeted in cFS05 were aligned with the 41 nt WT sequence of each undesired SNP and the 20 nt immediately upstream and downstream in order to find potential matches between target protospacers and the sequence around undesired SNP using Geneious (RRID:SCR_010519).

3 Results

3.1 Target-AID proof-of-concept: *pyrE* knockout in *Clostridium autoethanogenum*

The Target-AID design was validated with the plasmid vFS36_TA_pyrE (Figure 2), customised to introduce a premature STOP codon in the *pyrE* gene of *C. autoethanogenum*, which is necessary for pyrimidine synthesis and confers resistance to 5-FOA when knocked out (Ng et al., 2013; Minton et al., 2016). Only 70% of the 816 transconjugant colonies induced with theophylline survived when transferred on 5-FOA, but all the fifteen 5-FOA-resistant colonies screened exhibited the expected C388T mutation which

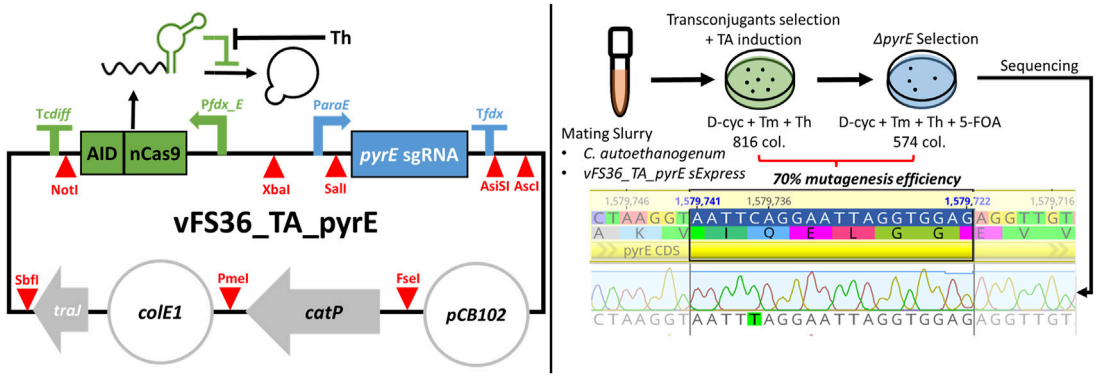


FIGURE 2
Target-AID proof of concept. (Left) Schematic of the plasmid vFS36_TA_pyrE, based on the pMTL83151 backbone. (Right) Knock-out of *pyrE* in *Clostridium autoethanogenum* using Target-AID. 816 colonies were obtained on the transconjugants selection plate with 5 mM Th; after transfer onto a plate complemented with 5 mM 5-FOA, only 574 colonies remained. The *pyrE* loci of fifteen 5-FOA-resistant colonies were then screened by Sanger sequencing, and all exhibited the desired C388T mutation. Blue highlighted region = 20 nt targeted protospacer; TA = Target-AID; D-cyc = D-cycloserine; Tm = Thiamphenicol; Th = Theophylline; 5-FOA = 5-Fluoroorotic acid; col. = colonies.Symbols inspired from SBOL visual (Baig et al., 2021).

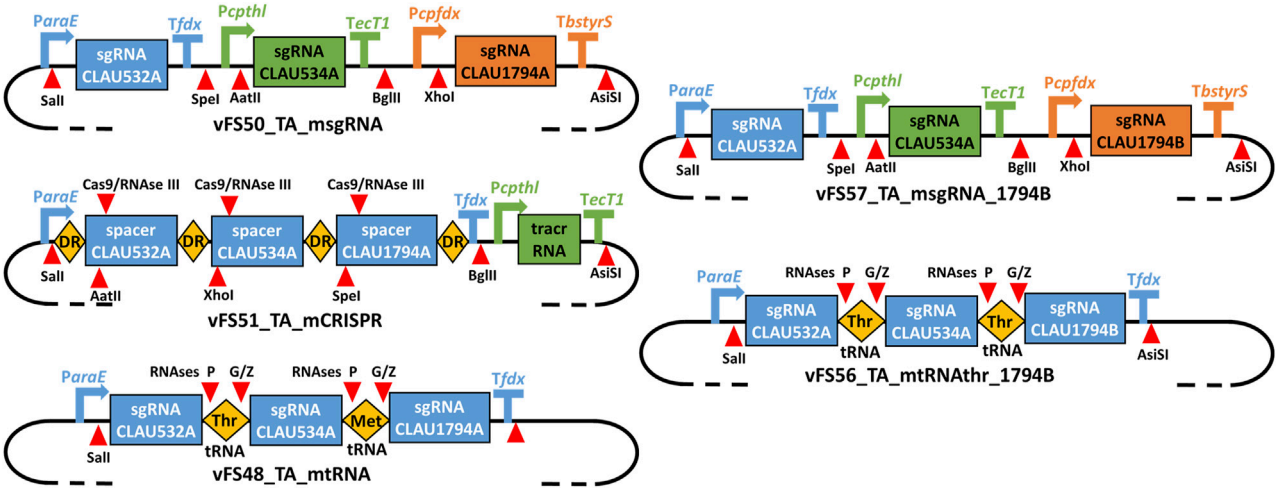


FIGURE 3
Schematic of the gRNA expression cassettes of different multiplex Target-AID plasmids. vFS50_TA_msgRNA has three separate sgRNA transcriptional units; vFS51_TA_mCRISPR has two transcriptional units expressing a CRISPR array and a tracrRNA, respectively; vFS48_TA_mtRNA, has a single transcriptional unit expressing an array of sgRNA-tRNA fusions All are derived from vFS36_TA_pyrE. vFS57_TA_msgRNA_1794B is derived from vFS50_TA_msgRNA and only differs in the protospacer sequence targeting CLAU_1794 (CLAU1794B instead of CLAU1794A). vFS56_TA_mtRNAthr_1794B is derived from vFS48_TA_mtRNA but uses twice the same tRNA (Thr-TGT-1-1 tRNA) and targets the CLAU1794B protospacer instead of CLAU1794A. DR = Direct repeat; tracrRNA = trans-activating CRISPR RNA. Symbols inspired from SBOL visual (Baig et al., 2021).

results in a premature STOP codon (TAG) (Figure 2). This preliminary experiment confirmed that our basic Target-AID construct and mutagenesis protocol was functional in *C. autoethanogenum*.

3.2 Multiplexing designs

The msgRNA, mCRISPR and mtRNA multiplexing strategies were respectively tested by the plasmids vFS50_TA_msgRNA vFS51_TA_mCRISPR and vFS48_TA_mtRNA (Figure 3). Because only the last 20 nt at the 3′-end of the CRISPR spacer were shown to be necessary for gRNA targeting, but the native *S. pyogenes* CRISPR array is composed of ~30 nt spacers (Jinek et al., 2012), each 20 nt protospacer sequence of vFS51_TA_mCRISPR was complemented with a 6 nt restriction site and 4 random nt at its 5′-end. Three non-selectable and non-essential genes (CLAU_532, CLAU_534 and CLAU_1794) coding for alcohol dehydrogenases were picked as arbitrary targets in *C. autoethanogenum*, and protospacers which could produce a STOP codon after a C-to-T

TABLE 3 Sequencing of five *Clostridium autoethanogenum* colonies conjugated with different Target-AID constructs and induced on theophylline. vFS48 = vFS48_TA_mtRNA; vFS50 = vFS50_TA_msgRNA; vFS51 = vFS51_TA_mCRISPR; vFS56 = vFS56_TA_mTHRtRNA; vFS57 = vFS57_TA_msgRNA_CA1794B; vFS52 = vFS52_TA_CA532A; vFS53 = vFS53_TA_CA534A; vFS54 = vFS54_TA_CA1794A; and vFS58 = vFS58_TA_CA1794B. The targeted codon is bolded and mutated bases are capitalized. Fractions indicate mixed reads for this base. The sequence of CLAU532A is oriented in the antisense direction, meaning that TTA, TCA, and CTA alleles would all results in a STOP codon (TAA, TGA and TAG, respectively).

Protospacer	Allele	Sequence (20 nt)	Mutagenesis efficiency (over 5 colonies)								
			Multiplex constructs					Monoplex constructs			
			vFS48	vFS50	vFS51	vFS56	vFS57	vFS52	vFS53	vFS54	vFS58
CLAU532A	WT	ct**cca**gtcaggtgttgtgca	0	1/5	4/5	0	0	0			
	A.1	ct**TTa**gtcaggtgttgtgca	0	1/5	0	0	1/5	0			
	A.2	ct **T/C** **T/C** **a**gtcaggtgttgtgca	0	3/5	0	0	0	5/5			
	A.3	ct **T/C** **ca**gtcaggtgttgtgca	5/5	0	1/5	3/5	3/5	0			
	A.4	ct**Tca**gtcaggtgttgtgca	0	0	0	2/5	0	0			
	A.5	ct**T** **T/C** **a**gtcaggtgttgtgca	0	0	0	0	1/5	0			
CLAU534A	WT	agcc**caa**tgtctagctggga	5/5	0	5/5	3/5	0		2/5		
	A.1	agTT**Taa**tgtctagctggga	0	0	0	0	0		0		
	A.2	agTc**caa**tgtctagctggga	0	0	0	0	1/5		1/5		
	A.3	ag **T/C** c**caa**tgtctagctggga	0	0	0	2/5	2/5		2/5		
	A.4	ag **T/C** **T/C** **T/C** **aa**tgtctagctggga	0	5/5	0	0	0		0		
	A.5	ag **T/C** **T/C** **caa**tgtctagctggga	0	0	0	0	1/5		0		
	A.6	agTT**caa**tgtctagctggga	0	0	0	0	1/5		0		
CLAU1794A	WT	aaa**caa**gcaattgttccgtt	5/5	5/5	5/5					5/5	
	A.1	aaa**Taa**gcaattgttccgtt	0	0	0					0	
CLAU1794B	WT	atca**caa**tgtttagcaggta				3/5	3/5				1/5
	A.1	atTa**Taa**tgtttagcaggta				0	0				0
	A.2	atTa**caa**tgtttagcaggta				0	1/5				2/5
	A.3	at **T/C** a**caa**tgtttagcaggta				2/5	1/5				2/5

mutation were identified for each gene (labelled CLAU532A, CLAU534A, and CLAU1794A, respectively).

In addition to these three multiplex constructs, three "monoplex" controls were also assembled to estimate the individual effectiveness of each individual sgRNA independently from its multiplexing system (Table 3). These are the vectors vFS52_TA_CA532A, vFS53_TA_CA534A, and vFS54_TA_CA1794A, all directly derived from vFS36_TA_pyrE with the *pyrE* protospacer sequence replaced by a protospacer sequence of their respective target gene.

3.3 Most Target-AID mutants exhibit mixed genotypes

No mutants could be obtained from any of the gRNAs targeting the protospacer CLAU1794A (Table 3). However, CLAU532A and CLAU534A were successfully mutated by their respective monoplex control and vFS50_TA_msgRNA. Unfortunately, all but two of the twenty-three mutated loci showed some level of mixed trace, revealing the presence of WT cells in the colony. In most cases, the WT trace was dominant over the mutated one. Only one colony transformed with a multiplex construct (vFS50_TA_msgRNA) showed a pure colony with the desired mutation for a protospacer (CLAU532A.A1). Fortunately, it also showed a mixed peak for the CLAU534A protospacer (CLAU534A.A4). That colony was thus re-streaked on a YTF + D-cyc + Tm + Th plate to isolate pure colonies. A single pure colony with the desired CLAU532A.A1 and CLAU534A.A1 C-to-T mutations could be identified after screening five of these re-streaked colonies (cFS15). Albeit with poor efficiency, this is evidence that at least two loci can be mutated with premature STOP codon in a single mutagenesis step using Target-AID.

A subsequent multiplex mutagenesis attempt was made with a different CLAU_1794 protospacer (CLAU1794B) with the plasmid vFS57_TA_msgRNA_1794B (Figure 3) and its monoplex control vFS58_TA_CA1794B. This time, the last protospacer could be targeted, albeit only two out of five colonies showed a mutated CLAU1794B protospacer (four for the monoplex control), both

TABLE 4 Sequencing of five *Clostridium autoethanogenum* colonies conjugated with different multiplex Target-AID constructs and induced on theophylline. vFS75_mTA-UGI-GLVA is a msgRNA Target-AID construct fused with a UGI and GLVA domains at its C-terminus. vFS94_mTA-trsgRNA is a msgRNA Target-AID construct with 18 nt spacers instead of 20 nt. vFS57_TA-msgRNA_CA1794B is a standard msgRNA Target-AID construct (without UGI-GLVA tag and with 20 nt spacers) that targets the same protospacers as vFS75_mTA-UGI-GLVA and vFS94_mTA-trsgRNA. The targeted codon is bolded and mutated bases are capitalized. Fractions indicate mixed reads for this base. The sequence of CLAU532A is oriented in the antisense direction, meaning that TTA, TCA, and CTA alleles would all results in a STOP codon (TAA, TGA and TAG, respectively). (*) Poor quality reads were removed from analysis.

Protospacer	Allele	Sequence (20 nt)	Mutagenesis efficiency (over 5 colonies)		
			Redesigned		Control
			vFS75_mTA-UGI-GLVA	vFS94_mTA-trsgRNA	vFS57_TA-msgRNA_CA1794B
CLAU532A	WT	ct**cca**gtcaggtgttgtgca	4/4*	0	0/5
	A.2	ct T/C T/C **a**gtcaggtgttgtgca	0	0	5/5
	A.3	ct T/C **ca**gtcaggtgttgtgca	0	4/5	0
	A.6	Tt T/C T/C **a**gtcaggtgttgtgca	0	1/5	0
CLAU534A	WT	agcc**caa**tgtctagctggga	3/5	5/5	0
	A.2	agTc**caa**tgtctagctggga	2/5	0	0
	A.5	ag T/C T/C **caa**tgtctagctggga	0	0	1/5
	A.7	agT T/C **caa**tgtctagctggga	0	0	3/5
	A.8	agT T/C T/C **aa**tgtctagctggga	0	0	1/5
CLAU1794B	WT	atca**caa**tgtttagcaggta	5/5	5/5	1/5
	A.3	at T/C a**caa**tgtttagcaggta	0	0	4/5

showed a mixed trace, and none had the desired C-to-T mutation which would have produced a STOP codon. This confirmed that the last sgRNA cassette was functional and that the previous failure to mutate CLAU1794A was not due to any multiplex system, but to a defective protospacer sequence. Unfortunately, the CLAU534A.A1 or CLAU534A.A4 allele could not be found among the five screened vFS57_TA_msgRNA_1794B colonies.

3.4 Arrays of tRNA-sgRNA fusions can be used to express multiple sgRNAs

The initial mtRNA and the mCRISPR multiplexing strategies both failed to mutate the CLAU534A protospacer, even though this could be achieved by the msgRNA strategy and the monoplex control targeting CLAU534A (vFS53_TA_CA534A). The mCRISPR strategy only managed to mutate one of the five colonies screened for CLAU532A mutations (Table 3). Interestingly the mtRNA strategy managed to mutate all five of them. This indicated that the Thr-TGT-1-1 tRNA had not interfered with the function of the CLAU532A sgRNA. The mCRISPR strategy was abandoned at this stage. However, we hypothesized that replacing the fMet-CAT-1-1 tRNA by another copy of the Thr-TGT-1-1 tRNA might rescue the mtRNA strategy.

While multiplex mutagenesis with the vFS56_TA_mtRNAthr_1794B (Figure 3), which only uses the Thr-TGT-1-1 tRNA, did not yield the expected C-to-T mutation in either CLAU534A or CLAU1794B, it did successfully target all three protospacers and even yielded two pure CLAU532A.A4 alleles with the desired C-to-T mutation (Table 3).

3.5 UGI-LVA and truncated sgRNAs do not improve mutagenesis efficiency

Next, we investigated whether fusing a UGI (with GLVA degradation tag) to Target-AID would improve mutagenesis efficiency in all protospacers, or if truncating the protospacer region of the sgRNAs targeting CLAU534A and CLAU1794B would shift their editing window to edit the cytosine in position −16 from the PAM more favourably. Unfortunately, while the initial vFS57_TA-msgRNA_CA1794B construct once again managed to edit all three loci with almost 100% efficiency (although without pure mutant and only one correctly edited base out of ten in position −16 for CLAU534A and CLAU1794B), the addition of UGI-LVA downstream at the C-terminus of Target-AID in the plasmid vFS75_mTA-UGILVA seemed only to harm editing efficiency (only 2 pure mutations for CLAU534A and no other mutation across the 12 remaining reads), and truncating the sgRNAs targeting CLAU534A and CLAU1794B from 20 nt to 18 nt in the plasmid vFS94_mTA-trsgRNA seemed to completely abolish mutagenesis of their respective protospacers (Table 4).

3.6 Target-AID-NG has a vastly superior targeting space

Besides CLAU534A, CLAU1794A and CLAU1794B, there were no other Target-AID-compatible protospacers that could potentially result in a premature STOP codon in CLAU_534 or CLAU_1794. This revealed an inherent weakness of Target-AID

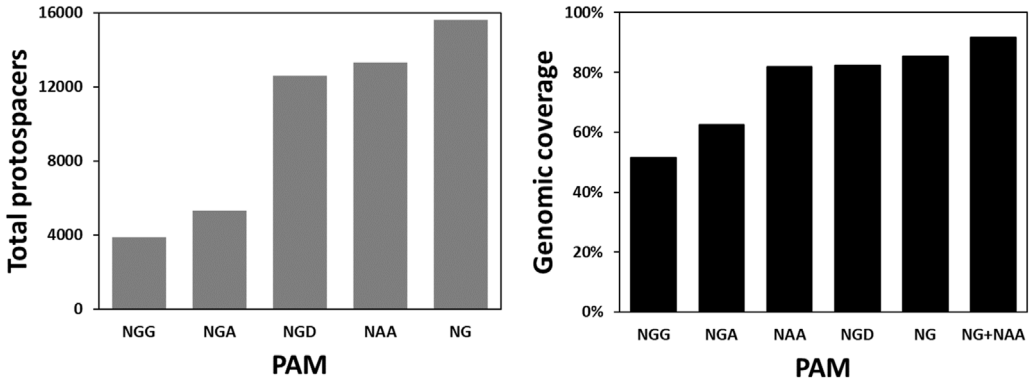


FIGURE 4
Targeting space and genomic coverage of Target-AID in *C. autoethanogenum* as a function of different PAM sequences. The targeting space is defined here as the total number of protospacers in the first 75% of their respective CDS which can lead to a STOP codon when the cytosines within the nucleotides −19 to −16 from the PAM are mutated into thymines. We define genomic coverage as the proportion of CDS's in the genome which have at least one of these protospacers.

mutagenesis: very few protospacers are available for each gene, and many genes cannot be targeted at all. To quantify the problem, the total number of Target-AID protospacers which could produce STOP codons in the first 75% of any gene in *C. autoethanogenum* were modelled and compared across several published SpCas9 variants (Supplementary Table S6) which exploit non-canonical PAMs (Figure 4). The proportion of *C. autoethanogenum* genes which could theoretically be disrupted by Target-AID using these protospacers for each PAM (Figure 4) was also measured. We named this parameter the genomic coverage of Target-AID.

The bioinformatic analysis undertaken revealed that the conventional Target-AID, with its NGG PAM, could potentially introduce STOP codons in only 3,895 protospacers for a genomic coverage of merely 51.64% in *C. autoethanogenum*. NG PAMs (Nishimasu et al., 2018), on the other hand, would give access to 15,602 protospacers which would theoretically allow 85.32% of *C. autoethanogenum*'s CDS to be inactivated. Also promising is the NAA PAM of Cas9-iSpymac (Chatterjee et al., 2020), which can target 13,293 protospacers for a genomic coverage of 81.81%. Together, Target-AID-NG and Target-AID-iSpymac would be able to knock out 91.78% *C. autoethanogenum*'s CDS. However, since both CLAU_532 and CLAU_534 happen to be among the 18.19% of CDS that Cas9-iSpymac could not knock-out as part of a Target-AID base editor, Target-AID-NG was chosen for a last multiplex mutagenesis attempt of CLAU_532, CLAU_534 and CLAU_1794 with the msgRNA strategy.

3.7 UGI without LVA tag does not improve mutagenesis efficiency of Target-AID-NG in *Clostridium autoethanogenum*

The vectors vFS72_mTA-NG and vFS103_mTA-NG-UGI_NoLVA were assembled to introduce nonsense mutations in CLAU_532, CLAU_534 and CLAU_1794 using Target-AID-NG and Target-AID-NG-UGI without GLVA tag (Table 5). Our model showed that Target-AID-NG could target five protospacers in CLAU_532, three in CLAU534, and seven in CLAU1794. This allowed the selection of protospacers where all the cytosines in the editing window would result in a nonsense mutation if mutated to thymines, with a preference for protospacers with cytosines in position −18 rather than −16 from the PAM. This time, with a UGI fusion but without GLVA tag, mutants were obtained at roughly the same rate as the Target-AID-NG construct without UGI (80% and 100% for CLAU532A, 0% and 20% for CLAU534B, and 80% and 60%, respectively). Only two traces out of 28 reflected pure mutants for one locus; all the other mutated loci were mixed with WT genotype (often predominantly WT). After one round of re-streak, no CLAU532A mutant could be isolated from the two colonies exhibiting the pure CLAU1794C.A.1 genotype.

3.8 Target-AID-NG is also suitable for duplex mutagenesis in *Clostridium autoethanogenum*

In a last attempt to mutate all three targeted genes in the same strain, we assembled the vector vFS74_TA_NG_msgRNA_1794DFG to target three of the four remaining CLAU_1794 protospacers targetable by Target-AID-NG to produce a nonsense mutation. This vector was not using a UGI fusion and was using the msgRNA strategy. It was conjugated into the plasmid-free double-mutant *C. autoethanogenum* strain (cFS04) that had been previously engineered with vFS50_TA_msgRNA. As described in Table 6, the protospacer CLAU1794D failed to deliver any mutations in the five colonies screened, but CLAU1794F and CLAU1794G showed 100% of mutagenesis efficiency. This time, only three out of ten traces showed mixed peaks, but all five pure CLAU1794F mutations affected a cytosine which did not result in the introduction of a STOP codon. Out of the three protospacers, only CLAU1794G resulted in nonsense mutations in CLAU_1794.

TABLE 5 Sequencing of five *Clostridium autoethanogenum* colonies conjugated with different Target-AID-NG constructs and induced on theophylline. vFS72 = vFS72_mTA-NG; vFS103 = vFS103_mTA-NG-UGI_NoLVA. The targeted codon is bolded and mutated bases are capitalized. Fractions indicate mixed reads for this base. The sequence of CLAU532A and CLAU1794C is oriented in the antisense direction, meaning that TTA, TCA, and CTA alleles would all results in a STOP codon (TAA, TGA and TAG, respectively). (*) Poor quality reads were removed from analysis.

Protospacer	Allele	Sequence (20 nt)	Mutagenesis efficiency (over 5 colonies)	
			vFS72_mTA-NG	vFS103_mTA-NG-UGI_NoLVA
CLAU532A	WT	ct**cca**gtcaggtgttgtgca	1/5	0
	A.2	ct $\frac{T}{c}$ $\frac{T}{c}$ **a**gtcaggtgttgtgca	3/5	0
	A.3	ct $\frac{T}{c}$ **ca**gtcaggtgttgtgca	1/5	3/3*
	A.8	Tt $\frac{T}{c}$ $\frac{T}{c}$ **a**gtcaggtgttgtgca	0	0
CLAU534B	WT	agg**cag**aaggcacaatttgt	5/5	4/5
	A.1	agg $\frac{T}{c}$ **ag**aaggcacaatttgt	0/5	1/5
CLAU1794C	WT	at**cca**atctggcccaaattc	1/5	2/5
	A.1	at**TTa**atctggcccaaattc	2/5	0
	A.2	at**T** $\frac{T}{c}$ **a**atctggcccaaattc	2/5	0
	A.5	at $\frac{T}{c}$ **ca**atctggcccaaattc	0	3/5

TABLE 6 Sequencing of five *Clostridium autoethanogenum* colonies conjugated with vFS74_TA_NG_msgRNA_1794DFG and induced on theophylline. The targeted codon is bolded and mutated bases are capitalized. Fractions indicate mixed reads for this base. The sequence of CLAU1794G is oriented in the antisense direction, meaning that TTA, TCA, and CTA alleles would all result in a STOP codon (TAA, TGA and TAG, respectively).

Protospacer	Allele	Sequence	Mutagenesis efficiency (over 5 colonies)
			vFS74_mTA-NG_CA1794DFG
CLAU1794D	WT	Aga**caa**aaagctaaatttgt	5/5
	A.1	Aga**Taa**aaagctaaatttgt	0/5
CLAU1794F	WT	tca**caa**tgtttagcaggtat	0
	A.1	tca**Taa**tgtttagcaggtat	0
	A.2	tTa**caa**tgtttagcaggtat	5/5
CLAU1794G	WT	g**cca**tacagctcctgtttta	0
	A.1	g**TTa**tacagctcctgtttta	1/5
	A.2	g**Tca**tacagctcctgtttta	1/5
	A.3	g $\frac{T}{c}$ **ca**tacagctcctgtttta	3/5

3.9 Fifteen off-target mutations were identified after two rounds of Target-AID and target-AID-NG multiplex mutagenesis

The resulting triple-mutant strain was finally validated by whole genome sequencing (NCBI accession PRJNA956560). It revealed twelve off-target single-nucleotide polymorphisms (SNP) (Supplementary Table S7): six exhibiting the canonical Target-AID C-to-T or G-to-A mutation, six within 50 nt of a putative off-target protospacer identified with Cas-Offinder (RRID: SCR_023390) (Zhao et al., 2017), and one which could be manually aligned with the protospacer CLAU1794G using Benchling (RRID: SCR_013955) sequence alignment tool (Supplementary Table S8). One 15 nt duplication event was also found, in addition to two regions of the genomes where the sequencing coverage abruptly dropped to 0 and thus seem to have been lost by the cell (Table 8). One of these deleted regions was only 23 nt long (but inside one of the nine 16S RNA loci), while the other region was 12 kb-long and comprised a putative off-target protospacer relatively similar to its on-target, with only four mismatches and a 2 nt gap. Altogether, fifteen off-target mutations were thus identified in our strain that underwent two consecutive rounds of multiplex Target-AID (then Target-AID-NG) mutagenesis. Importantly, the gene coding for UDG was among the thirteen genes lost within the 12 kb deleted region (Table 8; Supplementary Table S9 for detailed list of deleted genes).

TABLE 7 Summary of off-target SNP and short duplication events identified in *Clostridium autoethanogenum* after two rounds of multiplex Target-AID (then Target-AID-NG) mutagenesis, and their putative association with targeted protospacers. (*) Found by standard alignment with the protospacers targeted in this strain (CLAU532A, CLAU534A, CLAU1794A, CLAU1794D, CLAU1794F, CLAU1794G) using a standard alignment tool (Geneious (RRID:SCR_010519)), instead of Cas-Offinder (RRID: SCR_023390) (Zhao et al., 2017).

Locus	Position	WT allele	Alternate allele	Putative protospacer	Mismatch (nt)	Gap (nt)	CAS-OFFINDER position (nt)
CLAU_0577	660931	C	T	CLAU1794G	9	1	660941
CLAU_0894	995262	C	T	N/A			
CLAU_1083	1205885	C	A	CLAU532A	9	0	1205894
CLAU_1087	1210587	C	A	N/A			
CLAU_1384	1525791	G	A	CLAU534A	6	2	1525752
CLAU_1943	2136051	T	C	N/A			
CLAU_2237	2449821	A	C	CLAU1794G	9	0	2449824*
CLAU_2237	2450037	G	T	CLAU1794D	6	2	2450048
CLAU_2344	2570391	TCCTGAAAGGACATCA	TCCTGAAAGGACATC ACCTGAAAGGACATCA	N/A			
N/A	2994056	G	A	N/A			
CLAU_3637	4014166	C	T	N/A			
CLAU_3830	4203465	T	G	CLAU534A	8	2	4203417
CLAU_3935	4318262	G	A	CLAU534A	7	2	4318245

TABLE 8 Summary of gene deletion events identified in *Clostridium autoethanogenum* after two rounds of multiplex Target-AID (then Target-AID-NG) mutagenesis, and their putative association with targeted protospacers. Putative off-target protospacers identified with Cas-Offinder (RRID: SCR_023390) (Zhao et al., 2017).

Coverage = 0					
Start position	Size (nt)	Putative protospacers	Mismatches (nt)	Gap (nt)	Cas-offinder position (nt)
2360221	12053	1794G	4	2	2372289
3873128	23	N/A			

4 Discussion

Our results show that, although Target-AID and Target-AID-NG can be used for multiplexed targeted point mutagenesis in *C. autoethanogenum*, they have serious drawbacks. While mutagenesis efficiency is high enough to reliably isolate mutants without selection markers, the resulting colonies are often mixed, which requires additional re-streaking steps to isolate pure mutants. Mixed colonies have been reported previously with several base editors in different organisms (Banno et al., 2018; Li et al., 2019; Liu et al., 2019). We hypothesize that they result from single cells that initially survived Target-AID mutagenesis by undergoing BER or by replicating instead of mutating their non-edited strand (Figure 1.VI, 1.VIII). From there, one of the cell lineages from the colony can mutate and produce the expected mutation, while the second lineage of cells continues to avoid mutagenesis by a variety of mechanisms, for example, by mutating key components of the Target-AID plasmid. The same process could produce mixed colonies when more than one cytosines are present in the editing window: each cell lineage in the same colony would acquire immunity from Target-AID by mutating a different cytosine. Protospacers with multiple cytosines in the editing window should thus be avoided to minimize the risk of obtaining mixed colonies.

Targeting cytosines in position −16 or −17 from the PAM, while possible, was rarely successful—especially when other cytosines were present in positions −18 or −19. Accordingly, only protospacers with a single cytosine in position −18 or −19 from the PAM should be used. This is consistent with the literature (Banno et al., 2018; Li et al., 2019; Nishida et al., 2016), but it severely restricts the already limited range of available protospacers in *C. autoethanogenum*. As illustrated in this work, the use of base editors with alternative PAM recognition domains such as Target-AID-NG or Target-AID-iSpymac can, however, greatly facilitate the identification of optimal protospacer targets.

This feature of Target-AID-NG was exploited to test three protospacers at once for the same target gene, in the hope that at least one would work. Indeed, out of the ten targeted protospacers (*pyrE* protospacer, CLAU532A, CLAU534A-B, and CLAU1794A-B-C-D-F-G), two did not work at all (CLAU1794A, and CLAU1794D)

and only four reliably introduced nonsense mutations (*pyrE* protospacer, CLAU532A, CLAU1794C, and CLAU1794G). Targeting the same gene with several protospacers at once might thus still have been a good use of multiplexing and Target-AID, if it had not also been associated with such a high rate of off-target mutagenesis (including a major 12 kb deletion). Consequently, even though we showed multiplex Target-AID mutagenesis was achievable in *C. autoethanogenum*, we cannot recommend its use as a standard practice, even just to screen several protospacers for the same target gene. Using parallel monoplex constructs is indeed less likely to result in off-target mutagenesis for the same number of targeted genes.

It is, however, difficult to assert with confidence which off-target mutations are a direct consequence of Target-AID mutagenesis. Some probably occurred randomly and were only selected through the many rounds of re-streaking; others might have been the indirect consequence of the loss of UDG, a key enzyme of the BER DNA repair pathway. The loss of UDG itself, nonetheless, was likely selected for by Target-AID, as UDG is a putative inhibitor of CBEs mutagenesis and thus a liability for the cells which express Target-AID. We hypothesize that UDG was lost in a random recombination event (potentially triggered by an off-target nick from Target-AID). *Δudg* cells should lose the BER pathway, which might have made them more susceptible to Target-AID mutagenesis. Accordingly, the *Δudg* mutants must have been over-represented in the colonies that survived mutagenesis. After the MMR pathway mutated the non-edited strand, in the absence of UDG, the repair of the edited strand from a "U" to a "T" might have been achieved solely through DNA replication.

The loss of UDG indirectly validates the strategy of using an UGI fusion to improve the effectiveness of CBEs, but it also exposes the gene coding for UDG as a mutational hotspot to look out for during Target-AID (or CBE) mutagenesis. This result highlights the importance of whole genome sequencing and rigorous complementation studies in any genome editing experiment, including base editing. Interestingly, in our hands, truncated sgRNA spacers, Target-AID-UGI and Target-AID-UGI-GLVA protein fusions did not result in higher mutagenesis efficiencies in *C. autoethanogenum* (there was even a marked decrease of mutagenesis efficiency when the GLVA-tag was present or truncated sgRNA spacers were used).

Here, multiplexing gRNA transcription using the native CRISPR array of *S. pyogenes* was not successful. In hindsight, it might be due to the addition of a restriction site and four extra nt upstream of each 20 nt spacer to facilitate cloning. Future iterations of this multiplexing strategy would be advised to comprise of 30 nt spacers fully homologous to their target. The tracrRNA could also be flanked with self-cleaving ribozymes to exclude any interference from its promoter and terminator in its RNA structure.

While an investigation of whether the tRNA-sgRNA array had successfully been processed into individual tRNAs and sgRNAs molecules was not undertaken, the array succeeded in targeting multiple protospacers almost as effectively as three separate transcriptional units. Given that the monoplex controls also failed to produce the desired nonsense mutation, the absence of correct CLAU534A and CLAU1794B mutants in either multiplexing strategy can be attributed to the protospacers themselves, and not the multiplexing method. An array of tRNA-sgRNA fusions has several advantages over an array of separate sgRNA transcriptional units. Notably, the reduced size of the sgRNA expression cassette (821 nt for mtRNA *versus* 1,146 nt for msgRNA) and its simplicity and scalability: only one promoter and terminator are needed to express an arbitrary number of sgRNA-tRNA fusions. Beyond Cas9 mutagenesis, tRNA fusions which exploit prokaryotic tRNA maturation mechanisms to process polycistronic RNA into individual molecules could be used to express any RNA-based synthetic biology tool (Altman, 1975; Apirion & Miczak, 1993; Gegenheimer & Apirion, 1981; Green et al., 2014; Lee et al., 2018; Li et al., 2005; Li & Deutscher, 2002; Mackie, 2013; Minagawa et al., 2004; Mörl & Marchfelder, 2001; Ow & Kushner, 2002; Sekiya et al., 1979). This strategy could be improved by identifying more prokaryotic tRNAs which are compatible with sgRNAs arrays, to avoid repeating the same tRNA sequence in the same sgRNA array.

Finally, Target-AID is an exceedingly easy system to build, that does not require any PCR amplification step: just swapping the 20 nt protospacer in a single two-parts Gibson or Golden-Gate assembly between a cut vector and a synthesised DNA oligo is sufficient to create a vector capable of editing a different target. The absence of homology cassette makes it uniquely straightforward to multiplex, especially with polycistronic RNA systems such as the mtRNA or mCRISPR strategies illustrated in this study. However, the time gained during design and assembly can be quickly lost again if mixed colonies need to be re-streaked (especially for slow-growing organisms like *C. autoethanogenum*). Initial screening of mutants is also more difficult than with standard homology-directed knockouts (Seys et al., 2020), as a colony PCR cannot readily identify the desired SNPs through simple gel electrophoresis.

5 Conclusion

This study is valuable because it illustrates the strengths and weaknesses of the base editor Target-AID-NG in the *Clostridium* genus and exemplifies the use of prokaryotic tRNAs to process a synthetic polycistronic RNA.

Like other base editors, Target-AID-NG offers several undeniable advantages besides bypassing the need for a functional HR or NHEJ DNA repair pathway. It is straightforward to assemble and use in a standard mutagenesis workflow and can easily be multiplexed. However, it suffers from a restricted pool of optimal and/or functional protospacers, as well as a propensity to generate mixed colonies of WT and mutated cells. Critically, like any other genome editing tool to date, Target-AID-NG is still not precise or innocuous enough to ignore the possibility of off-target mutations and the necessity of complementation studies. As illustrated in this study, particular attention should be paid to the *udg* homolog of the targeted organism, as its loss might be selected by Target-AID-NG. With whole-genome sequencing technology becoming more affordable every year, it should become a standard step to any mutant characterisation.

Fortunately, base editing is a quickly evolving field (Anzalone et al., 2020; Kantor et al., 2020). While Target-AID or Target-AID-NG in their current form might find some niche applications in contexts where conventional HR- or NHEJ-mediated mutagenesis methodologies are impossible, the obstacles encountered in the present study make it difficult to recommend as a tool for mainstream knock-out experiments in

C. autoethanogenum when other methods are available. Other base editors such as BE4 (Gaudelli et al., 2017; Gehrke et al., 2018; Li et al., 2019) or Target-AID-iSPymac (Chatterjee et al., 2020) might have marginally improved performances while conserving its simplicity of design and potential for multiplexing. Alternatively, the recently developed prime editor (Anzalone et al., 2019), that uses a reverse-transcriptase fused to nCas9 to introduce custom small mutations independently from the target sequence, might hold the greatest promise for a multiplex and scarless point mutagenesis tool in *Clostridium*.

Data availability statement

The datasets presented in this study can be found in online repositories. This data can be found here: [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA956560]

Author contributions

Conceptualisation: FS, NM, LQ, CH, and SY; choice of CLAU_532,534,1794 targets: TM; Investigation, methodology: FS; modelling of total protospacers and genomic coverage: CT-A; writing—original draft preparation, FS; writing—review and editing, CH, CT-A, NM, QL, and SY. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Biotechnology and Biological Sciences Research Council (grant numbers BB/L013940/1, BB/W01453X/1).

Acknowledgments

FS and CT-A acknowledge the receipt of a studentship from the University of Nottingham, Synthetic Biology Research Centre (SBRC) Doctoral Training Program. All authors acknowledge Graeme Fox from the Deepseq next-generation sequencing facility of the University of Nottingham for the whole genome sequencing and SNP analysis.

Conflict of interest

The 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.

The handling editor MK declared a past collaboration with the author NM.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2023.1211197/full#supplementary-material

References

Adiego-Pérez, B., Randazzo, P., Daran, J. M., Verwaal, R., Roubos, J. A., Daran-Lapujade, P., et al. (2019). Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS Microbiol. Lett.* 366 (8), fnz086–19. doi:10.1093/femsle/fnz086

Altman, S. (1975). Biosynthesis of transfer RNA in Escherichia coli. *Cell* 4 (1), 21–29. doi:10.1016/0092-8674(75)90129-4

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Available at: http://www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc.

Anzalone, A. v., Koblan, L. W., and Liu, D. R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* 38(7), 824–844. doi:10.1038/s41587-020-0561-9

Anzalone, A. v., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576 (7785), 149–157. doi:10.1038/s41586-019-1711-4

Apirion, D., and Miczak, A. (1993). RNA processing in prokaryotic cells. *BioEssays* 15(2), 113–120. doi:10.1002/bies.950150207

Baig, H., Fontanarossa, P., McLaughlin, J., Scott-Brown, J., Vaidyanathan, P., Gorochowski, T., et al. (2021). Synthetic biology open language visual (SBOL visual) version 3.0. *J. Integr. Bioinforma.* 18 (3), 20210013. doi:10.1515/jib-2021-0013

Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H., and Kondo, A. (2018). Deaminase-mediated multiplex genome editing in Escherichia coli. *Nat. Microbiol.* 3 (4), 423–429. doi:10.1038/s41564-017-0102-6

Bourgade, B., Minton, N. P., and Islam, M. A. (2021). Genetic and metabolic engineering challenges of C1-gas fermenting acetogenic chassis organisms. *Oxf. Univ. Press* 45 (2), fuab008. doi:10.1093/femsre/fuab008

Cañadas, I. C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., and Minton, N. P. (2019). RiboCas: A Universal CRISPR-based editing tool for Clostridium. *ACS Synth. Biol.* 8 (6), 1379–1390. doi:10.1021/acssynbio.9b00075

Chan, P. P., and Lowe, T. M. (2016). GtRNAdb 2.0: An expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* 44 (D1), D184–D189. doi:10.1093/nar/gkv1309

Chan, P. P., and Lowe, T. M. (2009). GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* 37, D93–D97. doi:10.1093/nar/gkn787

Chatterjee, P., Lee, J., Nip, L., Koseki, S. R. T., Tysinger, E., Sontheimer, E. J., et al. (2020). A Cas9 with PAM recognition for adenine dinucleotides. *Nat. Commun.* 11 (1), 2474. doi:10.1038/s41467-020-16117-8

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10 (2), giab008. doi:10.1093/gigascience/giab008

Davis, M. W., and Jorgensen, E. M. (2022). *Frontiers in Bioinformatics*, 2. doi:10.3389/fbinf.2022.818619ApE, A plasmid editor: A freely available DNA Manipulation and Visualization Program

Dong, F., Xie, K., Chen, Y., Yang, Y., and Mao, Y. (2017). Polycistronic tRNA and CRISPR guide-RNA enables highly efficient multiplexed genome engineering in human cells. *Biochem. Biophysical Res. Commun.* 482 (4), 889–895. doi:10.1016/j.bbrc.2016.11.129

Eid, A., Alshareef, S., and Mahfouz, M. M. (2018). CRISPR base editors: Genome editing without double-stranded breaks. *Biochem. J.* 475 (11), 1955–1964. doi:10.1042/BCJ20170793

Fackler, N., Heijstra, B. D., Rasor, B. J., Brown, H., Martin, J., Ni, Z., et al. (2021). Stepping on the gas to a Circular Economy: Accelerating development of Carbon-negative chemical production from gas fermentation. *Annu. Rev. Chem. Biomol. Eng.* 12 (1), 439–470. doi:10.1146/annurev-chembioeng-120120-021122

Fukui, K. (2010). DNA mismatch repair in eukaryotes and bacteria. *J. Nucleic Acids* 2010, 1–16. doi:10.4061/2010/260512

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al. (2017). Programmable base editing of A●T to G●C in genomic DNA without DNA cleavage. *Nature* 551 (7681), 464–471. doi:10.1038/nature24644

Gegenheimer, P., and Apirion, D. (1981). Processing of procaryotic ribonucleic acid. *Microbiol. Rev.* 45 (4), 502–541. doi:10.1128/mr.45.4.502-541.1981

Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Zeng, J., Bauer, D. E., et al. (2018). An apobec3a-cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* 36 (10), 977–982. doi:10.1038/nbt.4199

Green, A. A., Silver, P. A., Collins, J. J., and Yin, P. (2014). Toehold switches: De-novo-designed regulators of gene expression. *Cell* 159 (4), 925–939. doi:10.1016/j.cell.2014.10.002

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R., and Hofacker, I. L. (2008). The Vienna RNA Websuite. *Nucleic Acids Res.* 36, W70–W74. doi:10.1093/nar/gkn188

Huang, H., Chai, C., Li, N., Rowe, P., Minton, N. P., Yang, S., et al. (2016). CRISPR/Cas9-Based efficient genome editing in Clostridium ljungdahlii, an autotrophic gas-fermenting bacterium. *ACS Synth. Biol.* 5 (12), 1355–1361. doi:10.1021/acssynbio.6b00044

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), 1085. doi:10.1186/s12864-015-2287-5

Jiang, W., Feng, S., Huang, S., Yu, W., Li, G., Yang, G., et al. (2018). BE-PLUS: A new base editing tool with broadened editing window and enhanced fidelity. *Cell Res.* 28 (8), 855–861. doi:10.1038/s41422-018-0052-4

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in adaptive bacterial immunity. *Science* 337 (6096), 816–821. doi:10.1126/science.1225829

Kantor, A., McClements, M., and Maclaren, R. (2020). Crispr-cas9 dna base-editing and prime-editing. In *Int. J. Mol. Sci.* 21 (17), 6240. doi:10.3390/ijms21176240

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. doi:10.1038/nature17946

Kreuger, F. (2012). *Trim Galore:: A wrapper tool around Cutadapt and FastQC*. Available at: https://www.Bioinformatics.Babraham.Ac.Uk/Projects/Trim_galore/.

Lee, Y. J., Kim, S. J., and Moon, T. S. (2018). Multilevel Regulation of bacterial gene expression with the combined STAR and antisense RNA system. *ACS Synth. Biol.* 7 (3), 853–865. doi:10.1021/acssynbio.7b00322

Li, G. M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Res.* 18 (1), 85–98. doi:10.1038/cr.2007.115

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26 (5), 589–595. doi:10.1093/bioinformatics/btp698

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324

Li, Q., Seys, F. M., Minton, N. P., Yang, J., Jiang, Y., Jiang, W., et al. (2019). CRISPR–Cas9 D10A nickase-assisted base editing in the solvent producer Clostridium beijerinckii. *Biotechnol. Bioeng.* 116 (6), 1475–1483. doi:10.1002/bit.26949

Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., et al. (2018). Base editing with a Cpf1-cytidine deaminase fusion. *Nat. Biotechnol.* 36 (4), 324–327. doi:10.1038/nbt.4102

Li, Z., and Deutscher, M. P. (2002). RNase E plays an essential role in the maturation of Escherichia coli tRNA precursors. *Rna* 8 (1), 97–109. doi:10.1017/S1355838202014929

Li, Z., Gong, X., Joshi, V. H., and Li, M. (2005). Co-evolution of tRNA 3′ trailer sequences with 3′ processing enzymes in bacteria. *RNA* 11 (5), 567–577. doi:10.1261/rna.7287505

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 (3), 335–344. doi:10.1038/s41587-021-01195-w

Liu, Z., Shan, H., Chen, S., Chen, M., Song, Y., Lai, L., et al. (2019). Efficient base editing with expanded targeting scope using an engineered Spy-mac Cas9 variant. *Cell Discov.* 5 (Issue 1), 58. doi:10.1038/s41421-019-0128-4

Lu, H., Zhang, Q., Yu, S., Wang, Y., Kang, M., Han, S., et al. (2022). Optimization of CRISPR/Cas9-based multiplex base editing in Corynebacterium glutamicum. *Shengwu Gongcheng Xuebao/Chinese J. Biotechnol.* 38 (2), 780–795. doi:10.13345/J.CJB.210109

Luo, Y., Ge, M., Wang, B., Sun, C., Wang, J., Dong, Y., et al. (2020). CRISPR/Cas9-deaminase enables robust base editing in Rhodobacter sphaeroides 2.4.1. *Microb. Cell Factories* 19 (1), 93–15. doi:10.1186/s12934-020-01345-w

Mackie, G. A. (2013). RNase E: At the interface of bacterial RNA processing and decay. *Nat. Rev. Microbiol.* 11 (1), 45–57. doi:10.1038/nrmicro2930

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* 17 (1), 10. doi:10.14806/ej.17.1.200

Marx, V. (2018). Base editing a CRISPR way. *Nat. Methods* 15 (10), 767–770. doi:10.1038/s41592-018-0146-4

Minagawa, A., Takaku, H., Takagi, M., and Nashimoto, M. (2004). A novel Endonucleolytic mechanism to generate the CCA 3′ Termini of tRNA molecules in Thermotoga maritima. *J. Biol. Chem.* 279 (15), 15688–15697. doi:10.1074/jbc.M313951200

Minton, N. P., Ehsaan, M., Humphreys, C. M., Little, G. T., Baker, J., Henstra, A. M., et al. (2016). A roadmap for gene system development in Clostridium. *Anaerobe* 41, 104–112. doi:10.1016/j.anaerobe.2016.05.011

Modrich, P. (2006). Mechanisms in eukaryotic mismatch repair. *J. Biol. Chem.* 281 (41), 30305–30309. doi:10.1074/jbc.R600022200

Molla, K. A., and Yang, Y. (2019). CRISPR/Cas-Mediated base editing: Technical Considerations and practical applications. *Trends Biotechnol.* 37 (10), 1121–1142. doi:10.1016/J.TIBTECH.2019.03.008

Mörl, M., and Marchfelder, A. (2001). The final cut: The importance of tRNA 3′-processing. *EMBO Rep.* 2 (Issue 1), 17–20. doi:10.1093/embo-reports/kve006

Ng, Y. K., Ehsaan, M., Philip, S., Collery, M. M., Janoir, C., Collignon, A., et al. (2013). Expanding the Repertoire of gene tools for precise Manipulation of the Clostridium difficile genome: Allelic Exchange using pyrE alleles. *PLoS ONE* 8 (2), e56051. doi:10.1371/journal.pone.0056051

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353 (6305), aaf8729. doi:10.1126/science.aaf8729

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* 361 (6408), 1259–1262. doi:10.1126/science.aas9129

Ow, M. C., and Kushner, S. R. (2002). Initiation of tRNA maturation by RNase E is essential for cell viability in E. coli. *Genes and Dev.* 16 (9), 1102–1115. doi:10.1101/gad.983502

Rice, P., Longden, I., and Bleasby, A. (2000). Emboss: The European molecular biology open Software suite. *Trends Genet. TIG* 16 (6), 276–277. doi:10.1016/S0168-9525(00)02024-2

Sekiya, T., Contreras, R., Takeya, T., and Khorana, H. G. (1979). Total synthesis of a tyrosine suppressor transfer RNA gene. XVII. Transcription, *in vitro*, of the synthetic gene and processing of the primary transcript to transfer RNA. *J. Biol. Chem.* 254 (13), 5802–5816. doi:10.1016/s0021-9258(18)50483-x

Seys, F. M., Humphreys, C. M., Tomi-Andrino, C., Qi, L., Millat, T., Yang, S., et al. (2023). *Target-AID off-target detection in C. autoethanogenum. NCBI. Version 1*. Bioproject PRJNA956560.

Seys, F. M., Rowe, P., Bolt, E. L., Humphreys, C. M., and Minton, N. P. (2020). A gold standard, CRISPR/Cas9-based complementation strategy reliant on 24 nucleotide bookmark sequences. *Genes* 11 (4), 458. doi:10.3390/genes11040458

Spampinato, C. P., Gomez, R. L., Galles, C., and Lario, L. D. (2009). From bacteria to plants: A compendium of mismatch repair assays. *Mutat. Research/Reviews Mutat. Res.* 682 (2–3), 110–128. doi:10.1016/j.mrrev.2009.07.001

van der Auwera, G., and O'Connor, B. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. 1st Edition. Available at: https://www.Oreilly.Com/Library/View/Genomics-in-the/9781491975183/.

Wang, L., Xue, W., Yan, L., Li, X., Wei, J., Chen, M., et al. (2017). Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Res.* 27 (10), 1289–1292. doi:10.1038/cr.2017.111

Wang, Y., Liu, Y., Liu, J., Guo, Y., Fan, L., Ni, X., et al. (2018). Macbeth: Multiplex automated Corynebacterium glutamicum base editing method. *Metab. Eng.* 47, 200–210. doi:10.1016/j.ymben.2018.02.016

Wang, Y., Liu, Y., Zheng, P., Sun, J., and Wang, M. (2020). Microbial base editing: A Powerful Emerging technology for Microbial genome engineering. *Trends Biotechnol.* 39, 165–180. doi:10.1016/j.tibtech.2020.06.010

Williams, J. S., and Kunkel, T. A. (2014). Ribonucleotides in DNA: Origins, repair and consequences. *DNA Repair* 19, 27–37. doi:10.1016/j.dnarep.2014.03.029

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

Woods, C., Humphreys, C. M., Tomi-Andrino, C., Henstra, A. M., Köpke, M., Simpson, S. D., et al. (2022). Required gene Set for autotrophic growth of *Clostridium autoethanogenum*. *Appl. Environ. Microbiol.* 88 (7), e0247921. doi:10.1128/aem.02479-21

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci.* 112 (11), 3570–3575. doi:10.1073/pnas.1420294112

Zhang, Y., Wang, J., Wang, Z., Zhang, Y., Shi, S., Nielsen, J., et al. (2019). A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in Saccharomyces cerevisiae. *Nat. Commun.* 10 (1), 1053. doi:10.1038/s41467-019-09005-3

Zhao, C., Zheng, X., Qu, W., Li, G., Li, X., Miao, Y.-L., et al. (2017). CRISPR-Offinder: A CRISPR guide RNA design and off-target searching tool for user-defined protospacer adjacent motif. *Int. J. Biol. Sci.* 13 (12), 1470–1478. doi:10.7150/ijbs.21312

Zhigang, W., Smith, D. G., and Mosbaugh, D. W. (1991). Overproduction and characterization of the uracil-DNA glycosylase inhibitor of bacteriophage PBS2. *Gene* 99 (1), 31–37. doi:10.1016/0378-1119(91)90030-F

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31 (13), 3406–3415. doi:10.1093/nar/gkg595

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Methods
PUBLISHED 29 August 2023
DOI 10.3389/fbioe.2023.1217811

OPEN ACCESS

EDITED BY
Anindya Bandyopadhyay,
Reliance Industries, India

REVIEWED BY
Nicholas R. Sandoval,
Tulane University, United States
Rahul Badhwar,
Reliance Industries, India

*CORRESPONDENCE
Ian Blaby,
ikblaby@lbl.gov

RECEIVED 05 May 2023
ACCEPTED 21 August 2023
PUBLISHED 29 August 2023

CITATION
Simirenko L, Cheng J-F and Blaby I (2023), gRNA-SeqRET: a universal tool for targeted and genome-scale gRNA design and sequence extraction for prokaryotes and eukaryotes.
Front. Bioeng. Biotechnol. 11:1217811.
doi: 10.3389/fbioe.2023.1217811

gRNA-SeqRET: a universal tool for targeted and genome-scale gRNA design and sequence extraction for prokaryotes and eukaryotes

Lisa Simirenko[1], Jan-Fang Cheng[1,2] and Ian Blaby[1,2]*

[1]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, [2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

High-throughput genetic screening is frequently employed to rapidly associate gene with phenotype and establish sequence-function relationships. With the advent of CRISPR technology, and the ability to functionally interrogate previously genetically recalcitrant organisms, non-model organisms can be investigated using pooled guide RNA (gRNA) libraries and sequencing-based assays to quantitatively assess fitness of every targeted locus in parallel. To aid the construction of pooled gRNA assemblies, we have developed an *in silico* design workflow for gRNA selection using the gRNA Sequence Region Extraction Tool (gRNA-SeqRET). Built upon the previously developed CCTop, gRNA-SeqRET enables automated, scalable design of gRNA libraries that target user-specified regions or whole genomes of any prokaryote or eukaryote. Additionally, gRNA-SeqRET automates the bulk extraction of any regions of sequence relative to genes or other features, aiding in the design of homology arms for insertion or deletion constructs. We also assess *in silico* the application of a designed gRNA library to other closely related genomes and demonstrate that for very closely related organisms Average Nucleotide Identity (ANI) > 95% a large fraction of the library may be of relevance. The gRNA-SeqRET web application pipeline can be accessed at https://grna.jgi.doe.gov. The source code is comprised of freely available software tools and customized Python scripts, and is available at https://bitbucket.org/berkeleylab/grnadesigner/src/master/ under a modified BSD open-source license (https://bitbucket.org/berkeleylab/grnadesigner).

KEYWORDS

CRISPR, gRNA, design tool, batch DNA extraction, CRISPR screen

Introduction

CRISPR-based genome editing has rapidly become the targeted engineering technology of choice due to its programmability, scalability and near universal application (Jinek et al., 2012; Wiedenheft et al., 2012; Jiang et al., 2013; Ran et al., 2013). The core machinery enabling editing comprises a CRISPR-associated protein (Cas; an endonuclease) and a short guide RNA (a fusion of a variable, target specific sequence and a Cas-specific stem-loop forming sequence required for CRISPR nuclease maturation). Programmability is achieved by complementarity of this variable region to the target DNA, which for Cas binding and double-strand cleavage must be juxtaposed to a Cas-specific sequence termed the protospacer adjacent motif (PAM) (Mojica et al., 2009). This requirement, which for *Streptococcus pyogenes* Cas9, the first to be discovered and the most widely used, is 5′ NGG 3′, constitutes the only limitation in targeting DNA. However, even here alternate Cas

endonuclease's afford flexibility due to different PAMs (albeit with different activities; for example, Cas12a with a PAM of 5′ YTN 3′, induces a 5′ overhang double strand cleavage) (Zetsche et al., 2015; Yamano et al., 2016).

Initial DNA-editing exploited the double-strand cutting induced by Cas9 followed by low efficiency non-homologous end joining (NHEJ) repair mechanisms in eukaryotes yielding loss-of-function mutants. In the absence of NHEJ, a DNA fragment comprising regions of homology on either side of the targeted cut site allows homologous recombination repair to generate a scarless mutation in the genome. Alternatively, CRISPR interference or activation (CRISPRi/a) can be employed to modulate transcription without inducing a break in the genome (Qi et al., 2013; Liu et al., 2019). As the technology has matured additional applications have been developed furthering CRISPR's editing utility (Pickar-Oliver and Gersbach, 2019; Liu et al., 2021; Nakamura et al., 2021; Wang and Doudna, 2023).

Many web-based and downloadable computational tools are available for gRNA design. These tools typically provide the user with 20 nucleotide sequences flanking a PAM for targeting the specified loci [(Naito et al., 2015; Doench et al., 2016; Concordet and Haeussler, 2018) and reviewed in (Wilson et al., 2018; Alipanahi et al., 2022)]. Automated design tools are particularly useful for the experimental design of constructs involving many, or genome-scale, targets such as needed for CRISPR-screening (Bock et al., 2022), or Perturb-Seq (Dixit et al., 2016). However, many preexisting tools are limited to single or pre-computed model organisms, or, where user-provided genomes can be provided, are specifically optimized for prokaryote genome architecture (i.e., input sequence format does not allow for structural annotations such as multiple chromosomes or specifying intron/exons coding/non-coding regions) (Poudel et al., 2022). CCTop and CHOPCHOP, for example, include the genomes for many organisms (Stemmer et al., 2015; Labun et al., 2019), and additional genomes can be requested by email. Other tools cater to specific communities or groups of organism (Peng and Tarleton, 2015; He et al., 2021).

To overcome these limitations, and to enable universal, organism-agnostic design, we developed the guide RNA Sequence Extraction Tool (gRNA-SeqRET), which is built upon the previously published tool CCTop. CCTop's standalone version was specifically selected due to its open-source licensing, allowing further development, and the options it provides for design. gRNA-SeqRET allows users to create their own accounts where genomes can be uploaded and securely saved. Designs are scoped by entering a series of criteria into the website, and the data packaged and piped into CCTop. Once the job is complete, the results are accessible for download from the tool's website. gRNA-SeqRET has two main advantages over other tools: 1) compatible with any user-provided input genome files in GenBank and GFF formats, allowing universal design for both prokaryote and eukaryote genome structures; and 2) functionality to bulk extract specified target DNA regions for scalable repair template design.

Methods

gRNA-SeqRET employs Flask and jQuery for the web-based user interface (UI) and a PostgreSQL database which maintains track of the user's genomes and submissions. The tool is composed in Python 3, and utilizes the following open source applications for the indicated tasks: CCTop standalone (Stemmer et al., 2015)—generates the complete list of potential gRNAs for a given pre-processed genome file ranked by predicted cutting score, and a FASTA file with the extracted target region(s); BowTie v1.3.0 (Langmead et al., 2009)—generates the indexes needed by CCTop; The ViennaRNA Package (Lorenz et al., 2011)—generates RNA folding predictions used to evaluate the gRNAs in the CCTop output; BioPython (Cock et al., 2009)—Bio.SeqIO is used to parse genome sequence from GenBank files and convert the sequence to FASTA format; GFFutils v0.11.1 (https://daler.github.io/gffutils/index.html)—provides methods for creating an SQLite database (which contains the processed genome files) from a GFF and searching features annotated in GFF files; GFFtools-GX (https://github.com/vipints/GFFtools-GX)—converts GenBank annotations to GFF3 format; Cromwell—Workflow engine for automating the back-end pipeline (https://cromwell.readthedocs.io/en/stable/).

Results and Discussion

gRNA-SeqRET is compatible with any prokaryote or eukaryote genome

The goal of gRNA-SeqRET is to provide an intuitive web-based user interface for the design of gRNA sequences and custom sequence extraction from user-provided genomes. Figure 1 provides overview schematics of the general tool workflow, focusing on both the processing of the uploaded genome data and the subsequent extraction and guide RNA designs.

The gRNA-SeqRET application accepts either FASTA or GenBank file types for a given organism's genome. New users of gRNA-SeqRET are required to register an account, which serves two purposes. Firstly, this allows uploaded genomes to be securely maintained on the server (genome files need only be uploaded once, and no files are accessible to other investigators). Secondly, this approach enables asynchronous use of the tool, by which an uploaded genome can be saved and processing (which takes many minutes) can be performed in the background. Files can be uploaded in compressed format (.zip or .gz), though must only contain a single file. If a FASTA file is uploaded the user will be prompted to additionally provide annotations in the format of a GFF3 file (Figure 1A; Table 1). The GFF format is preferred due to the complexities of converting GenBank to GFF, especially for eukaryotes. The uploaded files are processed in order to generate the required input files for CCTop: Bowtie indices, Exon and Gene BED and json files. Specifically, indices are generated using bowtie-build, and the BED files are created using a python script provided with CCTop's standalone source code. Additionally, a searchable SQLite database is created using GFFUtils, which contains the annotations extracted from the uploaded GenBank or GFF file. This is used to generate JSON files that list the annotated genes or locus IDs and the features (e.g., CDS, exons, *etc.*) enabling the population of genome feature menus (see below).

Processing newly uploaded genomes typically takes around a minute for a small prokaryotic genome, and approximately 5–10 min for a large complex eukaryote (e.g., a plant) genome.

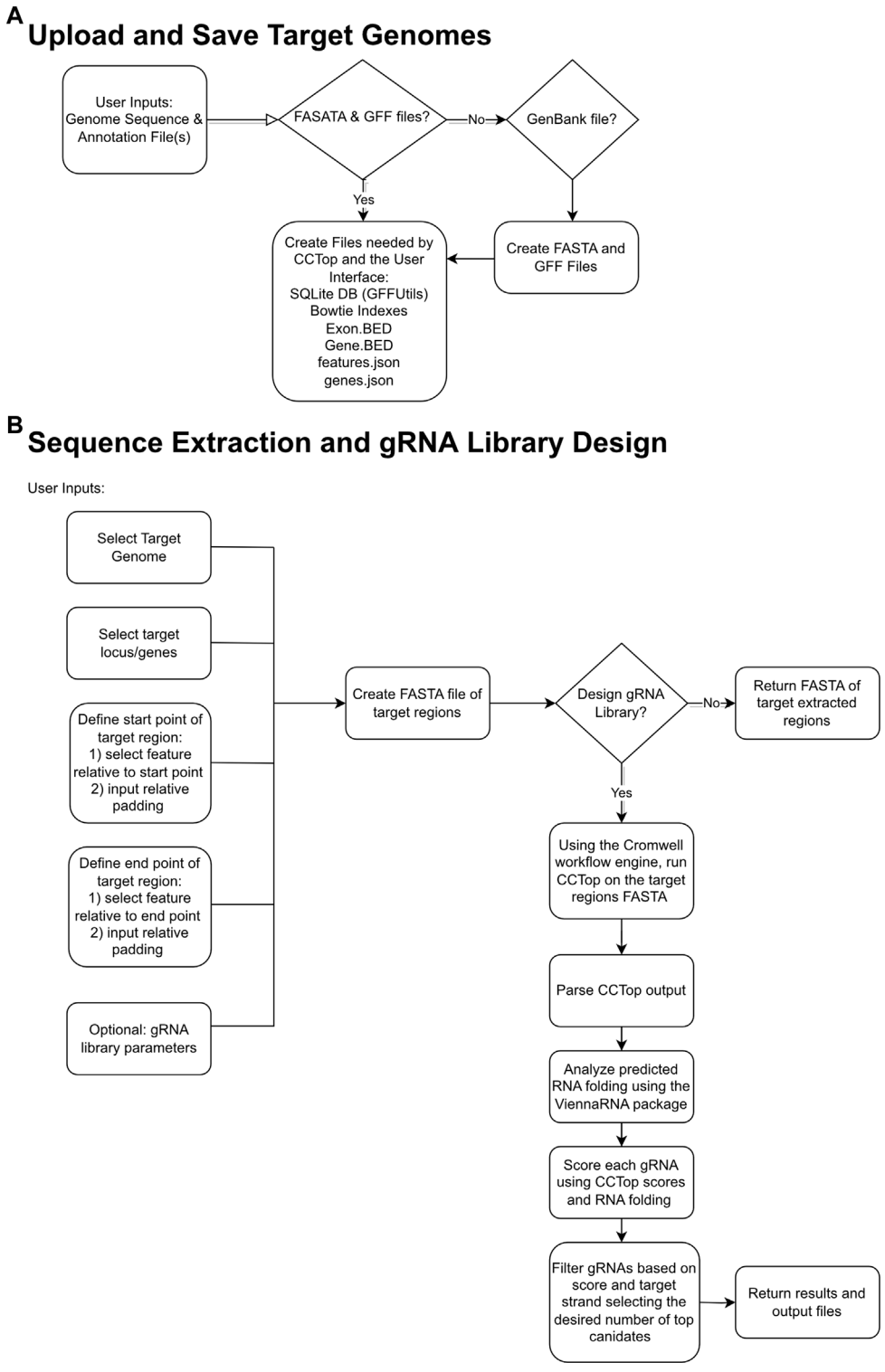


FIGURE 1
Flowcharts demonstrating relationships between input files, logic choices and outputs for gRNA-SeqRET. Schemas are shown for saving target genomes **(A)** and sequence extraction and gRNA design **(B)**, where diamond boxes represent conditions and rounded rectangles represent tasks and files.

Once generated, the processed files are preserved in a user's individual account and are inaccessible to other users. Consequently, this step needs to be performed only once per genome, and once complete, is ready for initiating designs. All processed genomes will be available in the "Your Genomes" page of gRNA-SeqRET with the status confirmed as complete (Figure 2).

TABLE 1 Input files, parameters and results for gRNASeq-RET.

Item (URL)	Field	Description, page location, default parameters and input options
Input file(s) (https://grna.jgi.doe.gov/save_genome.html)	Genome data	Requisite genome data providing sequence data and feature coordinates. Can be provided as either a GenBank or both a FASTA and GFF (GFF is preferred, especially for eukaryotes)
Target region selection (https://grna.jgi.doe.gov/create_step1.html)	Design name	Unique user provided name that will become the job ID
	Target genome	Uploaded, processed genomes can be selected from the dropdown menu
	Target regions	Individual, or groups of loci in the menu detected from the input annotations can be copied to select, or left blank for genome-scale targeting
	Limit regions	Target regions can be limited to coding, non-coding or both
	Start and end point definitions	Defines the number of base pairs upstream and downstream of a feature in the genome file
	Guide RNA design	Opens options for gRNA design for the specified target regions
	PAM	Specifies the Cas-specific protospacer adjacent motif (PAM)
	Scaffold sequence	Defines the scaffold sequence; default is the canonical hybrid scaffold and contributes to the folding score generated by gRNA-SeqRET
	Target site length	Specifies the spacer region length; default is 20
	Max mismatches offsite targets	Specifies the maximum number of mismatches that will be considered an off-site target and will be excluded
	No. gRNAs	Specifies the number of guide sequences returned to the output file per target region. Note, unlike CCTop, which returns all possible guides, gRNASeqRET will only output this specified number, ranked by CCTop's predicted cutting score (default is 3)
	Target strand	Optionally filter results based on whether the gRNAs are on the coding or non-coding strand
Job results (https://grna.jgi.doe.gov/design_list)		Lists all complete and running jobs

gRNA-SeqRET automates feature batch extraction and guide RNA design

Once preprocessing is complete, designs and sequence extractions can be performed on genomes uploaded by selecting the Create New Design tool in the top left corner of the menu (Figures 1B, 2). After entering a name for the design, which will become the job title, and selecting a genome from the drop-down menu, the user will specify the loci to be targeted. All locus IDs identified in the provided genome input files will be automatically populated in the menu on the right of the page; from here, individual loci can be selected or pasted, or if left blank all loci will be applied for genome-scale targeting.

Next, the precise regions to target must be provided, enabling batch programming for 1) extraction of genomic sequence data within defined regions and 2) design of gRNA sequences within these defined areas (described in Table 1). Specific use-case examples for various options are described below. Target region definitions are enabled at the nucleotide-level resolution relative to defined features in the input files by entering details into subsequent menus (Figures 1B, 2). This is done by firstly stipulating whether coding, non-coding or both coding and non-coding regions should be included, and secondly by entering the number of bases upstream or downstream of feature reference points to target. The nature of features (defined as reference points in gRNA-SeqRET menus) available in these menus is dependent on those provided in the genome input file; for prokaryotes this will generally be limited to coding sequences (CDSs) and non-coding sequences, but eukaryote GFF files may provide broader options, such as introns and exon coordinates, depending on the extent of structural annotations in the uploaded annotations file. The tool warns if incompatible inputs are provided (for example, an error message will appear and the submit button will be disabled if in the first section the default "coding regions only" is selected as well as a region upstream of the loci coding region). To provide maximal flexibility of region selection, a user can define the number of bases up- or downstream of two independent reference points, constituting the start and end points of the target regions within the genome. Bases upstream of a feature are designated as being negative, and positive numbers indicate the number of bases downstream of this feature (e.g., a start codon). A dynamically generated schematic illustrating the selected area aids the user by updating in real time as entries are made (Figure 2).

This sequence selection feature has multiple utilities, as it enables genome-scale, batch extraction of defined regions. One example would be for homology arm design, as required for CRISPR repair template design, which can then be bulk-downloaded as a FASTA file. Additionally, defining sequence regions also serves to specify the regions with which to target gRNA design, such as targeting specifically upstream (but not so far upstream as to enter the next 5' coding genome feature) of a gene for promoter-region CRISPRa guides, as described below. Furthermore, batch extraction of defined regions of sequence has

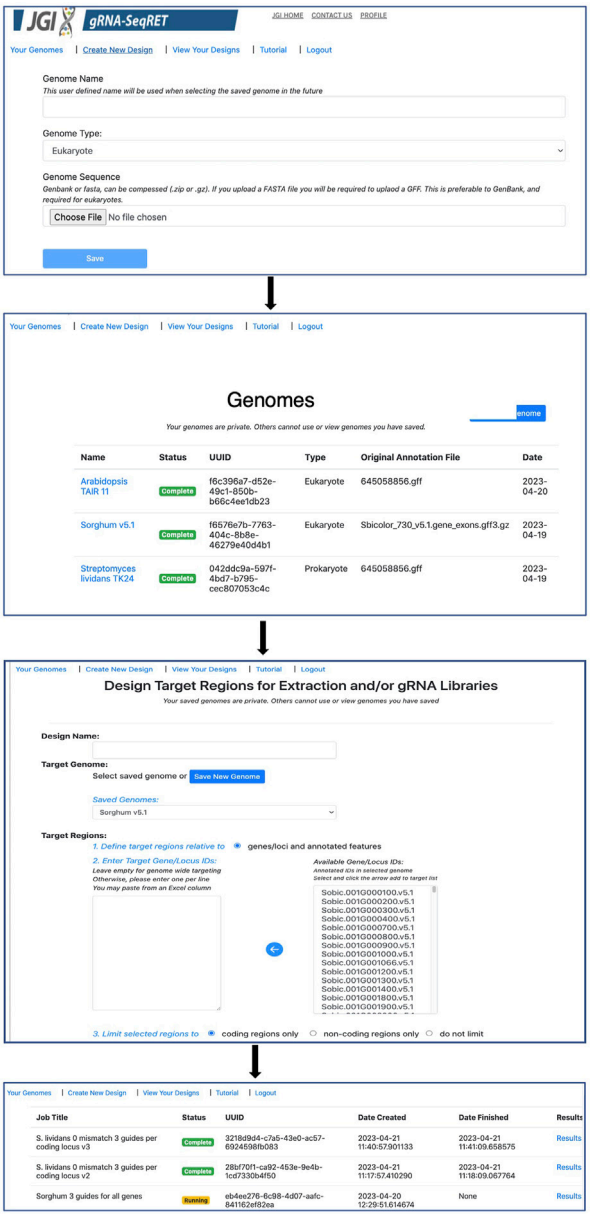


FIGURE 2
Walkthrough of the gRNA-SeqRET design process.

functions beyond CRISPR, for example, defining homology arms for homologous recombination (HR)-mediated gene disruptions, or for bulk extraction of sequence upstream of coding regions for promoter libraries. At this stage the defined sequences can be downloaded as a FASTA file, and/or the tool can proceed to gRNA design.

If the gRNA design option is selected, additional fields are made available (Figure 2). Several of these parameters, including the specific PAM, scaffold sequence, target site length and number of allowed sequence mismatches are necessary inputs for CCTop, and detailed information is provided in the publication describing that tool (Stemmer et al., 2015) and summarized in Table 1. Beyond CCTop prerequisites, gRNA-SeqRET additionally asks for the desired number of gRNAs per target and a preference for targeted strand. These last two fields provide the criteria serving to limit the complete CCTop output to just those matching the user's requirements; i.e., the number of guides specified in the "Number of guide RNAs per gene/custom feature" input box. For example, entering "3" in this field will yield the top 3 gRNAs as determined by CCTop's predicted cutting score and RNA folding predictions for the protospacer and scaffold. Clicking "submit" executes the job, which can range from seconds for low numbers (e.g., <10) of target loci to several minutes for genome-scale (i.e., all loci) in prokaryotic genomes, to several hours for genome-scale jobs in very large eukaryotic genomes (a submission comprising design of 3 guide RNAs for every coding region of the Araport11 annotation of the *Arabidopsis thaliana* genome (Cheng et al., 2017) ran ~12 h). The "View Your Designs" page lists all completed and running jobs and provides a results link to the job output (Figure 2). Clicking this link provides a summary of the job parameters as well as the output files for downloading. "Results" provides a comma separated values (CSV) file containing all guide sequences, providing a unique name (the locus name appended by the CCTop designation), the start and end chromosomal coordinates, strand orientation, sequence and specific PAM. "Target regions" provides a FASTA output of all targeted sequences, and the Report contains a list of loci where the desired number of gRNAs was not found, which can be used to run another design round with an altered sequence targeting criteria if desired. Clicking on "All output files (.tar.gz)" will download an archive of all files described above, as well as a file called "scoring.log" which contains all statistics and predicted cutting scores associated with each gRNA. Output files are maintained within the user's individual account for 3 months. Additionally, users can make their designs public by clicking the "Make Publicly Available" option. Clicking this will open a new page where fields describing the purpose of the library and the target organism NCBI taxonomic ID can be completed, and the designs will be accessible to all users in the "Public Designs" link.

Applicability of a given gRNA library to other closely related genomes

Cost reductions and availability of high-variant oligonucleotide pools and high throughput sequencing has enabled the construction of gRNA libraries to become powerful and increasingly accessible approaches for rapid gene function interrogation (Schwartz et al., 2019; Bock et al., 2022; Cooper et al., 2022; Shi et al., 2022; Trivedi et al., 2023). Nevertheless, the library assembly and sequencing-based quality control (ensuring representation of all variants and possible skews in the population) can represent a significant endeavor and investment. Since a single constructed cloned library yields sufficient material for many tens of independent experiments, we wondered how applicable a given gRNA library would be to other closely related species. To address this, we first used gRNA-SeqRET to design a genome-scale gRNA library targeting all coding regions in the Actinomycetes *Streptococcus lividans* TK24. gRNA-SeqRET was employed to report 3 gRNAs per target region, resulting in 22,641 total spacer designs (Supplementary Table S1). We turned to the Integrated Microbial Genomes and Microbiomes (IMG) (Chen et al., 2023)

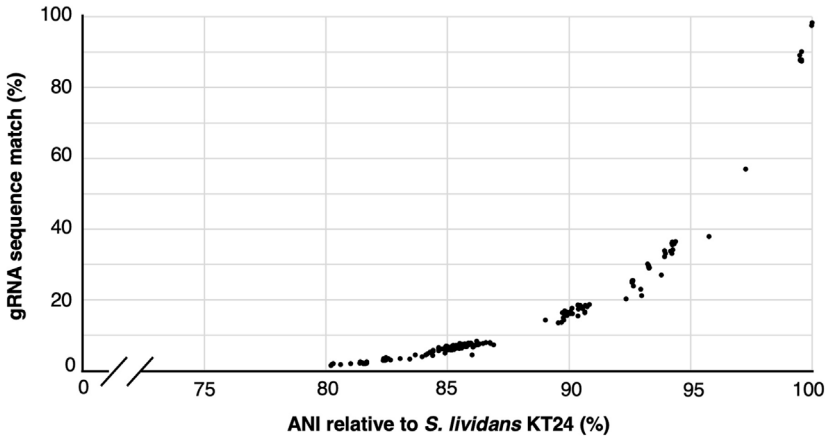


FIGURE 3
gRNA sequence applicability reduces rapidly with genome distance. Correlation between the percentage of 22,741 gRNAs designed to target *S. lividans* TK24 that align perfectly to 192 related organisms showing an average nucleotide identify (ANI) of ≥80%.

database and queried for all genomes characterized with an average nucleotide identity (ANI) ≥80% relative to *S. lividans* TK24, yielding 192 genomes (Supplementary Table S2). While an ANI is a statistic generated from the complete genome sequence, and we specifically searched for gRNA sequences in coding regions only, this approach enabled us to retrieve highly similar genome sequences. We then searched the genomes of each of these 192 microbes for perfect matches to each of the ~22 thousand 20mer sequences (plus PAM; Supplementary Table S3), providing a sense as to how applicable a given gRNA library could be to alternate closely related organisms. Unsurprisingly, guide sequence alignment reduces dramatically with genome distance (Figure 3). While the fraction of guide sequences with perfect alignment remains high (~90%) for genomes with very high similarity (i.e., ANI >98%), this drops to ~45% in genomes with and ANI of 95% (Figure 3; Supplementary Table S3). Nevertheless, since organisms belonging to the same species typically exhibit an ANI of 95%, this analysis demonstrates a pooled gRNA library designed to one organism could be of value to strains within a species, and perhaps to other closely related species (Konstantinidis and Tiedje, 2005), although this will depend on the specific species. Worth noting though is with increased taxonomic distance and genetic drift, off site targeting may also increase and confound the data. Additionally, we found that many of the 192 genomes contained several hundred occurrences of multiple matches per guide sequence, indicating a perfect sequence alignment in an off-target location (Supplementary Table S3).

Future directions

During our internal use of the present version of gRNA-SeqRET, we have identified several areas we are considering for future enhancement. The first of these builds upon the possible re-use of guide RNAs in multiple organisms, by adding an option to provide multiple genomes so that the application can filter out possible gRNAs that occur in both genomes but without off-target events. Another is to enable a user to provide custom annotations allowing regions to be targeted or extracted in addition to those defined in the genome input file. Finally, we may develop an option to target specific alleles in polyploid genomes.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

LS, J-FC, and IB conceived of the tool. LS designed and built the software and performed analysis. LS, J-FC, and IB tested and refined the software. IB wrote the manuscript with input and comments from J-FC and LS. All authors contributed to the article and approved the submitted version.

Funding

This work has been supported by the DOE Joint Genome Institute (http://jgi.doe.gov) by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through Contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

Acknowledgments

The authors would like to thank Nicolas Grosjean for helpful feedback on the manuscript, Rekha Seshadri for technical assistance with IMG, as well as Ben Shen, Cameron Currie and Ashley Shade for access to their unpublished genomic data.

Conflict of interest

The 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.2023.1217811/full#supplementary-material

References

Alipanahi, R., Safari, L., and Khanteymoori, A. (2022). CRISPR genome editing using computational approaches: a survey. *Front. Bioinforma.* 2, 1001131. doi:10.3389/fbinf.2022.1001131

Bock, C., Datlinger, P., Chardon, F., Coelho, M. A., Dong, M. B., Lawson, K. A., et al. (2022). High-content CRISPR screening. *Nat. Rev. Methods Prim.* 2 (1), 8. doi:10.1038/s43586-021-00093-4

Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., et al. (2023). The IMG/M data management and analysis system v. 7: content updates and new features. *Nucleic Acids Res.* 51 (1), D723–D732. doi:10.1093/nar/gkac976

Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., and Town, C. D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 89 (4), 789–804. doi:10.1111/tpj.13415

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25 (11), 1422–1423. doi:10.1093/bioinformatics/btp163

Concordet, J.-P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic acids Res.* 46 (1), W242–W245. doi:10.1093/nar/gky354

Cooper, Y. A., Guo, Q., and Geschwind, D. H. (2022). Multiplexed functional genomic assays to decipher the noncoding genome. *Hum. Mol. Genet.* 31 (1), R84–R96. doi:10.1093/hmg/ddac194

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., et al. (2016). Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167 (7), 1853–1866. e17. doi:10.1016/j.cell.2016.11.038

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34 (2), 184–191. doi:10.1038/nbt.3437

He, C., Liu, H., Chen, D., Xie, W. Z., Wang, M., Li, Y., et al. (2021). CRISPR-cereal: a guide RNA design tool integrating regulome and genomic variation for wheat, maize and rice. *Plant Biotechnol. J.* 19 (11), 2141–2143. doi:10.1111/pbi.13675

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31 (3), 233–239. doi:10.1038/nbt.2508

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science* 337 (6096), 816–821. doi:10.1126/science.1225829

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci.* 102 (7), 2567–2572. doi:10.1073/pnas.0409727102

Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic acids Res.* 47 (1), W171–W174. doi:10.1093/nar/gkz365

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10 (3), R25–R10. doi:10.1186/gb-2009-10-3-r25

Liu, G., Lin, Q., Jin, S., and Gao, C. (2021). The CRISPR-Cas toolbox and gene editing technologies. *Mol. Cell* 82, 333–347. doi:10.1016/j.molcel.2021.12.002

Liu, Y., Wan, X., and Wang, B. (2019). Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria. *Nat. Commun.* 10 (1), 3693. doi:10.1038/s41467-019-11479-0

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., et al. (2011). ViennaRNA package 2.0. *Algorithms Mol. Biol.* 6, 26–14. doi:10.1186/1748-7188-6-26

Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155 (3), 733–740. doi:10.1099/mic.0.023960-0

Naito, Y., Hino, K., Bono, H., and Ui-Tei, K. (2015). CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* 31 (7), 1120–1123. doi:10.1093/bioinformatics/btu743

Nakamura, M., Gao, Y., Dominguez, A. A., and Qi, L. S. (2021). CRISPR technologies for precise epigenome editing. *Nat. Cell Biol.* 23 (1), 11–22. doi:10.1038/s41556-020-00620-7

Peng, D., and Tarleton, R. (2015). EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens. *Microb. genomics* 1 (4), e000033. doi:10.1099/mgen.0.000033

Pickar-Oliver, A., and Gersbach, C. A. (2019). The next generation of CRISPR–Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* 20 (8), 490–507. doi:10.1038/s41580-019-0131-5

Poudel, R., Rodriguez, L. T., Reisch, C. R., and Rivers, A. R. (2022). GuideMaker: software to design CRISPR-cas guide RNA pools in non-model genomes. *GigaScience* 11, giac007. doi:10.1093/gigascience/giac007

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152 (5), 1173–1183. doi:10.1016/j.cell.2013.02.022

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8 (11), 2281–2308. doi:10.1038/nprot.2013.143

Schwartz, C., Cheng, J. F., Evans, R., Schwartz, C. A., Wagner, J. M., Anglin, S., et al. (2019). Validating genome-wide CRISPR-Cas9 function improves screening in the oleaginous yeast Yarrowia lipolytica. *Metab. Eng.* 55, 102–110. doi:10.1016/j.ymben.2019.06.007

Shi, H., Doench, J. G., and Chi, H. (2022). CRISPR screens for functional interrogation of immunity. *Nat. Rev. Immunol.* 23, 363–380. doi:10.1038/s41577-022-00802-4

Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., and Mateo, J. L. (2015). CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS One* 10 (4), e0124633. doi:10.1371/journal.pone.0124633

Trivedi, V., Ramesh, A., and Wheeldon, I. (2023). Analyzing CRISPR screens in non-conventional microbes. *J. Industrial Microbiol. Biotechnol.* 50, kuad006. doi:10.1093/jimb/kuad006

Wang, J. Y., and Doudna, J. A. (2023). CRISPR technology: a decade of genome editing is only the beginning. *Science* 379 (6629), eadd8643. doi:10.1126/science.add8643

Wiedenheft, B., Sternberg, S. H., and Doudna, J. A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482 (7385), 331–338. doi:10.1038/nature10886

Wilson, L. O., O'Brien, A. R., and Bauer, D. C. (2018). The current state and future of CRISPR-Cas9 gRNA design tools. *Front. Pharmacol.* 9, 749. doi:10.3389/fphar.2018.00749

Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I. M., Li, Y., et al. (2016). Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* 165 (4), 949–962. doi:10.1016/j.cell.2016.04.003

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163 (3), 759–771. doi:10.1016/j.cell.2015.09.038

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Mini Review
PUBLISHED 31 August 2023
DOI 10.3389/fbioe.2023.1215466

OPEN ACCESS

EDITED BY
Anindya Bandyopadhyay,
Reliance Industries, India

REVIEWED BY
Debojyoti Chakraborty,
Council of Scientific and Industrial Research (CSIR), India
Changtian Pan,
Zhejiang University, China

*CORRESPONDENCE
Xi Yu,
xyu@must.edu.mo
Chengliang Yin,
chengliangyin@163.com
Chunying Zhu,
zcy20182985@126.com

†These authors share first authorship

RECEIVED 02 May 2023
ACCEPTED 15 August 2023
PUBLISHED 31 August 2023

CITATION
Li X, Zhu S, Zhang X, Ren Y, He J, Zhou J, Yin L, Wang G, Zhong T, Wang L, Xiao Y, Zhu C, Yin C and Yu X (2023), Advances in the application of recombinase-aided amplification combined with CRISPR-Cas technology in quick detection of pathogenic microbes.
Front. Bioeng. Biotechnol. 11:1215466.
doi: 10.3389/fbioe.2023.1215466

Advances in the application of recombinase-aided amplification combined with CRISPR-Cas technology in quick detection of pathogenic microbes

Xiaoping Li[1,2†], Shuying Zhu[2†], Xinling Zhang[2†], Yanli Ren[3†], Jing He[2], Jiawei Zhou[2], Liliang Yin[2], Gang Wang[2], Tian Zhong[1], Ling Wang[1], Ying Xiao[1,4], Chunying Zhu[5]*, Chengliang Yin[1]* and Xi Yu[1,4]*

[1]Faculty of Medicine, Macau University of Science and Technology, Avenida Wai Long Taipa, Macau, 999078, China, [2]Key Laboratory of Pollution Exposure and Health Intervention of Zhejiang Province, Shulan International Medical College, Zhejiang Shuren University, Hangzhou, Zhejiang Province, 310015, China, [3]State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang Province, 310003, China, [4]Guangdong-Hong Kong-Macau Joint Laboratory for Contaminants Exposure and Health, Guangzhou, Guangdong Province, 510006, China, [5]Clinical Psychology Department, The Affiliated Hospital of Hangzhou Normal University, Hangzhou, Zhejiang Province, 310005, China

The rapid diagnosis of pathogenic infections plays a vital role in disease prevention, control, and public health safety. Recombinase-aided amplification (RAA) is an innovative isothermal nucleic acid amplification technology capable of fast DNA or RNA amplification at low temperatures. RAA offers advantages such as simplicity, speed, precision, energy efficiency, and convenient operation. This technology relies on four essential components: recombinase, single-stranded DNA-binding protein (SSB), DNA polymerase, and deoxyribonucleoside triphosphates, which collectively replace the laborious thermal cycling process of traditional polymerase chain reaction (PCR). In recent years, the CRISPR-Cas (clustered regularly interspaced short palindromic repeats-associated proteins) system, a groundbreaking genome engineering tool, has garnered widespread attention across biotechnology, agriculture, and medicine. Increasingly, researchers have integrated the recombinase polymerase amplification system (or RAA system) with CRISPR technology, enabling more convenient and intuitive determination of detection results. This integration has significantly expanded the application of RAA in pathogen detection. The step-by-step operation of these two systems has been successfully employed for molecular diagnosis of pathogenic microbes, while the single-tube one-step method holds promise for efficient pathogen detection. This paper provides a comprehensive review of RAA combined with CRISPR-Cas and its applications in pathogen detection, aiming to serve as a valuable reference for further research in related fields.

KEYWORDS

recombinase-aided amplification, clustered regularly interspaced short palindromic repeats associated proteins, pathogenic microbes, quick detection, specific diagnosis

1 Introduction

With the continuous development of the Chinese social economy, animal husbandry has become more and more large-scale, intensive, and specialized, and the occurrence of animal epidemics has also increased, which has seriously hindered the sustainable and healthy development of Chinese animal husbandry. At the same time, most animal pathogens can also cause human diseases, such as *Brucella* (Abedi et al., 2020), rabies virus (Fisher et al., 2018), Toxoplasma gondii (Elsheikha et al., 2021), etc. People can become infected by touching infected animals or eating pathogen-containing meat. Therefore, establishing rapid and effective diagnostic methods is an important prerequisite to preventing and controlling the spread of pathogenic microbes.

Currently, microscopic detection technology and serological detection technology are two conventional pathogen detection technologies. However, the operation of microscopic inspection technology is complex and vulnerable to human factors. Although serological detection technology has the advantages of high sensitivity, rapidity, and intuition, it also has the disadvantage that infection cannot be detected at an early stage because antibodies need several days to produce (Lee et al., 2020a). With the rapid development of molecular biology technology, some new molecular biology detection technologies have emerged and have been widely used (Umesha and Manukumar, 2018). RAA is a new molecular biological detection technology. Compared to traditional PCR technology, it does not need expensive experimental equipment and consumables and has a wider application range and easier operation. At present, loop-mediated isothermal amplification (LAMP), recombinase polymerase amplification (RPA), and RAA are several mature isothermal nucleic acid detection technologies (Yan et al., 2014). They are isothermal, specific, and efficient, among which RAA is the only backup technology that is possible to carry out *in vivo* nucleic acid amplification technology (Gui et al., 2022).

RAA is a new isothermal nucleic acid rapid amplification technology. Rapid amplification of DNA or RNA can be achieved at a lower temperature (generally 37°C). The reaction speed is fast and the amplification product can be obtained in 30 min. This technology must use four core substances, namely recombinase, SSB, DNA polymerase, and deoxyribonucleoside triphosphate (dNTP). These four substances can replace the chain-breaking thermal cycle process of ordinary PCR. CRISPR-Cas as a system was originally found and whereafter was taken as a useful tool for genome editing and nucleic acid diagnosis owing to its reliability, sensitivity, and specificity. Multiple types of CRISPR-Cas systems including Cas9, Cas12a, Cas12b, Cas13, and Cas14 are successively developed for bacteria, viruses, and parasite detections. In addition, RAA can be combined with CRISPR-Cas and other new detection technologies to make pathogen detection more convenient and efficient, and its portable detection instrument also provides the possibility of on-site detection (Zhang et al., 2022a).

RAA (Recombinase Polymerase Amplification) plays a crucial role in amplifying the initial signals by increasing the concentration of the template for the Cas proteins. In diagnostics, two key players are Cas12a and Cas13a, which are known for their collateral cleavage activity (Kaminski et al., 2021; Karmakar et al., 2022). One of the notable advancements in Cas12a-based diagnostic systems is the DNA Endonuclease-Targeted CRISPR Trans Reporter (DETECTR). DETECTOR uses Cas12a, which recognizes the target dsDNA and indiscriminately cuts through nearby single-stranded DNA. In DETECTOR, Cas12a is guided to the target dsDNA by a complementary guide RNA. Once Cas12a binds to the correct target, it collateral cleaves ssDNA reporter molecules that are fused with a quencher and a fluorophore. The indiscriminate cleavage leads to the separation of the quencher from the fluorophore, which can be detected through the fluorescence signal. The sensitivity of DETECTR can be further enhanced by combining it with RPA/RAA preamplification. This technique has demonstrated successful detection of human papillomavirus and SARS-CoV-2 (Broughton et al., 2020).

Another powerful diagnostic tool harnessing collateral cleavage activity is the Specific High-sensitivity Enzymatic Reporter unLOCKing (SHERLOCK), which is based on the Cas13a protein (Gootenberg et al., 2017). Similar to Cas12a, Cas13a exhibits indiscriminate cleavage of nearby single-stranded RNA upon recognizing the target RNA. SHERLOCK has been used for the sensitive detection of disease-causing bacteria such as *M. tuberculosis* and *K. pneumoniae*, as well as viruses like SARS-CoV-2 (Xiang et al., 2020). The SHERLOCK method involves isothermal amplification of collected samples using RPA/RAA. Subsequently, an engineered Cas13a, a guide RNA, an RNA reporter, and a reporter probe containing a short oligonucleotide separating a fluorophore and a quencher are added to the sample. Once Cas13a binds to its specific target, it initiates the indiscriminate splicing of nearby RNA, including the reporter. This leads to the separation of the reporter and the quencher, generating distinct signals within the sample (Patchsung et al., 2020; An et al., 2021).

This review paper aims to provide an overview of the principle, characteristics, derivative technologies, and applications of RAA combined with CRISPR-Cas technology in the detection of pathogenic microbes. It seeks to serve as a valuable reference for related research in this field.

2 Exploration of the RAA technical principle and conditions

2.1 The replication mechanism

RAA technology mainly uses recombinase, single-strand binding protein, and DNA polymerase to amplify the target gene. First, under an isothermal 37°C condition, the recombinase (from bacteria or fungi) and the primer DNA were closely combined to form a primer polymer; The next step is the template unwinding process. When the primer specifically recognizes the complementary sequence, the template DNA will unwind with the help of recombinase; Then, under the action of dNTP and energy, DNA polymerase completes the chain extension to form a new chain. The whole reaction is usually carried out at 37°C for 30 min, and a large number of amplification products can be obtained. From the working principle of RAA, it can be seen that once amplification conditions are fulfilled, the reaction products will increase exponentially. But how can RAA achieve real-time detection at this time? The real-time detection incorporates a fluorogenic probe besides forward and reverse primers, and the reaction initiates based on the cleavage of probe at an abasic site (Tetrahydrofuran, THF)

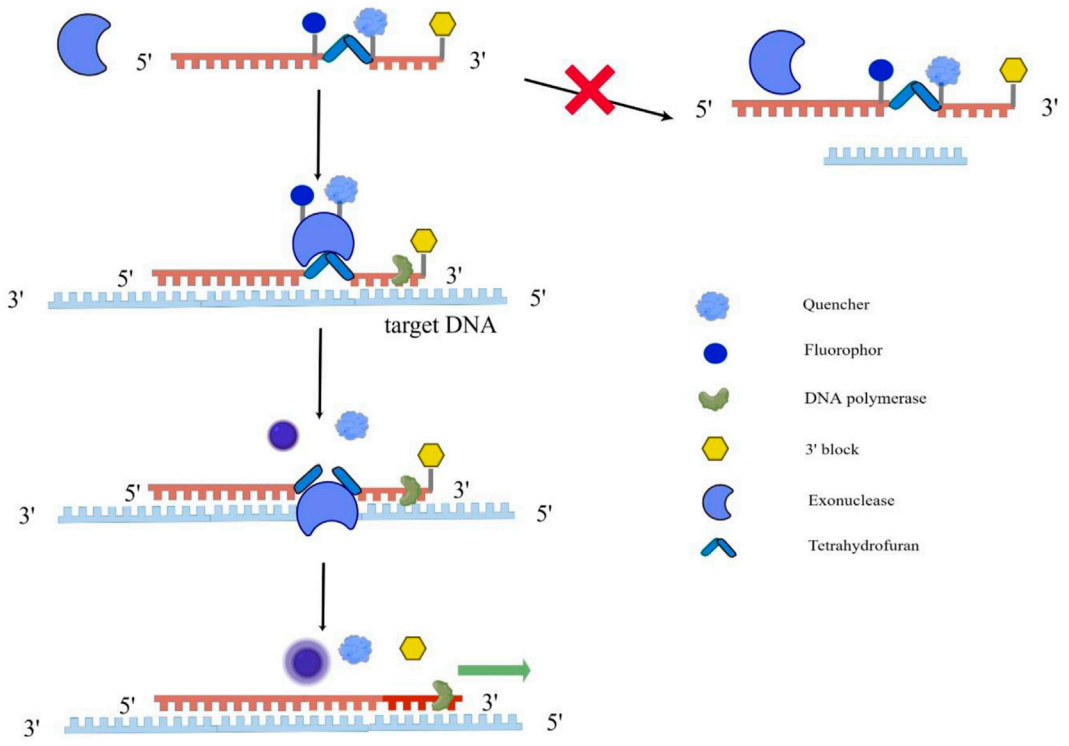


FIGURE 1
Schematic diagram of RAA implementation for real-time detection. This figure was drawn by Figdraw. The single fluorogenic probe includes fluorophor, quencher and 3′block, and exonuclease cannot recognize the single fluorogenic probe. As the probe binds to the target DNA, exonuclease recognizes tetrahydrofuran on the probe abasic site and cleavage them, allowing the fluorophor and quencher to separate and fluoresce. At the same time, DNA polymerase unblocks the 3′block and completes the chain extension.

between the fluorophor and the quencher with exonuclease (Figure 1) (Fan et al., 2020; Mao et al., 2022).

2.2 The design of RAA primers

Certain principles should be followed when designing the RAA primer, and it is different from the design of conventional PCR primers. First, the RAA primer's length is generally 30–35 nt. Too short of a primer will affect the combination of primer and recombinase. Too long primers are easy to form primer dimers, leading to an increase in false positive rate. In addition, when designing RAA primers, it is better not to have a palindrome sequence, a continuous single-base repeat sequence, and an internal secondary structure region in the sequence, and the melting temperature (Tm value) is not an important factor to be considered. Because the amplification efficiency of different primers is different, it is usually necessary to design a series of candidate primers for an amplified gene, and finally select the best primer pair through experimental optimization (Xia et al., 2022).

2.3 The template for the RAA reaction

In the RAA reaction system, the length of the target gene is generally 100 to 200 bp; to ensure rapid and effective amplification of the template for RAA, the maximum length should not exceed 500 bp, because the long target fragment will reduce the rate of the whole reaction.

3 Advantages and disadvantages of RAA technology

The RAA assay, as a novel isothermal amplification technology, has been widely employed for detecting various pathogens. Unlike the RPA assay, the RAA assay utilizes three core enzymes: recombinase UvsX, DNA polymerase, and SSB (Xue et al., 2020a). Wang et al. (2020) successfully developed an RAA assay for the detection of Orf virus (ORFV). This assay demonstrates a rapid detection time of 30 min and a detection limit of 10 copies per reaction. Importantly, it exhibits no cross-reactivity with other common DNA viruses. Consequently, the RAA assay proves to be a fast, sensitive, and specific approach for clinical testing of ORFV. In comparison to LAMP, RAA is simpler to conduct, requiring only a pair of primers, operating at a lower temperature range of 37°C–42°C, and yielding shorter run times of less than 30 min. In contrast, LAMP necessitates four or six primers, operates at a higher temperature of 65°C, and has a longer run time (Pang and Long, 2023). The advantage of RAA technology is that the reaction can be amplified at the optimized 37°C or room temperature, the target amplification product can be obtained in 30 min, and the product grows exponentially without any auxiliary heating equipment, which is particularly suitable for on-site

detection of pathogens. Furthermore, the primers in the RAA reaction system must be strictly complementary to the template, and the length is 30–35 nt, which ensures the precision and specificity of the RAA method.

However, because the sensitivity of RAA is very high, it is easy to cause false positive results, so cross-contamination should be avoided throughout the process. Another disadvantage of RAA, it is easy to produce nonspecific products in the reaction process, which is mainly due to the short length of the target gene (Jaroenram and Owens, 2014). In addition, the formation of primer dimers is caused by excessive primers and specific interactions between primer molecules (Kellner et al., 2019). Currently, there are significant challenges associated with RAA. Firstly, achieving on-site sample pre-treatment, such as nucleic acid extraction, is difficult. This step is crucial to obtain high-quality nucleic acid templates for amplification. Overcoming this challenge is essential for enabling the use of RAA in various settings where immediate sample processing is required. Secondly, an urgent technical challenge that needs to be addressed is how to achieve multi-target isothermal amplification detection under single-tube closed-tube conditions with high sensitivity and specificity. The ability to detect multiple targets simultaneously within a single reaction is highly desirable for efficient and comprehensive pathogen identification. Without a solution to this challenge, the application range of RAA may be limited. Addressing these issues is critical to expanding the application range of RAA and maximizing its potential as a powerful diagnostic tool. Ongoing research and development efforts are focused on overcoming these limitations to further enhance the utility and effectiveness of RAA in practical settings.

4 Derivative technologies based on RAA technology

The target gene product of traditional RAA technology is mainly judged by gel electrophoresis, but with the rapid development of this technology, traditional RAA technology can be combined with other new technologies, including the real-time fluorescent RAA, RT-RAA, RAA auxiliary CRISPR-Cas13a/Cas12a method, which determines detection results more convenient and intuitive. At the same time, it has greatly expanded the application of this technology in pathogen detection.

4.1 Real-time fluorescent RAA

Real-time fluorescent RAA technology is a combination of the basic reaction system of RAA and the fluorescent probe system, which can monitor the amplification results of target gene fragments in real-time and fully meet the needs of on-site detection. In 2022, Wang et al. (2022b) used real-time fluorescent RAA to rapidly detect Senecavirus A. The sensitivity of this method is the same as that of the real-time fluorescent quantitative PCR method, and there is no cross-reaction with other types of respiratory viruses. Furthermore, RAA technology can also be used together with a fluorometer to mark different target genes with different fluorescence colors in the same reaction, which can form a set of RAA multiple fluorescence real-time monitoring systems, while other isothermal nucleic acid amplification technology or Nest PCR technology cannot achieve this effect (Seo et al., 2022).

4.2 RT-RAA

Reverse transcriptase recombinase-aided amplification (RT-RAA) technology is based on RAA technology. Its principle is that first the pathogen's RNA is reverse transcribed into cDNA, and then recombinase was used to replace the high-temperature denatured quantitative PCR (qPCR) to complete the template chain breaking process. Then, under the action of the SSB, the DNA polymerase completes the chain extension, and finally, the new chain can be generated. Chen et al. (2018a) used RT-RAA technology to rapidly detect different subgroups of respiratory syncytial virus (RSV) and the results were consistent with those of real-time qPCR (RT-qPCR). In addition, this technology can also be combined with lateral flow immunoassay (LFIA) technology to make the determination of the result more intuitive. In 2020, Xiong et al. (2020) successfully established a rapid and intuitive detection method for dengue virus (DENV) based on RT-RAA and LFIA. This method only needs to react at 37°C for 20 min, then it can be visually observed by LFD in 3 min. It has the advantages of strong specificity and high sensitivity and is especially suitable for the accurate detection of DENV when laboratory resources are scarce.

4.3 RAA combined with CRISPR-Cas13a/Cas12a method (RAA-Cas13a/Cas12a)

In recent years, clustered regularly interspaced short palindromic repeats associated proteins (CRISPR-Cas), as a genome engineering tool that revolutionizes life science, has received extensive attention in the fields of biotechnology, agriculture, and medicine (Hille et al., 2018). CRISPR-associated protein 13a (Cas13a or C2c2) is a class VI-A protein, which can be used not only for gene editing but also for the detection of pathogenic nucleic acids (Gootenberg et al., 2017; Myhrvold et al., 2018). CRISPR-associated protein 12a (Cas12a or Cpf1) is a kind of RNA-guided enzyme that can specifically bind and cut target DNA and is an important part of the adaptive bacterial immune system. Cas12a protein, known as Cpf1, is widely utilized in genome editing due to its ability to trigger double-stranded DNA breakage at specific locations (Chen et al., 2018a). Bai et al. (2019) combined RAA technology with CRISPR-Cas to establish a diagnosis system of RAA-Cas12a, which can quickly and easily detect the African swine fever virus. Furthermore, Chen et al. (2022) successfully established a quick, sensitive, and specific detection method for SARS-CoV-2 using RAA technology and the CRISPR-Cas detection system. The sensitivity of the RAA-Cas13a method can reach 10 copies of RNA molecules/reaction (Li et al., 2022). It summarizes all important reports about the application of RAA combined with CRISPR-Cas12a/Cas13a in the quick detection of pathogenic microbes (Table 1) (Li et al., 2021; Wang et al., 2022c; Zhang et al., 2022b; Chen et al., 2022; Fang et al., 2022; Qian et al., 2022; Wei et al., 2022; Xing et al., 2023; Zhao et al., 2023).

TABLE 1 The application of RAA combined with CRISPR-Cas12a/13a technology in the quick detection of pathogenic microbes.

RAA- Cas12a/ Cas13a	Signal approach	Pathogens	Limit of detection (copies/mL)	Time (min)	References
RAA-Cas12a	Fluorescence	*Norovirus*	1.0×10^{2}	40	Qian et al. (2022)
RAA-Cas12a	Fluorescence	*Methicillin-resistant Staphylococcus aureus*	1.0×10^{4}	60	Wang et al. (2022b)
RAA-Cas12a	finger-actuated microfluidic biosensor (FA-MB)	*Bacillus cereus, Pseudomonas aeruginosa, Salmonella typhimurium,* et al.	2.5 × 105	60	Xing et al. (2023)
RAA-Cas12a	Electrochemical biosensor (E-CRISPR)	*L. monocytogenes*	1.3×10^{4}	45	Li et al. (2021)
RAA-Cas12a	Centrifugal microfluidics	*SARS-CoV-2*	1.0×10^{3}	30	Chen et al. (2022)
RAA-Cas12a	Fluorescence	*Escherichia coli O157:H7*	2.7×10^{5}	30	Fang et al. (2022)
RAA-Cas12a	Fluorescence	*Monkeypox virus*	1.0×10^{3}	75	Zhao et al. (2023)
RAA-Cas13a	Fluorescence	*Encephalomyocarditis virus*	1.0×10^{3}	60	Wei et al. (2022)
RAA-Cas13a	fluorescent visualizers	*SARS-CoV-2*	1.0×10^{4}	60	Zhang et al. (2022b)

5 Application of RAA technology

Since the advent of RAA technology, it has been widely used in the detection of pathogenic microbes, including bacteria (Zhang et al., 2017), viruses (Shen et al., 2019), parasites (Lin et al., 2022), and other pathogens. Furthermore, this technology can also be applied to SNP and cancer detection measure (Duan et al., 2018).

5.1 Bacteria detection

Salmonella is a kind of non-spore Gram-negative *bacillus*, distributed throughout the world and mainly parasitizes human and animal intestines, posing a serious threat to human and animal health (Yin and Zhou, 2018). *Salmonella* laboratory diagnosis methods rely mainly on microbial culture and serological detection technology from blood or bone marrow (Crump et al., 2015). With the continuous development of molecular biological technology, Zhang et al. (2017) successfully established a molecular biological method for the rapid detection of *salmonella* by RAA. The whole reaction can completely amplify a large number of target genes within 20 min, and it has high sensitivity and good specificity. The lower limit for *salmonella* detection is 102 copies/μL. There was no cross-reaction with *Escherichia coli* and *Shigella*. This method provides powerful technical support for the molecular diagnosis of *Salmonella*.

5.2 Virus detection

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first discovered in Wuhan, China. The 99 patients admitted to Wuhan Jinyintan Hospital were the first confirmed patients reported with Corona Virus Disease 2019 (COVID-19) (Chen et al., 2020). The harm of SARS-CoV-2 manifests itself primarily in systemic infection of many animals and human respiratory tract infections, such as severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) (Drosten et al., 2003; Zaki et al., 2012; Yin and Wunderink, 2018). Clinical manifestations of COVID-19 are flu-like symptoms such as fever, cough, fatigue, dyspnea, and gastrointestinal symptoms (Huang et al., 2020). In 2022, Yao et al. (2022) designed two sets of RT-RAA primers and probes based on the ORF1ab gene and the N gene, and successfully established a rapid RT-RAA detection method for SARS-CoV-2. This method is specific and sensitive. It has no cross-reaction with other respiratory pathogens and has the same sensitivity as RT-qPCR. This study provides a simple and reliable method for the detection of SARS-CoV-2. This method is suitable for diagnostic laboratories with insufficient equipment and provides technical support for clinical treatment and disease control in grass-roots hospitals in the future (Li et al., 2022). The avian influenza virus (AIV) is an RNA virus that belongs to the influenza A virus causing zoonotic diseases. In the past 20 years, avian influenza has been ravaging the world, and new AIV has also been discovered (Yang et al., 2023), becoming a global public health problem and attracting high attention from the international community (Domanska-Blicharz et al., 2023; Tahmo et al., 2023). Wang et al. (2022a) established a sensitive and specific detection method for H5 subtype AIV with the RT-RAA method for the first time. The results showed that H5 subtype AIV detection by the RT-RAA method could produce positive signals in 20 min, and the whole reaction could be completed in about 30 min. While RT-qPCR takes at least 2 h. Furthermore, the cost of RT-RAA detection is almost half that of RT-qPCR. This is a comparison of clinical applications between the RT-RAA method and the RT-qPCR method (Table 2).

5.3 Parasite detection

Malaria is a serious global vector disease caused by plasmodium. Plasmodium is a kind of single-celled protozoa, mainly distributed in tropical and subtropical regions, covering more than 90 countries and regions around the world, with 224 million infected people, most of whom are children (Lee

TABLE 2 Comparison of clinical applications between RT-RAA and RT-qPCR.

Compare items	RT-RAA	RT-qPCR	References
Principle differences	SSB keeps the template in a single-stranded notch state and nucleic acid amplification is uninterrupted at 37°C.	Reamplification required chain denaturation at 94°C.	Zhang et al. (2017)
Reaction process	In the whole reaction process, RNA can be directly used as a template to be added to the amplification system, and reverse transcription and amplification are carried out simultaneously.	For RNA amplification, it is necessary to perform reverse transcription reaction first, which is to reverse transcribe RNA into cDNA, and then use cDNA as template for nucleic acid amplification.	Li et al. (2022)
Reaction temperature	The reaction can be carried out at a constant temperature ranging from 37°C to 42°C.	Three temperatures are required for continuous circulation, such as 95°C, 72°C, 55°C, and the temperature must be precisely controlled.	Shen et al. (2019)
Reaction time	20 min.	At least 120 min.	Chen et al. (2020)
Detection equipment	Small fluorescent detector.	RT-qPCR instrument.	Yao et al. (2022)
Detection site	On site.	PCR special laboratory.	Tang et al. (2021)
Detection personnel	On-site personnel.	Medical professional.	Lee et al. (2020b)

et al., 2020b; Naserrudin et al., 2022). The early detection of malaria parasites is mainly by microscopy, but this method is time-consuming, laborious, and subject to human factors. Since the advent of RAA technology, this method has been widely used in the pathogen detection (Allen et al., 2022). In 2022, Lin et al. (2022) used the search in the NCBI database to obtain multiple genomic sequences of 18SrDNA from malaria parasites, designed universal primers to detect malaria parasites, and successfully established a detection method for malaria parasites using RAA. It can be seen that this technology provides a powerful technical means for effective prevention and control of parasites.

6 Discussion

In recent years, there has been an increasing number of studies that have integrated the RPA or RAA system with the CRISPR-Cas system. Cas13a protein, renowned for its collateral cleavage activity, serves as a key component in diagnostics. RPA/RAA aids in amplifying the initial signals by enhancing the template concentration for Cas13a. The sequential implementation of these two systems is employed for the molecular diagnosis of pathogenic microbes (Kirchner and Schneider, 2015). Among them, RAA technology uses reverse transcriptase (RT), recombinase, single-stranded DNA binding proteins (SSB), DNA polymerase, etc. It can amplify the target gene by millions or even tens of millions of times in 30 min under isothermal conditions (37°C). Thus, the mixed system 'RPA/RAA and CRISPR-Cas13a' can achieve the goal of high specificity and quick detection (Gootenberg et al., 2017). However, at present, the detection of pathogenic microbes only achieves the step-by-step operation of RAA and CRISPR-Cas13a: first, the RAA system is used to rapidly isothermal amplify the nucleic acid of pathogenic microbes, and then the CRISPR-Cas13a system is used to quickly identify and detect products amplified by RAA. The nonintegrated detection method has been applied to the quantitative detection of adenovirus, *Staphylococcus aureus*, *Shigella*, and *Vibrio cholerae* (Xue et al., 2020b; Fan et al., 2021; Tang et al., 2021). It is precise because of the two-step operation of RAA and CRISPR-Cas13a (after the nucleic acid of pathogenic microbes has been amplified by RAA, it is necessary to open the cap and transfer the solution to the CRISPR-Cas13a tube for detection), which is prone to pollution. It is necessary to optimize the performance parameters of the two mixed systems of RAA and CRISPR-Cas13a, which can integrate the two tubes into one tube (single tube). This single-tube one-step method is similar to the one-pot SHERLOCK of Zhang Feng's team (Figure 2), which can implement the universal detection of pathogenic microbes at various points (Joung et al., 2020).

7 Summary and outlook

Pathogenic microbes in nature are varied and mutate rapidly, which is one of the main factors causing food and public health safety problems. In recent years, the infection of viruses, bacteria, and parasites not only seriously threatens human health, but also hinders the development of the Chinese aquaculture industry. Therefore, there is a need to develop a more quick and accurate method to detect pathogenic zoonotic microbes (Rees et al., 2021). Traditional pathogen detection relies mainly on microscopic detection technology, but this technology has high requirements for operators. With the continuous development of molecular biology, some new molecular biological technologies are emerging. Among them, RAA technology is a relatively mature isothermal amplification technology, which has the advantages of simple operation, high sensitivity, strong specificity, and quick detection and can be used for on-site detection (Dinnes et al., 2022). Therefore, this technology is widely used to quickly detect viruses, bacteria, and parasites. With the continuous development and improvement of RAA technology, especially the improvement of the purity of some enzymes and proteins in the two systems (RAA and CRISPR-Casl3a) and the optimization of some conditions and parameters of the mixed system, it could be possible to establish a

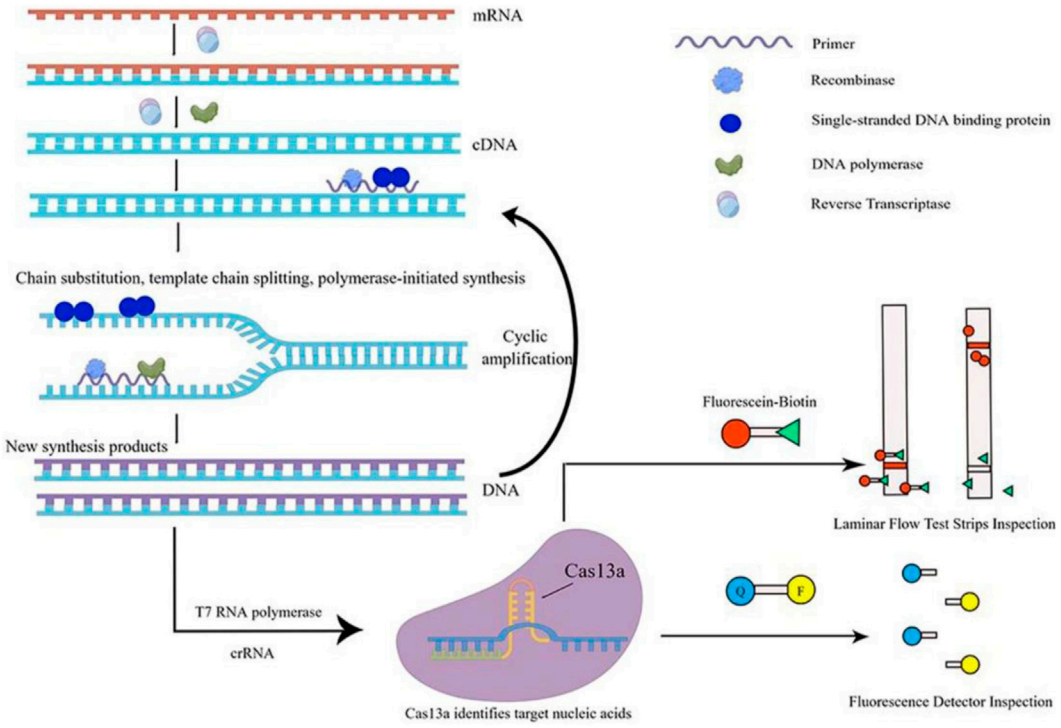


FIGURE 2
Schematic diagram of single-tube one-step method. This figure was drawn by Figdraw. The specific principle of single-tube one-step detection of RNA viruses is as follows: in the nucleic acid amplification system under constant temperature conditions, firstly, dNTP is used as substrate material, mRNA is used as a template, and under the action of reverse transcriptase, a cDNA single-strand complementary to the RNA template is synthesized to form DNA-RNA heterozygote. In this DNA-RNA heterozygote, RNA is specifically degraded by reverse transcriptase, and then dNTP is used as the substrate, the first strand of cDNA is used as a template, and under the action of DNA polymerase, the second strand of cDNA is synthesized, and finally the double-stranded DNA molecule is formed, that is, the DNA synthesis process guided by RNA is completed. Then, this double-stranded DNA molecule is used as the template DNA. When the primer finds the complementary sequence that perfectly matches the template DNA, with the help of recombinase, the double-stranded structure of the template DNA is opened, while SSB stabilizes the displaced DNA chain, and the extension of the strand is completed under the action of DNA polymerase to form a new complementary strand of DNA. The T7 RNA polymerase transcribed the amplified product into RNA, while the Cas13a protein specifically bound the target nucleic acid under the guidance of crRNA, the protein structure was changed, and was converted into ribonuclease, which was non-specific to cut RNA. The incidental cutting activity could cut the introduced externally derived FQ-coated ssRNA probe. The fluorescence signal can be amplified rapidly to detect the target nucleic acid quickly.

single-tube one-step method for quick detection of various viruses. This innovative single-tube one-step method technology and its derivative technology will have broad application prospects in the detection of pathogenic microbes (Gootenberg et al., 2018; Tao et al., 2020).

Author contributions

XL, SZ, and XZ: Framework construction, article writing, project administration, and literature collection. YR: Literature search and framework construction. JH and JZ: Collation of data. LY, GW, TZ, LW, and YX: Review and editing. CZ and CY: writing-review and editing, supervision. XY: Review and editing, resource, funding. All authors contributed to the article and approved the submitted version.

Funding

This study was financially supported by the Science and Technology Development Fund, Macau SAR (Grant 0024/2022/A), the Science and Technology Planning Project of Guangdong Province (2020B1212030008), study on the role and molecular mechanism of HuD abnormal expression in the development and progression of non-small cell lung cancer (Y202147513), and 2022 National Student Innovation and Entrepreneurship Training Program Project (Item Number: 202211842047). Key Laboratory of Biomarkers and *In Vitro* Diagnosis Translation of Zhejiang province (KFJJ2023005). These funders did not participate in the designing, performing or reporting in the current study.

Acknowledgments

XL is a Ph.D. student in public Health at the Faculty of Medicine, Macau University of Science and Technology.

Conflict of interest

The 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

References

Abedi, A. S., Hashempour-Baltork, F., Alizadeh, A. M., Beikzadeh, S., Hosseini, H., Bashiry, M., et al. (2020). The prevalence of Brucella spp. in dairy products in the Middle East region: a systematic review and meta-analysis. *Acta Trop.* 202, 105241. doi:10.1016/j.actatropica.2019.105241

Allen, E. N., Wiyeh, A. B., and McCaul, M. (2022). Adding rapid diagnostic tests to community-based programmes for treating malaria. *Cochrane Database Syst. Rev.* 9 (9), Cd009527. doi:10.1002/14651858.cd009527.pub3

An, B., Zhang, H., Su, X., Guo, Y., Wu, T., Ge, Y., et al. (2021). Rapid and sensitive detection of Salmonella spp. using CRISPR-cas13a combined with recombinase polymerase amplification. *Front. Microbiol.* 12, 732426. doi:10.3389/fmicb.2021.732426

Bai, J., Lin, H., Li, H., Zhou, Y., Liu, J., Zhong, G., et al. (2019). Cas12a-Based on-site and rapid nucleic acid detection of african swine fever. *Front. Microbiol.* 10, 2830. doi:10.3389/fmicb.2019.02830

Broughton, J. P., Deng, X., Yu, G., Fasching, C. L., Servellita, V., Singh, J., et al. (2020). CRISPR-Cas12-based detection of SARS-CoV-2. *Nat. Biotechnol.* 38 (7), 870–874. doi:10.1038/s41587-020-0513-4

Chen, J. S., Li, X. N., Li, G. X., Zhao, L., Duan, S. X., Yan, T. F., et al. (2018a). Use of a rapid reverse-transcription recombinase aided amplification assay for respiratory syncytial virus detection. *Diagn Microbiol. Infect. Dis.* 90 (2), 90–95. doi:10.1016/j.diagmicrobio.2017.10.005

Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., et al. (2018b). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Sci. (New York, NY)* 360 (6387), 436–439. doi:10.1126/science.aar6245

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, China: a descriptive study. *Lancet (London, Engl.* 395 (10223), 507–513. doi:10.1016/s0140-6736(20)30211-7

Chen, Y., Zong, N., Ye, F., Mei, Y., Qu, J., and Jiang, X. (2022). Dual-CRISPR/Cas12a-Assisted RT-RAA for ultrasensitive SARS-CoV-2 detection on automated centrifugal microfluidics. *Anal. Chem.* 94 (27), 9603–9609. doi:10.1021/acs.analchem.2c00638

Crump, J. A., Sjölund-Karlsson, M., Gordon, M. A., and Parry, C. M. (2015). Epidemiology, clinical presentation, laboratory diagnosis, antimicrobial resistance, and antimicrobial management of invasive Salmonella infections. *Clin. Microbiol. Rev.* 28 (4), 901–937. doi:10.1128/cmr.00002-15

Dinnes, J., Sharma, P., Berhane, S., van Wyk, S. S., Nyaaba, N., Domen, J., et al. (2022). Rapid, point-of-care antigen tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst. Rev.* 7 (7), Cd013705. doi:10.1002/14651858.cd013705.pub3

Domanska-Blicharz, K., Opolska, J., Lisowska, A., and Szczotka-Bochniarz, A. (2023). Bacterial and viral rodent-borne infections on poultry farms. An attempt at a systematic review. *J. Vet. Res.* 67 (1), 1–10. doi:10.2478/jvetres-2023-0012

Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., et al. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348 (20), 1967–1976. doi:10.1056/nejmoa030747

Duan, S., Li, G., Li, X., Chen, C., Yan, T., Qiu, F., et al. (2018). A probe directed recombinase amplification assay for detection of MTHFR A1298C polymorphism associated with congenital heart disease. *Biotechniques* 64 (5), 211–217. doi:10.2144/btn-2018-2010

Elsheikha, H. M., Marra, C. M., and Zhu, X. Q. (2021). Epidemiology, pathophysiology, diagnosis, and management of cerebral toxoplasmosis. *Clin. Microbiol. Rev.* 34 (1), e00115–e00119. doi:10.1128/cmr.00115-19

Fan, G., Zhang, R., He, X., Tian, F., Nie, M., Shen, X., et al. (2021). RAP: a novel approach to the rapid and highly sensitive detection of respiratory viruses. *Front. Bioeng. Biotechnol.* 9, 766411. doi:10.3389/fbioe.2021.766411

Fan, X., Li, L., Zhao, Y., Liu, Y., Liu, C., Wang, Q., et al. (2020). Clinical validation of two recombinase-based isothermal amplification assays (RPA/RAA) for the rapid detection of african swine fever virus. *Front. Microbiol.* 11, 1696. doi:10.3389/fmicb.2020.01696

Fang, T., Shen, J., Xue, J., Jiang, Y., Guo, D., Yang, J., et al. (2022). Sensitive and rapid detection of *Escherichia coli* O157:H7 from beef samples based on recombinase aided amplification assisted CRISPR/Cas12a system. *J. AOAC Int.* 106 (1), 156–164. doi:10.1093/jaoacint/qsac101

Fisher, C. R., Streicker, D. G., and Schnell, M. J. (2018). The spread and evolution of rabies virus: conquering new frontiers. *Nat. Rev. Microbiol.* 16 (4), 241–255. doi:10.1038/nrmicro.2018.11

Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J., and Zhang, F. (2018). Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Sci. (New York, NY)* 360 (6387), 439–444. doi:10.1126/science.aaq0179

Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., et al. (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Sci. (New York, NY)* 356 (6336), 438–442. doi:10.1126/science.aam9321

Gui, L., Zhao, Y., Xu, D., Li, X., Luo, J., Zhou, W., et al. (2022). Quick detection of *Carassius auratus* herpesvirus (CaHV) by recombinase-aid amplification lateral flow dipstick (RAA-LFD) method. *Front. Cell Infect. Microbiol.* 12, 981911. doi:10.3389/fcimb.2022.981911

Hille, F., Richter, H., Wong, S. P., Bratovič, M., Ressel, S., and Charpentier, E. (2018). The biology of CRISPR-cas: backward and forward. *Cell* 172 (6), 1239–1259. doi:10.1016/j.cell.2017.11.032

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet (London, Engl.* 395 (10223), 497–506. doi:10.1016/s0140-6736(20)30183-5

Jaroenram, W., and Owens, L. (2014). Recombinase polymerase amplification combined with a lateral flow dipstick for discriminating between infectious Penaeus stylirostris densovirus and virus-related sequences in shrimp genome. *J. Virol. Methods* 208, 144–151. doi:10.1016/j.jviromet.2014.08.006

Joung, J., Ladha, A., Saito, M., Kim, N. G., Woolley, A. E., Segel, M., et al. (2020). Detection of SARS-CoV-2 with SHERLOCK one-pot testing. *N. Engl. J. Med.* 383 (15), 1492–1494. doi:10.1056/nejmc2026172

Kaminski, M. M., Abudayyeh, O. O., Gootenberg, J. S., Zhang, F., and Collins, J. J. (2021). CRISPR-based diagnostics. *Nat. Biomed. Eng.* 5 (7), 643–656. doi:10.1038/s41551-021-00760-7

Karmakar, S., Das, P., Panda, D., Xie, K., Baig, M. J., and Molla, K. A. (2022). A detailed landscape of CRISPR-Cas-mediated plant disease and pest management. *Plant Sci.* 323, 111376. doi:10.1016/j.plantsci.2022.111376

Kellner, M. J., Koob, J. G., Gootenberg, J. S., Abudayyeh, O. O., and Zhang, F. (2019). SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nat. Protoc.* 14 (10), 2986–3012. doi:10.1038/s41596-019-0210-2

Kirchner, M., and Schneider, S. (2015). CRISPR-cas: from the bacterial adaptive immune system to a versatile tool for genome engineering. *Angew. Chem. Int. Ed. Engl.* 54 (46), 13508–13514. doi:10.1002/anie.201504741

Lee, C. Y., Lin, R. T. P., Renia, L., and Ng, L. F. P. (2020a). Serological approaches for COVID-19: epidemiologic perspective on surveillance and control. *Front. Immunol.* 11, 879. doi:10.3389/fimmu.2020.00879

Lee, C. Y., Puig, H., Nguyen, P. Q., Angenent-Mari, N. M., Donghia, N. M., McGee, J. P., et al. (2020b). Ultrasensitive CRISPR-based diagnostic for field-applicable detection of Plasmodium species in symptomatic and asymptomatic malaria. *Proc. Natl. Acad. Sci. U. S. A.* 117 (41), 25722–25731. doi:10.1073/pnas.2010196117

Li, F., Ye, Q., Chen, M., Zhou, B., Zhang, J., Pang, R., et al. (2021). An ultrasensitive CRISPR/Cas12a based electrochemical biosensor for Listeria monocytogenes detection. *Biosens. Bioelectron.* 179, 113073. doi:10.1016/j.bios.2021.113073

Li, X., Zhang, H., Zhang, J., Song, Y., Shi, X., Zhao, C., et al. (2022). Diagnostic accuracy of CRISPR technology for detecting SARS-CoV-2: a systematic review and meta-analysis. *Expert Rev. Mol. Diagn* 22 (6), 655–663. doi:10.1080/14737159.2022.2107425

Lin, H., Zhao, S., Liu, Y., Shao, L., Ye, Y., Jiang, N., et al. (2022). Rapid visual detection of plasmodium using recombinase-aided amplification with lateral flow dipstick assay. *Front. Cell Infect. Microbiol.* 12, 922146. doi:10.3389/fcimb.2022.922146

Mao, L., Ying, J., Selekon, B., Gonofio, E., Wang, X., Nakoune, E., et al. (2022). Development and characterization of recombinase-based isothermal amplification assays (RPA/RAA) for the rapid detection of monkeypox virus. *Viruses* 14 (10), 2112. doi:10.3390/v14102112

Myhrvold, C., Freije, C. A., Gootenberg, J. S., Abudayyeh, O. O., Metsky, H. C., Durbin, A. F., et al. (2018). Field-deployable viral diagnostics using CRISPR-Cas13. *Sci. (New York, NY)* 360 (6387), 444–448. doi:10.1126/science.aas8836

Naserrudin, N. A., Hassan, M. R., Jeffree, M. S., Culleton, R., Hod, R., and Ahmed, K. (2022). A systematic review of asymptomatic plasmodium knowlesi infection: an emerging challenge involving an emerging infectious disease. *Malar. J.* 21 (1), 373. doi:10.1186/s12936-022-04339-8

Pang, F., and Long, Q. (2023). Recent advances in diagnostic approaches for orf virus. *Appl. Microbiol. Biotechnol.* 107 (5-6), 1515–1523. doi:10.1007/s00253-023-12412-8

Patchsung, M., Jantarug, K., Pattama, A., Aphicho, K., Suraritdechachai, S., Meesawat, P., et al. (2020). Clinical validation of a Cas13-based assay for the detection of SARS-CoV-2 RNA. *Nat. Biomed. Eng.* 4 (12), 1140–1149. doi:10.1038/s41551-020-00603-x

Qian, W., Huang, J., Wang, T., Fan, C., Kang, J., Zhang, Q., et al. (2022). Ultrasensitive and visual detection of human norovirus genotype GII.4 or GII.17 using CRISPR-Cas12a assay. *Virology J.* 19 (1), 150. doi:10.1186/s12985-022-01878-z

Rees, E. M., Minter, A., Edmunds, W. J., Lau, C. L., Kucharski, A. J., and Lowe, R. (2021). Transmission modelling of environmentally persistent zoonotic diseases: a systematic review. *Lancet Planet Health* 5 (7), e466–e478. doi:10.1016/s2542-5196(21)00137-6

Seo, S. B., Hwang, J. S., Kim, E., Kim, K., Roh, S., Lee, G., et al. (2022). Isothermal amplification-mediated lateral flow biosensors for *in vitro* diagnosis of gastric cancer-related microRNAs. *Talanta* 246, 123502. doi:10.1016/j.talanta.2022.123502

Shen, X. X., Qiu, F. Z., Shen, L. P., Yan, T. F., Zhao, M. C., Qi, J. J., et al. (2019). A rapid and sensitive recombinase aided amplification assay to detect hepatitis B virus without DNA extraction. *BMC Infect. Dis.* 19 (1), 229. doi:10.1186/s12879-019-3814-9

Tahmo, N. B., Wirsiy, F. S., Nnamdi, D. B., Tongo, M., Lawler, J. V., Broadhurst, M. J., et al. (2023). An epidemiological synthesis of emerging and re-emerging zoonotic disease threats in Cameroon, 2000-2022: a systematic review. *IJID Reg.* 7, 84–109. doi:10.1016/j.ijregi.2022.12.001

Tang, Y., Gao, L., Feng, W., Guo, C., Yang, Q., Li, F., et al. (2021). The CRISPR-Cas toolbox for analytical and diagnostic assay development. *Chem. Soc. Rev.* 50 (21), 11844–11869. doi:10.1039/d1cs00098e

Tao, D., Liu, J., Nie, X., Xu, B., Tran-Thi, T. N., Niu, L., et al. (2020). Application of CRISPR-cas12a enhanced fluorescence assay coupled with nucleic acid amplification for the sensitive detection of african swine fever virus. *ACS Synth. Biol.* 9 (9), 2339–2350. doi:10.1021/acssynbio.0c00057

Umesha, S., and Manukumar, H. M. (2018). Advanced molecular diagnostic techniques for detection of food-borne pathogens: current applications and future challenges. *Crit. Rev. Food Sci. Nutr.* 58 (1), 84–104. doi:10.1080/10408398.2015.1126701

Wang, S., Li, Y., Zhang, F., Jiang, N., Zhuang, Q., Hou, G., et al. (2022c). Reverse transcription recombinase-aided amplification assay for H5 subtype avian influenza virus. *Virology J.* 19 (1), 129. doi:10.1186/s12985-022-01807-0

Wang, S., Zhou, L., Ge, X., Han, J., Guo, X., Chen, Y., et al. (2022a). Development of a VP2-based real-time fluorescent reverse transcription recombinase-aided amplification assay to rapidly detect Senecavirus A. *Transbound. Emerg. Dis.* 69 (5), 2828–2839. doi:10.1111/tbed.14435

Wang, Y., Cui, Y., Yu, Z., Li, Y., Bai, C., Sun, P., et al. (2020). Development of a recombinase-aided amplification assay for detection of orf virus. *J. Virological Methods* 280, 113861. doi:10.1016/j.jviromet.2020.113861

Wang, S., Liang, X., Xu, J., Nan, L., Liu, F., Duan, G., et al. (2022b). Rapid and ultrasensitive detection of methicillin-resistant *Staphylococcus aureus* based on CRISPR-cas12a combined with recombinase-aided amplification. *Front. Microbiol.* 13, 903298. doi:10.3389/fmicb.2022.903298

Wei, N., Xiong, J., Ma, J., Ye, J., Si, Y., and Cao, S. (2022). Development of efficient, sensitive, and specific detection method for Encephalomyocarditis virus based on CRISPR/Cas13a. *J. Virol. Methods* 309, 114592. doi:10.1016/j.jviromet.2022.114592

Xia, W., Chen, Y., Ding, X., Liu, X., Lu, H., Guo, C., et al. (2022). Rapid and visual detection of type 2 porcine reproductive and respiratory syndrome virus by real-time fluorescence-based reverse transcription recombinase-aided amplification. *Viruses* 14 (11), 2526. doi:10.3390/v14112526

Xiang, X., Qian, K., Zhang, Z., Lin, F., Xie, Y., Liu, Y., et al. (2020). CRISPR-cas systems based molecular diagnostic tool for infectious diseases and emerging 2019 novel coronavirus (COVID-19) pneumonia. *J. Drug Target.* 28 (7-8), 727–731. doi:10.1080/1061186x.2020.1769637

Xing, G., Shang, Y., Wang, X., Lin, H., Chen, S., Pu, Q., et al. (2023). Multiplexed detection of foodborne pathogens using one-pot CRISPR/Cas12a combined with recombinase aided amplification on a finger-actuated microfluidic biosensor. *Biosens. Bioelectron.* 220, 114885. doi:10.1016/j.bios.2022.114885

Xiong, Y., Luo, Y., Li, H., Wu, W., Ruan, X., and Mu, X. (2020). Rapid visual detection of dengue virus by combining reverse transcription recombinase-aided amplification with lateral-flow dipstick assay. *Int. J. Infect. Dis.* 95, 406–412. doi:10.1016/j.ijid.2020.03.075

Xue, G., Li, S., Zhang, W., Du, B., Cui, J., Yan, C., et al. (2020a). Reverse-Transcription recombinase-aided amplification assay for rapid detection of the 2019 novel coronavirus (SARS-CoV-2). *Anal. Chem.* 92 (14), 9699–9705. doi:10.1021/acs.analchem.0c01032

Xue, G., Li, S., Zhao, H., Yan, C., Feng, Y., Cui, J., et al. (2020b). Use of a rapid recombinase-aided amplification assay for Mycoplasma pneumoniae detection. *BMC Infect. Dis.* 20 (1), 79. doi:10.1186/s12879-019-4750-4

Yan, L., Zhou, J., Zheng, Y., Gamson, A. S., Roembke, B. T., Nakayama, S., et al. (2014). Isothermal amplified detection of DNA and RNA. *Mol. Biosyst.* 10 (5), 970–1003. doi:10.1039/c3mb70304e

Yang, X. Y., Gong, Q. L., Li, Y. J., Ata, E. B., Hu, M. J., Sun, Y. Y., et al. (2023). The global prevalence of highly pathogenic avian influenza A (H5N8) infection in birds: A systematic review and meta-analysis. *Microb. Pathog.* 176, 106001. doi:10.1016/j.micpath.2023.106001

Yao, X. H., Hu, D. H., Fu, S. H., Li, F., He, Y., Yin, J. Y., et al. (2022). A reverse-transcription recombinase-aided amplification assay for the rapid detection of the wuxiang virus. *Biomed. Environ. Sci.* 35 (8), 746–749. doi:10.3967/bes2022.096

Yin, Y., and Wunderink, R. G. (2018). MERS, SARS and other coronaviruses as causes of pneumonia. *Respirology* 23 (2), 130–137. doi:10.1111/resp.13196

Yin, Y., and Zhou, D. (2018). Organoid and enteroid modeling of Salmonella infection. *Front. Cell Infect. Microbiol.* 8, 102. doi:10.3389/fcimb.2018.00102

Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., and Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367 (19), 1814–1820. doi:10.1056/nejmoa1211721

Zhang, T., Liu, X., Yang, X., Liu, F., Yang, H., Li, X., et al. (2022a). Rapid on-site detection method for white spot syndrome virus using recombinase polymerase amplification combined with lateral flow test strip technology. *Front. Cell Infect. Microbiol.* 12, 889775. doi:10.3389/fcimb.2022.889775

Zhang, T., Ge, X., Shen, F., Qiao, J., Zhang, Y., and Li, H. (2022b). Diagnostic efficiency of RPA/RAA integrated CRISPR-cas technique for COVID-19: aa systematic review and meta-analysis. *PLoS One* 17 (10), e0276728. doi:10.1371/journal.pone.0276728

Zhang, X., Guo, L., Ma, R., Cong, L., Wu, Z., Wei, Y., et al. (2017). Rapid detection of Salmonella with recombinase aided amplification. *J. Microbiol. Methods* 139, 202–204. doi:10.1016/j.mimet.2017.06.011

Zhao, F., Wang, P., Wang, H., Liu, S., Sohail, M., Zhang, X., et al. (2023). CRISPR/Cas12a-mediated ultrasensitive and on-site monkeypox viral testing. *Anal. Methods* 15 (17), 2105–2113. doi:10.1039/d2ay01998a

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 17 October 2023
DOI 10.3389/fbioe.2023.1272811

OPEN ACCESS

EDITED BY
Anindya Bandyopadhyay,
International Maize and Wheat Improvement Center, Mexico

REVIEWED BY
Yechun Wang,
Independent Researcher, NJ, United States
Fengjiao Xin,
Institute of Food Science and Technology (CAAS), China
Anjanabha Bhattacharya,
Maharashtra Hybrid Seeds, India

*CORRESPONDENCE
Susan C. Roberts,
scroberts@wpi.edu

RECEIVED 04 August 2023
ACCEPTED 09 October 2023
PUBLISHED 17 October 2023

CITATION
Brzycki Newton C, Young EM and Roberts SC (2023), Targeted control of supporting pathways in paclitaxel biosynthesis with CRISPR-guided methylation.
Front. Bioeng. Biotechnol. 11:1272811.
doi: 10.3389/fbioe.2023.1272811

Targeted control of supporting pathways in paclitaxel biosynthesis with CRISPR-guided methylation

Cassandra Brzycki Newton, Eric M. Young and Susan C. Roberts*

Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, MA, United States

Introduction: Plant cell culture biomanufacturing is rapidly becoming an effective strategy for production of high-value plant natural products, such as therapeutic proteins and small molecules, vaccine adjuvants, and nutraceuticals. Many of these plant natural products are synthesized from diverse molecular building blocks sourced from different metabolic pathways. Even so, engineering approaches for increasing plant natural product biosynthesis have typically focused on the core biosynthetic pathway rather than the supporting pathways.

Methods: Here, we use both CRISPR-guided DNA methylation and chemical inhibitors to control flux through the phenylpropanoid pathway in *Taxus chinensis*, which contributes a phenylalanine derivative to the biosynthesis of paclitaxel (Taxol), a potent anticancer drug. To inhibit PAL, the first committed step in phenylpropanoid biosynthesis, we knocked down expression of PAL in *Taxus chinensis* plant cell cultures using a CRISPR-guided plant DNA methyltransferase (NtDRM). For chemical inhibition of downstream steps in the pathway, we treated *Taxus chinensis* plant cell cultures with piperonylic acid and caffeic acid, which inhibit the second and third committed steps in phenylpropanoid biosynthesis: cinnamate 4-hydroxylase (C4H) and 4-coumaroyl-CoA ligase (4CL), respectively.

Results: Knockdown of PAL through CRISPR-guided DNA methylation resulted in a profound 25-fold increase in paclitaxel accumulation. Further, through the synergistic action of both chemical inhibitors and precursor feeding of exogenous phenylalanine, we achieve a 3.5-fold increase in paclitaxel biosynthesis and a similar reduction in production of total flavonoids and phenolics. We also observed perturbations to both activity and expression of PAL, illustrating the complex transcriptional co-regulation of these first three pathway steps.

Discussion: These results highlight the importance of controlling the metabolic flux of supporting pathways in natural product biosynthesis and pioneers CRISPR-guided methylation as an effective method for metabolic engineering in plant cell cultures. Ultimately, this work demonstrates a powerful method for rewiring plant cell culture systems into next-generation chassis for production of societally valuable compounds.

KEYWORDS

plant cell culture, secondary metabolism, metabolic engineering, paclitaxel, CRISPR

1 Introduction

Plants produce an incredible variety of unique compounds with bioactive properties that have many industrial uses as pesticides, food additives, cosmetics, and pharmaceuticals (Zhu et al., 2021). However, sustainable industrial production of these plant natural products is often challenging. Complex stereochemistry limits yields from chemical synthesis, while lengthy or uncharacterized biosynthetic pathways make it difficult or impossible to produce certain plant natural products heterologously (Zhang et al., 2022). Plant cell culture (PCC) is therefore becoming an attractive option for sustainable biomanufacturing of complex and high-value plant secondary metabolites, including small molecule therapeutics, vaccine adjuvants, and nutraceuticals. One key challenge with optimizing plant cell systems for production of a single metabolite is that plants often have a multitude of different cooperating and competing secondary metabolic pathways whose interactions must be considered in the context of metabolic engineering. However, these complex interactions are often overlooked in favor of solely focusing on manipulating the pathway of interest. To enable more complete understanding and control of secondary metabolism and achieve the highest possible yields of products of interest, approaches for manipulating cooperating and competing metabolic pathways are needed.

One example of a PCC chassis where manipulating cooperating and competing pathways is vital is *Taxus chinensis,* which along with other species in the *Taxus* genus produces a chemotherapeutic drug called paclitaxel as a natural product (Wilson and Roberts, 2012). Production of this drug using *Taxus* PCC is one of the most wildly successful commercializations of PCC technology to date, with the world's largest supplier of paclitaxel (Phyton Biotech) using a PCC process (Wilson and Roberts, 2012). The biosynthetic pathway for synthesis of paclitaxel is highly complex—it is a cyclic diterpenoid that is synthesized from the terpene precursors IPP and DMAPP via a branching pathway that includes 19 putative enzymatic steps, 15 of which have been characterized (Sanchez-Muñoz et al., 2020). While much effort has been devoted to studying paclitaxel biosynthesis, comparatively little work has been done to understand how paclitaxel biosynthesis interacts with other secondary metabolic pathways that may compete for cellular resources (McKee et al., 2021). Thus, to improve yields of paclitaxel from PCC, it is vital to understand and manipulate not only paclitaxel biosynthesis, but also the many other metabolic pathways that make up primary and secondary metabolism.

Paclitaxel biosynthesis is hypothesized to interact with phenylpropanoid biosynthesis, as both pathways utilize the aromatic amino acid phenylalanine as a common precursor (McKee et al., 2021). In paclitaxel biosynthesis, α-phenylalanine is converted to β-phenylalanine via the enzyme phenylalanine aminomutase (PAM), which is then ligated to acetyl-CoA and attached to the taxane backbone via the enzyme BAPT (Sanchez-Muñoz et al., 2020). Phenylalanine is also a key building block used in synthesis of thousands of different phenylpropanoid compounds, including flavonoids and phenolics, which play roles as pigments, signaling molecules, and antioxidants, as well as lignin, a major component of the plant cell wall (Liu et al., 2021). The first committed step in synthesis of all phenylpropanoid-derived compounds is the conversion of α-phenylalanine to cinnamic acid by the enzyme phenylalanine ammonia-lyase (PAL) (Liu et al., 2021; McKee et al., 2021), which is known to play a vital role in controlling the division of carbon flux from primary to secondary metabolism in many plant species (Kim and Hwang, 2014; Feduraev et al., 2020; Liu et al., 2021). PAL is an especially critical enzyme because of its important role in controlling flux to lignin biosynthesis, which affects plant cell wall architecture and susceptibility to infestations and infections (Pant et al., 2021; Pant and Huang, 2022). Cinnamic acid is then converted to coumaric acid via the enzyme cinnamate 4-hydroxylase (C4H), which is then converted to coumaroyl-CoA via the enzyme 4-coumaroyl-CoA ligase (4CL) (Li et al., 2015; Liu et al., 2021; McKee et al., 2021). Cinnamic acid, coumaric acid, and coumaroyl-CoA all serve as precursors that can ultimately feed into phenylpropanoid biosynthesis through complex, branching pathways (Liu et al., 2021). Due to the involvement of the phenylpropanoid pathway in producing lignin, in many plants there is a substantial amount of carbon flux that is routed to producing phenylalanine (Yoo et al., 2013). Therefore, rerouting even a small fraction of this phenylalanine away from phenylpropanoid metabolism and toward paclitaxel biosynthesis could have the potential to significantly improve paclitaxel yields.

It is well-established that elicitation of *Taxus* spp. PCCs with methyl jasmonate or other elicitors causes general upregulation of secondary metabolism, including both taxane biosynthesis and phenylpropanoid biosynthesis (Bulgakov et al., 2011; Li et al., 2012; Zhou et al., 2019). While these studies indicate that these pathways have some degree of co-regulation, the exact effect of targeted inhibition of phenylpropanoid metabolism on taxane biosynthesis remains inconclusive. One previous study in *Taxus* found that inhibition of phenylpropanoid biosynthesis with either cinnamic acid or α-aminooxyacetic acid (AOA) significantly decreased paclitaxel biosynthesis (Brincat et al., 2002), while another study suggests no significant relationship between PAL activity and taxane biosynthesis (Bamneshin et al., 2022). Despite these results, patents on *Taxus* plant cell culture technology report that inhibitors of phenylpropanoid metabolism (in particular, cinnamic acid derivatives) can be used to increase paclitaxel production (Bringi et al., 2012). To further understand the relationship amongst these cooperating pathways, we studied the effect of inhibition of three different steps in early-stage phenylpropanoid biosynthesis (PAL, C4H, and 4CL) on taxane metabolism.

Pathway inhibition can be accomplished using two general classes of methods: genetic methods that manipulate gene expression and non-genetic methods that manipulate enzyme activity through chemical inhibition. Manipulation of metabolism using non-genetic methods can be advantageous for non-model organisms because these methods can be used to inhibit enzymes for which the exact gene sequence is not known for a particular species. While the gene for PAL has been identified in multiple *Taxus* species (Dong et al., 2016; Bamneshin et al., 2022), genes encoding the following two downstream steps in phenylpropanoid biosynthesis (C4H and 4CL) have yet to be discovered, necessitating this approach. As most current research on phenylpropanoid biosynthesis in PCCs has been centered on chemical inhibition of PAL (Brincat et al., 2002; Su et al., 2015; Andi et al., 2019; Bamneshin et al., 2022), we instead studied the effects of inhibiting the

downstream enzymes C4H and 4CL. Piperonylic acid is one of the most potent irreversible inhibitors of C4H and has been used effectively to inhibit phenylpropanoid biosynthesis in BY-2 tobacco cell suspensions (Schalk et al., 1998). Caffeic acid is an alternative substrate of 4CL (*11*) that has been shown to decrease 4CL enzymatic activity through competitive inhibition (Chen et al., 2013) and has been used to manipulate flux to lignin biosynthesis in soybean plants (Bubna et al., 2011). Neither of these inhibitors have been studied to date in *Taxus* PCC.

In addition to chemical inhibition, we also sought to manipulate the phenylpropanoid pathway using a genetic approach. Two of the most widely used methods for inhibiting gene expression are RNA interference (RNAi) and CRISPR interference (CRISPRi) (Brzycki et al., 2021). RNAi causes degradation of mRNA transcripts, thus preventing translation (Fire et al., 1998), while CRISPRi prevents transcription of a target sequence through steric hindrance (Larson et al., 2013). A third, emerging approach involves controlling gene expression epigenetically through linking catalytically dead Cas9 (dCas9) proteins to catalytic domains that add or remove epigenetic markers, such as DNA methylation and histone acetylation (Brzycki et al., 2021). For example, these effector domains could include DRM methyltransferases, which methylate DNA (Papikian et al., 2019), TET family enzymes, which demethylate DNA (Gallego-Bartolomé et al., 2018), and p300 histone acyltransferases, which participate in histone acetylation (Hilton et al., 2015). Similar to CRISPRi, these approaches regulate gene expression on the transcriptional level, while RNAi regulates gene expression on the translational level. However, CRISPR-based epigenetic engineering has the additional advantage of being able to directly manipulate specific epigenetic markers that are known to play roles in regulation of biological processes, such as cell aging and differentiation (Brzycki et al., 2021).

While CRISPR-mediated epigenetic engineering has been used effectively to control gene expression in the model plant *Arapidopsis thaliana* (Papikian et al., 2019), it has not yet been explored as a tool for metabolic engineering in plant cell culture (Perez-Matas et al., 2023). This approach could be particularly advantageous for engineering PCCs because methylation has been established as a key regulator of gene expression that increases as plant cell lines are continuously subcultured and removing DNA methylation has been associated with rescuing gene expression in epigenetically silenced plant cell lines (Fu et al., 2012; Tyunin et al., 2012). Therefore, developing approaches for specifically manipulating DNA methylation rather than transcription more generally may be advantageous for PCC systems. To this end, we sought to explore a novel technique for manipulating phenylpropanoid metabolism through CRISPR-mediated targeted DNA methylation of PAL, the first committed step in phenylpropanoid biosynthesis and the only phenylpropanoid pathway enzyme currently characterized in *Taxus chinensis* (Dong et al., 2016). Through studying the effects of inhibiting these first three steps in phenylpropanoid biosynthesis, we aim to gain a more complete understanding of the relationship between phenylpropanoid and taxane metabolism than through manipulating PAL alone.

Here, we explore genetic and non-genetic inhibition of the first three committed steps in phenylpropanoid biosynthesis (PAL, C4H, and 4CL) to elucidate its co-regulation with taxane biosynthesis. Using chemical inhibitors and precursor feeding with phenylalanine, we find that we can achieve up to a 3.5-fold increase in paclitaxel accumulation. The increase in paclitaxel production observed with chemical inhibition is amplified when cultures are fed with exogenous phenylalanine, indicating that phenylalanine availability likely plays a regulatory role in directing flux between phenylpropanoid and taxane biosynthesis. Unexpectedly, we also found that accumulation of the taxane pathway intermediates 10-deacetylbaccatin III and baccatin III also increased significantly when phenylpropanoid metabolism was disrupted, despite these steps being upstream of where the phenylalanine side chain attaches to the taxane backbone. We propose that this differential accumulation is due to the involvement of phenylpropanoid metabolism in biosynthesis of plant hormones that regulate the stress response, specifically salicylic acid. We also show that CRISPR-mediated DNA methylation upstream of the PAL gene disrupts phenylpropanoid metabolism and consequently increases production of paclitaxel by up to 25-fold. This provides a proof-of-concept for the application of CRISPR tools to PCCs and demonstrates that targeted repression of pathways can be a powerful tool for manipulating metabolism. These results provide insights on the complex interactions between phenylpropanoid biosynthesis and paclitaxel biosynthesis in *Taxus* species and more generally demonstrate the efficacy of several novel genetic and non-genetic approaches for manipulating flux between competing metabolic pathways in PCC.

2 Materials and methods

2.1 Plant cell line initiation and maintenance

Taxus chinensis cell line 48.82A.3s was initiated from tree 48–82*A at the Arnold Arboretum of Harvard University (Cambridge, MA). Healthy light green needle samples were surface-sterilized by incubating for 10 min at room temperature in a 10% bleach solution and rinsed 3 times with sterile water. Needles were cut lengthwise with a sterile scalpel and plated on solid callus initiation media containing 20 g/L sucrose (Caisson Labs), 4 g/L Gelzan (Phytotechnology Laboratories), 3.21 g/L Gamborg B5 Basal Medium (Caisson Labs), 0.5 g/L casamino acids (Fisher BioReagents), 2.4 mg/L picloram (TCI America), and 2 μg/L 6-benzylaminopurine (Sigma Life Sciences), pH 5.5. Plates were incubated at 23°C in the dark for approximately 4–6 weeks, or until the formation of friable callus. Callus tissue was gently separated from the needles using a sterile spatula and transferred to freshly prepared callus initiation media every 4 weeks until sufficient biomass was obtained to transfer into suspension cultures. Approximately 3 g of wet biomass was inoculated into 50 mL of *Taxus* growth medium (20 g/L sucrose, 0.5 mg/L naphthaleneacetic acid, 22.5 μg/L 6-benzylaminopurine, 3.21 g/L Gamborg B5 Basal Medium, pH 5.5) supplemented with a ×20 sterile-filtered antioxidant solution (2.5 g/L ascorbic acid, 2.5 g/L citric acid, 14.6 g/L L-glutamine) at 5% v/v. Cultures were transferred to an incubator-shaker at 23°C and 125 rpm and maintained in suspension via biweekly subculturing as previously described (Kolewe et al., 2011).

2.2 Phenylalanine precursor feeding and inhibitor experiments

For experiments with phenylalanine, piperonylic acid, and caffeic acid, ×20 antioxidant solutions were prepared as described in Section 2.1 with the addition of 20 mM phenylalanine, 2 mM piperonylic acid (PA), and/or 2 mM caffeic acid (CA). At the beginning of a new subculture cycle, *Taxus chinensis* 48.82A.3s cultures were subcultured into growth media supplemented with these antioxidants and incubated at 23°C and 125 rpm for 21 days. 1 mL well-mixed culture samples were then taken for measurement of taxanes, cinnamic acid, and coumaric acid via UPLC, phenolic and flavonoid assays, gene expression via RT-qPCR, and PAL activity. Samples for UPLC and phenolic and flavonoid assays were dried overnight in an evaporative centrifuge and then frozen at −80°C. Samples for gene expression were allowed to sit at room temperature for several minutes, or until cells settled to the bottom of the microcentrifuge tube. The supernatant was removed using a pipette and samples were then immediately frozen at −80°C. Samples for the PAL activity assay were processed immediately (see Section 2.5).

2.3 Quantification of taxanes, coumaric acid, and cinnamic acid

Taxanes (including paclitaxel, baccatin III, 10-deacetyltaxol, cephalomannine, and 10-deacetylbaccatin III) were extracted from 1 mL of well-mixed culture and quantified using a previously described method (Naill and Roberts, 2004). Dried cell samples stored at −80°C were resuspended in 1 mL acidified methanol (0.01% acetic acid in methanol), broken up with a metal spatula, then sonicated three times for 20 min per cycle in a sonication bath. When the pellets were evenly and sufficiently pulverized to homogenous sand-like granules, the non-dissolved cell matter was separated from the methanol extract by centrifugation at 20,000 ×g for 10 min. This extract was evaporated using an evaporative centrifuge and samples were then resuspended in 100 µL 25/35/40% (v/v) methanol/acetonitrile/water. Taxanes were quantified using a Waters Acquity UPLC H-Class system equipped with a 1.7 µm Acquity UPLC BEH C18 column. Separation was performed using a 10 µL injection with a gradient of 30%–80% acetonitrile in water over 4 min at a flow rate of 0.35 mL/min. Peak areas at 228 nm were used to determine the concentration of each taxane.

For cinnamic acid and coumaric acid, the same samples were run using a gradient HPLC method utilizing 0.5% phosphoric acid in water (mobile phase A) and acetonitrile (mobile phase B) adapted for use in UPLC from Wan Mazlina Md et al. (2019). Separation was performed using a 1 µL injection with a gradient of 95%–20% mobile phase A over 1.97 min at a flow rate of 0.5 mL/min. Peak areas at 280 nm were used to determine concentrations of cinnamic acid and peak areas at 310 nm were used to determine concentrations of coumaric acid. All taxane, cinnamic acid, and coumaric acid peaks were identified using retention time, UV spectra, and co-chromatography with commercially available standards.

2.4 Total phenolic and flavonoid assays

Spectrophotometric assays for total phenolics and flavonoids were performed based on the procedure described in McKee et al. (2021). Evaporated cell samples stored at −80°C were resuspended in 1 mL acidified methanol (0.01% acetic acid in methanol), broken up with a metal spatula, then sonicated three times for 20 min per cycle in a sonication bath. When the pellets were evenly and sufficiently pulverized to homogenous sand-like granules, the non-dissolved cell matter was separated from the methanol extract by centrifugation at 20,000 ×g for 10 min.

The phenolic content per sample was determined using the Folin-Ciocalteu reagent assay (Ainsworth and Gillespie, 2007). 120 µL methanol extract was combined with 200 µL Folin-Ciocalteu reagent (0.2 N) and 960 µL sodium carbonate (700 mM) in a microcentrifuge tube. After 10 min of incubation at room temperature, any precipitant was pelleted via centrifugation at 20,000 ×g for 1 min 200 µL of supernatant was transferred to a 96-well plate in triplicate, and the absorbance was read at 750 nm using the Synergy H1 Hybrid Multi-Mode Reader (BioTek). Absorbance values were converted to concentrations using gallic acid as a phenolic compound standard.

The flavonoid content per sample was determined using an aluminum chloride-based assay (Chang et al., 2002). 100 µL methanol extract and 200 µL water were combined with 300 µL $NaNO_2$ (6 g/L) in a microcentrifuge tube. After 1 min of incubation at room temperature, 300 µL $AlCl_3\square 6H_2O$ (22 g/L) was added. After 2 min of incubation at room temperature, 300 µL of NaOH (0.8 M) was added to the microcentrifuge tube. 200 µL of supernatant was transferred to a 96-well plate in triplicate, and the absorbance was read at 490 nm using the Synergy H1 Hybrid Multi-Mode Reader (BioTek). Absorbance values were converted to concentrations using catechin as a flavonoid compound standard.

2.5 PAL activity assays

Phenylalanine ammonia-lyase (PAL) activity was quantified in samples using a previously established protocol with several modifications (Sykłowska-Baranek et al., 2012). Fresh well-mixed culture samples in Safe-Lock microcentrifuge tubes (Eppendorf) were incubated at room temperature for several minutes, or until the cells settled to the bottom of the tube. Culture media was removed using a pipette tip and cells were washed twice with phosphate buffered saline (PBS) to remove any residual phenylalanine from the culture media. Following this, 1 mL assay buffer (50 mM Tris-HCl pH 8.0, 0.8 mM β-mercaptoethanol, 1% w/v polyvinylpyrrolidone) was added to each sample, along with the contents of one 2 mm BashingBead lysis tube (Zymo Research). Samples were homogenized in a bullet blender (Next Advance) at speed 8 for two 5 min cycles, then centrifuged at 20,000 ×g for 3 min. The following components were then transferred to a fresh microcentrifuge tube: 250 µL supernatant from cell lysate, 125 µL water, 250 µL Tris-HCl (pH 8.0), 125 µL L-phenylalanine (50 mM). Tubes were inverted several times to mix and incubated at room temperature for 1 h, following which 200 µL from each tube was transferred to a 96 well plate in triplicate and the absorbance was read at 290 nm. Enzyme activity was normalized to the cell line not treated with phenylalanine, PA, or CA.

2.6 Construction of binary vectors and transgenic Agrobacterium strains

All cloning was performed using NEB 10-beta competent *E. coli* (New England Biolabs) according to manufacturer's instructions. Templates for construction of the no guide RNA control (ng) and PAL knockdown (PALg1) vectors were obtained from Addgene. Addgene #115488 was used as the no guide RNA control with no modifications. For generation of the PAL knockdown vector, the guide RNA sequence of Addgene #115486 was modified to 5′-CAC AGGCTTAAGCACCACCC-3′ using Gibson assembly. Guide RNA was designed in Benchling to target the 5′UTR of the *Taxus chinensis* PAL gene (Genbank accession #KF713533). Both ng and PALg1 constructs were then transformed into electrocompetent *Agrobacterium tumefaciens* EHA105 (GoldBio) according to manufacturer's instructions.

2.7 Transformation and maintenance of transgenic *Taxus chinensis* cell lines

Transgenic *Taxus chinensis* 48.82A.3s cell lines were generated using a transformation protocol based on that described in Wilson et al. (2018). Glycerol stocks of *Agrobacterium tumefaciens* strains used for transformation were streaked on YEP plates (10 g/L yeast extract, 20 g/L peptone, 20 g/L agar) with 50 μg/mL kanamycin and rifampicin and grown for 2 days at 30°C, or until the formation of colonies. For each transformation, a single colony was used to inoculate 25 mL of YEP with 50 μg/mL kanamycin and rifampicin in a 125 mL flask. Flasks were incubated in a shaker-incubator at 30°C and 220 rpm until turbid, about 24–48 h. The OD_{600} of each culture was measured and cultures were pelleted by centrifugation at 4,000 ×g for 10 min. The supernatant was removed and cultures were resuspended to OD_{600} = 0.6 in 50 mL *Taxus* growth media (20 g/L sucrose, 0.5 mg/L naphthaleneacetic acid, 22.5 μg/L 6-benzylaminopurine, 3.21 g/L Gamborg B5 Basal Medium, pH 5.5) supplemented with a ×20 anti-necrotic solution (2.5 g/L ascorbic acid, 2.5 g/L citric acid, 14.6 g/L L-glutamine, 1 g/L L-cysteine) at 5% v/v. Acetosyringone (final concentration 200 μM) was then added to each culture and cultures were incubated at 30°C and 220 rpm for 30 min.

Healthy *Taxus chinensis* 48.82A.3s cultures on day 7–10 of a subculture cycle (mid-late log phase; approximately 15–20 mL packed cell volume per 50 mL culture) were sheared approximately 25 times with a 10 mL serological pipette to reduce aggregate size and improve infiltration. Cultures were then transferred to 50 mL conical tubes and allowed to settle without centrifugation for 10 min. Once there was a clear phase separation between cells and media, the supernatant was carefully removed using a serological pipette and replaced with the prepared *Agrobacterium* culture. The *Taxus-Agrobacterium* co-culture was then transferred back into a 125 mL flask and incubated at 23°C and 125 rpm for approximately 48 h. Following co-incubation, the co-culture was transferred into a 50 mL conical tube and cells were allowed to settle for 10 min, or until there was a clear phase separation between plant cells and *Agrobacterium* culture. *Agrobacterium* culture was removed using a serological pipette and replaced with an equal volume of *Taxus* growth media with anti-necrotic solution and 300 mg/L cefotaxime. Flasks were swirled approximately 10–20 times to mix and then poured into a conical tube to settle again. This media washing step was repeated 2 more times or until the supernatant appeared clear. The washed and resuspended *Taxus* culture was then transferred back into a 125 mL flask and incubated at 23°C and 125 rpm overnight to ensure complete eradication of *Agrobacterium*.

After overnight incubation, cultures were vacuum-filtered through a sterile Buchner funnel lined with Miracloth for approximately 30 s, or until biomass was lightly dried. The resulting biomass was then transferred to *Taxus* growth media plates (20 g/L sucrose, 0.5 mg/L naphthaleneacetic acid, 22.5 μg/L 6-benzylaminopurine, 3.21 g/L Gamborg B5 Basal Medium, 4 g/L Gelzan, pH 5.5) with anti-necrotic solution and 300 mg/L cefotaxime using a sterile spatula. Plates were wrapped with Parafilm and allowed to incubate at 23°C in the dark until friable callus growth was observed; approximately 4–6 weeks. Approximately 3 g of callus was then transferred into 125 mL flasks containing 50 mL of *Taxus* growth media with anti-necrotic solution, 300 mg/L cefotaxime, and 10 μg/mL hygromycin (selectable marker). Transgenic cultures were sheared approximately 50 times to break up cellular aggregates, then allowed to incubate at 23°C and 125 rpm for 21 days before taking culture samples for taxanes, DNA methylation, and gene expression. Samples for taxanes were evaporated overnight in an evaporative centrifuge as described in Section 2.2 and then analyzed for taxane content as described in Section 2.3. Samples for DNA and PAL gene expression were processed and stored as described for the gene expression samples in Section 2.2.

2.8 Estimation of PAL DNA methylation via MSRE-qPCR

To determine the approximate percentage of DNA methylation of the region targeted by the guide RNA, we used methylation-sensitive restriction enzyme qPCR (MSRE-qPCR). Genomic DNA from ng and PALg1 cell lines was extracted from frozen cell samples using the Wizard High MW DNA Extraction Kit (Promega) following the procedure for isolation from plant tissue. Instead of homogenization with liquid nitrogen, homogenization was performed using a Bullet Blender as described in Section 2.6, following which the procedure was performed according to manufacturer's specifications. 100 ng of genomic DNA was then digested with the restriction enzyme McrBC (New England BioLabs) at 37°C for 4 h. Undigested controls were incubated with the same volume of nuclease-free water instead of restriction enzyme under the same conditions.

The digested genomic DNA and undigested controls were then amplified using qPCR. PrimeTime qPCR Probe Assays (Integrated DNA Technologies) were designed using IDT's PrimerQuest design tool to target the same region as the guide RNA and contain at least 2 CpG sites (Supplementary Table S1). qPCR was then performed on an Applied Biosystems 7500 Real-Time PCR System (Thermo Fisher) using IDT's PrimeTime 2x Gene Expression Master Mix. The difference in C_t values between the digested DNA and undigested control for each sample was used to determine the approximate percentage of DNA methylation of the targeted

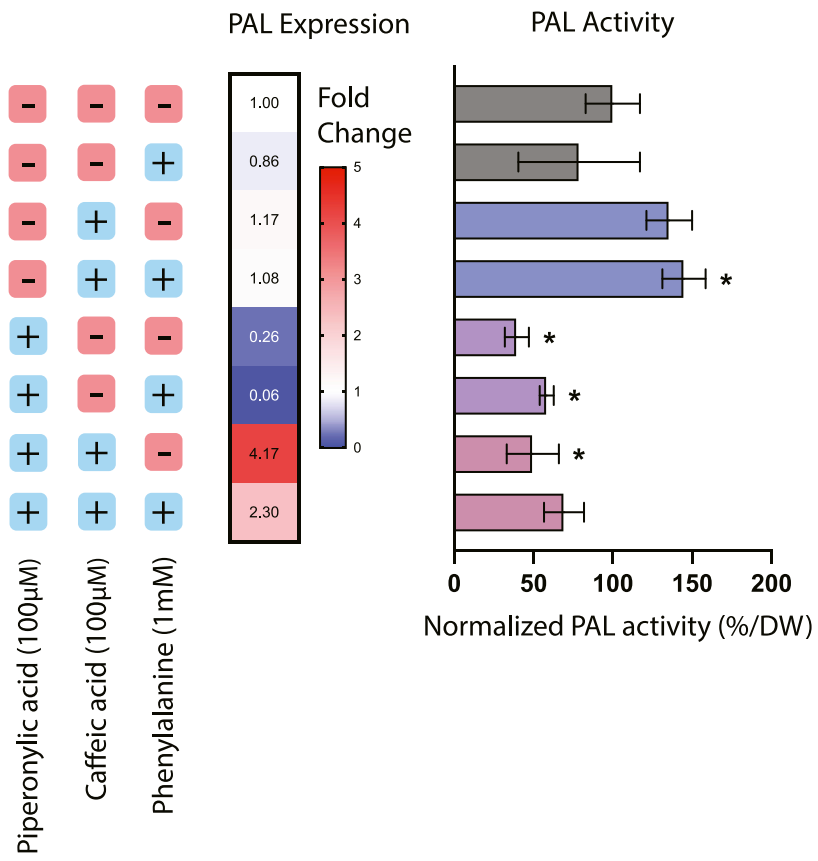


FIGURE 1
Phenylalanine ammonia-lyase (PAL) expression and activity in *Taxus chinensis* cell lines treated with phenylpropanoid pathway inhibitors piperonylic acid (PA) and caffeic acid (CA). PAL expression was measured using RT-qPCR. PAL activity was assessed using an enzyme activity assay and normalized by biomass dry weight. All values are normalized relative to the cell line not treated with PA or CA and without the addition of exogenous phenylalanine. Data represent biological triplicates ± standard deviation; *$p < 0.05$.

region. Due to the lack of a fully methylated control, a no template control was used instead to simulate fully methylated and thus completely digested DNA.

2.9 Quantification of PAL gene expression

For determination of PAL differential gene expression for both the inhibitor experiments (Section 2.2) and the experiments with transgenic cell lines (Section 2.7), total RNA was extracted using a Quick-RNA Plant Kit (Zymo Research). Cell samples were homogenized using a Bullet Blender as previously described in Section 2.5, following which RNA extraction was performed according to manufacturer's instructions. Immediately following RNA extraction, cDNA was synthesized using approximately 1 µg of total RNA using the ProtoScript II First Strand cDNA Synthesis Kit (New England Biolabs). Expression of PAL was then quantified using an Applied Biosystems 7500 Real-Time PCR System (Thermo Fisher). PrimeTime qPCR Probe Assays (Integrated DNA Technologies) were designed for the PAL gene as well as the housekeeping gene (actin) using the IDT PrimerQuest design tool (Supplementary Table S1). HEX/ZEN/Iowa Black FQ probes were used for the housekeeping gene and FAM/ZEN/Iowa Black FQ probes were used for PAL, to enable in-well multiplexing. The $\log_2$fold change in expression of PAL relative to the respective control for each experiment was calculated using the double delta C_t method.

3 Results

3.1 Disruption of phenylpropanoid biosynthesis with chemical inhibitors affects PAL expression and activity

To begin to investigate the effects of phenylpropanoid pathway inhibition on taxane biosynthesis, we first turned to chemical inhibition of the pathway. Here, *Taxus chinensis* suspension cultures were treated with 100 µM piperonylic acid (PA) to inhibit C4H, 100 µM caffeic acid (CA) to inhibit 4CL, or 100 µM of both compounds. Notably, the genes coding for these enzymes have not yet been identified in *Taxus* species, necessitating the use of chemical inhibition rather than genetic manipulation. These inhibitors were tested with or without the presence of 1 mM exogenous phenylalanine to evaluate the effect of an increased

precursor pool on flux to either phenylpropanoid or taxane metabolism.

Treatment with phenylpropanoid pathway inhibitors had a significant effect on both expression and activity of PAL, the first committed step in the phenylpropanoid biosynthesis pathway from the precursor phenylalanine (Figure 1). The addition of PA alone resulted in a 60.5% decrease in PAL activity as well as a corresponding decrease in PAL expression of around 76%, suggesting that inhibition of C4H also affects regulation of this entry point into the pathway. This result is consistent with previous work in *Nicotiana tabacum* where silencing of C4H resulted in a decrease in both PAL activity and levels of downstream phenylpropanoid compounds, likely through product inhibition from cinnamic acid accumulation (Blount et al., 2000).

For the cell lines treated with CA or both CA and PA, effects of these inhibitors on PAL activity were decoupled from their effects on PAL expression. Treatment with CA alone resulted in an increase in PAL activity of around 35.6% but no significant effect on PAL expression, while treatment with both PA and CA together resulted in a 50.4% decrease in PAL activity but around a 4-fold increase in PAL expression. This discrepancy suggests that in these instances, transcription-level regulation is not necessarily coupled with protein-level regulation. For instance, in the case of treatment with both CA and PA, while PAL activity is decreased due to the putative action of C4H reducing PAL activity through product inhibition, it is possible that gene expression of PAL is actually increasing due to transcription-level regulatory mechanisms seeking to compensate for the decreased conversion of phenylalanine. One further consideration is also the fact that many plants have multiple copies of the PAL gene, ranging from 4 copies in the model plant *Arapidopsis thaliana* to 26 copies in the polyphenol-rich tomato (Liu et al., 2023). While PAL activity would be detectable from all possible PAL isoforms, it is likely that expression of multiple possible isoforms would not all be quantifiable via qPCR without designing multiple different probes. Thus, while to date only one copy of the PAL gene has been identified in *Taxus chinensis* (Genbank accession #KF713533), it is reasonable to assume that there may likely be other copies that have not yet been characterized that may be subject to different regulation than the specific copy of PAL targeted by these qPCR probes. More detailed mining for PAL isoforms in the *Taxus chinensis* genome would be needed to determine the number of copies of PAL and elucidate their regulation.

3.2 Treatment with piperonylic acid decreases biosynthesis of total flavonoids and phenolics

Next, to evaluate the effects of chemical inhibition on phenylpropanoid pathway products, the two early-stage phenylpropanoid precursors cinnamic acid and coumaric acid were quantified via UPLC and total phenolics and flavonoids were estimated using colorimetric assays (Figure 2). The addition of exogenous phenylalanine without the presence of inhibitors resulted in both a decrease in cinnamic acid accumulation and an increase in coumaric acid accumulation, indicating that under conditions of excess phenylalanine bioavailability, C4H is more active in converting cinnamic acid to coumaric acid. Since under these circumstances PAL activity was slightly decreased (78.8% that of the control) and PAL expression is largely unchanged, this suggests that the accumulation of coumaric acid is likely primarily due to regulation at C4H rather than PAL. Further, treatment with PA alone resulted in a significantly higher accumulation of cinnamic acid as expected due to inhibition of C4H, but also a slightly elevated level of coumaric acid. One potential explanation for this phenomenon is that through initial inhibition of coumaric acid formation via inhibition of C4H, the enzyme TAL (which often works in tandem with PAL) became more active in converting tyrosine to coumaric acid due to the lack of production of coumaric acid by 4CL.

Cultures treated with CA alone did not accumulate significantly higher levels of coumaric acid, which was unexpected due to the proposed mechanism of caffeic acid inhibiting conversion of coumaric acid to coumaroyl-CoA via 4CL. It is possible that this concentration of caffeic acid alone was insufficient to inhibit 4CL via product inhibition for the entire length of the experiment because caffeic acid is consumed as a substrate in the reaction. In contrast, the same concentration of PA provided sufficient inhibition because it is hypothesized to irreversibly bind to its substrate and is known to be active in concentrations as low at 10 μM (Nkomo et al., 2022). However, with both CA and PA treatment we observed a dramatic increase in coumaric acid accumulation as well as a smaller increase in cinnamic acid accumulation, despite the decreased PAL activity observed for these treatments. This indicates that CA is in fact inhibiting coumaric acid conversion, at least when combined with treatment with PA, and suggests that there may be some emergent effects on the pathway from treatment with both inhibitors simultaneously.

While treatment with phenylalanine, CA, and PA resulted in complex and sometimes unexpected effects on PAL activity and accumulation of the earlier pathway intermediates, the effect on accumulation of total phenolics and flavonoids was quite straightforward. All treatments with PA resulted in a statistically significant decrease in both total flavonoids and total phenolics (approximately a 2- to 5-fold decrease, depending on the treatment), while phenylalanine feeding alone or treatment with CA had no effect on accumulation of either of these metabolite classes. This decrease in metabolite accumulation for the cell lines treated with PA is consistent with the decreases in PAL activity observed in Figure 1 and previously characterized effects of PA on downstream phenylpropanoid metabolite accumulation (Nkomo et al., 2022). Interestingly, while treatment with PA and CA resulted in dramatic increases in coumaric acid accumulation, there was still an overall decrease in the total accumulation of both phenolics and flavonoids despite elevated levels of these early pathway intermediates. This suggests that there are likely additional regulatory mechanisms downstream of coumaric acid that are affected by the presence of these inhibitors, illustrating the complexity and level of cross-talk in regulation of this pathway.

3.3 Inhibition of phenylpropanoid metabolism with chemical inhibitors increases production of paclitaxel and its precursors

To elucidate the relationship between taxane and phenylpropanoid biosynthesis, we evaluated the effect of

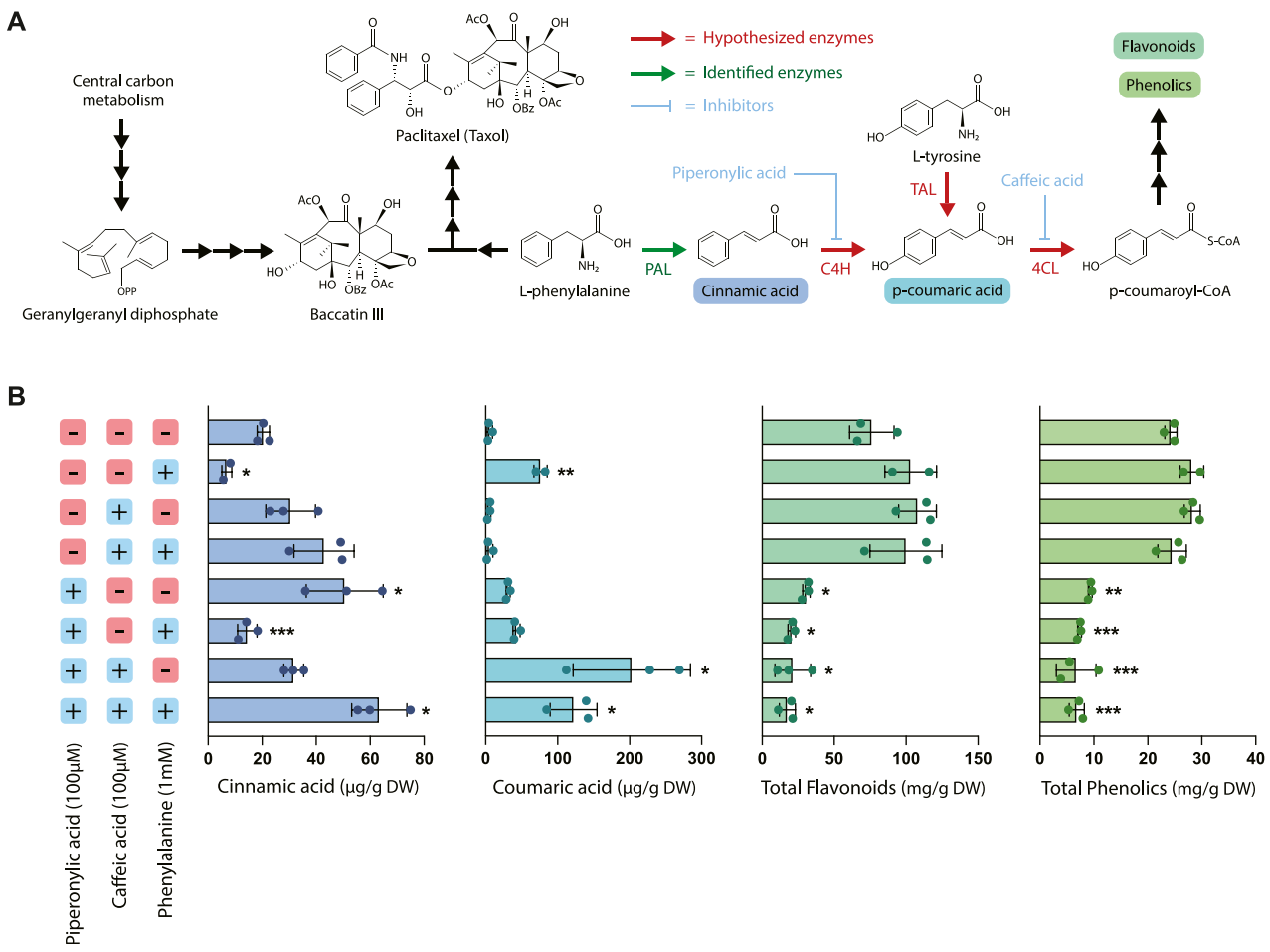


FIGURE 2
Effect of piperonylic acid (PA), caffeic acid (CA) and phenylalanine on phenylpropanoid pathway intermediates, flavonoids, and phenolics. **(A)** Pathway diagram illustrating the interactions between phenylpropanoid biosynthesis and taxane biosynthesis. Abbreviations: (PAL) phenylalanine ammonia-lyase, (C4H) cinnamate 4-hydroxylase, (TAL) tyrosine ammonia-lyase, (4CL) 4-coumarate-CoA ligase. **(B)** Concentrations of cinnamic acid, coumaric acid, total flavonoids, and total phenolics in *Taxus chinensis* cell cultures treated with 100 μM piperonylic acid, 100 μM caffeic acid, and/or 100 mM phenylalanine normalized by biomass dry weight. Statistical significance is relative to untreated control (top row); $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

phenylalanine precursor feeding and phenylpropanoid pathway inhibition with CA and PA on taxane accumulation (Figure 3). In particular, we measured the accumulation of five different taxanes: 10-deacetylbaccatin III (10-DAB), baccatin III, and paclitaxel, which occur in the canonical paclitaxel biosynthesis pathway, as well as the two taxane side products 10-deacetyltaxol (10-DAT) and cephalomannine. These taxanes are two of dozens of "paclitaxel impurities," which are compounds closely related to paclitaxel whose biosynthetic pathways are not yet characterized. While not much is currently known about biosynthesis of these compounds, they are possibly derived from taxane pathway intermediates like 10-DAB and baccatin III (Croteau et al., 2006; Li et al., 2017). Through studying their accumulation patterns, we aim to gain insight into their biosynthetic pathways and pathway regulation.

Most notably, treatment with PA significantly increased paclitaxel accumulation, with the largest increase (approximately a 3.5-fold increase over the control) observed in cultures treated with all three compounds (phenylalanine, PA, and CA). This increase in paclitaxel accumulation was strongly inversely correlated with the decreased PAL activity and phenolic/flavonoid accumulation observed in the cell lines treated with PA and interestingly, not PAL expression. This strongly suggests that repression of phenylpropanoid biosynthesis through decreased activity of PAL and/or C4H reroutes phenylalanine toward synthesis of other natural products, such as taxanes. 10-DAT production appeared to be roughly correlated with paclitaxel production, with the exception of the cultures treated with all three compounds that instead produced a large amount of cephalomannine. Additionally, 10-DAT production was consistently slightly increased with the addition of phenylalanine, while cephalomannine production did not follow this same trend. This suggests that perhaps 10-DAT is synthesized from either paclitaxel itself or another late-stage intermediate that occurs after the addition of the phenylalanine side chain to baccatin III, since production of

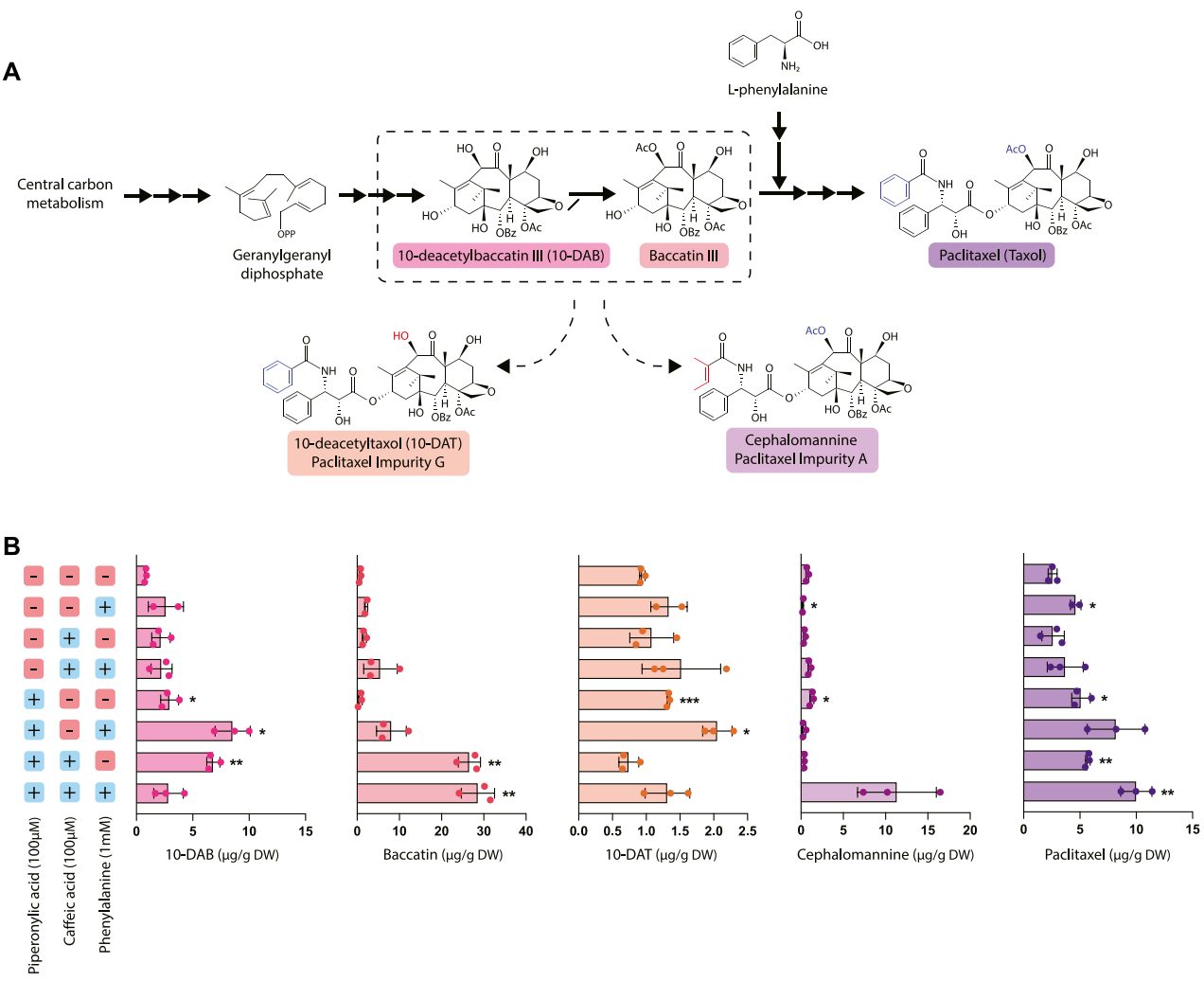


FIGURE 3
Effect of piperonylic acid (PA), caffeic acid (CA) and phenylalanine on taxane biosynthesis. **(A)** Simplified pathway depicting biosynthesis of paclitaxel from the precursor geranylgeranyl diphosphate. 10-deacetyltaxol (10-DAT) and cephalomannine are taxane derivatives synthesized from currently unknown pathways likely derived from paclitaxel biosynthetic pathway intermediates **(B)** Concentrations of 10-DAB, baccatin III, 10-DAT, cephalomannine, and paclitaxel in *Taxus chinensis* cell cultures treated with 100 μM piperonylic acid, 100 μM caffeic acid, and/or 100 mM normalized by biomass dry weight. Statistical significance is relative to untreated control (top row); $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

10-DAT and paclitaxel is possibly co-regulated. Conversely, since cephalomannine is not affected by the addition of phenylalanine in the same way that 10-DAT and paclitaxel are, it may be synthesized from a different, earlier precursor.

Unexpectedly, inhibition of phenylpropanoid metabolism not only affected paclitaxel accumulation, but also the accumulation of other taxanes upstream of the point in the pathway where phenylalanine attaches to form the side chain of paclitaxel (after the synthesis of baccatin III). Treatment with PA generally resulted in a significant increase in 10-DAB accumulation with the exception of the cell line treated with all three compounds (phenylalanine, CA, and PA). This line had significantly increased accumulation of both paclitaxel and cephalomannine, suggesting that rather than accumulating high levels of 10-DAB, it was instead converted to these downstream products. Further, cell lines treated with both CA and PA had profoundly increased levels of baccatin III accumulation—nearly 30 times higher than that of the control. These data suggest that while treatment with CA alone was insufficient to affect flux to taxane biosynthesis, the emergent effect of PA and CA together is more effective at directing flux than PA alone. Additionally, this points to DBAT, the gene controlling conversion of 10-DAB to baccatin, as an important gene involved in controlling taxane biosynthesis. This is consistent with previous data establishing DBAT as a putative rate-controlling step that is known to be subject to complex transcriptional regulation (Zhang et al., 2011; Li et al., 2013; Chen et al., 2022).

3.4 Repression of PAL by CRISPR-mediated DNA methylation specifically upregulates paclitaxel biosynthesis

While chemical inhibition of metabolism is advantageous for certain circumstances, we also wanted to pioneer novel approaches

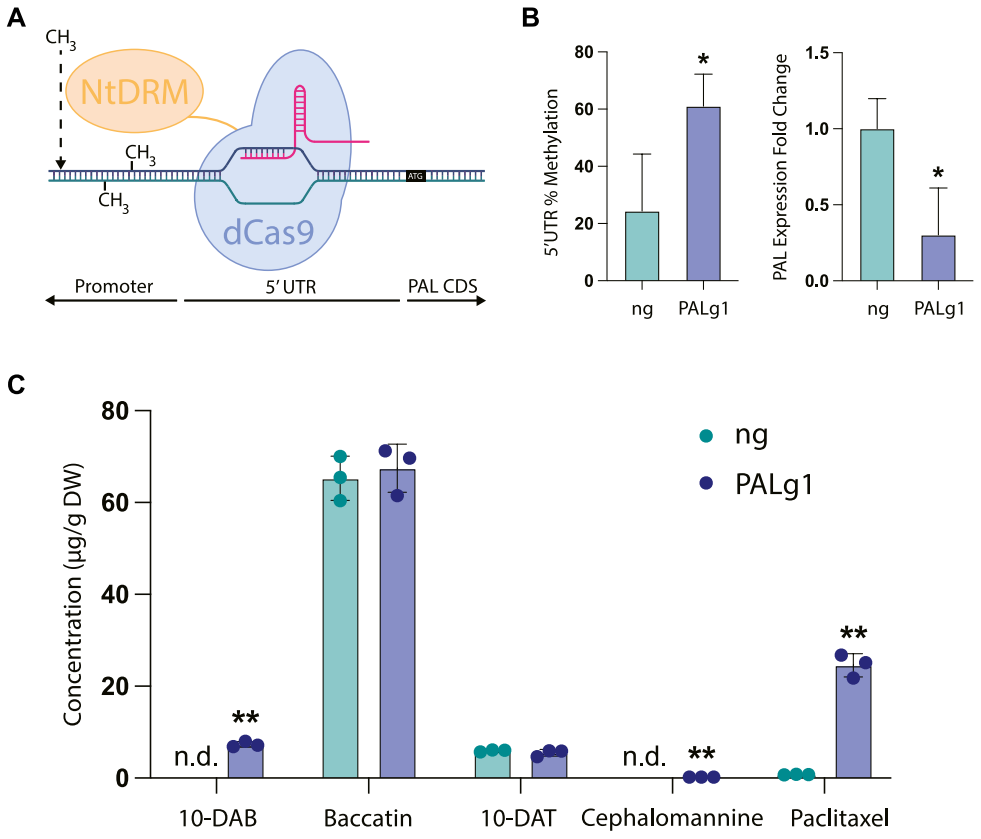


FIGURE 4
CRISPR/dCas9-NtDRMcd targeted methylation of PAL in *Taxus chinensis*. **(A)** Schematic of CRISPR/dCas9-NtDRMcd mechanism of action. Guide RNA was designed to allow dCas9 to bind to the 5′ untranslated region (5′UTR) of the PAL gene to enable targeted DNA methylation by the NtDRMcd domain. **(B)** Approximate DNA methylation of the PAL 5′UTR measured by MSRE-qPCR and PAL gene expression measured by RT-qPCR for the no guide RNA control (ng) and PAL methylated (PALg1) cell lines. **(C)** Concentrations of 10-DAB, baccatin III, 10-DAT, cephalomannine, and paclitaxel in engineered ng and PALg1 cell lines normalized by biomass dry weight. Statistical significance represents difference between ng and PALg1 cell lines; $*p < 0.05$, $**p < 0.01$.

for targeted genetic manipulation of metabolism. In particular, while transient overexpression and knockdown of genes has been used previously in *Taxus* cell culture for gene characterization (Zhang et al., 2015; Zhang et al., 2018; Chen et al., 2021), we wanted to study the effects of repressing phenylpropanoid biosynthesis in a stably transformed cell line representative of what might be used in an industrial process (Wilson et al., 2018). There have been some efforts to increase paclitaxel titers in *Taxus* plant cell culture through pathway overexpression (Zhang et al., 2011), but CRISPR-based repression of competing pathways has not yet been explored to our knowledge. To this end, we chose to knock down the PAL gene using a CRISPR-based approach, since this was the only gene in the phenylpropanoid biosynthetic pathway that had specifically been characterized in *Taxus chinensis*. Due to the importance of the PAL gene in producing natural products vital to plant cell health (such as lignin, which is vital for formation of the plant cell wall), we chose to use targeted DNA methylation as an approach to modestly decrease, but not completely knock down expression of this gene.

As a template, we chose to use a dCas9-SunTag system for engineering DNA methylation in the model plant *Arabidopsis thaliana* developed by the Jacobsen lab (Papikian et al., 2019). This work adapted the dCas9-SunTag system for use in engineering DNA methylation in plants by using the catalytic domain of a DNA methyltransferase from *Nicotiana tabacum* (NtDRMcd) as well as plant-specific promoters and terminators, which resulted in *Arabidopsis* plants that had heritable and stable epialleles. The original work found that in general, guide RNAs placed near the transcription start site of the targeted gene tended to be more effective, so we designed our guide RNA to bind to the 5′ untranslated region (5′UTR) of the PAL gene, with the 3′ end of the guide RNA placed 6 base pairs upstream of the transcription start site. We hypothesized that targeting the gRNA to the 5′UTR would broadly methylate both the 5′UTR and the unidentified promoter region upstream, since in the original manuscript the authors found that the methyltransferase was able to methylate a large region of DNA spanning up to approximately 2 kb around the guide RNA. Finally, the absence of a sequenced *Taxus chinensis* genome at the time of these experiments also necessitated the use of a no guide RNA control, since we did not want to design a scrambled guide RNA control that had unpredictable targeting activity.

In our stably transformed cell lines, we observed a significant increase in DNA methylation of the region targeted by the guide RNA from approximately 24% in the no guide RNA control to around 61% in the PAL knockdown line PALg1 (Figure 4B). This

was coupled with a significant decrease in PAL gene expression, with expression in the PALg1 line approximately 30% that of the ng control. These results confirm that our guide RNA is targeting the desired region, and that this increased DNA methylation is resulting in changes in gene expression. While we did not specifically evaluate DNA methylation of the PAL promoter region since it has not yet been characterized, from these data we can assume that methylation of the 5′UTR was sufficient to cause gene repression and/or methylation of the promoter region was similarly increased as expected from the original study.

As expected, we also observed differential accumulation of taxanes in the PALg1 and ng *Taxus chinensis* cell lines (Figure 4C). While with the chemical inhibitors (Figure 3B) we observed approximately a 3.5-fold increase in paclitaxel accumulation for the cell line treated with all three compounds (phenylalanine, PA, and CA) relative to the control, here we saw approximately a 25-fold increase in paclitaxel accumulation from 0.76 μg/g DW in the ng cell line to 24.5 μg/g DW in the PALg1 cell line. Additionally, 10-DAB and cephalomannine were not detected in the ng line but detected in small amounts in the PALg1 line, while there was no difference in 10-DAT or baccatin production between the two lines. This suggests that by directly inhibiting PAL, we can achieve a more specific increase in yield of paclitaxel when compared to chemically inhibiting downstream steps in the pathway, which resulted in a more significant increase in the yield of other taxane intermediates and side products.

4 Discussion

Plant cell culture systems are becoming increasingly popular for large-scale production of pharmaceuticals, supplements, and other complex plant natural products of social interest. However, a major challenge with improving yield of products of interest in plant cell culture is the number of complex and often intertwined metabolic pathways, each subject to their own intricate pathway regulation. In the case of *Taxus* species, the enzyme PAL acts as a bridge between two competing pathways: phenylpropanoid metabolism and taxane metabolism, which both utilize phenylalanine as a precursor. Often in the context of metabolic engineering we solely consider manipulating the pathway of interest, but here, we show that manipulating cooperating or competing pathways can be of equal or even greater importance in plant cell systems. In this work, we demonstrated that we can achieve up to a 25-fold increase in paclitaxel accumulation by downregulating phenylpropanoid metabolism at multiple different steps in the pathway. Additionally, we uncovered several interesting insights about phenylpropanoid biosynthesis in *Taxus* species and its relationship with taxane biosynthesis.

First, we found that treatment with piperonylic acid, an irreversible chemical inhibitor of C4H (the second step in phenylpropanoid biosynthesis after PAL) resulted in both a decrease in accumulation of total phenolics and flavonoids and an increase in paclitaxel biosynthesis. This was also correlated with a decrease in PAL activity, indicating that phenylpropanoid biosynthesis and taxane biosynthesis likely compete for metabolic flux despite the fact that they are often co-regulated. While the effect of piperonylic acid on phenolic and flavonoid biosynthesis did not change significantly when CA or phenylalanine was added, there was a clear increase in paclitaxel accumulation when all three compounds were used in tandem. This indicates that while inhibition of C4H using PA plays an important regulatory role in controlling synthesis of phenolics and flavonoids, there are likely other uncharacterized effects on pathway regulation that are more important for paclitaxel production.

The fact that manipulating phenylpropanoid metabolism also affected metabolism of early taxane intermediates can potentially be explained by considering the interplay of phenylpropanoid metabolism with the plant stress response. In nature, production of secondary metabolites such as paclitaxel is often dynamically upregulated in response to stress, and the effect of elicitors on induction of taxane biosynthesis in *Taxus* plant cell culture is an extremely well-studied phenomenon (Yukimune et al., 1996; Zhang et al., 2000; Li et al., 2012; Lenka et al., 2015; Zhang et al., 2015). In particular, it is likely that treatment with PA could increase production of salicylic acid and reactive oxygen species, inducing the stress response and thus upregulating production of taxanes. Salicylic acid is a plant signaling hormone also commonly used as an eliciting agent that is generally accepted to be synthesized through two possible pathways in plants: through the enzyme isochorismate synthase (ICS) or from cinnamic acid synthesized by PAL (Lefevere et al., 2020; Liu et al., 2022). Since treatment with PA increased levels of cinnamic acid in our cell lines (Figure 2B), it is possible that this also upregulated biosynthesis of salicylic acid, which is known to be a positive regulator of taxane biosynthesis (Wang et al., 2007; Sarmadi et al., 2018). This would be consistent with previous work that found that treatment with PA increased markers associated with oxidative stress and oxygen scavenging enzyme activity in chia seedlings, prompting increased production of signaling hormones like SA that are involved in protection from oxidative damage (Nkomo et al., 2022).

In addition, at least one transcription factor has been identified that upregulates taxane biosynthesis specifically in response to salicylic acid (Chen et al., 2021). This salicylic acid-responsive transcription factor was found to bind specifically to the DBAT promoter, resulting in a large increase in DBAT expression. When transiently overexpressed, this transcription factor induced activation of most taxane biosynthetic pathway genes, indicating that in addition to directly upregulating DBAT, it likely upregulates these other pathway genes through indirect mechanisms. This general upregulation of taxane biosynthetic pathway genes would explain the effects of PA on increasing not only paclitaxel production, but also the production of the precursors 10-DAB and baccatin.

Second, while PA and CA are inhibitors of the second and third steps in phenylpropanoid biosynthesis (C4H and 4CL, respectively), they also had off-target effects on both PAL activity and expression. This indicates that at least these initial steps in the pathway are subject to complex transcriptional regulation and likely form a tight regulatory unit, as has been reported in other plants (Koopmann et al., 1999; Zhang et al., 2020). There was also a more complex relationship than expected between PAL activity and expression, suggesting distinct regulation at the transcriptional level and the protein level that will require more in-depth study to fully deconvolute. Interestingly, we also found that while treatment with PA resulted in variable accumulation of cinnamic acid and coumaric acid when combined with other treatments, there was a

very consistent decrease in total flavonoid and phenolic accumulation in these lines. This suggests that perhaps there are some additional downstream effects of C4H inhibition not captured within this study that may be responsible for downregulating production of these end products. In future work, it would be useful to identify the effects of C4H enzyme inhibition on additional phenylpropanoid biosynthetic pathway intermediates occurring later in the pathway in order to identify the mechanism of this clear decrease in total phenolics and flavonoids.

We also demonstrated that repression of PAL using a CRISPR/dCas9-based tool was highly effective at specifically increasing flux to paclitaxel production rather than precursors or side products, illustrating the efficacy of targeted engineering of competing pathways for products of interest. Interestingly, the patterns of metabolite accumulation in these transgenic cell lines (Figure 4) were notably different than those observed with the non-transgenic cell lines in Figure 3, where there was comparatively more 10-DAB and less baccatin in all cell lines. This can potentially be explained by two different mechanisms.

First, the differences in metabolite accumulation between the lines treated with chemical inhibitors and the transgenic cell lines may be a byproduct of transformation. *Agrobacterium*-based transformation has been shown to result in changes to the transcriptome in several plant species and in particular, upregulation of genes associated with hormone signaling and stress (Duan et al., 2017; Wang et al., 2023). This could contribute to the differential taxane accumulation patterns in transformed *Taxus* cell cultures compared to untransformed ones. Furthermore, *Taxus* cell cultures are known to produce phenylpropanoid pathway derivatives in response to stress, so if there is a higher amount of flux through the phenylpropanoid biosynthetic pathway due to the stress associated with transformation, this could explain the large increase in paclitaxel observed in the PALg1 cell line since there is more metabolic flux available to be rerouted to paclitaxel production.

Secondly, in the manuscript we adapted this tool from, the authors noted that there was a slightly increased level of global DNA methylation throughout the genome in their transgenic plants due to non-specific methylation activity of NtDRM (Papikian et al., 2019). This could potentially explain the lower levels of paclitaxel accumulation in the ng cell line compared to untransformed Taxus cell lines (Figure 3C) if some of the later-stage genes in the paclitaxel biosynthetic pathway are non-specifically methylated. It is also important to note that a BLAST search of the guide RNA against the *Taxus chinensis* genome confirmed that this guide RNA had a single target, so it is unlikely that these differences are due to off-target activity of the guide RNA. However, it is difficult to discern which effects are due to hypermethylation and which are simply due to the stress and other metabolic changes associated with transformation; further genomics and transcriptomic work would be needed to deconvolute these effects.

Additionally, despite the fact that many plants have multiple copies of the PAL gene, we found here that methylating presumably only a single copy of the gene was sufficient to drastically increase paclitaxel production by over 25-fold. This could indicate that *Taxus* species only have a single copy of the PAL gene, but that would be highly unusual among plant species, especially for a plant that produces many natural products dependent on phenylalanine as a precursor. At present, the lack of knowledge of the number of copies of PAL in the *Taxus* genome limits in-depth interpretation of these results. As genomic and transcriptomic resources for non-model plants such as *Taxus* species become more abundant, conclusively determining the number of copies of the PAL gene in the genome and multiplexing guide RNAs to target multiple different copies could help further inform how downregulation of PAL affects taxane biosynthesis. This could also be used to more specifically tune the degree of knockdown of PAL to prevent potential deleterious effects on cell wall architecture, due to the importance of PAL in controlling flux to lignin biosynthesis.

More broadly, this work provides a proof-of-concept for the use of epigenetic engineering strategies to manipulate metabolism in plant cell culture and opens the door for more in-depth studies of secondary metabolism. In this work, we revealed some of the intricacies in regulation of the phenylpropanoid biosynthetic pathway, but there are likely many more complex interactions that have yet to be discovered. Using transcriptomics to study *Taxus* cell lines where the phenylpropanoid pathway was inhibited using both genetic and chemical methods could help further elucidate the coregulation of taxane and phenylpropanoid biosynthesis. Additionally, by identifying genes that are coregulated with PAL, we could likely identify more genes involved in the phenylpropanoid biosynthetic pathway in *Taxus* species, enabling further genetic manipulation of the pathway. These transcriptomic studies could be combined with enzyme activity assays validated for different phenylpropanoid pathway enzymes (such as TAL) to understand the contributions of regulation on both the transcriptional and protein levels.

Another natural extension of this work would be to study the efficacy of gene activation through targeted demethylation of genes that are proposed to be downregulated through increased DNA methylation, such as BAPT (Sanchez-Muñoz et al., 2018). This tool could potentially be used to recover secondary metabolite production in continuously subcultured plant cell lines and thus improve their useful lifetime and productivity. However, applying this more broadly in plant cell culture systems would require more extensive epigenomics work in order to identify loci that are both known to be methylated and play important roles in controlling plant metabolism. Ultimately, we envision using these tools as a complement to other CRISPR-based technologies (such as CRISPRi or CRISPRa) and traditional overexpression-based pathway engineering approaches. Notably, this targeted DNA methylation tool required minimal adaptation from its original intended chassis (*Arapidopsis thaliana*) to a non-model plant (*Taxus chinensis*), likely due to the broad host suitability of the promoters used. This shows that other similar tools developed for model plants may be easily transferrable to non-model plants such as *Taxus chinensis*, which could rapidly speed the development of more modern engineering strategies for these chassis.

In conclusion, this work uniquely investigated the effect of manipulating phenylpropanoid metabolism on taxane biosynthesis in *Taxus* plant cell culture using both chemical inhibition and CRISPR-based gene repression. To optimize production of compounds of interest in organisms like plants that produce such a diverse array of natural products, our results demonstrate that it is imperative to manipulate not only the pathway of interest, but also competing or cooperating secondary metabolic pathways. By combining repression of side pathways using the tools developed here with activation of desired pathways through metabolic engineering, we could likely achieve unprecedented levels of control over flux through plant secondary metabolism. Through the application

of these novel tools for manipulating metabolism, we have the potential to rationally optimize medicinal plant cell systems as next-generation chassis for the production of societally valuable compounds.

Scope statement

Plant cell culture biomanufacturing is rapidly becoming an effective strategy for production of high-value plant natural products, such as chemotherapeutics (paclitaxel), vaccine adjuvants (QS-21), and nutraceuticals (cacao, coffee, echinacea). Despite the rapid growth of these technologies, strategies for metabolic engineering of plant cell cultures remain underdeveloped and hyper-focused on engineering the primary metabolic pathway of interest. Here, we show that manipulating phenylpropanoid biosynthesis in *Taxus* species using CRISPR-guided DNA methylation is highly effective at increasing taxane production—increasing paclitaxel accumulation by over 25-fold. While CRISPR-based technologies have been used to engineer well-studied plants, here we describe one of the first applications of CRISPR to a non-model plant cell culture system. Furthermore, this work illustrates that manipulating DNA methylation of key secondary metabolic genes is a useful tool to consider for repression of metabolism along with other more established technologies such as CRISPRi. This demonstrates the power and potential of these tools to manipulate metabolism in plant cell systems for metabolic engineering applications and paves the way for further rational cell line engineering in this rapidly growing field.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

CB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing–original draft, Writing–review and editing. EY: Data curation, Methodology, Resources, Supervision, Writing–review and editing. SR: Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Supervision, Writing–review and editing.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors thank Michelle McKee for her assistance with initial experimental design and preliminary work on the inhibitor experiments with piperonylic acid and caffeic acid.

Conflict of interest

The 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2023.1272811/full#supplementary-material

References

Ainsworth, E. A., and Gillespie, K. M. (2007). Estimation of total phenolic content and other oxidation substrates in plant tissues using Folin–Ciocalteu reagent. *Nat. Protoc.* 2, 875–877. doi:10.1038/nprot.2007.102

Andi, S. A., Gholami, M., Ford, C. M., and Maskani, F. (2019). The effect of light, phenylalanine and methyl jasmonate, alone or in combination, on growth and secondary metabolism in cell suspension cultures of Vitis vinifera. *J. Photochem. Photobiol. B Biol.* 199, 111625. doi:10.1016/j.jphotobiol.2019.111625

Bamneshin, M., Mirjalili, M. H., Naghavi, M. R., Cusido, R. M., and Palazón, J. (2022). Gene expression pattern and taxane biosynthesis in a cell suspension culture of Taxus baccata L. subjected to light and a phenylalanine ammonia lyase (PAL) inhibitor. *J. Photochem. Photobiol. B Biol.* 234, 112532. doi:10.1016/j.jphotobiol.2022.112532

Blount, J. W., Korth, K. L., Masoud, S. A., Rasmussen, S., Lamb, C., and Dixon, R. A. (2000). Altering expression of cinnamic acid 4-hydroxylase in transgenic plants provides evidence for a feedback loop at the entry point into the phenylpropanoid pathway. *Plant Physiol.* 122, 107–116. doi:10.1104/pp.122.1.107

Brincat, M. C., Gibson, D. M., and Shuler, M. L. (2002). Alterations in taxol production in plant cell culture via manipulation of the phenylalanine ammonia lyase pathway. *Biotechnol. Prog.* 18, 1149–1156. doi:10.1021/bp0256115

Bringi, V., Kadkade, P., Prince, C., and Roach, B. (2012). *Phyton holdings*. United States: LLC.

Brzycki, C. M., Young, E. M., and Roberts, S. C. (2021). in *Exploring plant cells for the production of compounds of interest*. Editor S. Malik (Cham: Springer International Publishing), 1–37.

Bubna, G. A., Lima, R. B., Zanardo, D. Y. L., dos Santos, W. D., Ferrarese, M. d. L. L., and Ferrarese-Filho, O. (2011). Exogenous caffeic acid inhibits the growth and enhances the lignification of the roots of soybean (Glycine max). *J. Plant Physiology* 168, 1627–1633. doi:10.1016/j.jplph.2011.03.005

Bulgakov, V. P., Tchernoded, G. K., Veselova, M. V., Fedoreyev, S. A., Muzarok, T. I., and Zhuravlev, Y. N. (2011). Catechin production in cultured cells of Taxus cuspidata and Taxus baccata. *Biotechnol. Lett.* 33, 1879–1883. doi:10.1007/s10529-011-0632-6

Chang, C.-C., Yang, M.-H., Wen, H.-M., and Chern, J.-C. (2002). Estimation of total flavonoid content in propolis by two complementary colometric methods. *J. food drug analysis* 10. doi:10.38212/2224-6614.2748

Chen, H. C., Song, J., Williams, C. M., Shuford, C. M., Liu, J., Wang, J. P., et al. (2013). Monolignol pathway 4-coumaric acid:coenzyme A ligases in Populus trichocarpa: novel specificity, metabolic regulation, and simulation of coenzyme A ligation fluxes. *Plant Physiol.* 161, 1501–1516. doi:10.1104/pp.112.210971

Chen, L., Wu, L., Yang, L., Yu, H., Huang, P., Wang, Y., et al. (2022). TcJAV3-TcWRKY26 cascade is a missing link in the jasmonate-activated expression of taxol biosynthesis gene DBAT in Taxus chinensis. *Int. J. Mol. Sci.* 23, 13194. doi:10.3390/ijms232113194

Chen, Y., Zhang, H., Zhang, M., Zhang, W., Ou, Z., Peng, Z., et al. (2021). Salicylic acid-responsive factor TcWRKY33 positively regulates taxol biosynthesis in Taxus chinensis in direct and indirect ways. *Front. Plant Sci.* 12, 697476. doi:10.3389/fpls.2021.697476

Croteau, R., Ketchum, R. E., Long, R. M., Kaspera, R., and Wildung, M. R. (2006). Taxol biosynthesis and molecular genetics. *Phytochem. Rev.* 5, 75–97. doi:10.1007/s11101-005-3748-2

Dong, Y.-s., Fu, C., Su, P., Xu, X., Yuan, J., Wang, S., et al. (2016). Mechanisms and effective control of physiological browning phenomena in plant cell cultures. *Physiol. Plant.* 156, 13–28. doi:10.1111/ppl.12382

Duan, K., Willig, C. J., De Tar, J. R., Spollen, W. G., and Zhang, Z. J. (2017). Transcriptomic analysis of Arabidopsis seedlings in response to an agrobacterium-mediated transformation process. *Mol. Plant-Microbe Interactions®* 31, 445–459. doi:10.1094/mpmi-10-17-0249-r

Feduraev, P., Skrypnik, L., Riabova, A., Pungin, A., Tokupova, E., Maslennikov, P., et al. (2020). Phenylalanine and tyrosine as exogenous precursors of Wheat (*Triticum aestivum* L.) secondary metabolism through PAL-associated pathways. *Plants* 9, 476. doi:10.3390/plants9040476

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811. doi:10.1038/35888

Fu, C., Li, L., Wu, W., Li, M., Yu, X., and Yu, L. (2012). Assessment of genetic and epigenetic variation during long-term Taxus cell culture. *Plant Cell. Rep.* 31, 1321–1331. doi:10.1007/s00299-012-1251-y

Gallego-Bartolomé, J., Gardiner, J., Liu, W., Papikian, A., Ghoshal, B., Kuo, H. Y., et al. (2018). Targeted DNA demethylation of the Arabidopsis genome using the human TET1 catalytic domain. *Proc. Natl. Acad. Sci. U. S. A.* 115, E2125–e2134. doi:10.1073/pnas.1716945115

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., et al. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 33, 510–517. doi:10.1038/nbt.3199

Kim, D. S., and Hwang, B. K. (2014). An important role of the pepper phenylalanine ammonia-lyase gene (PAL1) in salicylic acid-dependent signalling of the defence response to microbial pathogens. *J. Exp. Bot.* 65, 2295–2306. doi:10.1093/jxb/eru109

Kolewe, M. E., Henson, M. A., and Roberts, S. C. (2011). Analysis of aggregate size as a process variable affecting paclitaxel accumulation in Taxus suspension cultures. *Biotechnol. Prog.* 27, 1365–1372. doi:10.1002/btpr.655

Koopmann, E., Logemann, E., and Hahlbrock, K. (1999). Regulation and functional expression of cinnamate 4-hydroxylase from parsley. *Plant Physiol.* 119, 49–56. doi:10.1104/pp.119.1.49

Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* 8, 2180–2196. doi:10.1038/nprot.2013.132

Lefevere, H., Bauters, L., and Gheysen, G. (2020). Salicylic acid biosynthesis in plants. *Front. Plant Sci.* 11, 338. doi:10.3389/fpls.2020.00338

Lenka, S. K., Nims, N. E., Vongpaseuth, K., Boshar, R. A., Roberts, S. C., and Walker, E. L. (2015). Jasmonate-responsive expression of paclitaxel biosynthesis genes in Taxus cuspidata cultured cells is negatively regulated by the bHLH transcription factors TcJAMYC1, TcJAMYC2, and TcJAMYC4. *Front. Plant Sci.* 6, 115. doi:10.3389/fpls.2015.00115

Li, B.-J., Wang, H., Gong, T., Chen, J. J., Chen, T. J., Yang, J. L., et al. (2017). Improving 10-deacetylbaccatin III-10-β-O-acetyltransferase catalytic fitness for Taxol production. *Nat. Commun.* 8, 15544. doi:10.1038/ncomms15544

Li, S. T., Zhang, P., Zhang, M., Fu, C. h., Zhao, C. f., Dong, Y. s., et al. (2012). Transcriptional profile of Taxus chinensis cells in response to methyl jasmonate. *BMC Genomics* 13, 295. doi:10.1186/1471-2164-13-295

Li, S., Zhang, P., Zhang, M., Fu, C., and Yu, L. (2013). Functional analysis of a WRKY transcription factor involved in transcriptional activation of the DBAT gene in Taxus chinensis. *Plant Biol. (Stuttg)* 15, 19–26. doi:10.1111/j.1438-8677.2012.00611.x

Li, Y., Kim, J. I., Pysh, L., and Chapple, C. (2015). Four isoforms of Arabidopsis 4-coumarate:CoA ligase have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant Physiol.* 169, 2409–2421. doi:10.1104/pp.15.00838

Liu, A., Zhu, Y., Wang, Y., Wang, T., Zhao, S., Feng, K., et al. (2023). Molecular identification of phenylalanine ammonia lyase-encoding genes EfPALs and EfPAL2-interacting transcription factors in Euryale ferox. *Front. Plant Sci.* 14, 1114345. doi:10.3389/fpls.2023.1114345

Liu, J., Qiu, G., Liu, C., Li, H., Chen, X., Fu, Q., et al. (2022). Salicylic acid, a multifaceted hormone, combats abiotic stresses in plants. *Life (Basel)* 12, 886. doi:10.3390/life12060886

Liu, W., Feng, Y., Yu, S., Fan, Z., Li, X., Li, J., et al. (2021). The flavonoid biosynthesis network in plants. *Int. J. Mol. Sci.* 22, 12824. doi:10.3390/ijms222312824

McKee, M. C., Wilson, S. A., and Roberts, S. C. (2021). The interface amongst conserved and specialized pathways in non-paclitaxel and paclitaxel accumulating Taxus cultures. *Metabolites* 11, 688. doi:10.3390/metabo11100688

Naill, M. C., and Roberts, S. C. (2004). Preparation of single cells from aggregated Taxus suspension cultures for population analysis. *Biotechnol. Bioeng.* 86, 817–826. doi:10.1002/bit.20083

Nkomo, M., Gokul, A., Ndimba, R., Badiwe, M., Keyster, M., and Klein, A. (2022). Piperonylic acid alters growth, mineral content accumulation and reactive oxygen species-scavenging capacity in chia seedlings. *AoB PLANTS* 14, plac025. doi:10.1093/aobpla/plac025

Pant, S., and Huang, Y. (2022). Genome-wide studies of PAL genes in sorghum and their responses to aphid infestation. *Sci. Rep.* 12, 22537. doi:10.1038/s41598-022-25214-1

Pant, S. R., Irigoyen, S., Liu, J., Bedre, R., Christensen, S. A., Schmelz, E. A., et al. (2021). Brachypodium phenylalanine ammonia lyase (PAL) promotes antiviral defenses against Panicum mosaic virus and its satellites. *mBio* 12, e03518-20. doi:10.1128/mbio.03518-20

Papikian, A., Liu, W., Gallego-Bartolomé, J., and Jacobsen, S. E. (2019). Site-specific manipulation of Arabidopsis loci using CRISPR-Cas9 SunTag systems. *Nat. Commun.* 10, 729. doi:10.1038/s41467-019-08736-7

Perez-Matas, E., Hidalgo-Martinez, D., Escrich, A., Alcalde, M. A., Moyano, E., Bonfill, M., et al. (2023). Genetic approaches in improving biotechnological production of taxanes: an update. *Front. Plant Sci.* 14, 1100228. doi:10.3389/fpls.2023.1100228

Sanchez-Muñoz, R., Bonfill, M., Cusidó, R. M., Palazón, J., and Moyano, E. (2018). Advances in the regulation of *in vitro* paclitaxel production: methylation of a Y-patch promoter region alters BAPT gene expression in Taxus cell cultures. *Plant Cell. Physiology* 59, 2255–2267. doi:10.1093/pcp/pcy149

Sanchez-Muñoz, R., Perez-Mata, E., Almagro, L., Cusido, R. M., Bonfill, M., Palazon, J., et al. (2020). A novel hydroxylation step in the taxane biosynthetic pathway: a new approach to paclitaxel production by synthetic biology. *Front. Bioeng. Biotechnol.* 8, 410. doi:10.3389/fbioe.2020.00410

Sarmadi, M., Karimi, N., Palazón, J., Ghassempour, A., and Mirjalili, M. H. (2018). The effects of salicylic acid and glucose on biochemical traits and taxane production in a Taxus baccata callus culture. *Plant Physiology Biochem.* 132, 271–280. doi:10.1016/j.plaphy.2018.09.013

Schalk, M., Cabello-Hurtado, F., Pierrel, M. A., Atanassova, R., Saindrenan, P., and Werck-Reichhart, D. (1998). Piperonylic acid, a selective, mechanism-based inactivator of the trans-cinnamate 4-hydroxylase: a new tool to control the flux of metabolites in the phenylpropanoid pathway. *Plant Physiol.* 118, 209–218. doi:10.1104/pp.118.1.209

Su, H., Jiang, H., and Li, Y. (2015). Effects of PAL and ICS on the production of total flavonoids, daidzein and puerarin in Pueraria thomsonii Benth. suspension cultures under low light stress. *J. Plant Biochem. Biotechnol.* 24, 34–41. doi:10.1007/s13562-013-0233-7

Sykłowska-Baranek, K., Pietrosiuk, A., Naliwajski, M. R., Kawiak, A., Jeziorek, M., Wyderska, S., et al. (2012). Effect of l-phenylalanine on PAL activity and production of naphthoquinone pigments in suspension cultures of Arnebia euchroma (Royle) Johnst. *Vitro Cell. Dev. Biol. Plant* 48, 555–564. doi:10.1007/s11627-012-9443-2

Tyunin, A. P., Kiselev, K. V., and Zhuravlev, Y. N. (2012). Effects of 5-azacytidine induced DNA demethylation on methyltransferase gene expression and resveratrol production in cell cultures of Vitis amurensis. *Plant Cell., Tissue Organ Cult. (PCTOC)* 111, 91–100. doi:10.1007/s11240-012-0175-0

Wan Mazlina Md, S., Rasdin, R., Nur Syamimi Md, L., Fatimah, S., and Salim, F. (2019). Determination and quantification of p-coumaric acid in pineapples (Ananas comosus) extracts using gradient mode RP-HPLC. *Pharmacogn. Res.* 11, 67. doi:10.4103/pr.pr_154_18

Wang, W., Guo, J., Ma, J., Wang, Z., Zhang, L., Wang, Z., et al. (2023). Comprehensive transcriptomic and metabolic profiling of agrobacterium-tumefaciens-infected immature Wheat embryos. *Int. J. Mol. Sci.* 24, 8449. doi:10.3390/ijms24098449

Wang, Y.-D., Wu, J.-C., and Yuan, Y.-J. (2007). Salicylic acid-induced taxol production and isopentenyl pyrophosphate biosynthesis in suspension cultures of Taxus chinensis var. mairei. *Cell. Biol. Int.* 31, 1179–1183. doi:10.1016/j.cellbi.2007.03.038

Wilson, S. A., Keen, P., McKee, M. C., Raia, N., Van Eck, J., and Roberts, S. C. (2018). Development of an Agrobacterium-mediated transformation method for Taxus suspension cultures. *Vitro Cell. Dev. Biol. - Plant* 54, 36–44. doi:10.1007/s11627-017-9876-8

Wilson, S. A., and Roberts, S. C. (2012). Recent advances towards development and commercialization of plant cell culture processes for the synthesis of biomolecules. *Plant Biotechnol. J.* 10, 249–268. doi:10.1111/j.1467-7652.2011.00664.x

Yoo, H., Widhalm, J. R., Qian, Y., Maeda, H., Cooper, B. R., Jannasch, A. S., et al. (2013). An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. *Nat. Commun.* 4, 2833. doi:10.1038/ncomms3833

Yukimune, Y., Tabata, H., Higashi, Y., and Hara, Y. (1996). Methyl jasmonate-induced overproduction of paclitaxel and baccatin III in Taxus cell suspension cultures. *Nat. Biotechnol.* 14, 1129–1132. doi:10.1038/nbt0996-1129

Zhang, C. H., Mei, X. G., Liu, L., and Yu, L. J. (2000). Enhanced paclitaxel production induced by the combination of elicitors in cell suspension cultures of Taxus chinensis. *Biotechnol. Lett.* 22, 1561–1564. doi:10.1023/a:1005684901329

Zhang, J., Hansen, L. G., Gudich, O., Viehrig, K., Lassen, L. M. M., Schrübbers, L., et al. (2022). A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature* 609, 341–347. doi:10.1038/s41586-022-05157-3

Zhang, J., Tuskan, G. A., Tschaplinski, T. J., Muchero, W., and Chen, J. G. (2020). Transcriptional and post-transcriptional regulation of lignin biosynthesis pathway genes in populus. *Front. Plant Sci.* 11, 652. doi:10.3389/fpls.2020.00652

Zhang, M., Jin, X., Chen, Y., Wei, M., Liao, W., Zhao, S., et al. (2018). TcMYC2a, a basic helix-loop-helix transcription factor, transduces JA-signals and regulates taxol biosynthesis in Taxus chinensis. *Front. Plant Sci.* 9, 863. doi:10.3389/fpls.2018.00863

Zhang, M., Li, S., Nie, L., Chen, Q., Xu, X., Yu, L., et al. (2015). Two jasmonate-responsive factors, TcERF12 and TcERF15, respectively act as repressor and activator of tasy gene of taxol biosynthesis in Taxus chinensis. *Plant Mol. Biol.* 89, 463–473. doi:10.1007/s11103-015-0382-2

Zhang, P., Li, S. T., Liu, T. T., Fu, C. H., Zhou, P. P., Zhao, C. F., et al. (2011). Overexpression of a 10-deacetylbaccatin III-10 β-O-acetyltransferase gene leads to increased taxol yield in cells of Taxus chinensis. *Plant Cell., Tissue Organ Cult. (PCTOC)* 106, 63–70. doi:10.1007/s11240-010-9894-2

Zhou, T., Luo, X., Yu, C., Zhang, C., Zhang, L., Song, Y. b., et al. (2019). Transcriptome analyses provide insights into the expression pattern and sequence similarity of several taxol biosynthesis-related genes in three Taxus species. *BMC Plant Biol.* 19, 33. doi:10.1186/s12870-019-1645-x

Zhu, X., Liu, X., Liu, T., Wang, Y., Ahmed, N., Li, Z., et al. (2021). Synthetic biology of plant natural products: from pathway elucidation to engineered biosynthesis in plant cells. *Plant Commun.* 2, 100229. doi:10.1016/j.xplc.2021.100229

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Review
PUBLISHED 26 October 2023
DOI 10.3389/fbioe.2023.1267826

OPEN ACCESS

EDITED BY
Kuan Shiong Khoo,
Yuan Ze University, Taiwan

REVIEWED BY
Jianhua Fan,
East China University of Science and Technology, China
Lim Wei,
University of Technology Petronas, Malaysia
Cheah Wai Yan,
Universiti Kebangsaan Malaysia, Malaysia

*CORRESPONDENCE
Bhaskar Bhadra,
bhaskar.bhadra@ril.com

RECEIVED 27 July 2023
ACCEPTED 11 October 2023
PUBLISHED 26 October 2023

CITATION
Dhokane D, Shaikh A, Yadav A, Giri N, Bandyopadhyay A, Dasgupta S and Bhadra B (2023), CRISPR-based bioengineering in microalgae for production of industrially important biomolecules.
Front. Bioeng. Biotechnol. 11:1267826.
doi: 10.3389/fbioe.2023.1267826

CRISPR-based bioengineering in microalgae for production of industrially important biomolecules

Dhananjay Dhokane, Arshi Shaikh, Anu Yadav, Nandinee Giri, Anindya Bandyopadhyay, Santanu Dasgupta and Bhaskar Bhadra*

Synthetic Biology Group, Reliance Industries Ltd., Navi Mumbai, India

Microalgae, as photosynthetic organisms, have the potential to produce biomolecules for use in food, feed, cosmetics, nutraceuticals, fuel, and other applications. Faster growth rates and higher protein and lipid content make microalgae a popular chassis for many industrial applications. However, challenges such as low productivity and high production costs have limited their commercialization. To overcome these challenges, bioengineering approaches such as genetic engineering, metabolic engineering, and synthetic biology have been employed to improve the productivity and quality of microalgae-based products. Genetic engineering employing genome editing tools like CRISPR/Cas allows precise and targeted genetic modifications. CRISPR/Cas systems are presently used to modify the genetic makeup of microalgae for enhanced production of specific biomolecules. However, these tools are yet to be explored explicitly in microalgae owing to some limitations. Despite the progress made in CRISPR-based bioengineering approaches, there is still a need for further research to optimize the production of microalgae-based products. This includes improving the efficiency of genome editing tools, understanding the regulatory mechanisms of microalgal metabolism, and optimizing growth conditions and cultivation strategies. Additionally, addressing the ethical, social, and environmental concerns associated with genetic modification of microalgae is crucial for the responsible development and commercialization of microalgae-based products. This review summarizes the advancements of CRISPR-based bioengineering for production of industrially important biomolecules and provides key considerations to use CRISPR/Cas systems in microalgae. The review will help researchers to understand the progress and to initiate genome editing experiments in microalgae.

KEYWORDS

bioactives, biomolecules, bioengineering, CRISPR/Cas9, microalgae, synthetic biology

Introduction

Microalgae as photosynthetic cell factories are gaining attraction for several industrial applications due to their rapid growth rate, higher photosynthetic efficiency, ability to produce higher biomass, and complex metabolites (Kumar et al., 2020). Owing to the taxonomic and inherent biochemical diversity among microalgal species, production of bioactive molecules for applications in pharmaceuticals, nutraceuticals, personal care, dietary supplements, biofuels, medicine, food, feed, etc. are being continuously

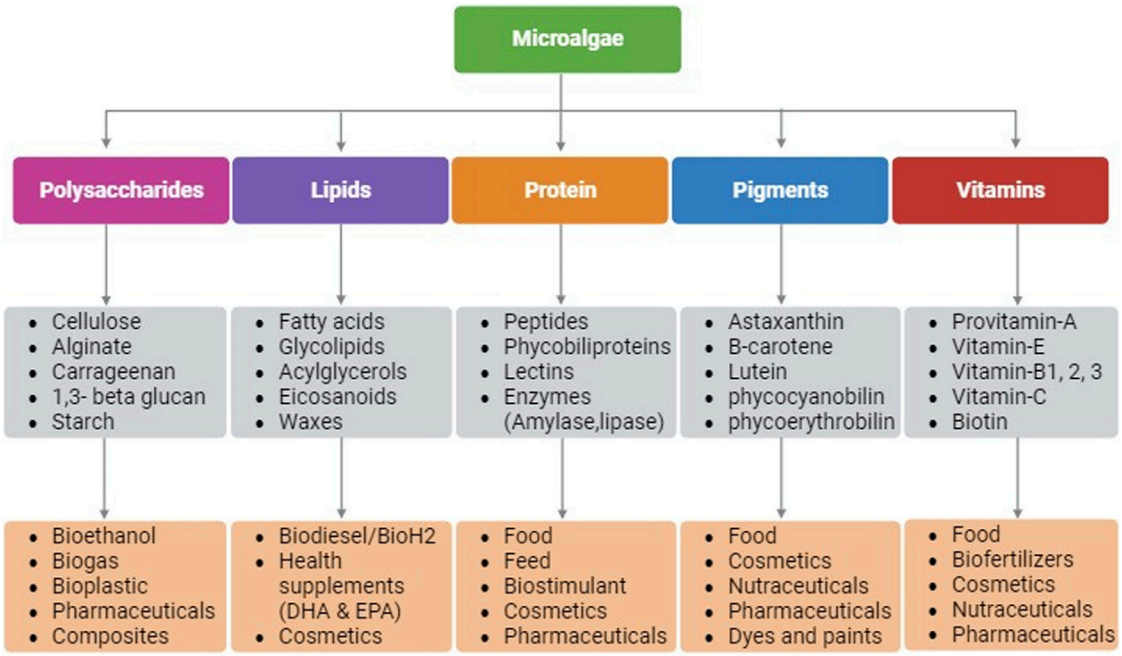


FIGURE 1
Schematic representation of biomolecules that are produced in microalgae and their applications. Created with BioRender.com.

investigated (Kumar et al., 2020). Microalgae synthesize compounds such as proteins, lipids, polysaccharides, pigments, and vitamins that have diverse applications (Figure 1) (Dolganyuk et al., 2020). These compounds are known to possess anticancer, anti-inflammatory, antimicrobial, and antioxidant properties making microalgae a suitable host for pharmaceutical and nutraceutical applications. In addition, microalgae are exploited for biotechnological applications such as, for production of recombinant proteins/peptides, monoclonal antibodies, and vaccines (Khavari et al., 2021). Moreover, their lipid accumulating ability (20%–70% cell dry weight) make microalgae a promising candidate for biofuels and/or nutraceuticals depending upon their fatty acid composition (Kumar et al., 2020).

The increase in emissions of greenhouse gases (GHGs) in the environment due to extensive use of fossil fuels for energy has resulted in climate change, worldwide. Carbon dioxide (CO_2) contributes to approximately 60% of total GHGs emissions globally (Solaymani, 2019). The levels of CO_2 in the environment are continuously increasing and are impacting every form of life on the planet earth. A report from National Oceanic and Atmospheric Administration (NOAA) global monitoring lab shows that in the year 2022, the global average CO_2 level was 417.06 ppm, hitting a new record. Increasing CO_2 levels in the atmosphere is raising the temperature every year. NOAA also observed that in 2021, CO_2 alone was responsible for about two-third of the total heating influence of all human-produced greenhouse gases (https://gml.noaa.gov/ccgg/trends/). To mitigate the adverse effects of CO_2 emissions, lot of efforts are put forth globally to either reduce emissions or to sequester CO_2. The United Nations has mandated achieving carbon neutrality by 2050. Carbon neutrality is balancing carbon emissions with carbon absorption to achieve a net zero carbon emissions in atmosphere. Carbon sequestration is the key process in achieving carbon neutrality. Globally, numerous efforts are concentrated on developing various carbon capture technologies; however, with little success due to complex economics (Sadvakasova et al., 2023). Therefore, biological sequestration of carbon would more likely be the method of choice to fix atmospheric CO_2 sustainably.

Microalgae are phototrophic organisms that possess carbon accumulation and metabolic capabilities for CO_2 sequestration (Sadvakasova et al., 2023). Microalgae can fix CO_2 and convert solar energy into biomass at higher efficiencies (8%–10%) than land plants (Benedetti et al., 2018). They can serve as means of natural carbon sink and help in the reduction of CO_2 burden globally. It has been reported that microalgae photosynthetically capture approximately 100 gigatons of CO_2 per annum (Peters et al., 2017). This clearly demonstrates that microalgae are invaluable resources for sequestering atmospheric CO_2 and generating biomass at industrial scale. To produce 1 kg of microalgal biomass, approximately 1.83 kg of CO_2 is captured and fixed (Sarwer et al., 2022). The microalgal biomass acts as an excellent sustainable feedstock for renewable energy and valuable products for diverse industrial applications. The microalgal biomass consist of high value biomolecules that are in high market demand for applications in nutraceuticals, cosmetics, pharmaceuticals, food, and feed.

Despite the numerous industrial applications of microalgae, there are many challenges associated with exploiting microalgae to their fullest potential. The major challenge is to improve the microalgal productivity and product titers to make the production processes commercially viable. However, altering only the growth conditions does not guarantee improved productivities and titers of

targeted products (Sreenikethanam et al., 2022) Genetic engineering has potential to alter the genetic constitution of a strain, thereby generating strains with higher growth rates, better photosynthetic efficiencies, and higher product titers. Genetic engineering using genome editing tools like CRISPR/Cas facilitates precise and targeted genetic manipulations. CRISPR/Cas systems are employed in microalgae for targeted gene knockouts and knockins, multiplex gene targeting, and modulation of gene expression, to produce industrially relevant strains (Jeong et al., 2023). CRISPR interventions in microalgae has generated strains with improved lipid and pigment contents, triacylglycerol productivity, lipid accumulation, efficient CO_2 sequestration capabilities (Asadian et al., 2022; Lee et al., 2022; Song et al., 2022). Likely, CRISPR/Cas systems will facilitate rapid development of strains with higher lipid productivities for biodiesel production, higher pigment contents for applications in food, cosmetics and nutraceuticals, biomolecules of interest and high performing strains with industrially important traits. This review provides a holistic understanding of the progress that has been made in algal genome editing, how CRISPR/Ca9 systems has been utilized, for generating high performing algal strains for various industrial applications.

Microalgae as a rich source of bioactives

Microalgae are excellent sources of diverse pigments, including carotenoids, xanthophylls, and phycobiliproteins (Saide et al., 2021). Phycobiliproteins are accessory light-harvesting pigment complexes of the algal photosynthetic machinery. They are made up of apoproteins (α, β, and γ subunits) and chromophores (phycocyanobilin, phycoerythrobilin, phycourobilin, and phycoviolobilin) linked by thioether bond and are found in red microalgae (e.g., *Porphyridium* spp. and *A. platensis*) (Ji et al., 2023). These are brightly colored water-soluble proteins and have numerous applications in food, cosmetic and biomedical industry. These proteins are known to act as antioxidants, boost the immune system, possibly lower the risk of heart diseases, prevent cancers, and protects against age-related diseases such as multiple sclerosis, and cataract (Hamouda and El-Naggar, 2021).

Fucoxanthin, a xanthophyll has shown potential in food, feed, health, and cosmetics applications. It exerts an antiobesity activity by modulating the increase of reactive oxygen species (ROS) and by down regulating lipid metabolism genes (Saide et al., 2021). Fucoxanthin has shown to reduce plasmatic and hepatic triglyceride concentrations and positively influences cholesterol-regulating enzymes, such as 3-hydroxy-3-methylglutaryl coenzyme A reductase and acyl-coenzyme A (D'Orazio et al., 2012). Microalgal-derived β-carotene is used as "natural" food additive with *Dunaliella salina* containing the highest amount and is used commercially in food, feed, supplements, and cosmetics applications. β-carotene exhibits various therapeutic potential, such as reduced risk of disease via modulating cell signaling pathways, antioxidant activity, nutritional value as provitamin A, peroxyl radicals scavenging activity, skin protective effects, and restoring hepatic enzymes (e.g., catalase, peroxidase, and superoxide dismutase) activity to protect vital organs against xenobiotic and other damages (Saide et al., 2021; Ampofo and Abbey, 2022). Astaxanthin exhibits great antioxidant activity and is commercially produced from microalgae *Haematococcus* spp. for use as colorant in feed industry, dietary supplement, in the prevention of some human pathologies, such as skin UV-mediated photo-oxidation, inflammatory processes, diabetes, and even cancer. Microalgal lutein protects cells from ROS damage under stress conditions and has potential role in preventing or ameliorating age-related macular degeneration, prevention of certain cancers, and for the protection of skin from UV-induced damage. Lutein has been extensively used as a feed additive and food coloration agent in industry (Saide et al., 2021; Ampofo and Abbey, 2022).

Microalgae are naturally abundant in phenolic compounds that are involved in various physiological processes. These compounds range from simple aromatic rings to more complex molecules and comprising flavonoids, phenolic acids, tannins, lignans or coumarins (Del Mondo et al., 2022). Polyphenols from microalgae exhibit extensive beneficial biological properties, including antioxidant, anticancer, antimicrobial, anti-inflammatory, antidiabetic, antiviral, and cardioprotective activities. *Tetraselmis suecica*, *Isochrysis* spp., *Chlorella vulgaris*, and *Phaeodactylum tricornutum* are the industrially relevant microalgae having highest antioxidant capacities and can be explored as potential new sources of natural antioxidants (Li et al., 2007; Hajimahmoodi et al., 2010; Saide et al., 2021).

Algal polysaccharides are mainly explored because of their bioactive properties for pharmaceutical, nutraceutical and biomedical applications (Patel et al., 2022). These polysaccharides are used as stabilizers and emulsifiers in food, feed, and pharmaceuticals. The microalgal polysaccharides are known to possess antiviral, antibacterial, antioxidant, anti-inflammatory, and immunomodulatory activities (de Jesus Raposo et al., 2014). In pharmaceutical and biomedical industry, polysaccharides like alginates, fucoidans, ulvans, carrageenans, and chitin are continuously explored. Polysaccharides from microalgae offer numerous advantages over synthetic polymers such as, safety, stability, hydrophilicity, biocompatibility, biodegradability, chemical modifiability, and biocompatibility. These properties enable them to be used for wide range of applications like, preparation of pharmaceutical materials, drug release agents, and plasma substitutes (Patel et al., 2022).

Microalgae are the factories for valuable lipids with approximately 25% of their dry weight, for example, fatty acids, polar lipids, oxylipins, and steroids, with promising applications in nutraceutical, pharmaceutical, cosmeceutical, and biofuel sectors. Microalgal lipids can be used in the prevention and treatment of several human pathologies, including anticancer, antioxidant, and anti-inflammatory activities, as well as treatment of diabetes. *Chlorella* spp. is found to have the most lipid content and is exploited commercially (Saide et al., 2021). Microalgae forms the best resource for production of third and fourth generation of biofuels with the ability to produce biodiesel, bioethanol, biohydrogen, etc. (Saide et al., 2021; Ampofo and Abbey, 2022). Higher lipid accumulation in their biomass along with great environmental adaptability for growth has a major advantage over other feedstock for microalgal biofuel production (Ampofo and Abbey, 2022). The major classes of microalgae suitable for

biofuel production include Bacillariophyceae, Chlorophyceae, Eustigmatophyte, Chrysophyceae, Haptophyceae (Prymnesiophyceae), and Cyanophyceas, of which Chlorophyceae group is the most favorable for biodiesel production (Saide et al., 2021). Generally, most microalgal species have approximately 30% lipid content of their dry biomass with some species having even more lipid content, such as *Nannochloris* spp. (56%), *Chlorella* spp. (53%), and *Neochloris oleoabundans* (65%). However, growth rate decreases with higher oil-producing strains (Saide et al., 2021; Ampofo and Abbey, 2022). This lipid extracted from microalgae is then used for biofuel production. Additionally, neutral lipids or triacylglycerols are the major stored forms of lipids which can be esterified with C16 and C18 profiles and are proven to be ideal for biofuel production (Ampofo and Abbey, 2022).

Microalgae can produce phytosterols that have been used as additives in food products and have gained attention due to their reduced cholesterol concentration in blood, thereby preventing cardiovascular disorders (Saide et al., 2021; Ampofo and Abbey, 2022). Some species, such as *Isochrysis galbana*, *Nannochloropsis* spp., and *Phaeodactylum tricornutum*, have phytosterol content ranging from 7 to 34 g/kg (Ryckebosch et al., 2014); *Pavlova lutherie*, *Tetraselmis* spp. M8 and *Nannochloropsis* spp. BR2 may have phytosterol ranging 0.4%–2.6% dry weight, while 5.1% dry weight of phytosterol could be achieved for *P. lutherie* (Ahmed et al., 2015).

Microalgae are rich source of sustainable proteins, consisting of upto 70% of total protein content based on dry weight basis. *Arthrospira* spp., *Chlorella* spp., *Aphanizomenon* spp., and *Nostoc* spp., are known to have very high protein content. Microalgal proteins contain all the essential amino acids and possess balanced total amino acid profiles (Lucakova et al., 2022). Microalgal proteins are known to have unique physicochemical and technofunctional properties that can withstand diverse and harsh environmental conditions. Therefore, microalgal proteins are continually explored for new food and feed formulations (Acquah et al., 2021). Microalgal proteins and peptides have different bioactivities such as, antioxidant, anticancer, antihypertensive, antiatherosclerotic, anti-UV radiation, and antiosteoporosis (Saide et al., 2021; Ampofo and Abbey, 2022). Dermochlorella, an oligopeptide purified from the microalgae *Chlorella vulgaris*, has been widely used for skin treatments in the biomedical field (Martins et al., 2014).

CRISPR/Cas as a genome engineering tool

CRISPR/Cas is presently emerging as one of the most promising genome editing tools for various organisms. The CRISPR locus was first discovered in the genome of *E. coli* in 1987 (Ishino et al., 2018) as a form of adaptive immunity against invading foreign DNA sequences of various viruses. It consists of an array of direct repeated sequences interspersed with short spacer sequences. The short spacer sequences are transcribed and processed into crRNAs (or guide RNAs) that in turn bind to effector nucleases (Cas proteins). This Cas ribonucleoprotein complex is then directed to the target DNA sequence (i.e., the sequence complementary to guide RNA) and owing to the nuclease activity of Cas protein, the target DNA is cleaved into shorter DNA fragments (Jeon et al., 2017; Ishino et al., 2018).

The ability of Cas nuclease to initiate double stranded breaks (DSBs) in the genomic DNA led to its usage as a genome editing tool in eukaryotic cells in 2013 (Cong et al., 2013; Mali et al., 2013). The DSBs created by Cas proteins are subsequently repaired using nonhomology mediated end joining machinery of the host cells, thereby creating indel (insertion/deletion) mutations at the target site. These indel mutations often result in creating knockout mutants of the target gene.

Strikingly, Cas nucleases are not the first endonucleases to be used as a genome editing tool. Zinc finger endonucleases as well as TALEN endonuclease have also been used to create targeted knockout mutants (Bibikova et al., 2002; Gaj et al., 2013; Zhang et al., 2013). However, the considerable surge in the popularity of the CRISPR/Cas system as a genome editing tool can be attributed to the fact that guide RNAs provide high degree of accuracy and flexibility to target any part of host genomic DNA sequence as per the requirement. In addition, with each passing day, different Cas proteins with varying degree of endonuclease activity are being discovered in various organisms (Kim et al., 2017a; Kim et al., 2017b). As a result, Cas endonucleases having better accuracy (with less off-target endonuclease activity) can be used for enhanced "targeted genome editing." Till date, Cas9 isolated from *Streptococcus pyogenes* is the most used nuclease for CRISPR/Cas mediated genome editing (Ishino et al., 2018). CRISPR/Cas tool is used to create targeted knockout mutants, integrations of gene(s) at desired location within the genomes, replacement of nonfunctional alleles, transcriptional regulations, and many more applications. Although CRISPR-based tools open a plethora of opportunities to selectively edit the genomes of various species, the technique in microalgae as compared to other organisms is still in its infancy.

CRISPR-based multiplexed gene knockout and knockin has been explored to manipulate target genes and CRISPR interference or activation (CRISPRi/a) has enabled the modulation of complex metabolic pathways and regulatory networks. Here we have listed the CRISPR-based bioengineering efforts applied in several algal species for the production of biomolecules of industrial importance (Table 1).

Chlamydomonas reinhardtii

Chlamydomonas reinhardtii is an ideal model organism for bioengineering due to its well-characterized genome, ease of cultivation, and the applicability of extensive genetic tools for its modification. In 2014, Jiang et al. (2014) reported the first successful application of CRISPR/Cas9 in *C. reinhardtii*, demonstrating the feasibility of targeted gene editing in this organism. Since then, numerous studies have been conducted to further optimize and improve the CRISPR/Cas9 based editing in *C. reinhardtii* (Shin et al., 2016; Baek et al., 2018; Dhokane et al., 2020). Owing to the toxicity of constitutively expressed Cas9 protein and off-targeted mutation(s) associated with vector driven expression of Cas9, Shin et al. (2016) improved the editing efficiency by delivering gRNA-Cas9 ribonucleoproteins (RNPs) by 100-fold. We have provided the workflow for editing of target genes in microalgae using gRNA-Cas9 RNP complex (Figure 2).

TABLE 1 Examples of CRISPR-based bioengineering in different microalgal species to produce industrially important biomolecules.

Strain	Target gene modified	Type of modification	Type of nuclease	Outcomes	References
Chlamydomonas reinhardtii					
CC-4349	*CpFTSY*, zeaxanthin epoxidase (*ZEP*)	Knockout	Cas9	Mutant strain capable of constitutively producing 13-fold zeaxanthin with improved photosynthetic productivity, greater biomass accumulation observed under high light growth conditions	Baek et al. (2016)
CC-400	phosphoenolpyruvate carboxylase isoform 1 (*CrPEPC1*)	CRISPRi	dCas9	*CrPEPC1* downregulated strains showed increase in lipid content by 28.5% dry cell weight basis compared to WT	Kao and Ng (2017)
CC-4349	zeaxanthin epoxidase (*ZEP*)	Knockout	Cas9	Mutant strain exhibited significantly higher zeaxanthin content (56-fold) and productivity (47-fold) than the WT without the reduction in lutein level	Baek et al. (2018)
CC-4349	phospholipase A2 (*PLA2*)	Knockout	Cas9	Lipid productivities of phospholipase A2 knockout mutants increased by approximately 64.25%	Shin et al. (2019)
CC-4349	*ELT*	Knockout	Cas9	Higher lipid accumulation in mutant strain compared to WT and noticeable shift in fatty acid composition with an increase of approximately 27.2% in the C18:1 proportion making it suitable for biofuel production	Nguyen et al. (2020)
CC-4349	Zeaxanthin epoxidase (*ZEP*), lycopene epsilon cyclase (*LCYE*)	Knockout	Cas9	Zeaxanthin yield of double knockout mutant (*ZEP* and *LCYE*) strains were 60% higher. Zeaxanthin yield reported was (5.24 mg/L) compared to single ZEP mutant	Song et al. (2020)
CC-4349	ADP-glucose pyrophosphorylase (*AGP*), zeaxanthin epoxidase (*ZEP*)	Knockout	Cas9	Double knockout strains (*AGP* and *ZEP*) demonstrated 81% higher oil productivity along with same zeaxanthin and lutein concentrations as that of WT	Song et al. (2022)
UVM11	Bacterial *phytase* gene	knockin	Cas9	Showed site specific integration of bacterial *phytase* gene which clearly demonstarted that complex recombinant proteins and novel synthetic biomolecules can be successfully produced in food grade organism *C. reinhardtii*	Zadabbas Shahabadi et al. (2023)
Phaeodactylum tricornutum					
CCMP2561	*CryP*	Knockout	Cas9	Knockout showed increase in fucoxanthin content by 1.29-fold compared to WT	Yang et al. (2022)
Pt1 wild type (WT)	*ptACSL3*	Knockout	Cas9	*PtACSL3* mutants showed altered FA profiles in two galactoglycerolipids and phosphatidylcholine (PC) with significantly reduced distribution of 16:0 and 16:1	Hao et al. (2022)
CCAP 1055/1, & CCMP2712	Violaxanthin de-epoxidase (*VDL2*) and zeaxanthin epoxidase (*ZEP*1)	Knockout	Cas9	Demonstrated the role of two genes in Fucoxanthin biosynthesis. Knockout mutants of *VDL2* and *ZEP*1 did not produce fucoxanthin	Bai et al. (2022)

(Continued on following page)

TABLE 1 (*Continued*) Examples of CRISPR-based bioengineering in different microalgal species to produce industrially important biomolecules.

Strain	Target gene modified	Type of modification	Type of nuclease	Outcomes	References
P. tricornutum Bohlin	enoyl CoA hydratase (*PtECH*)	CRISPRi	dCas9	The *ECH* knockdown mutants exhibited an enhanced lipid accumulation relative to WT. The mutants showed higher photosynthetic efficiency, but impaired growth compared to WT	Guo et al. (2023)
Nannochloropsis spp.					
N. gaditana CCMP1894	*ZnCys*	Attenuation and Knockout	Cas9	Attenuated mutants of *ZnCys* demonstrated double the strain's lipid productivity (~5.0 g/m^2/d) compared to WT (~2.5 g/m^2/d) and retained the ability to grow and fix CO_2 at levels nearly equivalent to those of the WT strain	Ajjawi et al. (2017)
N. gaditana CCMP1894	Acyl-CoA oxidase (*Aco1*)	Knockout and Knockin	Cas9	Knockout mutant strains showed double the lipid productivity compared to WT	Verruto et al. (2018)
N. salina CCMP1776	Three cellulose synthase genes (cesA1, cesA2, cesA4)	Knockout	Cas9	Cell wall thickness and cellulose content were reduced in *cesA1* mutant but not in *cesA*2 or *cesA*4 cells. *CesA*1 mutation resulted in a reduction of chrysolaminarin and neutral lipid contents, by 66.3% and 37.1%, respectively, but increased the soluble protein content by 1.8-fold. Thinned cell wall cells were susceptible to mechanical stress, resulting in a 1.7-fold enhancement of lipid extractability	Jeong et al. (2020)
N. gaditana CCMP526	beta-glucan synthase (*BGS*), transglycosylase (*TGS*)	Knockout	Cas9	The generated knockout mutants showed ~5-fold lower accumulation of soluble carbohydrate (β-1,3-glucose oligomers) following nitrogen starvation compared to WT, without any observed growth defect. The *TGS* knockout mutants showed 25%–40% (dry cell weight) increases in total fatty acids	Vogler et al. (2021)
N. salina CCMP1776	Δ12-fatty acid desaturase (*FAD12*)	Knockin	Cas9	The targeted knockin mutants of *FAD12* showed four-fold higher production of linoleic acid and 1.5-fold increase in eicosapentaenoic acid	Ryu et al. (2021)
N. oceanica IMET1	bZIP-family regulator *NobZIP77*	Knockout	Cas9	The *NobZIP77* knockout mutants under nitrogen deprivation showed similar growth like WT, with 3 times more triacylglycerol productivities	Zhang et al. (2022)
Chlorella spp.					
C. vulgaris FSP-E	Omega-3 fatty acid desaturase (*fad*3)	Knockout	Cas9	Knockout mutants of *fad*3 showed 46% (w/w) higher lipid content over WT	Lin and Ng (2020)
C. sorokiniana UTEX 1602	—	CRISPRa/i [Adaptive Single Guide Assisted Regulation DNA (ASGARD)]	dCas9	Gene regulation via CRISPRa-VP64 (CRISPRa) enhanced the protein contents up to 60% (w/w) of dry cell weight, where the highest concentration was 570 mg/L, while CRISPRi-KRAB (CRISPRi) with ASGARD increased protein content to 65% and lipid formed in the range of 150–250 mg/L (WT: 150 mg/L)	Lin et al. (2022)

(Continued on following page)

TABLE 1 (*Continued*) Examples of CRISPR-based bioengineering in different microalgal species to produce industrially important biomolecules.

Strain	Target gene modified	Type of modification	Type of nuclease	Outcomes	References
Parachlorella kessleri NIES-2152	Three genes, calcium-dependent membrane targeting (9934_t), duplicated mannanases 1 (8741_t), plastidic ATP/ADP translocase (9067_t)	Knockout	Cas9	The knockout mutants of plastidic ATP/ADP translocase showed >30% more lipid productivity compared to WT strain under dark light cycle	Kasai et al. (2023)
Tetraselmis spp.					
Tetraselmis spp. KCTC12432BP	ADP-glucose pyrophosphorylase (AGP)	Knockout	Cas9	*AGP* mutants showed enhanced lipid production. Fatty acids in the *AGP* mutants increased by 274% and 314% compared to WT. The mutants showed 2.3 and 2.7-fold higher lipid productivity than the WT	Chang et al. (2020)
Porphyridium purpureum					
P. purpureum CCMP 1328	Chlorophyll synthase (*CHS*1)	Knockout	Cas9	*CHS*1 mutants produced significantly higher phycoerythrin compared to WT	Jeon et al. (2021)

In 2016, Shin et al. developed a high-throughput CRISPR/Cas9 system for *C. reinhardtii*, which enabled the editing of multiple gene targets simultaneously. This system was used to disrupt genes involved in the regulation of lipid metabolism, demonstrating the potential for using CRISPR/Cas9 for metabolic engineering in *C. reinhardtii* (Shin et al., 2016). The potential applications of gene knockout by CRISPR/Cas9 have been explored for modifying the biochemical pathways for accumulation of product(s) of interest.

Baek et al. (2016) showed the successful applicability of DNA-free RNA guided engineered nucleases RNPs by creating specific knockouts of *CpFTSY* gene with a 0.56% efficiency and zeaxanthin epoxidase (*ZEP*) gene with 0.46% efficiency. Baek et al. (2018) continued this study and knocked out *ZEP* gene in higher carotenoid production strain of *C. reinhardtii* CC-4349, resulting in 56-fold higher zeaxanthin levels without reduction in lutein levels. Feeding this mutant to hens also resulted in fortified eggs with higher lutein (2-fold) and zeaxanthin (2.2-fold) content (Baek et al., 2018).

In 2017, Kao and Ng tested the applicability of vector derived CRISPRi to downregulate the *CrPEPC1* gene, successfully manipulating the carbon flux to increase lipid production, and indicating that CRISPRi can be used for modulating the expression of target gene(s) in *C. reinhardtii* to improve desired traits (Kao and Ng, 2017).

In another study, Baek et al. reported the use of CRISPR/Cas9 system to edit the *C. reinhardtii* genome to increase the lipid content in cells. The authors targeted the genes encoding acyl-CoA diacylglycerol acyltransferase (*DGAT*) and phospholipid diacylglycerol acyltransferase (*PDAT*), which are involved in lipid biosynthesis and demonstrated that CRISPR/Cas9 system can be used to generate targeted mutations that increase the lipid content of *C. reinhardtii* (Baek et al., 2018).

Shin et al. (2019) used the RNP mediated CRISPR/Cas9 technology to increase the lipid content via knockout of the phospholipase A2 (*PLA2*) gene, increasing the diacylglycerol pool, and higher accumulation of triacylglycerol, without significantly effecting the growth of the cells.

An increase in lipid productivity was also observed by Nguyen et al. (2020) after the knockout of the *ELT* (Cre01.g000300) gene encoding an enzyme in lipid catabolism, resulting in 28.52% increased total lipid and a shift in the fatty acid composition with an increase of approximately 27.2% in the C18:1 proportion.

Song et al. (2020) used RNP mediated CRISPR/Cas9 technology to produce highly purified zeaxanthin by blocking the synthesis of lutein via knocking *ZEP* gene and α-branch of lycopene epsilon cyclase (*LCYE*) gene, leading to 60% higher zeaxanthin yield than parental strain. In a further study, Song et al. (2022) generated a double mutant by knocking out *ZEP* and the ADP-glucose pyrophosphorylase (*AGP*) genes to accumulate lutein (2.93 mg/g), zeaxanthin (3.12 mg/g), and lipids (450.09 mg/g) in an N-deprived condition, achieving an 81% higher oil productivity with increased macular pigment productivity.

Freudenberg et al. (2022) knocked out the key enzymes of polyamine biosynthesis pathway to study the accumulation of putrescine, demonstrating the crucial role of ornithine decarboxylase 1 (*ODC1*) as the limiting factor in putrescine accumulation. The group identified that overexpression of ornithine decarboxylases and functional knockout of amine oxidase 2 (*AMX*2) for preventing putrescine degradation resulted in a 10-fold increase in cellular putrescine titers and yielded 200 mg/L.

Overall, the use of CRISPR/Cas genome editing technology has greatly advanced the ability to manipulate the *Chlamydomonas* genome in a precise and efficient manner. These advancements have enabled researchers to study the functions of individual genes and engineer strains with desired traits for various applications.

Phaeodactylum tricornutum

Phaeodactylum tricornutum, a marine diatom, has gained attention due to its unique characteristics, including rapid growth, potential to produce a variety of valuable bioactive compounds (e.g., pigments, lipids, and polysaccharides), and the concomitant assimilation of

10.3389/fbioe.2023.1267826

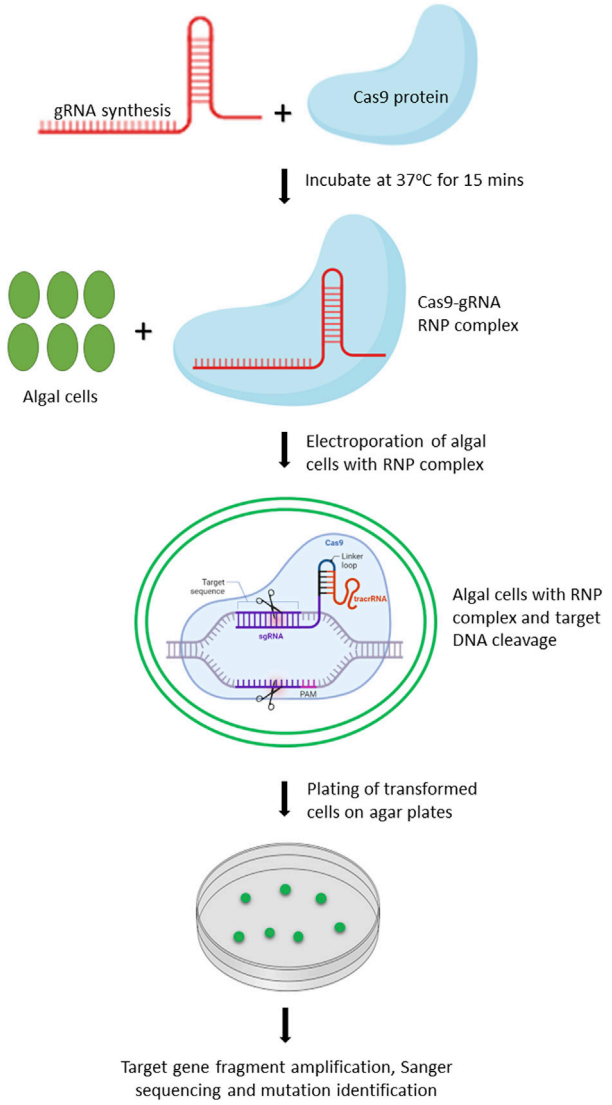


FIGURE 2
Schematic representation of workflow for editing target genes in microalgae by delivery of Cas9-gRNA RNP complex. Created with BioRender.com.

carbon dioxide. The lipid biosynthesis pathways of *P. tricornutum* have been extensively investigated and the species have been commercially exploited as a source of high-value carotenoids (fucoxanthin), omega-3 long-chain polyunsaturated fatty acids (LC-PUFA) and other lipids. These traits make it an attractive candidate for various biotechnological applications.

The CRISPR/Cas9 tool, for the first time, was successfully applied to *P. tricornutum* in 2016 (Nymark et al., 2016). The knockout mutants of chloroplast signal recognition particle 54 (*CPSRP54*) gene, a member of the chloroplast signal recognition particle pathway, were generated with a knockout efficiency of 31% (Nymark et al., 2016). After that, CRISPR-based gene knockout (Russo et al., 2018; Serif et al., 2018) and knockin (Moosburner et al., 2020) have been successfully reported in *P. tricornutum*.

In 2018, Russo et al. (2018) developed an episomal plasmid carrying Cas9 and transformed it via conjugation to strengthen CRISPR toolbox for *P. tricornutum*. The transformation of an episomal plasmid allowed to avoid unwanted perturbations due to random integration in the genome and excluding the Cas9 activity when it was no longer required, thereby reducing the probability of obtaining off-target mutations. Later in 2020, Moosburner et al. (2020) created knockout mutants by transforming episomal plasmid carrying Cas9 via particle-bombardment transformation method. The authors developed and reported a protocol that could select mutants in less than 3 weeks, demonstrating the feasibility of generating CRISPR-based industrial strains in minimal time. Furthermore, Taparia et al. (2022) successfully constructed a CRISPR/Cas9 episome for multiplexed targeting and creation of marker-free edited genomes. The group constructed an efficient, compact, episome expressing Cas9 targeting Stramenopile-type lipid droplet protein (*StLDP*) gene. The reported knockout efficiency ranged from 6.7% to 13.8%. This study provides a protocol for modular assembly of a multiplexed genome-editing episome that uses RNA polymerase II promoters, which can easily transcribe long sgRNA arrays and make multiplexed gene editing feasible.

Serif et al. (2018) first successfully demonstrated multiplexed genome-editing by transforming Cas9-gRNA RNPs using a gene gun in *P. tricornutum*. Two endogenous genes, *ptUMPS* (5-fluoroorotic acid resistance) and *PtAPT* (2-fluoroadenine resistance) were knocked out with 65%–100% efficiency, clearly showing the feasibility of multiplexed gene editing in this species (Serif et al., 2018). It opened the way to study the functions of multiple gene family members. Similarly, Stukenberg et al. (2018) further showed the efficiency of vector derived CRISPR/Cas9 with biolistic transformation by mutating the *vtc2* and *Pho4*, observing the easiness of mono- and bi-allelic mutants and without any off targets in the genome. Later in 2021, Sharma et al. (2021) in *P. tricornutum* reported successful simultaneous knockout of five [light-harvesting complex (LHC)] homologous genes using two gRNAs and a high fidelity Cas9 nuclease in which four amino acids substitutions had been introduced compared to wild type (WT) Cas9 nuclease. This study clearly demonstrated that engineering Cas9 nuclease reduced off-target editing, indicating that the altered high fidelity Cas9 nuclease must be exploited for precise genome editing.

In a study, to investigate the function of cryptochrome *CryP* and its role in regulating fucoxanthin content, *CryP* gene was knocked out in *P. tricornutum* using CRISPR/Cas9. The authors reported that *CryP* knockout line demonstrated stable heredity after hundreds of generations. CryP functions as a blue light-sensitive protein that regulates the expression of genes encoding carotenoid biosynthesis enzymes and fucoxanthin chlorophyll *a*/c-binding proteins (FCPs). Upon knockout of the *CryP* gene, both fucoxanthin content and FCP levels in the *P. tricornutum* knockout line increased considerably compared to the WT (Yang et al., 2022).

Functional characterization of long-chain Acyl-CoA synthetases (LACS) isozymes by CRISPR/Cas9 knockouts of *ptACSL*1-5 genes was done by Hao et al. (2022). Their findings demonstrated the potential of generating gene knockout mutants with the mutation of large fragment deletion using multiplexed CRISPR/Cas9 and provided insights into the functions of LACS isozymes in lipid metabolism in the oleaginous microalgae (Hao et al., 2022).

CRISPR/Cas9 was also used to knockout genes in uracil, histidine, and tryptophan biosynthetic pathways using plasmid

vector expressing Cas9 and target specific gRNA (Slattery et al., 2020). Sequencing of mutants indicated that editing events are characterized by the occurrence of large deletions of approximately 2.7 kb centered on the editing site. This study provides new auxotrophic markers to easily select mutants, a viable alternative to traditionally used antibiotic selection markers, thus aiding the development of marker-free production strains. Llavero-Pasquina et al. (2022) used CRISPR/Cas9 to test the predicted function of genes containing thiamine pyrophosphate riboswitches. Knockout mutants of violaxanthin de-epoxidase (*VDL2*) and zeaxanthin epoxidase (*ZEP*1) were developed in *P. tricornutum* via CRISPR/Cas9, demonstrating their role in fucoxanthin biosynthesis pathways via xanthophyll cycle (Bai et al., 2022). This study clearly elucidates that CRISPR can be efficiently used to knockout one or multiple genes simultaneously and demonstrate their functions in different biosynthetic pathways.

Recently, *P. tricornutum* fucosyltransferase 1 (*PtFucT*1) which is located on the medial/trans-Golgi apparatus was knocked out using CRISPR/Cas9 (Xie et al., 2023). The knockout mutants of *PtFucT*1 demonstrated reduced algal growth, biomass, and photosynthetic efficiency. Golgi fucosyltransferase 1 (*PtFucT*1) revealed its important role in α-1,4-fucose modification of N-glycan. This study provided critical information to understand the mechanism of protein N-glycosylation modification and demonstrated *P. tricornutum* as an alternative, ecofriendly cell factory to produce biopharmaceuticals (Xie et al., 2023).

Understanding the regulatory mechanism of neutral lipid accumulation and degradation, which is mediated by lipid droplet-associated proteins, is important in improving lipid productivity. Hence, a knockout mutant of Stramenopile-type lipid droplet protein (*StLDP*) was generated by CRISPR/Cas9 based genome editing (Yoneda et al., 2023). The mutant strain showed a decrease in lipid droplet (LD) numbers per cell, an increase in LD size, and no alteration of neutral lipid content under nitrogen deficiency, clearly elucidating that *StLDP* acts as an LD scaffold protein.

Nannochloropsis spp.

Nannochloropsis spp. are the emerging industrial microalgae favored due to their robust growth performance at large scale, accumulation proteins, and triacylglycerol with high-value PUFAs (Wang et al., 2021). They serve as an outstanding research model for synthetic biology owing to their small genome size, simple gene structure, and accessibility to newly developed genetic tools. The resources for genomic, transcriptomic, proteomic, lipidomic, and physiological data are extensive in this species, and thus, can be exploited for targeted genome engineering (Wang et al., 2021).

First study on application of CRISPR/Cas9 in *Nannochloropsis* spp. was reported in 2016, by creating knockout mutants of *nitrate reductase* gene via plasmid-based expression of Cas9 (Wang et al., 2016). This demonstration of CRISPR/Cas9-based genome editing in *Nannochloropsis* opened the doors for *Nannochloropsis*-based biotechnological applications. Later, in 2017, the expression of a transcription regulator *ZnCys* in *Nannochloropsis gaditana* CCMP1894 was attenuated by insertion in the 5′UTR region using Cas9, which enabled to double the strain's lipid productivity while retaining its ability to grow and fix CO_2 at levels nearly equivalent to those of the WT strain, under dense semicontinuous culture (Ajjawi et al., 2017). This ability of CRISPR/Cas system to control the production of molecules of interest will more likely enable the commercialization of microalgal-derived bioproducts.

To broaden the CRISPR toolbox for *Nannochloropsis* spp., an episomal plasmid based Cas9 system was developed and transformed in *N. oceanica* CCMP1779 (Poliner et al., 2018). This system efficiently generated targeted mutations and allowed the loss of episomal DNA after the removal of selection pressure, resulting in marker-free nontransgenic engineered lines (Poliner et al., 2018). The generation of such non-transgenic mutant strains may face less regulatory challenges and more likely to meet acceptability in various markets.

Later in 2018, Verruto et al. (2018) reported unrestrained markerless trait stacking in *Nannochloropsis gaditana* through combined genome editing and marker recycling technologies. The authors demonstrated the proof-of-concept for generation of a markerless knockout in a gene encoding an acyl-CoA oxidase (*Aco*1) as well as the markerless recapitulation of a 2-kb insert in the *ZnCys* gene 5′-UTR that resulted in a doubling of lipid productivity in WT strain (Verruto et al., 2018). In the same study, they generated mutants that exhibit approximately 50% reduction in photosynthetic antennae size by markerless knockout of seven genes in the large light-harvesting complex gene family (Verruto et al., 2018).

In 2020, Jeong et al. (2020) used CRISPR/Cas9 to knockout three genes (*cesA*1, 2, and 4) involved in cellulose biosynthesis. Cell wall thickness and cellulose content were reduced in the *cesA*1 mutant, but not in *cesA*2 or *cesA*4. The *CesA*1 mutation resulted in a reduction of chrysolaminarin and neutral lipid contents by 66.3% and 37.1%, respectively, but increased the soluble protein content by 1.8-fold. Further, cells with a thinned cell wall were susceptible to mechanical stress, resulting in a 1.7-fold enhancement of lipid extractability. This study clearly demonstrated CRISPR tool could efficiently led to the production of chassis strain for different industrial applications.

In another study, Naduthodi et al. (2021) used Cas12a nuclease (variant of Cas9) in *Nannochloropsis oceanica* for precise modifications. They demonstrated RNPs and homology-directed repair (HDR) genome editing strategy to generate scarless and markerless mutants. The authors also developed an episomal plasmid based Cas12a system for efficiently introducing indels at the target site. They also reported the ease of multiplexed genome engineering using Cas12a. Furthermore, dCas9 and dfnCas12a based CRISPRi platform was reported for down regulating of target genes. Reduction of 85% in the transcript levels upon performing CRISPRi with dCas9 in *N. oceanica* was reported. Overall, these developments substantially accelerate genome engineering efforts in *N. oceanica* and potentially provide a general toolbox for improving other *Nannochloropsis* strains for industrial applications (Naduthodi et al., 2021).

A study was conducted to build a minimal genome for *Nannochloropsis using* CRISPR/Cas9 system by serially and precisely deleting large genome fragments of approximately 100 kb from its 30.01 Mb nuclear genome. The "non-essential" chromosomal regions based on minimal gene expression [low expression regions (LERs)] under N-replete and N-depleted

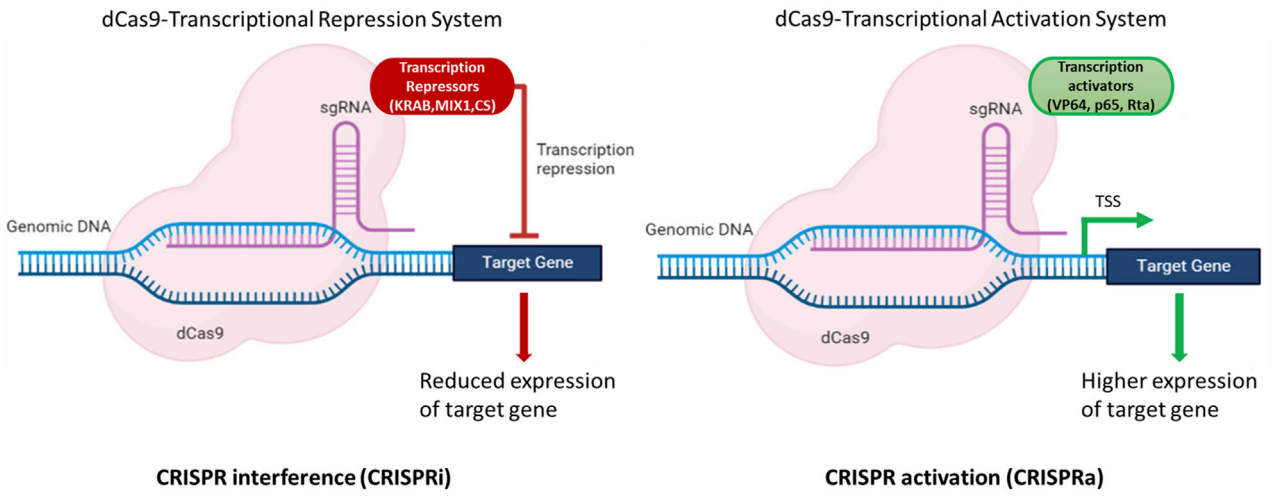


FIGURE 3
Illustration of dCas9 based expression modulation of target genes. The dCas9-gRNA complex fused with effectors, either transcriptional activators (VP64, p65, Rta) or repressors (KRAB, MIX1, CS) bind to the upstream regulatory elements and modulate the expression of target gene(s). CRISPR interference (CRISPRi) is used to repress the gene(s) of the competing pathways to channelize the flux towards the molecule(s) of interest. CRISPR activation (CRISPRa) is used to enhance the expression of the target gene(s) and hence aid in improving the productivities of target molecules of interest. Created with BioRender.com.

conditions were identified and deleted. The LER1 deletion (~110 kb deletion) and the LER1–LER2 serial deletion (~214 kb in total) showed essentially normal growth, lipid contents, fatty acid saturation levels, and photosynthesis, or, in the case of a LER1–LER2 double-deletion mutant, slightly higher growth and biomass productivity than the WT (Wang et al., 2021).

A transcriptional activation system based on CRISPR/dCas9 in *N. oceanica* IMET1 was constructed in a study by Wei et al. (2022). This construct (dCas9 protein fused with transcriptional activator VP64) could efficiently alter gene expression. The expression of *g1248* gene responsible for DNA or RNA methylation as a methyltransferase was increased by 2–6 fold at the transcriptional level. Furthermore, the growth and photosynthetic parameter (Fv/Fm) of mutants was increased by 23% and 12%, respectively, compared to WT under atmospheric CO_2 concentration (Wei et al., 2022). This study clearly demonstrated that the expression of target genes could be modulated using dCas9-based CRISPR system in *Nannochloropsis* spp. to achieve higher titers of molecules of interest. These strategies can aid in the development of a strain that can produce higher titers of biomolecules of interest either by modulating gene expression by CRISPRa or CRISPRi or by knocking out the genes of the competing pathways (Figure 3).

Chlorella spp.

Chlorella species have easier culture conditions, rapid growth rate, ability to thrive in varied and challenging environments, and produce diverse biomolecules. *Chlorella* spp. are rich in lipids, making it an excellent candidate for biofuel production. They are rich in nutritional compounds that include high protein, essential fatty acids, vitamins, and minerals content. They are widely used as a dietary supplement and functional food ingredient due to their potential health benefits, such as immune system support, detoxification, and antioxidant properties. However, insufficient genomic resources for *Chlorella* and low transformation efficiency are the major concerns for genetic improvement of these genus.

The first report on successful application of CRISPR-based gene editing in *Chlorell*a spp. was reported in 2020 (Lin and Ng, 2020). In this study, a plasmid vector harboring Cas9 and omega-3 fatty acid desaturase (*fad*3) gene specific gRNA was transformed in *Chlorella vulgaris* FSP-E, the knockout mutants of *fad*3 showed 46% (w/w) higher lipid content over WT (Lin and Ng, 2020). The study clearly demonstrated relevant empirical evidence of applying CRISPR-based genetic manipulation in *Chlorella* spp. for producing robust chassis strains of *Chlorella* for several industrial applications.

Knockout mutants of two genes, nitrate reductase (*NR*) and adenine phosphoribosyltransferase (*APT*), were generated in *C. vulgaris* UTEX395 using plasmid harboring Cas9 and target specific gRNAs and ribonucleoprotein complex (Kim et al., 2021). The mutants were negatively selected on potassium chlorate or 2-fluoroadenine clearly exhibiting the practicality of generating markerless mutants in *Chlorella* that might likely not fall under regulatory purview (Kim et al., 2021).

Lin et al. (2022) used CRISPRa/i system to modulate the gene expression in *Chlorella sorokiniana* UTEX 1602. They demonstrated that gene regulation via dCas9-VP64 (CRISPRa) increased the protein contents by approximately 60% (w/w), while that via dCas9-KRAB (CRISPRi) increased the protein content to 65%, and lipid accumulated in the range of 150–250 mg/L (WT: 150 mg/L). This study clearly opens new avenues in altering the expression levels of target genes, thus allowing improvement in the titer yields of target molecules.

Picochlorum spp.

Picochlorum genus have been explored to produce various biomolecules, including lipids, pigments, proteins, and bioactive

compounds. For example, lipids from *Picochlorum* spp. can be utilized to produce biofuels or high-value fatty acids. Pigments derived from these algae, such as chlorophylls and carotenoids, have applications in the food, cosmetic, and pharmaceutical industries. Additionally, *Picochlorum* spp. have the potential to produce recombinant proteins and enzymes. Green algae in *Picochlorum* genus can achieve rapid growth rates and thrive in intense light, high-temperature, and high-salt conditions.

In 2020, for the first, time CRISPR/Cas9 was applied in *Picochlorum celeri* to generate knockout mutants of two genes, nitrate reductase and carotenoid isomerase. The loss-of-function mutants were present in approximately 6% of transformants targeting nitrate reductase and in approximately 50% of transformants targeting carotenoid isomerase (Krishnan et al., 2020). This report provides evidence that CRISPR-based genome editing is feasible in *Picochlorum* spp. that can rapidly generate mutants, help in elucidating the functions of candidate genes, and precisely manipulate genes/metabolic pathways to produce robust production strains.

Key considerations for using CRISPR-based genome editing in microalgae

Transformation method

One of the major challenges is the lack of efficient delivery methods for genome editing components into microalgae cells. As with other eukaryotic cells, delivery of foreign nucleic acids into microalgae cells is challenging due to the presence of cell wall and low efficiency of transformation methods. Various delivery methods, such as electroporation, biolistic transformation, and *Agrobacterium*-mediated transformation, have been developed for microalgae (Radakovits et al., 2010), but their efficiency varies widely among different species and strains. Thus, well optimized and reproducible transformation methods need to be implemented to achieve high levels of targeted modifications.

Homologous recombination

Low efficiencies of homologous recombination impede precise targeted knockin of desired gene fragments in microalgae. Even after several optimizations, such as concentrations of DNA fragments, lengths of HDR cassette(s), and optimizing the transformation conditions, the success of homologous recombination has been very low in *C. reinhardtii* (Plecenikova et al., 2013) and *Phaeodactylum* (Moosburner et al., 2020). The codelivery of gRNA-Cas9 RNP complex along with a dsDNA repair template efficiently enhanced homologous recombination at the target site, resulting in a remarkable higher percentage of targeted integration clones (Naduthodi et al., 2019). Very recently, with optimizing the concentrations of gRNA and Cas9 protein and with 1000-bp homology arms, the targeted knockin efficiency was reported to be 15% in *C. reinhardtii* (Shahabadi et al., 2023). However, the efficiency of HDR is species-dependent, and it has been proposed that the homologous recombination efficiency in microalgae could more likely be enhanced by interfering with the DNA repair proteins Ku70, Ku80, DNA ligase IV, and other relevant genes (He et al., 2013). Hence, considering the native efficiency of homologous recombination in species of interest, several parameters need to be considered, such as concentration of HDR cassette, length of homology arms, site of integration, concentrations of gRNA and Cas9 nuclease, and the best optimized transformation method, before planning targeted knockin experiments.

Off-targets

Another major concern is the off-target mutations caused by genome editing tools. Although CRISPR/Cas has been shown to have high specificity, it can still induce unintended mutations at off-target sites (Hsu et al., 2013). This can be particularly problematic in microalgae, where the genomic organization and presence of repetitive sequences can increase the likelihood of off-target effects. The targeting specificity depends on how guide RNA(s) are designed to target the gene of interest, the type of nuclease, and the protospacer adjacent motif (PAM) recognition sequences. Different tools such as Cas-OFFinder, Benchling, CHOPCHOP, and CRISPOR are developed to design target specific gRNAs and predict off-target cleavage sites (Ghribi et al., 2020). Unavailability of high-quality whole genome sequence makes identification/prediction of the off-target sites challenging. Therefore, it is very important that gRNAs are designed carefully to minimize or have no-off targets, wherever possible.

Use of high-fidelity nucleases

In recent years, extensive studies have been conducted to improve the gene-editing specificity of the most popular Cas nucleases using different strategies. Non-rational strategies like directed evolution and rational strategies such as structure and/or function-guided protein engineering or combination of both are used for engineering high-fidelity Cas proteins and representative variants (Huang et al., 2022). HiFi Cas9, xCas9, SpartaCas, efSaCas9 (enhance-fidelity SaCas9), SpCas9 nickase, SpCas9-D1135E, eSpCas9, eSaCas9, and SaCas9-HF are some of the high-fidelity variants that have been discovered through these strategies (Huang et al., 2022). Engineered Cas9 variants, Cas9 homologs such as KKH SaCas9 (smaller than SpCas9) (Ran et al., 2013), and novel Cas9 protein such as C2C1 (Shmakov et al., 2015) have been successfully applied in genome editing with high efficiency and specificity. It was reported in a study that the DNA binding specificity of a catalytically inactive Cas9 mutant (dCas9) was sufficiently high in *E. coli*, yielding no detectable off-target transcriptional repression in the *E. coli* transcriptome (Qi et al., 2013). Apart from SpCas9 variants, other Cas proteins such as *Staphylococcus aureus* Cas9 (SaCas9), with different PAM requirements is now widely used for gene editing applications due to its high activity in eukaryotic systems (Wang et al., 2019). Recent advancements in the discovery and/or development of several HiFi variants of nuclease(s) suggest that the specificity of WT-nucleases can be enhanced to a greater extent more likely using protein engineering or evolution approaches. Use of such HiFi nucleases will greatly facilitate enhanced editing efficiencies.

Sequencing of CRISPR edited clones

It is well reported that editing might also occur at unintended loci if there is sequence complementarity with the designed gRNAs in the genome. Identifying these off-target sites using computational analysis and analyzing the off-target editing becomes very crucial to verify, if these edits lead to any unintended effects. Whole genome sequencing of the edited clone using next-generation sequencing (NGS) platforms is apparently the only way to discover off target editing. NGS technologies also aid in identification of allelic changes in CRISPR edited clones (Veeranagouda et al., 2018). A NGS method called CRISPR/Cas9 edited site sequencing (CRES-Seq) has been reported for efficient and high throughput screening of CRISPR/Cas9 edited clones. CRES-Seq aids in precise genotyping up to 96 CRISPR-Cas9-edited sites in a single MiniSeq (Illumina) run with an approximate sequencing cost of $6/clone (Veeranagouda et al., 2018). Alternatively, due to unavailability of funds, checking for off-target editing is done by performing sequencing of sequences with the highest probability of off-target effects (i.e., most like your CRISPR target region). The targeted CRISPR induced mutations can be confirmed by amplifying the target regions and sequencing the purified fragments using Sanger sequencing (Clement et al., 2020). While sequencing the fragment using Sanger method make sure, you sequence each purified fragment atleast twice, to ensure you do not have erroneous sequenced data. Moreover, always send purified PCR fragments of targeted regions for Sanger sequencing, as the purity of samples will determine the quality of sequenced data. It is always advisable to have an amplicon size of 600–700 bp for Sanger sequencing and oligos that are used for sequencing must bind at least 110 bp away from the cleavage sites, in case of multiple gRNAs used for targeting the same region.

Regulatory challenges

In addition to the technical limitations described in previous sections, the regulatory framework for the use of genome editing tools in microalgae is still evolving. In many countries, the use of genome editing tools in agriculture and biotechnology is subject to strict regulations and the safety and ethical implications of their use are still being debated. The rapid emergence and ongoing developments in genome editing demand a timely review and revision of the current definitions and regulations around genetically modified organisms (GMOs) (Spicer and Molnar, 2018). New strains of microalgae generated through gene editing, where the genetic manipulations are made only in the native genes are also considered under GMO umbrella in the European Union, hampering practical and commercial applications. In March 2022, the Ministry of Environment, Forest and Climate Change, Government of India issued an office memorandum for exemption of genome-edited plants that falls in SDN1 (site directed nuclease 1) and SDN2 categories, from the Indian GMO regulation (https://pib.gov.in/PressReleasePage.aspx?PRID=1871153#:~:text=Ministry%20of%20Environment%2C%20Forest%20and,Rules%201989%20of%20EPA%2C%201986). SDN1 and SDN2 includes alterations only within the organisms own genetic code. In the US, the US Department of Agriculture are imposing regulations on the product or the organism itself rather than the process, which tends to favor gene-edited organisms for their commercial applications (Spicer and Molnar, 2018). Therefore, the regulatory landscape for gene edited organisms needs to be clearly understood before initiating projects involving application of genome editing tools.

Conclusion

Microalgae are attractive candidates for bioengineering due to their high photosynthetic efficiency and higher growth rates. They are increasingly favored sustainable feedstock for food, feed, biofuels, and nutraceutical applications. Despite the apparent advantages, the major bottleneck in employing microalgae as an economically viable production platform is the high cost of biomass production. Genetic engineering has an ability to change the inherent potential of the strain for improved productivity by augmenting various genetic traits. CRISPR/Cas9 allows precise and targeted genetic manipulations literally at any loci in the genome. CRISPR-aided bioengineering has emerged as a powerful tool and is being widely applied for the genetic manipulation of microalgae for enhanced performance. CRISPR/Cas systems have been successfully used in several microalgal species, including *Chlamydomonas reinhardtii*, *Nannochloropsis* spp., *Chlorella* spp., *Phaeodactylum tricornutum* and many more. Overall, CRISPR-aided bioengineering has enormous potential for the development of microalgae-based bioproduction platforms to produce a wide range of value-added products. However, further research is needed to optimize the use of CRISPR/Cas tools in microalgae to overcome the associated challenges.

Author contributions

DD: Writing–original draft, Writing–review and editing. AS: Writing–original draft, Writing–review and editing. AY: Writing–original draft. NG: Writing–original draft, Writing–review and editing. AB: Writing–review and editing. SD: Writing–review and editing. BB: Writing–review and editing.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We thank Dr. Azam Sheikh for critically reviewing the manuscript and providing critical inputs.

Conflict of interest

Authors DD, AS, AY, NG, AB, SD, and BB were employed by the company Reliance Industries Ltd.

Publisher's note

References

Acquah, C., Ekezie, F.-G., and Udenigwe, C. C. (2021). Potential applications of microalgae-derived proteins and peptides in the food industry. *Cult. Microalgae Food Industry*, 97–126. doi:10.1016/b978-0-12-821080-2.00011-3

Ahmed, F., Zhou, W., and Schenk, P. M. (2015). *Pavlova lutheri* is a high-level producer of phytosterols. *Algal Res.* 10, 210–217. doi:10.1016/j.algal.2015.05.013

Ajjawi, I., Verruto, J., Aqui, M., Soriaga, L. B., Coppersmith, J., Kwok, K., et al. (2017). Lipid production in *Nannochloropsis gaditana* is doubled by decreasing expression of a single transcriptional regulator. *Nat. Biotechnol.* 35, 647–652. doi:10.1038/nbt.3865

Ampofo, J., and Abbey, L. (2022). Microalgae: bioactive composition, health benefits, safety and prospects as potential high-value ingredients for the functional food industry. *Foods* 11, 1744. doi:10.3390/foods11121744

Asadian, M., Saadati, M., Bajestani, F. B., Beardall, J., Abdolahadi, F., and Mahdinezhad, N. (2022). Knockout of *Cia5* gene using CRISPR/Cas9 technique in *Chlamydomonas reinhardtii* and evaluating CO_2 sequestration in control and mutant isolates. *J. Genet.* 101, 6–7. doi:10.1007/s12041-021-01350-x

Baek, K., Kim, D. H., Jeong, J., Sim, S. J., Melis, A., Kim, J.-S., et al. (2016). DNA-free two-gene knockout in *Chlamydomonas reinhardtii* via CRISPR-Cas9 ribonucleoproteins. *Sci. Rep.* 6, 30620–30627. doi:10.1038/srep30620

Baek, K., Yu, J., Jeong, J., Sim, S. J., Bae, S., and Jin, E. (2018). Photoautotrophic production of macular pigment in a *Chlamydomonas reinhardtii* strain generated by using DNA-free CRISPR-Cas9 RNP-mediated mutagenesis. *Biotechnol. Bioeng.* 115, 719–728. doi:10.1002/bit.26499

Bai, Y., Cao, T., Dautermann, O., Buschbeck, P., Cantrell, M. B., Chen, Y., et al. (2022). Green diatom mutants reveal an intricate biosynthetic pathway of fucoxanthin. *Proc. Natl. Acad. Sci.* 119, e2203708119. doi:10.1073/pnas.2203708119

Benedetti, M., Vecchi, V., Barera, S., and Dall'Osto, L. (2018). Biomass from microalgae: the potential of domestication towards sustainable biofactories. *Microb. Cell Fact.* 17, 173–218. doi:10.1186/s12934-018-1019-3

Bibikova, M., Golic, M., Golic, K. G., and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics* 161, 1169–1175. doi:10.1093/genetics/161.3.1169

Chang, K. S., Kim, J., Park, H., Hong, S.-J., Lee, C.-G., and Jin, E. (2020). Enhanced lipid productivity in *AGP* knockout marine microalga *Tetraselmis sp.* using a DNA-free CRISPR-Cas9 RNP method. *Bioresour. Technol.* 303, 122932. doi:10.1016/j.biortech.2020.122932

Clement, K., Hsu, J. Y., Caner, M. C., Joung, J. K., and Pinello, L. (2020). Technologies and computational analysis strategies for CRISPR applications. *Mol. Cell* 79, 11–29. doi:10.1016/j.molcel.2020.06.012

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi:10.1126/science.1231143

de Jesus Raposo, M. F., De Morais, A., and De Morais, R. (2014). Bioactivity and applications of polysaccharides from marine microalgae. *Polysaccharides bioactivity Biotechnol.*, 1683–1727. doi:10.1007/978-3-319-03751-6_47-1

Del Mondo, A., Sansone, C., and Brunet, C. (2022). Insights into the biosynthesis pathway of phenolic compounds in microalgae. *Comput. Struct. Biotechnol. J.* 20, 1901–1913. doi:10.1016/j.csbj.2022.04.019

Dhokane, D., Bhadra, B., and Dasgupta, S. (2020). CRISPR based targeted genome editing of *Chlamydomonas reinhardtii* using programmed Cas9-gRNA ribonucleoprotein. *Mol. Biol. Rep.* 47, 8747–8755. doi:10.1007/s11033-020-05922-5

Dolganyuk, V., Belova, D., Babich, O., Prosekov, A., Ivanova, S., Katserov, D., et al. (2020). Microalgae: a promising source of valuable bioproducts. *Biomolecules* 10, 1153. doi:10.3390/biom10081153

D'Orazio, N., Gemello, E., Gammone, M. A., De Girolamo, M., Ficoneri, C., and Riccioni, G. (2012). Fucoxantin: a treasure from the sea. *Mar. Drugs* 10, 604–616. doi:10.3390/md10030604

Freudenberg, R. A., Wittemeier, L., Einhaus, A., Baier, T., and Kruse, O. (2022). The spermidine synthase gene *SPD1*: a novel auxotrophic marker for *Chlamydomonas reinhardtii* designed by enhanced CRISPR/cas9 gene editing. *Cells* 11, 837. doi:10.3390/cells11050837

Gaj, T., Gersbach, C. A., and Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405. doi:10.1016/j.tibtech.2013.04.004

Ghribi, M., Nouemssi, S. B., Meddeb-Mouelhi, F., and Desgagné-Penix, I. (2020). Genome editing by CRISPR-Cas: a game change in the genetic manipulation of *Chlamydomonas*. *Life* 10, 295. doi:10.3390/life10110295

Guo, W., Gao, Y., Weng, Y., Chang, C., Huang, X., and Zhang, F. (2023). *Boosting lipid production in the diatom Phaeodactylum tricornutum by knockdown of the Enoyl CoA hydratase using CRISPR interference.*

Hajimahmoodi, M., Faramarzi, M. A., Mohammadi, N., Soltani, N., Oveisi, M. R., and Nafissi-Varcheh, N. (2010). Evaluation of antioxidant properties and total phenolic contents of some strains of microalgae. *J. Appl. Phycol.* 22, 43–50. doi:10.1007/s10811-009-9424-y

Hamouda, R. A., and El-Naggar, N. E.-A. (2021). "Cyanobacteria-based microbial cell factories for production of industrial products," in *Microbial cell factories engineering for production of biomolecules* (Cambridge: Academic Press), 277–302.

Hao, X., Chen, W., Amato, A., Jouhet, J., Maréchal, E., Moog, D., et al. (2022). Multiplexed CRISPR/Cas9 editing of the long-chain acyl-CoA synthetase family in the diatom *Phaeodactylum tricornutum* reveals that mitochondrial *ptACSL3* is involved in the synthesis of storage lipids. *New Phytol.* 233, 1797–1812. doi:10.1111/nph.17911

He, Y., Liu, Q., Shao, Y., and Chen, F. (2013). ku70 and ku80 null mutants improve the gene targeting frequency in *Monascus ruber* M7. *Appl. Microbiol. Biotechnol.* 97, 4965–4976. doi:10.1007/s00253-013-4851-8

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832. doi:10.1038/nbt.2647

Huang, X., Yang, D., Zhang, J., Xu, J., and Chen, Y. E. (2022). Recent advances in improving gene-editing specificity through CRISPR–Cas9 nuclease engineering. *Cells* 11, 2186. doi:10.3390/cells11142186

Ishino, Y., Krupovic, M., and Forterre, P. (2018). History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. *J. Bacteriol.* 200, 005800–617. doi:10.1128/jb.00580-17

Jeon, M. S., Han, S.-I., Jeon, M., and Choi, Y.-E. (2021). Enhancement of phycoerythrin productivity in *Porphyridium purpureum* using the clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9 ribonucleoprotein system. *Bioresour. Technol.* 330, 124974. doi:10.1016/j.biortech.2021.124974

Jeon, S., Lim, J.-M., Lee, H.-G., Shin, S.-E., Kang, N. K., Park, Y.-I., et al. (2017). Current status and perspectives of genome editing technology for microalgae. *Biotechnol. Biofuels* 10, 267–318. doi:10.1186/s13068-017-0957-z

Jeong, B., Jang, J., and Jin, E. (2023). Genome engineering via gene editing technologies in microalgae. *Bioresour. Technol.* 373, 128701. doi:10.1016/j.biortech.2023.128701

Jeong, S. W., HwangBo, K., Lim, J. M., Nam, S. W., Lee, B. S., Jeong, B., et al. (2020). Genetic impairment of cellulose biosynthesis increases Cell Wall fragility and improves lipid extractability from oleaginous alga *Nannochloropsis Salina*. *Microorganisms* 8, 1195. doi:10.3390/microorganisms8081195

Ji, L., Qiu, S., Wang, Z., Zhao, C., Tang, B., Gao, Z., et al. (2023). Phycobiliproteins from algae: current updates in sustainable production and applications in food and health. *Food Res. Int.* 21, 112737. doi:10.1016/j.foodres.2023.112737

Jiang, W., Brueggeman, A. J., Horken, K. M., Plucinak, T. M., and Weeks, D. P. (2014). Successful transient expression of Cas9 and single guide RNA genes in *Chlamydomonas reinhardtii*. *Eukaryot. Cell* 13, 1465–1469. doi:10.1128/ec.00213-14

Kao, P.-H., and Ng, I.-S. (2017). CRISPRi mediated phosphoenolpyruvate carboxylase regulation to enhance the production of lipid in *Chlamydomonas reinhardtii*. *Bioresour. Technol.* 245, 1527–1537. doi:10.1016/j.biortech.2017.04.111

Kasai, Y., Takagi, S., Ota, S., Ishii, K., Takeshita, T., Kawano, S., et al. (2023). *Development of efficient genetic transformation and genome editing systems, and the isolation of a CRISPR/Cas9-mediated high-oil mutant in the unicellular green alga* Parachlorella kessleri *strain NIES-2152.*

Khavari, F., Saidijam, M., Taheri, M., and Nouri, F. (2021). Microalgae: therapeutic potentials and applications. *Mol. Biol. Rep.* 48, 4757–4765. doi:10.1007/s11033-021-06422-w

Kim, E., Koo, T., Park, S. W., Kim, D., Kim, K., Cho, H.-Y., et al. (2017a). *In vivo* genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* 8, 14500. doi:10.1038/ncomms14500

Kim, H., Kim, S.-T., Ryu, J., Kang, B.-C., Kim, J.-S., and Kim, S.-G. (2017b). CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nat. Commun.* 8, 14406. doi:10.1038/ncomms14406

Kim, J., Chang, K. S., Lee, S., and Jin, E. (2021). Establishment of a genome editing tool using CRISPR-Cas9 in *Chlorella vulgaris* UTEX395. *Int. J. Mol. Sci.* 22, 480. doi:10.3390/ijms22020480

Krishnan, A., Cano, M., Burch, T. A., Weissman, J. C., and Posewitz, M. C. (2020). Genome editing using Cas9-RNA ribonucleoprotein complexes in the high-productivity marine alga *Picochlorum celeri*. *Algal Res.* 49, 101944. doi:10.1016/j.algal.2020.101944

Kumar, G., Shekh, A., Jakhu, S., Sharma, Y., Kapoor, R., and Sharma, T. R. (2020). Bioengineering of microalgae: recent advances, perspectives, and regulatory challenges for industrial application. *Front. Bioeng. Biotechnol.* 8, 914. doi:10.3389/fbioe.2020.00914

Lee, Y. Y., Park, R., Miller, S. M., and Li, Y. (2022). Genetic compensation of triacylglycerol biosynthesis in the green microalga *Chlamydomonas reinhardtiie*. *Plant J.* 111 (4), 1069–1080. doi:10.1111/tpj.15874

Li, H.-B., Cheng, K.-W., Wong, C.-C., Fan, K.-W., Chen, F., and Jiang, Y. (2007). Evaluation of antioxidant capacity and total phenolic content of different fractions of selected microalgae. *Food Chem.* 102, 771–776. doi:10.1016/j.foodchem.2006.06.022

Lin, J., Lin, W., and Ng, I. (2022). CRISPRa/i with Adaptive Single Guide Assisted Regulation DNA (ASGARD) mediated control of *Chlorella sorokiniana* to enhance lipid and protein production. *Biotechnol. J.* 17, 2100514. doi:10.1002/biot.202100514

Lin, W.-R., and Ng, I.-S. (2020). Development of CRISPR/Cas9 system in *Chlorella vulgaris* FSP-E to enhance lipid accumulation. *Enzyme Microb. Technol.* 133, 109458. doi:10.1016/j.enzmictec.2019.109458

Llavero-Pasquina, M., Geisler, K., Holzer, A., Mehrshahi, P., Mendoza-Ochoa, G. I., Newsad, S. A., et al. (2022). Thiamine metabolism genes in diatoms are not regulated by thiamine despite the presence of predicted riboswitches. *New Phytol.* 235, 1853–1867. doi:10.1111/nph.18296

Lucakova, S., Branyikova, I., and Hayes, M. (2022). Microalgal proteins and bioactives for food, feed, and other applications. *Appl. Sci.* 12, 4402. doi:10.3390/app12094402

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. doi:10.1126/science.1232033

Martins, A., Vieira, H., Gaspar, H., and Santos, S. (2014). Marketed marine natural products in the pharmaceutical and cosmeceutical industries: tips for success. *Mar. Drugs* 12, 1066–1101. doi:10.3390/md12021066

Moosburner, M. A., Gholami, P., McCarthy, J. K., Tan, M., Bielinski, V. A., and Allen, A. E. (2020). Multiplexed knockouts in the model diatom *Phaeodactylum* by episomal delivery of a selectable Cas9. *Front. Microbiol.* 11, 5. doi:10.3389/fmicb.2020.00005

Naduthodi, M. I. S., Mohanraju, P., Südfeld, C., D'Adamo, S., Barbosa, M. J., and Van Der Oost, J. (2019). CRISPR–Cas ribonucleoprotein mediated homology-directed repair for efficient targeted genome editing in microalgae *Nannochloropsis oceanica* IMET1. *Biotechnol. Biofuels* 12, 66–11. doi:10.1186/s13068-019-1401-3

Naduthodi, M. I. S., Sudfeld, C., Avitzigiannis, E. K., Trevisan, N., van Lith, E., Alcaide Sancho, J., et al. (2021). Comprehensive genome engineering toolbox for microalgae *Nannochloropsis oceanica* based on CRISPR-Cas systems. *ACS Synth. Biol.* 10, 3369–3378. doi:10.1021/acssynbio.1c00329

Nguyen, T. H. T., Park, S., Jeong, J., Shin, Y. S., Sim, S. J., and Jin, E. (2020). Enhancing lipid productivity by modulating lipid catabolism using the CRISPR-Cas9 system in *Chlamydomonas*. *J. Appl. Phycol.* 32, 2829–2840. doi:10.1007/s10811-020-02172-7

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6, 24951. doi:10.1038/srep24951

Patel, A. K., Vadrale, A. P., Singhania, R. R., Michaud, P., Pandey, A., Chen, S.-J., et al. (2022). Algal polysaccharides: current status and future prospects. *Phytochem. Rev.* 22, 1167–1196. doi:10.1007/s11101-021-09799-5

Peters, G. P., Le Quéré, C., Andrew, R. M., Canadell, J. G., Friedlingstein, P., Ilyina, T., et al. (2017). Towards real-time verification of CO_2 emissions. *Nat. Clim. Chang.* 7, 848–850. doi:10.1038/s41558-017-0013-9

Plecenikova, A., Mages, W., Andrésson, Ó. S., Hrossova, D., Valuchova, S., Vlcek, D., et al. (2013). Studies on recombination processes in two *Chlamydomonas reinhardtii* endogenous genes, *NIT*1 and *ARG*7. *Protist* 164, 570–582. doi:10.1016/j.protis.2013.05.004

Poliner, E., Pulman, J. A., Zienkiewicz, K., Childs, K., Benning, C., and Farré, E. M. (2018). A toolkit for *Nannochloropsis oceanica* CCMP 1779 enables gene stacking and genetic engineering of the eicosapentaenoic acid pathway for enhanced long-chain polyunsaturated fatty acid production. *Plant Biotechnol. J.* 16, 298–309. doi:10.1111/pbi.12772

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183. doi:10.1016/j.cell.2013.02.022

Radakovits, R., Jinkerson, R. E., Darzins, A., and Posewitz, M. C. (2010). Genetic engineering of algae for enhanced biofuel production. *Eukaryot. Cell* 9, 486–501. doi:10.1128/ec.00364-09

Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., et al. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154, 1380–1389. doi:10.1016/j.cell.2013.08.021

Russo, M. T., Cigliano, R. A., Sanseverino, W., and Ferrante, M. I. (2018). Assessment of genomic changes in a CRISPR/Cas9 *Phaeodactylum tricornutum* mutant through whole genome resequencing. *PeerJ* 6, e5507. doi:10.7717/peerj.5507

Ryckebosch, E., Bruneel, C., Termote-Verhalle, R., Muylaert, K., and Foubert, I. (2014). Influence of extraction solvent system on extractability of lipid components from different microalgae species. *Algal Res.* 3, 36–43. doi:10.1016/j.algal.2013.11.001

Ryu, A. J., Jeong, B., Kang, N. K., Jeon, S., Sohn, M. G., Yun, H. J., et al. (2021). Safe-harboring based novel genetic toolkit for *Nannochloropsis salina* CCMP1776: efficient overexpression of transgene via CRISPR/Cas9-mediated knock-in at the transcriptional hotspot. *Bioresour. Technol.* 340, 125676. doi:10.1016/j.biortech.2021.125676

Sadvakasova, A. K., Kossalbayev, B. D., Bauenova, M. O., Balouch, H., Leong, Y. K., Zayadan, B. K., et al. (2023). Microalgae as a key tool in achieving carbon neutrality for bioproduct production. *Algal Res.* 72, 103096. doi:10.1016/j.algal.2023.103096

Saide, A., Martínez, K. A., Ianora, A., and Lauritano, C. (2021). Unlocking the health potential of microalgae as sustainable sources of bioactive compounds. *Int. J. Mol. Sci.* 22, 4383. doi:10.3390/ijms22094383

Sarwer, A., Hamed, S. M., Osman, A. I., Jamil, F., Al-Muhtaseb, A. H., Alhajeri, N. S., et al. (2022). Algal biomass valorization for biofuel production and carbon sequestration: a review. *Environ. Chem. Lett.* 20, 2797–2851. doi:10.1007/s10311-022-01458-1

Serif, M., Dubois, G., Finoux, A.-L., Teste, M.-A., Jallet, D., and Daboussi, F. (2018). One-step generation of multiple gene knock-outs in the diatom *Phaeodactylum tricornutum* by DNA-free genome editing. *Nat. Commun.* 9, 3924. doi:10.1038/s41467-018-06378-9

Shahabadi, H. Z., Akbarzadeh, A., Ofoghi, H., and Kadkhodaei, S. (2023). Site-specific gene knock-in and bacterial phytase gene expression in *Chlamydomonas reinhardtii* via Cas9 RNP-mediated HDR. *Front. Plant Sci.* 14, 1150436. doi:10.3389/fpls.2023.1150436

Sharma, A. K., Nymark, M., Flo, S., Sparstad, T., Bones, A. M., and Winge, P. (2021). Simultaneous knockout of multiple LHCF genes using single sgRNAs and engineering of a high-fidelity Cas9 for precise genome editing in marine algae. *Plant Biotechnol. J.* 19, 1658–1669. doi:10.1111/pbi.13582

Shin, S.-E., Lim, J.-M., Koh, H. G., Kim, E. K., Kang, N. K., Jeon, S., et al. (2016). CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*. *Sci. Rep.* 6, 27810–27815. doi:10.1038/srep27810

Shin, Y. S., Jeong, J., Nguyen, T. H. T., Kim, J. Y. H., Jin, E., and Sim, S. J. (2019). Targeted knockout of phospholipase A2 to increase lipid productivity in *Chlamydomonas reinhardtii* for biodiesel production. *Bioresour. Technol.* 271, 368–374. doi:10.1016/j.biortech.2018.09.121

Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., et al. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell* 60, 385–397. doi:10.1016/j.molcel.2015.10.008

Slattery, S. S., Wang, H., Giguere, D. J., Kocsis, C., Urquhart, B. L., Karas, B. J., et al. (2020). Plasmid-based complementation of large deletions in *Phaeodactylum tricornutum* biosynthetic genes generated by Cas9 editing. *Sci. Rep.* 10, 13879. doi:10.1038/s41598-020-70769-6

Solaymani, S. (2019). CO_2 emissions patterns in 7 top carbon emitter economies: the case of transport sector. *Energy* 168, 989–1001. doi:10.1016/j.energy.2018.11.145

Song, I., Kim, J., Baek, K., Choi, Y., Shin, B., and Jin, E. (2020). The generation of metabolic changes for the production of high-purity zeaxanthin mediated by CRISPR-Cas9 in *Chlamydomonas reinhardtii*. *Microb. Cell Fact.* 19, 220–229. doi:10.1186/s12934-020-01480-4

Song, I., Kim, S., Kim, J., Oh, H., Jang, J., Jeong, S. J., et al. (2022). Macular pigment-enriched oil production from genome-edited microalgae. *Microb. Cell Fact.* 21, 27. doi:10.1186/s12934-021-01736-7

Spicer, A., and Molnar, A. (2018). Gene editing of microalgae: scientific progress and regulatory challenges in Europe. *Biol. (Basel)* 7, 21. doi:10.3390/biology7010021

Sreenikethanam, A., Raj, S., Gugulothu, P., and Bajhaiya, A. K. (2022). Genetic engineering of microalgae for secondary metabolite production: recent developments, challenges, and future prospects. *Front. Bioeng. Biotechnol.* 10, 836056. doi:10.3389/fbioe.2022.836056

Stukenberg, D., Zauner, S., Dell'Aquila, G., and Maier, U. G. (2018). Optimizing CRISPR/Cas9 for the diatom *Phaeodactylum tricornutum*. *Front. Plant Sci.* 9, 740. doi:10.3389/fpls.2018.00740

Taparia, Y., Dolui, A. K., Boussiba, S., and Khozin-Goldberg, I. (2022). Multiplexed genome editing via an RNA polymerase II promoter-driven sgRNA array in the diatom *Phaeodactylum tricornutum*: insights into the role of *StLDP*. *Front. Plant Sci.* 12, 784780. doi:10.3389/fpls.2021.784780

Veeranagouda, Y., Debono Lagneaux, D., Fournet, H., Thill, G., and Didier, M. (2018). CRISPR-Cas9-edited site sequencing (CRES Seq): an efficient and high throughput method for the selection of CRISPR-Cas9-edited clones. *Curr. Protoc. Mol. Biol.* 121, 1–31. doi:10.1002/cpmb.53

Verruto, J., Francis, K., Wang, Y., Low, M. C., Greiner, J., Tacke, S., et al. (2018). Unrestrained markerless trait stacking in *Nannochloropsis gaditana* through combined genome editing and marker recycling technologies. *Proc. Natl. Acad. Sci.* 115, E7015–E7022. doi:10.1073/pnas.1718193115

Vogler, B. W., Ashford, A., and Posewitz, M. C. (2021). CRISPR/Cas9 disruption of *glucan synthase* in *Nannochloropsis gaditana* attenuates accumulation of β-1, 3-glucose oligomers. *Algal Res.* 58, 102385. doi:10.1016/j.algal.2021.102385

Wang, D., Zhang, C., Wang, B., Li, B., Wang, Q., Liu, D., et al. (2019). Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nat. Commun.* 10, 4284. doi:10.1038/s41467-019-12281-8

Wang, Q., Gong, Y., He, Y., Xin, Y., Lv, N., Du, X., et al. (2021). Genome engineering of Nannochloropsis with hundred-kilobase fragment deletions by Cas9 cleavages. *Plant J.* 106, 1148–1162. doi:10.1111/tpj.15227

Wang, Y.-Z., Hallenbeck, P. C., Leite, G. B., Paranjape, K., and Huo, D.-Q. (2016). Growth and lipid accumulation of indigenous algal strains under photoautotrophic and mixotrophic modes at low temperature. *Algal Res.* 16, 195–200. doi:10.1016/j.algal.2016.03.017

Wei, L., Jiang, Z., and Liu, B. (2022). A CRISPR/dCas9-based transcription activated system developed in marine microalga *Nannochloropsis oceanica*. *Aquaculture* 546, 737064. doi:10.1016/j.aquaculture.2021.737064

Xie, X., Yang, J., Du, H., Chen, J., Sanganyado, E., Gong, Y., et al. (2023). Golgi fucosyltransferase 1 reveals its important role in α-1, 4-fucose modification of N-glycan in CRISPR/Cas9 diatom *Phaeodactylum tricornutum*. *Microb. Cell Fact.* 22, 1–8. doi:10.1080/13607863.2023.2240261

Yang, W., Zhou, L., Wang, J., Wang, L., Gao, S., and Wang, G. (2022). Knockout of a diatom cryptochrome by CRISPR/Cas9 causes an increase in light-harvesting protein levels and accumulation of fucoxanthin. *Algal Res.* 66, 102822. doi:10.1016/j.algal.2022.102822

Yoneda, K., Oishi, R., Yoshida, M., Matsuda, Y., and Suzuki, I. (2023). Stramenopile-type lipid droplet protein functions as a lipid droplet scaffold protein in the marine diatom *Phaeodactylum tricornutum*. *Plant Cell Physiol.* 64, 803–813. doi:10.1093/pcp/pcad040

Zadabbas Shahabadi, H., Akbarzadeh, A., Ofoghi, H., and Kadkhodaei, S. (2023). Site-specific gene knock-in and bacterial phytase gene expression in *Chlamydomonas reinhardtii* via Cas9 RNP-mediated HDR. *Front. Plant Sci.* 14, 1150436. doi:10.3389/fpls.2023.1150436

Zhang, P., Xin, Y., He, Y., Tang, X., Shen, C., Wang, Q., et al. (2022). Exploring a blue-light-sensing transcription factor to double the peak productivity of oil in *Nannochloropsis oceanica*. *Nat. Commun.* 13, 1664. doi:10.1038/s41467-022-29337-x

Zhang, Y., Zhang, F., Li, X., Baller, J. A., Qi, Y., Starker, C. G., et al. (2013). Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiol.* 161, 20–27. doi:10.1104/pp.112.205179

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 06 November 2023
DOI 10.3389/fbioe.2023.1296132

OPEN ACCESS

EDITED BY
Tae Seok Moon,
Engineering Biology Research Consortium, EBRC, United States

REVIEWED BY
Dongsoo Yang,
Korea University, Republic of Korea
Sang Jun Lee,
Chung-Ang University, Republic of Korea
Dae-Hee Lee,
Korea Research Institute of Bioscience and Biotechnology (KRIBB), Republic of Korea

*CORRESPONDENCE
Taek Soon Lee,
tslee@lbl.gov

RECEIVED 18 September 2023
ACCEPTED 16 October 2023
PUBLISHED 06 November 2023

CITATION
Kim J and Lee TS (2023), Enhancing isoprenol production by systematically tuning metabolic pathways using CRISPR interference in *E. coli*.
Front. Bioeng. Biotechnol. 11:1296132.
doi: 10.3389/fbioe.2023.1296132

Enhancing isoprenol production by systematically tuning metabolic pathways using CRISPR interference in *E. coli*

Jinho Kim[1,2] and Taek Soon Lee[1,2]*

[1]Joint BioEnergy Institute, Emeryville, CA, United States, [2]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

Regulation of metabolic gene expression is crucial for maximizing bioproduction titers. Recent engineering tools including CRISPR/Cas9, CRISPR interference (CRISPRi), and CRISPR activation (CRISPRa) have enabled effective knock-out, knock-down, and overexpression of endogenous pathway genes, respectively, for advanced strain engineering. CRISPRi in particular has emerged as a powerful tool for gene repression through the use of a deactivated Cas9 (dCas9) protein and target guide RNA (gRNA). By constructing gRNA arrays, CRISPRi has the capacity for multiplexed gene downregulation across multiple orthogonal pathways for enhanced bioproduction titers. In this study, we harnessed CRISPRi to downregulate 32 essential and non-essential genes in *E. coli* strains heterologously expressing either the original mevalonate pathway or isopentenyl diphosphate (IPP) bypass pathway for isoprenol biosynthesis. Isoprenol remains a candidate bioproduct both as a drop-in blend additive and as a precursor for the high-performance sustainable aviation fuel, 1,4-dimethylcyclooctane (DMCO). Of the 32 gRNAs targeting genes associated with isoprenol biosynthesis, a subset was found to vastly improve product titers. Construction of a multiplexed gRNA library based on single guide RNA (sgRNA) performance enabled simultaneous gene repression, yielding a 3 to 4.5-fold increase in isoprenol titer (1.82 $\pm$ 0.19 g/L) on M9-MOPS minimal medium. We then scaled the best performing CRISPRi strain to 2-L fed-batch cultivation and demonstrated translatable titer improvements, ultimately obtaining 12.4 $\pm$ 1.3 g/L isoprenol. Our strategy further establishes CRISPRi as a powerful tool for tuning metabolic flux in production hosts and that titer improvements are readily scalable with potential for applications in industrial bioproduction.

KEYWORDS

CRISPR interference, isoprenol, multiplexed CRISPRi arrays, mevalonate pathway, IPP-bypass pathway, fed-batch cultivation

1 Introduction

Microbial host engineering is a promising strategy for improving advanced biofuel production and increasing sustainability in the energy sector. Advanced biofuels are petrochemical analogues typically derived by microbial hosts grown on non-food based feedstocks (e.g., lignin, waste cooking oil, or syngas) with comparably low lifecycle greenhouse gas emissions (Keasling et al., 2021). To date, microbial ethanol remains one of a handful of first-generation biofuels to achieve commercialization owing in part to its metabolic simplicity as a byproduct of anaerobic fermentation. However, ethanol has

significant physicochemical drawbacks relative to petroleum fuels, such that biofuel research has shifted towards more energy-dense molecules, including fatty acid methyl esters, higher alcohols, polyketides, and terpenes (Baral et al., 2021; Keasling et al., 2021).

Improving microbial titer, rate, and yield of more favorable and often more complex biofuels on recalcitrant carbon sources remains an outstanding challenge in metabolic engineering owing largely to the complexities of metabolic networks. A primary goal of metabolic engineering is to reroute metabolic flux towards a desired pathway while reducing inhibitory flux imbalances (Kim et al., 2017). Traditionally, *Eschericha coli* has been engineered by knockout of endogenous genes associated with competing byproducts or pathways via λ-Red recombineering (Datsenko and Wanner, 2000; Lee and Kim, 2015). However, consecutive gene knockouts are typically irreversible, strictly limited to non-essential genes, static, and laborious to multiplex. A more recent strategy for parallel gene editing is multiplex automated genome engineering (MAGE), which enables simultaneous modification of multiple genomic locations, including mismatches, insertions, and deletions (Wang et al., 2009). However, MAGE demands that hosts are deficient in DNA mismatch-repair and the frequency of variants harboring multiple phenotype improving mutations is often much lower than that of single-mutation variants (Wang et al., 2009; Gallagher et al., 2014). Furthermore, these strategies are often complicated by the unintuitive and nonobvious interplay between genetic expression owing to regulatory elements, enzyme promiscuity, substrate toxicity, and their collective impacts on targeted production.

The recent discovery of clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein (Cas) has rapidly advanced precise RNA-guided genome engineering. Unlike previous genome engineering tools such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), CRISPR/Cas9 systems enable efficient deletion, insertion, or modification at a target locus (Cho et al., 2018). Modifications of the canonical CRISPR/Cas9 system have enabled trans-acting gene modulation, namely, CRISPR activation (CRISPRa) and CRISPR interference (CRISPRi) systems for effective upregulation and downregulation of target genes for advanced strain engineering, respectively (Chae et al., 2017; Tan and Prather, 2017). In particular, CRISPRi utilizes a deactivated Cas9 (dCas9) protein and a guide RNA (gRNA) to downregulate transcription of a target gene without knockout (Wu et al., 2015; Kim et al., 2016; Tian et al., 2019). Consequently, CRISPRi systems can reversibly and dynamically modulate expression of both non-essential and essential genes to elucidate their relative impacts on a given production pathway. Downregulation of selected genes using CRISPRi has contributed to significant bioproduct titer improvements, including dyes (Kim et al., 2016; Czajka et al., 2022; Cho et al., 2023), flavonoids (Tao et al., 2018), nutraceuticals (Wu et al., 2017), and biofuels (Tian et al., 2019; Wang et al., 2022).

One such biofuel, isoprenol (3-methyl-3-buten-1-ol), has arisen as a promising renewable intermediate for high-volume bioblendstocks (Foo et al., 2014). Isoprenol has attractive physicochemical properties including higher energy content, lower water miscibility, hygroscopicity, and volatility compared to ethanol. Isoprenol can be acetylated to generate isoprenyl acetate, an octane boosting fuel blend additive (Carruthers et al., 2023). Furthermore, isoprenol can be easily converted to isoprene (by catalytic dehydration). Isoprene is a well-known precursor of synthetic rubber and recently it is identified as a precursor for the high-performance jet fuel blend additive 1,4-dimethylcyclooctane (DMCO) (Rosenkoetter et al., 2019; Baral et al., 2021).

Isoprenol production in *Escherichia coli* may be accomplished via either the native methylerythritol 4-phosphate (MEP) pathway or heterologous mevalonate (MVA) pathway. Although metabolically distinct, both pathways ultimately generate isopentenyl diphosphate (IPP), a universal precursor to isoprenoid biosynthesis. Sequential dephosphorylation of IPP to isoprenol is accomplished by the promiscuous activity of one or more endogenous *E. coli* phosphatases (George et al., 2015). Many subsequent strategies have attempted to eliminate metabolic bottlenecks by tuning phosphatase and MVA pathway expression to improve isoprenol biosynthesis (George et al., 2015). However, IPP accumulation by the heterologous MVA pathway remained a significant bottleneck due to its inhibition of cell growth and deleterious impact on isoprenol production (George et al., 2018). To address IPP toxicity, a phosphomevalonate decarboxylase (PMD) enzyme with promiscuous activity towards mevalonate monophosphate (MVAP) was adapted to bypass the formation of IPP, thereby avoiding toxic accumulation, reducing pathway ATP consumption, and ultimately enhancing overall isoprenol titers by 2.4-fold compared to wild-type (Kang et al., 2017; 2019). Further optimization of this "IPP-bypass" pathway by expression of gene variants achieved titers of 3.7 g/L and 10.8 g/L isoprenol under batch and fed-batch conditions at scale, respectively, which is the highest titers reported so far (Kang et al., 2019; Tian et al., 2019).

CRISPRi has been successfully used to modulate transcription of genes in the original MVA isoprenol biosynthesis pathway to improve titers by 98% (Tian et al., 2019) and, more recently, to downregulate *E. coli* genes in pathways competing for isoprenol precursors (Wang et al., 2022). While these studies improved isoprenol titer, they used rich medium, which makes it difficult for industrial scale-up, reshapes the cellular metabolic landscape dramatically, and often yields disparate sets of beneficial CRISPRi target genes.

In this study, we harnessed CRISPRi methods to downregulate an expanded pool of endogenous genomic *E. coli* genes to improve isoprenol production titers in minimal medium via both the original MVA pathway and the IPP-bypass pathway. We then constructed multiplexed gRNA arrays leveraged for enhanced isoprenol biosynthesis with the best combinations yielding a 3 to 4.5-fold improvement in isoprenol via the IPP-bypass pathway. Finally, we show that the CRISPRi platform and the resultant titer improvements were scalable to 2-L bioreactors under fed-batch conditions, ultimately achieving 12.4 ± 1.3 g/L isoprenol in minimal medium (Figure 1). Broadly, our isoprenol titer improvements further establish CRISPRi as a powerful tool for microbial metabolic engineering.

2 Materials and methods

2.1 Plasmids and strains

The pdCas9-Marraffini plasmid (JBEI-18706) served as a template for our CRISPRi system (Bikard et al., 2013; Tian et al.,

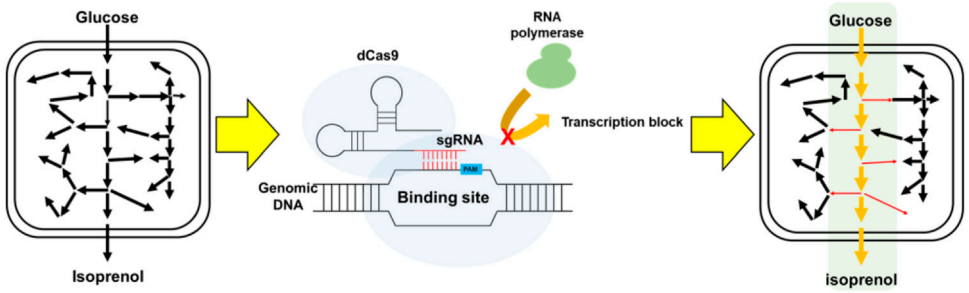


FIGURE 1
Schematic for the improvement of isoprenol biosynthesis by CRISPRi-mediated downregulation of competing pathways.

2019). The plasmid encodes a catalytically inactive dCas9 from *Streptococcus pyogenes* driven by an anhydrotetracycline (aTc) inducible promoter (P_{tet}) along with a nontargetting spacer (5′-TGAGACCAGTCTCGGAAGCTCAAAGGTCTC-3′). The spacer is flanked by BsaI cut sites for integration of gRNAs via golden gate assembly (NEBridge Golden Gate Assembly Kit, New England Biolabs, United Kingdom). Target sgRNAs were designed using Benchling (http://benchling.com) and screening for PAM sites that were 1) on the non-template DNA strand, 2) proximal to the start codon or promoter region, and 3) unlikely to cause off-target effects. Selected 30 bp sgRNAs were integrated under the constitutive promoter pJ23119 (Tian et al., 2019). Multiplexed gRNA arrays were then constructed with gRNAs flanked by repeat sequences (5′-GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC-3′).

E. coli DH1 was selected as a production strain owing to demonstrated high titer production of isoprenol via either the original MVA pathway (George et al., 2018) or the IPP-bypass pathway (Kang et al., 2019).

All plasmids and strains used in this study are listed in Table 1 with gRNA nucleotide sequences listed in Supplementary Table S1. The plasmids and strains were deposited and are publicly available in the JBEI Registry (http://public-registry.jbei.org). All results in this study were also deposited into the Experiment Data Depot (EDD, http://edd.jbei.org).

2.2 Batch production of isoprenol

All batch and fed-batch isoprenol production experiments were completed in M9-MOPS defined medium Briefly, M9-MOPS minimal medium contains M9 salts (33.9 g/L Na_2HPO_4, 15 g/L KH_2PO_4, 5 g/L NH_4Cl, and 2.5 g/L NaCl; Sigma-Aldrich) supplemented with 75 mM MOPS, 2 mM $MgSO_4$, 1 mg/L thiamine, 10 nM $FeSO_4$, 0.1 mM $CaCl_2$, and micronutrients (Korz et al., 1995) including $3*10^{-8}$ M $(NH_4)_6Mo_7O_{24}$, $4*10^{-6}$ M boric acid, $3*10^{-7}$ M $CoCl_2$, $1*10^{-7}$ M $CuSO_4$, $8*10^{-7}$ M $MnCl_2$, and $1*10^{-7}$ M $ZnSO_4$ with 20 g/L of glucose as the sole carbon source. Three antibiotics (100 mg/L carbenicillin, 30 mg/L chloramphenicol, and 100 mg/L kanamycin) were used for selection where necessary (Kang et al., 2019; Tian et al., 2019).

E. coli DH1 was transformed with either the original MVA or IPP-bypass pathway plasmids as well as a dCas9-gRNA plasmid via electroporation (2500 V, 5 ms) in 2 mm gap cuvettes (Bio Rad), recovered at 37°C for 1 hour in SOC medium, then plated on antibiotic for overnight outgrowth. A successful transformant was inoculated into Luria-Bertani (LB) liquid medium with antibiotic at 37°C overnight for each transformation. When preparing the glycerol stock, we inoculated a significant tiny single colony from the agar plate into LB culture medium, then the cell was cultured to ensure they did not exceed OD_{600} of 2 in the LB medium. Cultures were passaged 50-fold (v/v) in fresh M9-MOPS medium, grown overnight, and passaged again into fresh M9-MOPS medium to ensure adaptation from complex to minimal medium. Adapted cells were stored as frozen glycerol cryo stocks at −80°C. The OD_{600} of glycerol stock is approximately 2, and we uniformly conducted all experiments by inoculating this strain into fresh M9-MOPS minimal medium with an initial OD_{600} of 0.2 to minimize the genetic mutations.

Cells were recovered from glycerol cryo stocks as needed by direct inoculation into glass culture tubes containing 5 mL M9-MOPS medium supplemented with 20 g/L glucose and grown at 37°C overnight. Overnight cultures were diluted 50-fold (v/v) into 5 mL fresh M9-MOPS medium and grown for 12 h at 37°C. Finally, cells were inoculated at an optical density (OD; λ = 600 nm, 1 cm path length) of 0.2 into 5 mL fresh M9-MOPS medium. When culture OD_{600} was between 0.6 and 0.8, isoprenol biosynthesis and dCas9 expression were simultaneously induced by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 10 nM anhydrotetracycline (aTc), respectively. Induced cultures were then grown at 30°C for 72 h. Samples were taken at 48 h and 72 h to measure strain growth rates using OD_{600} (SpectraMax 384, Molecular Devices) and isoprenol titer (Focus GC-FID, Thermo Scientific).

2.3 Isoprenol production in fed-batch cultivation

As in batch experiments, cells from a frozen glycerol stock were inoculated into 2 mL M9-MOPS medium and grown for 12 h at 37°C, then diluted 50-fold into 5 mL fresh M9-MOPS medium for overnight growth. The seed culture was then passed into 100 mL M9-MOPS medium, which was grown for 8 h at 37°C and transferred to 900 mL fresh M9-MOPS medium in a 2-L bioreactor vessel at an OD_{600} of 0.3.

Fed-batch cultivation was conducted using a 2-L bioreactor (Sartorius BIOSTAT B plus) with control of dissolved oxygen

TABLE 1 List of base strains and plasmids used in this study.

Strains	Description	Reference
Isoprenol pathway strains		
E. coli DH1	F^- supE44 hsdR17 recA1 gyrA96 relA1 endA1 thi^{-1} lambda$^-$	Wild type
OriMVA	*E. coli* DH1 harboring JBEI-6829 and JPUB_004507	George et al. (2018)
IPP-By	*E. coli* DH1 harboring JBEI-9310 and JBEI-14559	Kang et al. (2019)
Plasmids	**Description**	**Reference**
Isoprenol pathway plasmids		
OriMVA-Upper	pLacUV5-atoB-HMGS_Sa-HMGR_Sa-MKco-PMKco, p15a, CmR	George et al. (2018)
OriMVA-Lower	pTrc99A-NudBR10-PMD_Sc, pBR322, AmR	George et al. (2018)
IPPBy-Upper	pLacUV5-atoB-HMGS_Sa-HMGR_Sa, p15a, CmR	Kang et al. (2019)
IPPBy-Lower	pTrc-PMD_R74H-R147K-M212Q, pBR322, AmR	Kang et al. (2019)
Single gRNA plasmids (pSC101 origin, KanR, dCas9 under pTet promoter, gRNA under pJ23119)		
Non-target Control (JBEI-18706)	pTet-dCas9-pJ23119-NT_Control, pSC101, KanR	Tian et al. (2019)
accA	pTet-dCas9-pJ23119-accA, pSC101, KanR	Tian et al. (2019)
ackA	pTet-dCas9-pJ23119-ackA, pSC101, KanR	Tian et al. (2019)
adhE	pTet-dCas9-pJ23119-adhE, pSC101, KanR	This study
arcC	pTet-dCas9-pJ23119-arcC, pSC101, KanR	Tian et al. (2019)
asnA	pTet-dCas9-pJ23119-asnA, pSC101, KanR	Tian et al. (2019)
citE	pTet-dCas9-pJ23119-citE, pSC101, KanR	This study
dacA	pTet-dCas9-pJ23119-dacA, pSC101, KanR	Tian et al. (2019)
deoB	pTet-dCas9-pJ23119-deoB, pSC101, KanR	Tian et al. (2019)
eutD	pTet-dCas9-pJ23119-eutD, pSC101, KanR	Tian et al. (2019)
fabH	pTet-dCas9-pJ23119-fabH, pSC101, KanR	This study
gldA	pTet-dCas9-pJ23119-gldA, pSC101, KanR	Tian et al. (2019)
ispA	pTet-dCas9-pJ23119-ispA, pSC101, KanR	Tian et al. (2019)
ldhA	pTet-dCas9-pJ23119-ldhA, pSC101, KanR	This study
mdh	pTet-dCas9-pJ23119-mdh, pSC101, KanR	Tian et al. (2019)
menA	pTet-dCas9-pJ23119-menA, pSC101, KanR	This study
mqo	pTet-dCas9-pJ23119-mqo, pSC101, KanR	Tian et al. (2019)
pgl	pTet-dCas9-pJ23119-pgl, pSC101, KanR	This study
poxB	pTet-dCas9-pJ23119-poxB, pSC101, KanR	This study
ppc	pTet-dCas9-pJ23119-ppc, pSC101, KanR	Tian et al. (2019)
ppsA	pTet-dCas9-pJ23119-ppsA, pSC101, KanR	This study
prpE	pTet-dCas9-pJ23119-prpE, pSC101, KanR	Tian et al. (2019)
pta	pTet-dCas9-pJ23119-pta, pSC101, KanR	Tian et al. (2019)
rssA	pTet-dCas9-pJ23119-rssA, pSC101, KanR	This study
sdhABCD	pTet-dCas9-pJ23119-sdhABCD, pSC101, KanR	This study
sucA	pTet-dCas9-pJ23119-sucA, pSC101, KanR	This study
sucB	pTet-dCas9-pJ23119-sucB, pSC101, KanR	This study

(Continued on following page)

TABLE 1 (*Continued*) List of base strains and plasmids used in this study.

Strains	Description	Reference
sucCD	pTet-dCas9-pJ23119-sucCD, pSC101, KanR	This study
thrC	pTet-dCas9-pJ23119-thrC, pSC101, KanR	Tian et al. (2019)
ubiA	pTet-dCas9-pJ23119-ubiA, pSC101, KanR	This study
yahl	pTet-dCas9-pJ23119-yahl, pSC101, KanR	Tian et al. (2019)
yhfW	pTet-dCas9-pJ23119-yhfW, pSC101, KanR	Tian et al. (2019)
yqeA	pTet-dCas9-pJ23119-yqeA, pSC101, KanR	Tian et al. (2019)
Double gRNA plasmids (pSC101 origin, KanR, dCas9 under pTet promoter, gRNA's under pJ23119)		
fabH-ubiA	pTet-dCas9-pJ23119-fabH-ubiA, pSC101, KanR	This study
adhE-menA	pTet-dCas9-pJ23119-adhE-menA, pSC101, KanR	This study
ldlhA-ubiA	pTet-dCas9-pJ23119-ldhA-ubiA, pSC101, KanR	This study
ldhA-fabH	pTet-dCas9-pJ23119-ldhA-fabH, pSC101, KanR	This study
ldhA-menA	pTet-dCas9-pJ23119-ldhA-menA, pSC101, KanR	This study
adhE-fabH	pTet-dCas9-pJ23119-adhE-fabH, pSC101, KanR	This study
ldhA-adhE	pTet-dCas9-pJ23119-ldhA-adhE, pSC101, KanR	This study
adhE-ubiA	pTet-dCas9-pJ23119-adhE-ubiA, pSC101, KanR	This study
fabH-menA	pTet-dCas9-pJ23119-fabH-menA, pSC101, KanR	This study
menA-ubiA	pTet-dCas9-pJ23119-menA-ubiA, pSC101, KanR	This study
ppsA-adhE	pTet-dCas9-pJ23119-ppsA-adhE, pSC101, KanR	This study
ppsA-fabH	pTet-dCas9-pJ23119-ppsA-fabH, pSC101, KanR	This study
ppsA-ldhA	pTet-dCas9-pJ23119-ppsA-ldhA, pSC101, KanR	This study
ppsA-menA	pTet-dCas9-pJ23119-ppsA-menA, pSC101, KanR	This study
ppsA-ubiA	pTet-dCas9-pJ23119-ppsA-ubiA, pSC101, KanR	This study
Triple gRNA plasmids		
adhE-fabH-ldhA	pTet-dCas9-pJ23119-adhE-fabH-ldhA, pSC101, KanR	This study
adhE-fabH-menA	pTet-dCas9-pJ23119-adhE-fabH-ldhA, pSC101, KanR	This study
adhE-fabH-ubiA	pTet-dCas9-pJ23119-adhE-fabH-ldhA, pSC101, KanR	This study
menA-ubiA-fabH	pTet-dCas9-pJ23119-menA-ubiA-fabH, pSC101, KanR	This study

(DO), temperature, and pH. DO and airflow were set to 30% and 1 VVM (volume of air per volume of liquid per minute), respectively, with temperature maintained at 30°C and pH held at 6.5 by supplementation with 25% ammonia water. The CRISPRi system was induced by addition of 10 nM aTc and isoprenol biosynthesis was induced by addition of 0.5 mM IPTG as in batch culture. For fed-batch mode, a mixture of 80 g/L glucose and 15 g/L ammonium chloride was supplied using a Watson-Marlow DU520 peristaltic pump. After the lag phase, the feeding flow rate was calculated following Korz's equation and increased every hour for a total of 6 h as described by Equation 1 (Korz et al., 1995; Kang et al., 2019).

$$m_s(t) = \left(\frac{\mu}{Y_{\frac{x}{s}}} + m\right) V_{t_F} * X_{t_F} * e^{\mu * (t - t_F)} \quad (1)$$

Here, $m_s(t)$ is the flow of glucose (g/hr) and μ represents the specific growth rate (hr^{-1}) of the *E. coli* strain. V_{tF} is the cultivation volume (L) and X_{tF} is biomass (g/L) at feeding time t_F (hr), while $Y_{x/s}$ is the yield of biomass on substrate ($g_{biomass}/g_{glucose}$) and m is the specific maintenance coefficient ($g_{biomass}/g_{glucose}$/hr). Glucose concentration was measured consistently using a glucose meter (CVSHealth) and high-performance liquid chromatography (HPLC) during exponential feeding. Following exponential feeding, a feed rate was selected that closely matched glucose consumption such that the glucose concentration was less than 1 g/L.

A 20% oleyl alcohol overlay was directly added to the vessel at the time of induction, serving both to extract isoprenol and to mitigate isoprenol evaporation due to bioreactor sparging. Antifoam B was added during the fed-batch phase as necessary.

2.4 Quantification of isoprenol

Isoprenol was quantified from both the aqueous culture medium as well as the organic oleyl alcohol overlay. For aqueous extraction, a 250 µL aliquot of cell culture was vigorously mixed with 250 µL of ethyl acetate containing 1-butanol (30 mg/L) internal standard by vortexing at 3000 RPM for 15 min. After vortexing, the mixture was centrifuged at 14,000 x g for 2 min to separate organic and aqueous phases. A 100 µL aliquot of the ethyl acetate was then diluted 5-fold into fresh ethyl acetate with internal standard for analysis using gas chromatography with flame ionization detection (GC-FID; Focus GC-FID, Thermo Scientific) with 1 µL injection volume into a DB-WAX column (15 m, 0.32 mm inner diameter, 0.25 µm film thickness, Agilent). The GC-FID oven temperature was initially set to 40°C, then ramped to 100 °C at a rate of 15 °C/min, ramped to 230°C at a rate of 40°C/min, and finally held at 230°C for 3 min.

Organic phase isoprenol concentration was determined by sampling 250 µL of the oleyl alcohol overlay and centrifuging at 14,000 x g for 5 min to separate any aqueous cell culture. Then 10 µL of the overlay was added to 990 µL ethyl acetate with an internal standard for GC-FID analysis. Isoprenol titers under fed-batch conditions with overlay were calculated on the basis of actual culture volume at the time of the sampling.

2.5 Quantification of glucose, acetate, and ethanol

High performance liquid chromatography (HPLC; Agilent Technologies) was performed to quantify glucose, ethanol, and acetate in aqueous cultures. Aqueous culture samples were centrifuged at 14,000 x g for 5 min to separate supernatant and cells. Following centrifugation, 100 µL of supernatant was filtered (0.45 µm centrifugal filter) and analyzed using an HPLC system equipped with an Aminex HPX-87H column (Bio-rad) and a refractive index detector. The sample tray temperature was set to 10°C and the column oven temperature was set to 65°C with a 0.005 M H_2SO_4 mobile phase flow rate of 600 µL/min. Data acquisition and analysis were performed via Agilent Chemstation.

3 Results

3.1 Preparing isoprenol production pathways

Isoprenol was produced by strains harboring either the original MVA pathway (George et al., 2015) or the IPP-bypass MVA pathway (Kang et al., 2019; 2017) as shown in Figure 2B. In brief, both pathway variants share 4 genes (*atoB*, *HMGS*, *HMGR*, and *MK*) which convert the central metabolite acetyl-CoA to mevalonate phosphate (MVAP) (Supplementary Figure S2). The original MVA pathway forms isopentenyl diphosphate (IPP) from MVAP by mevalonate phosphate kinase (PMK) and phosphomevalonate decarboxylase (PMD), and the IPP is hydrolyzed to isoprenol sequentially by native promiscuous phosphatases (NudB and other monophosphatase).

The accumulation of IPP is toxic to the production host, however, (George et al., 2018), and the IPP-bypass pathway was designed to produce isoprenol using promiscuous activity of PMD without forming IPP (Kang et al., 2016). In the IPP-bypass pathway strain, the expression of 4 genes of the MVA pathway (*atoB*, *HMGS*, *HMGR*, and *MK*) was controlled under the P_{lacUV5} on the plasmid 1 and the evolved PMD mutant (triple mutants, R74H-R147K-M212Q) (Kang et al., 2017) was overexpressed under P_{trc} on the plasmid 2 (Supplementary Figure S1).

3.2 Designing CRISPRi system and sgRNA library of target genes

In this study, we used the CRISPRi strategy to identify genes associated with improved isoprenol production by attenuating gene expression of the competing metabolic pathways. The dCas9-sgRNA system is a simple and effective method for partial or completed gene downregulation by blocking transcription. The SC101 origin and tetracycline-inducible promoter were employed for high efficiency stability of the plasmid and dCas9 expression (Tian et al., 2019). The use of the inducible promoter system has allowed for the selective inhibition of genes, even for those essential genes, and this work has also demonstrated that the effects of targeted gene suppression last for an extended period (at least up to 72 h). We designed to induce the heterologous isoprenol biosynthesis pathway concurrently with the CRISPRi induction, so that endogenous target genes in competing pathways are downregulated in parallel with the pathway overexpression for the isoprenol production.

To maximize the efficiency of the CRISPRi system, we designed the sgRNAs to target non-template DNA strands of the specific target gene after the protospacer adjacent motif (PAM) sequence (5′-NGG-3′) at the open reading frame (ORF) or promoter region as previously reported (Tian et al., 2019). We selected 32 genes as targets for downregulation as they competitively use precursors, cofactors or the intermediates of the MVA pathway (i.e., acetyl-CoA, pyruvate precursors, and cofactors) (Figure 2A). A subselection of these 32 candidate target genes have been downregulated by CRISPRi to improve isoprenol titer in a previous study using a complex or rich medium and only the original MVA pathway (Tian et al., 2019). First, we selected 15 gene targets that compete with isoprenol biosynthesis for acetyl-CoA and pyruvate precursors, including *accA* (acetyl-coenzyme A carboxyl transferase), *ackA* (acetate kinase), *adhE* (aldehyde alcohol dehydrogenase), *citE* (citrate lyase), *fabH* (3-oxoacyl-acyl-carrier protein synthase), *ldhA* (D-lactate dehydrogenase), *mdh* (malate dehydrogenase), *poxB* (pyruvate dehydrogenase), *ppc* (phosphoenolpyruvate carboxylase), *ppsA* (phosphoenolpyruvate synthase), *pta* (phosphate acetyltransferase), *sdhABCD* (succinate dehydrogenase), *sucA* (2-oxoglutarate dehydrogenase), *sucB* (2-oxoglutarate dehydrogenase), and *sucCD* (succinyl-CoA synthetase) (Wu et al., 2015; Kim et al., 2016; Tian et al., 2019). In addition, *asnA* (aspartate ammonia ligase), *gldA* (glycerol dehydrogenase), *pgl* (6-phosphogluconolactonase), and *prpE* (propionyl-CoA synthase) were also chosen as targets to enhance production of isoprenol as tested in the previous study (Tian et al., 2019). We added 3 genes, *ispA* (farnesyl diphosphate synthase), *menA* (1,4-dihydroxy-2-naphthoate octaprenyltransferase), and *ubiA* (4-hydroxybenzoate octaprenyltransferase) that compete for

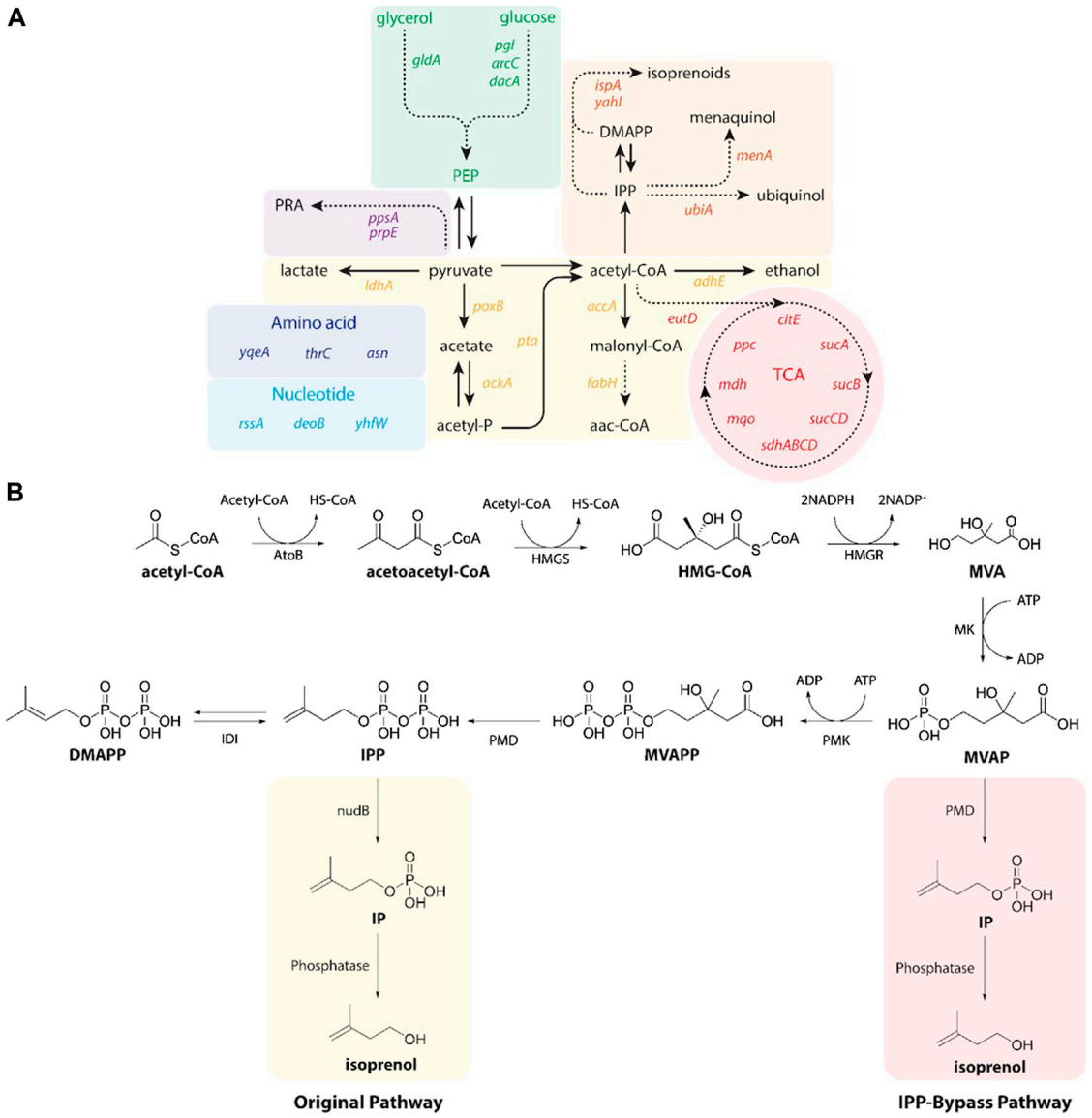


FIGURE 2
(A) A pathway map of gene knockdown targets for redirecting metabolic flux toward isoprenol production using CRISPRi. Pathways and relevant genes are colored as follows: green box, glycolysis; purple box, amino acids pathway; violet box, propionic acid pathway; light blue box, nucleotides pathway; yellow box, organic acids pathway; orange box, isoprenoids pathway; red box, TCA cycle pathway. PRA, propionic acid; PEP, phosphoenolpyruvate; TCA, citric acid cycle. **(B)** Original (yellow box) and IPP-bypass (red box) MVA pathways for isoprenol production. The MVA pathway converts acetyl-CoA to IPP in 6 enzymatic steps with subsequent isomerization to dimethylallyl diphosphate (DMAPP). Dephosphorylation of IPP and DMAPP by nudB, a promiscuous *E. coli* phosphatase, produces isoprenol and prenol, respectively. The IPP-bypass pathway was proposed in this study for direct decarboxylation of mevalonate monophosphate (MVAP) to isopentenyl monophosphate (IP) followed by dephosphorylation by an endogenous phosphatase. ADP, adenosine diphosphate; AtoB, acetoacetyl-CoA thiolase; ATP, adenosine triphosphate; DMAPP, dimethylallyl diphosphate; HMG-CoA, 3-hydroxy-3-methyl-glutaryl-CoA; HMGR, HMG-CoA reductase; HMGS, HMG-CoA synthase; IDI, isopentenyl diphosphate isomerase; IP, isopentenyl monophosphateIPP, isopentenyl diphosphate; MK, mevalonate kinase; MVA, mevalonate; MVAP, mevalonate monophosphate; MVAPP, mevalonate diphosphate; PMD, mevalonate diphosphate kinase; PMK, phosphomevalonate kinase.

IPP utilization (Kong and Lee, 2011; Zada et al., 2018) and selected additional 10 genes including *araC* (arabinose regulator), *dacA* (alanine carboxypeptidase), *deoB* (deoxyribose), *eutD* (phosphotransacetylase), *mqo* (quinone-oxidoreductase), *rssA* (sigma S regulator protein), *thrC* (threonine synthase), *yahI* (polypeptide carbamate kinase), *yhfW* (mutase), and *yqeA* (carbamate kinase) that were previously identified as knockout targets in the lycopene and carotenoid production via the MEP pathway (Alper et al., 2005; Alper and Stephanopoulos, 2008; Tian et al., 2019). A detailed metabolic map of these genes in their cellular context is depicted in Supplementary Figure S2.

3.3 Knockdown of single target genes in *E. coli* harboring the MVA pathway

When applying the CRISPRi system to downregulate metabolic genes and enhance isoprenol production, the selection of an

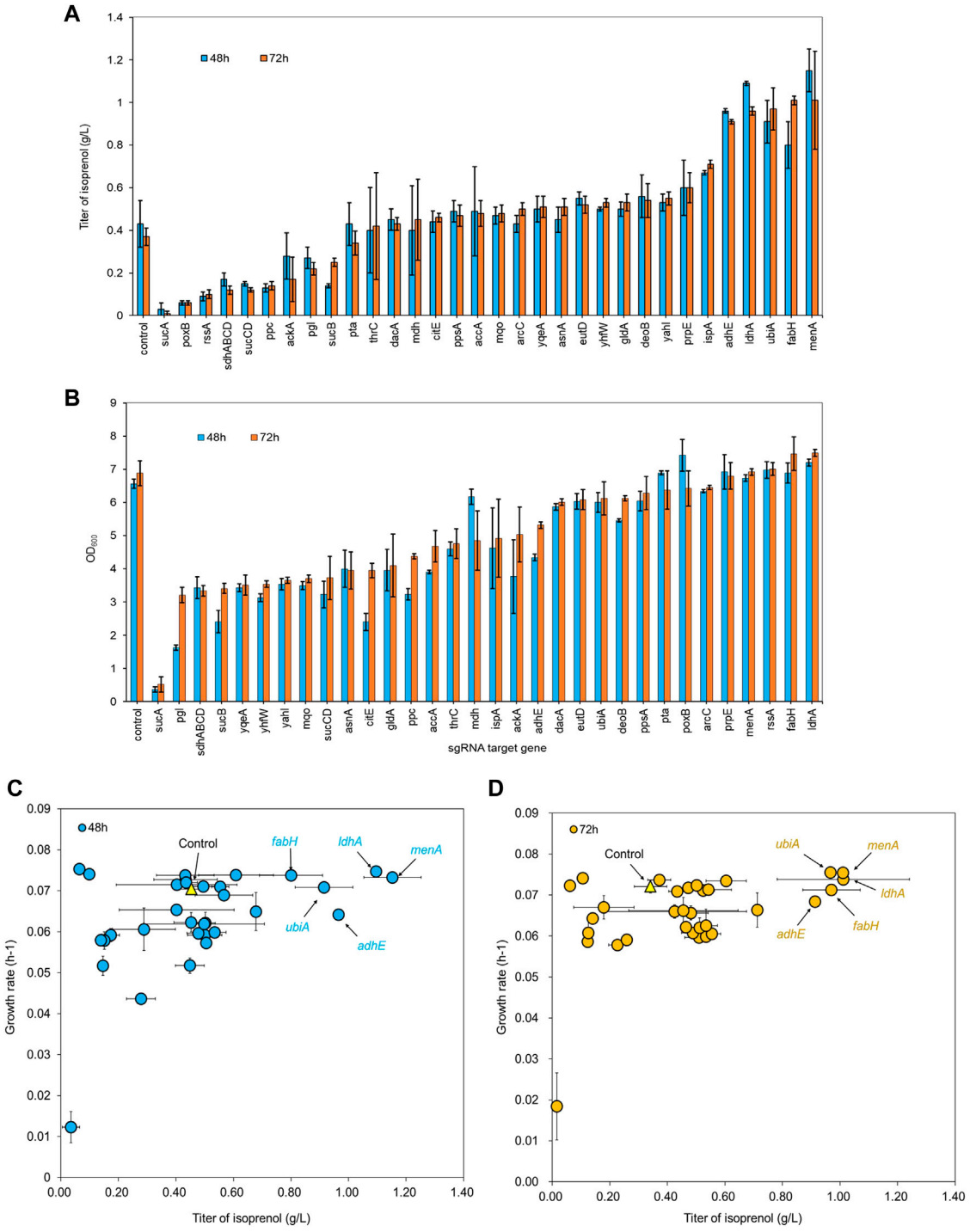


FIGURE 3
(A) Growth and **(B)** isoprenol production by strains harboring single gRNAs and the original MVA pathway. Batch cultures were grown in 5 mL M9-MOPS minimal medium supplemented with 20 g/L of glucose (n = 3). **(C)** Scatter plots of growth rate and isoprenol titer at 48 h and **(D)** 72 h show that selected gRNA targets were leveraged for high isoprenol production while maintaining a high growth rate relative to the control. The control harboring non-targeting gRNA plasmid JBEI-18706 is demarcated with a bright yellow triangle.

appropriate growth medium could be a pivotal determinant. Downregulating the expression of essential genes impedes the metabolic pathway. *E. coli*'s response to this, however, could be hampered by its nutrient-absorbing behavior if a complex rich medium is used for its growth (Liu et al., 2020). To minimize this effect, we used a minimal medium for all of our experiments.

We transformed a library of 32 sgRNA-dCas9 plasmids into *E. coli* DH1 strains harboring the original MVA pathway to explore

whether CRISPRi-mediated downregulation of selected target genes could improve isoprenol titer in minimal medium (Alper et al., 2005; Alper and Stephanopoulos, 2008; Kong and Lee, 2011; Wu et al., 2015; Kim et al., 2016; Zada et al., 2018; Tian et al., 2019). The growth rates of the strains were calculated to assess the relative burden of gene downregulation against a non-targeting control. Overall, the result showed that 22 out of 32 downregulation experiments for individual sgRNA targets resulted in higher isoprenol titers compared to the control strain harboring non-targeting gRNA plasmid JBEI-18706 (Figure 3A). Notably, sgRNAs targeting *adhE, ldhA, ubiA, fabH,* and *menA* yielded improved productions to 245.40%, 259.84%, 260.85%, 271.68%, and 271.94%, respectively, at isoprenol titers of 0.91 ± 0.01 g/L, 0.96 ± 0.02 g/L, 0.97 ± 0.10 g/L, 1.01 ± 0.02 g/L and 1.00 ± 0.23 g/L, respectively (Figure 3A).

Interestingly, the strains with the highest isoprenol titer harbor sgRNAs targeting orthogonal pathways: *ldhA* and *adhE* are involved in lactate and ethanol production, respectively, *fabH* is involved in fatty acid biosynthesis, and *ubiA* and *menA* utilize isoprenoid precursors for cellular respiration. Broadly speaking, these pathways are directly or indirectly involved in the consumption of the precursors used in the biosynthesis of isoprenol. For example, downregulation of *adhE, ldhA,* and *fabH* could reduce flux toward corresponding competing pathways and channel more precursor towards isoprenol biosynthesis. This is supported by the ethanol and lactate levels which were not detected in the culture medium when *adhE* and *ldhA* were downregulated while 0.8 ± 0.03 g/L of ethanol and 0.4 ± 0.03 g/L of lactate were detected in the control strain (Supplementary Figure S3).

Likewise, downregulation of *ldhA*, *adhE*, and *fabH* could also affect overall cellular metabolism and redox balance. These genes are involved in metabolic pathways that generate reducing equivalents (NADH or NADPH) or consume ATP, and the downregulation of these genes could result in altered metabolic fluxes, cellular redox state, and energy availability. It is important to note that unintended consequences of sgRNA-mediated gene downregulation could also be a contributing factor to the observed increase in isoprenol titer. Off-target effects or secondary effects on other cellular processes could indirectly impact isoprenol biosynthesis. Further studies, such as transcriptomic, proteomic, or metabolomic analyses, could provide a comprehensive understanding of the effects of *ldhA*, *adhE*, and *fabH* downregulation on cellular metabolism and isoprenol production.

In addition, *menA* and *ubiA* downregulation resulted in a significant increase in isoprenol titer. *MenA* is involved in the biosynthesis of menaquinone, an important cofactor in the electron transport chain, while *UbiA* is an essential enzyme in the ubiquinone biosynthesis pathway (Goodall et al., 2018). Both menaquinone and ubiquinone, which are vital components for cellular respiration and crucial to energy production, are derived from IPP, and downregulation of *menA* and *ubiA* could increase IPP availability towards isoprenol biosynthesis without compromising necessary function.

Conversely, sgRNAs targeting glycolysis pathway genes *rssA*, *ppc*, and *pgl* yielded a dramatic decrease in isoprenol production. In addition to decreasing isoprenol production, sgRNAs targeting TCA cycle genes *sucA*, *sucB*, *sucCD*, and *sdhABCD* also resulted in a significant growth inhibition (Figure 3B). This could be due to the repression of genes involved in sugar utilization, which may limit the availability of energy and precursor molecules for isoprenol biosynthesis. Alternatively, it's possible that these sgRNAs caused off-target effects or unintended consequences on cellular metabolism, leading to reduced isoprenol production, but more studies with extensive omics data will be needed to verify this.

In summary, the results suggest that sgRNAs targeting *adhE, ldhA, ubiA, fabH, and menA* provide a net benefit to isoprenol production via the MVA pathway and therefore may be valuable targets for further engineering efforts. On the other hand, those targeting TCA cycle and glycolysis pathway genes are strictly detrimental to the isoprenol titer improvement.

An overall comparison of the growth rates and isoprenol production at 48 h (Figure 3C) and 72 h (Figure 3D) across the 32 sgRNA experiments indicated these 5 genes with higher isoprenol titer and comparable growth rate to the control. We chose these five genes to serve as candidates for multiplexed gRNA arrays to elucidate potential synergistic effects to improve isoprenol titers.

3.4 Multiple gRNA arrays co-expression with the original MVA pathway

We created gRNA arrays by assembling five individual sgRNAs (*adhE, ldhA, ubiA, fabH, and menA*) into 10 double gRNA arrays (*adhE/ldhA*, *adhE/ubiA*, *adhE/fabH*, *adhE/menA*, *ldhA/ubiA*, *ldhA/fabH*, *ldhA/menA*, *ubiA/fabH*, *ubiA/menA*, and *fabH/menA*) to conduct a repression of multiple endogenous genes in *E. coli*.

In the 10 double gRNA arrays we tested, three (*menA/ubiA*, *adhE/menA* and *adhE/fabH*) resulted in 20%–30% improvement of isoprenol titers over their respective single sgRNAs, attaining isoprenol titers of 1.13 ± 0.06 g/L, 1.22 ± 0.03 g/L and 1.31 ± 0.06 g/L, respectively (Figure 4A). This suggests that certain sgRNA combinations can synergistically redirect metabolic flux towards isoprenol production. Interestingly, we also observed that simultaneous downregulation of double gRNAs significantly reduced *E. coli* growth rate compared to the control strain (Figure 4B). This suggests that simultaneous gene repression poses a metabolic cellular burden, negatively affecting their overall fitness and growth. Further investigation is warranted to better understand the underlying mechanisms and optimize the sgRNA expression and regulation for improved isoprenol production without compromising cellular growth. In contrast, the strain targeting *adhE/fabH* exhibited similar growth rate compared to the control strain, indicating that the repression of *adhE/fabH* genes does not impose a significant metabolic burden on the cells. This may explain why the *adhE/fabH* strain showed comparable growth and isoprenol production levels, making it a promising candidate for further optimization and scale-up studies.

Using the above results, we created gRNA arrays by assembling four individual sgRNAs (*adhE, ubiA, fabH, and menA*) into 4 triple gRNA arrays (*adhE/ubiA/fabH, adhE/ubiA/menA, adhE/ubiA/menA,* and *ubiA/fabH/menA*) to conduct multiple repression of endogenous genes in *E. coli* (Figure 4C). When the triple gRNAs were simultaneously downregulated, the growth rate of *E. coli* and the production of isoprenol were significantly reduced compared to those of double gRNAs strains even though they are still slightly higher than those of the control (Figure 5A). Interestingly, the strain with gRNA array targeting *adhE/menA/ubiA* did not grow (data not shown). Isoprenol titers of *adhE/fabH/menA, menA/ubiA/fabH, and adhE/*

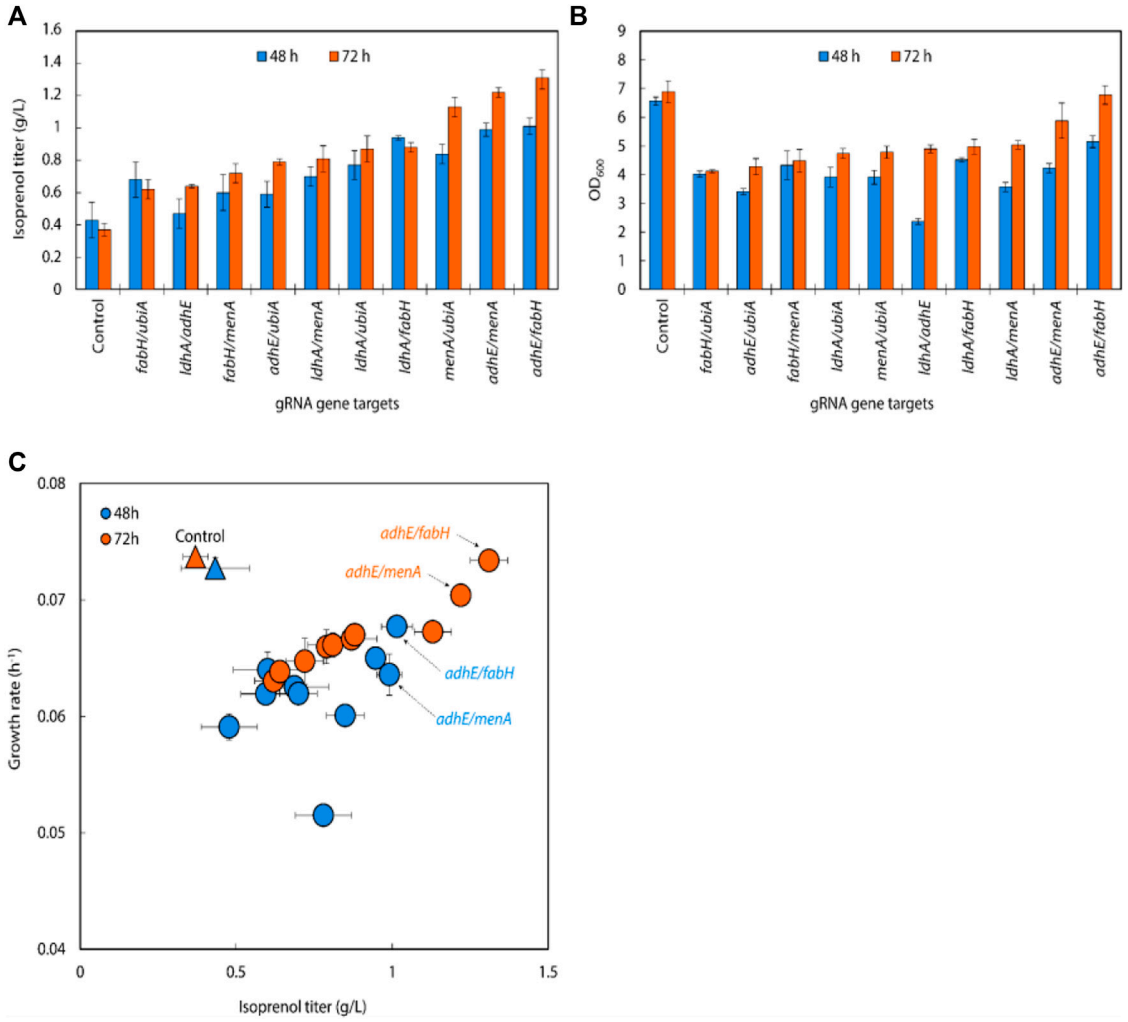


FIGURE 4
(A) Growth and **(B)** isoprenol production by strains harboring double gRNAs along with the original MVA pathway. Batch cultures were grown in 5 mL M9-MOPS minimal medium supplemented with 20 g/L of glucose (n = 3). **(C)** A scatter plot of growth rate and isoprenol titer at 48 h and 72 h and again show certain double gRNA strains significantly improved isoprenol production relative to the controls. The control at 48 h and 72 h is demarcated with appropriately colored triangles.

fabH/ubiA knock-down strains were 0.44 ± 0.09 g/L, 0.56 ± 0.09 g/L, and 0.62 ± 0.06 g/L, respectively (Figure 5A), with an approximately 50% reduction in growth compared to the control (Figure 5B).

The growth and isoprenol titer reduction of the triple gRNA strains relative to the control and the double gRNA strains that target the combination of the same set of genes (*adhE*, *fabH*, *menA*, and *ubiA*) suggests multiple gene downregulation in general is somewhat burdensome to cellular fitness. Further investigation could be conducted to elucidate the specific mechanism through which the simultaneous repression of multiple genes affects cellular fitness and isoprenol production.

3.5 Single gRNA library co-expression with the IPP-bypass pathway

Following the successful result of single, double, and triple gRNA arrays experiments in strains with the original MVA pathway, we transformed the same 32 sgRNAs library into *E. coli* DH1 strains with the IPP-bypass isoprenol production pathway and tested how the same gene knock down library behave in the slightly different isoprenol producing system. When co-expressing the IPP-bypass pathway and single gRNA library, 15 out of 32 sgRNAs harboring strains demonstrated higher isoprenol titers compared to the control with non-targeting gRNA (Figure 6A).

Similar to the results from the original MVA pathway, the IPP-bypass strains with sgRNAs targeting *menA*, *ubiA*, *ppsA*, *fabH,* and *adhE,* showed increases in isoprenol production by 136.93%, 138.52%, 146.62%, 173.24%, and 187.27%, with titers of 0.90 ± 0.22 g/L, 0.91 ± 0.18 g/L, 0.96 ± 0.10 g/L, 1.14 ± 0.17 g/L, and 1.23 ± 0.23 g/L, respectively (Figure 6A). The highest titer, however, was obtained from the strain with sgRNAs targeting *ldhA* with almost two-fold increased titer at 1.32 ± 0.31 g/L. Interestingly, this gene target did not show any titer improvement in the original MVA pathway experiment. We also found that the growth was similar to or slightly lower than that of the control strain when we

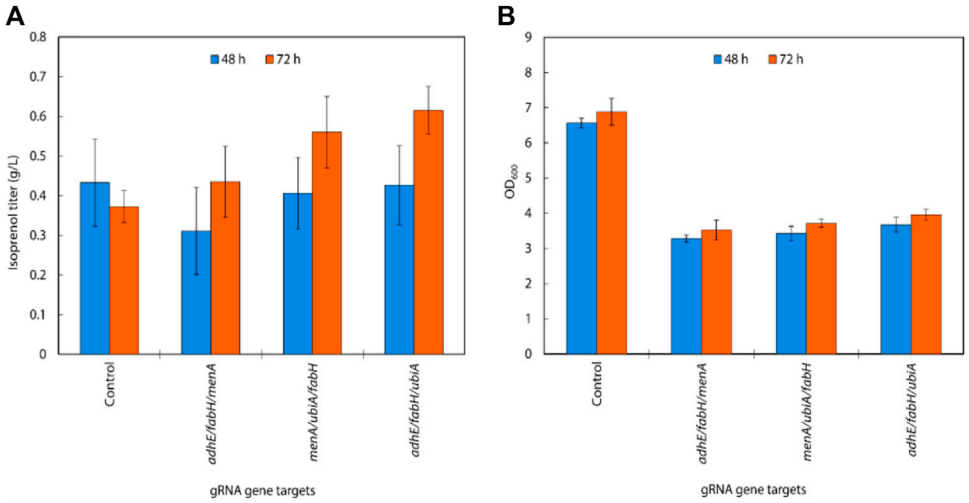


FIGURE 5
(A) Isoprenol production and **(B)** growth of strains harboring triple gRNAs along with the original MVA pathway grown in 5 mL M9-MOPS minimal medium supplemented with 20 g/L of glucose (n = 3).

downregulated *fabH* and *ldhA* expression using gRNAs (Figure 6B), while the growth of these strains was slightly higher than the control in the experiment with the original MVA pathway.

When downregulating glycolysis pathway and TCA cycle genes in the IPP-bypass pathway, we again observed growth inhibition and lower isoprenol titers confirming that genes on the glycolysis and the TCA cycle are not viable downregulation targets for isoprenol titer improvement. The reduction in growth of the strain upon downregulation of these genes implies that these pathways play a crucial role in the growth of *E. coli* and emphasizes the importance of maintaining the TCA cycle for survival before supplying the metabolic intermediates necessary for isoprenol production.

We generated a scatter plot of growth and isoprenol production of single gRNA arrays to construct the multiple gRNA arrays. Based on this analysis, the 6 genes (*adhE*, *ldhA*, *ppsA*, *ubiA*, *fabH*, and *menA*) were identified as a candidate group of sgRNAs with high isoprenol production and growth rate (Figures 6C,D).

3.6 Multiple gRNA arrays co-expression with the IPP-bypass pathway

Following the results of the single gRNA library on the IPP-bypass pathway strains, we assembled six individual gRNAs (*adhE, ldhA, ppsA, ubiA, fabH, and menA*) leveraged for higher isoprenol titer into all 15 possible double gRNA arrays (*adhE*/*ldhA*, *adhE*/*ppsA*, *adhE*/*ubiA*, *adhE*/*fabH*, *adhE*/*menA*, *ldhA*/*ppsA*, *ldhA*/*ubiA*, *ldhA*/*fabH*, *ldhA*/*menA*, *ppsA*/*ubiA*, *ppsA*/*fabH*, *ppsA*/*menA*, *ubiA*/*fabH*, *ubiA*/*menA*, and *fabH*/*menA*) to see whether multiplexed (i.e., double) endogenous gene repression could provide synergistic titer improvements in the IPP-bypass pathway strains. As shown in Figure 7A, the simultaneous targeting of *menA*/*ubiA*, *adhE*/*menA* and *adhE*/*fabH* resulted in significantly higher isoprenol titers than the control with non-targeting gRNA, achieving 1.31 ± 0.06 g/L, 1.32 ± 0.03 g/L and 1.61 ± 0.06 g/L, respectively. This production levels are higher than the previous production level increase by individual gene knockdown, and this result suggests that repression of *menA* and *ubiA*, *adhE* and either *menA* or *fabH* gene may more effectively redirect metabolic flux towards isoprenol production than when these genes are downregulated individually. The highest titer was achieved with a simultaneous downregulation of *adhE*/*fabH* and this titer improvement is attributed to synergistic effects of dual repression, resulting in a more pronounced enhancement of isoprenol production.

Interestingly, the results also show that downregulating the target genes of *adhE*/*fabH* and *adhE*/*menA* individually did not significantly affect the growth rate of *E. coli* compared to the control strain (Figure 7B). However, we observed a significant increase in the isoprenol titer when these genes were downregulated. These findings suggest that the *adhE* gene may not be essential for the growth of *E. coli* under the tested growth conditions. However, the *fabH* and *menA* are well known as essential for the growth of *E. coli* (Dhiman et al., 2019; Lai and Cronan, 2003; Beld et al., 2015; Rousset et al., 2021). This is consistent with previous research that has shown that these genes are involved in the production of fatty acid derivatives and menaquinone, respectively, which are important for cell membrane formation and energy production in bacteria. The significant increase in the isoprenol titer when these genes were downregulated suggests that *adhE*/*fabH* and *adhE*/*menA* may be competing for metabolic resources required for isoprenol production. By downregulating these genes individually, we may have reduced the metabolic burden on the cell, allowing more resources to be dedicated to isoprenol biosynthesis.

An analysis of double gRNA downregulation showed high isoprenol production by *adhE*/*fabH*. Based on comparison of growth rate and production titers (Figure 7C; Figure 6C), we then assembled three triple gRNA arrays that paired *adhE*/*fabH* with each of *ldhA*, *menA*, and *ubiA*.

Isoprenol production was significantly increased compared to the single and the double gRNAs strains with the constitutive sgRNAs with the best performing combination, *adhE*/*fabH*/*ldhA*, resulting in an isoprenol titer of 1.82 ± 0.19 g/L (Figures 8A,B). This

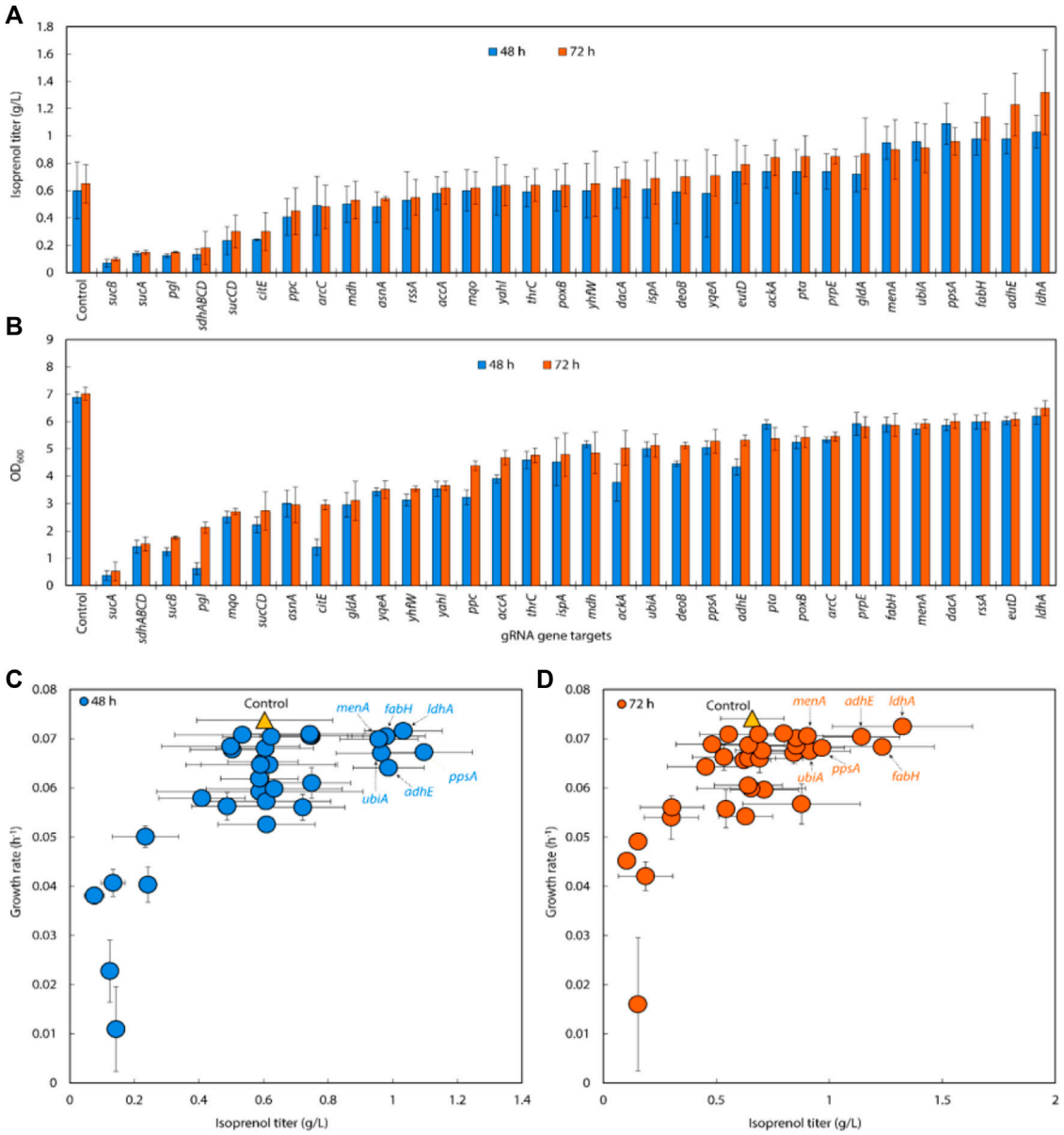


FIGURE 6
(A) Isoprenol production and **(B)** growth of strains harboring single gRNAs and the IPP-bypass pathway. Cultures were grown in 5 mL M9-MOPS minimal medium supplemented with 20 g/L of glucose (n = 3). **(C)** Scatter plots of growth rate and isoprenol titer at 48 h and **(D)** 72 h show that gene attenuation by certain gRNAs contributed to significant titer improvements, tracking well with observations from the original MVA pathway experiments. The control is demarcated with a yellow triangle.

juxtaposition suggests that acetyl-CoA accumulation is more crucial for enhancing isoprenol production in the IPP-bypass pathway than the accumulation of IPP.

As reported previously, the accumulation of IPP in the mevalonate pathway can be toxic to bacteria (George et al., 2018). By redirecting the flux of IPP away from the mevalonate pathway, *E. coli* was able to avoid the toxic effects of IPP accumulation and improved isoprenol production.

The increased production of isoprenol in the triple gRNA strains compared to the control strain and double gRNA strains indicates the potential of this multiplexed gRNAs approach for enhancing metabolic pathways in microbial cells.

3.7 Fed-batch cultivation of the multiple genes downregulated isoprenol strain

After screening the titer improvement by CRISPRi-mediated downregulation of genes in two isoprenol biosynthesis pathways, we selected the best performing strain, *E. coli* DH1 harboring the IPP-bypass pathway and triple gRNA targeting *adhE*, *ldhA*, and *fabH*, and scaled up the isoprenol production at a 2-L bioreactor under fed-batch conditions. Similar to what we observed in the 5 mL culture tube experiments, the strain harboring the CRISPRi system with the triple gRNAs array showed a slower growth rate and a higher maximum isoprenol production titer than the control strain

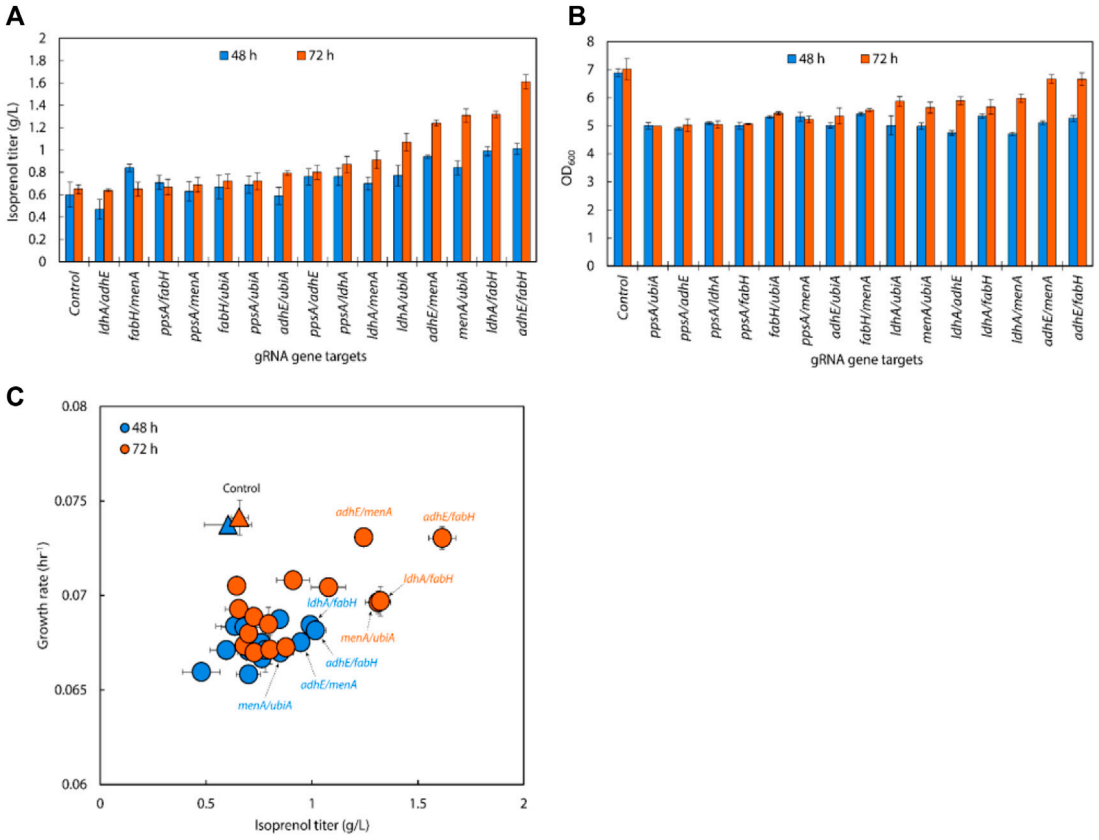


FIGURE 7
(A) Isoprenol production and **(B)** growth of strains harboring double gRNAs and the IPP-bypass pathway. Cultures were grown in 5 mL M9-MOPS minimal medium supplemented with 20 g/L of glucose (n = 3). **(C)** A scatter plot of growth rate and isoprenol titer at 48 h and 72 h show significant titer improvements by double gRNA harboring strains compared to the control. The control at 48 h and 72 h is demarcated with appropriately colored triangles.

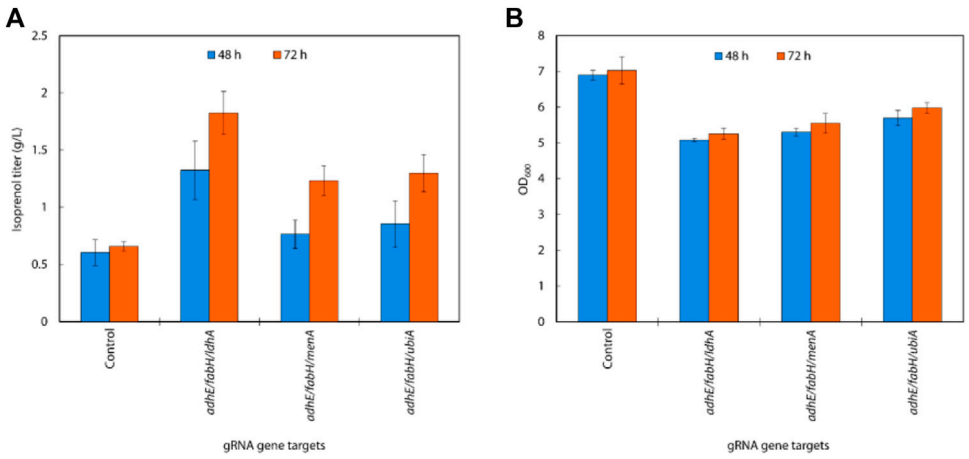


FIGURE 8
(A) Growth and **(B)** isoprenol production by strains harboring triple gRNAs and the IPP-bypass pathway grown in 5 mL M9-MOPS minimal medium supplemented with 20 g/L of glucose (n = 3).

harboring non-targeting gRNA plasmid JBEI-18706 (Figures 9A,B). In fed-batch cultivation of the control strain, the maximum cell growth was achieved at 96 h with an OD_{600} of 17.5 ± 0.5 and the maximum isoprenol production was achieved at 120 h with isoprenol titer of 10.4 ± 0.5 g/L (Figure 9A). The triple gRNAs array strain showed a slightly slower growth rate compared to the

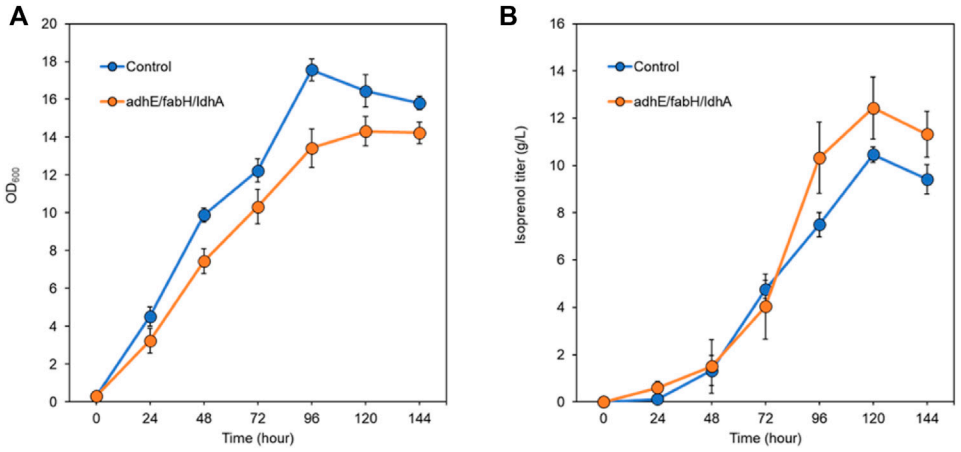


FIGURE 9
Isoprenol production and growth under fed-batch conditions by the control strain (harboring non-target gRNA plasmid (JBEI-18706)) and the triple gRNA strain targeting *adhA*, *fabH*, and *ldhA*. **(A)** growth of control and the strain harboring triple gRNA arrays. **(B)** production of isoprenol by the control strain and the strain harboring triple gRNA arrays. All strains include the IPP-bypass biosynthesis pathway plasmids. The fed-batch productions were performed in the 2-L bioreactors including M9-MOPS medium and 20% oleyl alcohol overlay at duplicate. Error bar represents standard deviation.

control strain, but about 20% increased isoprenol titer after 96 h with the highest titer of 12.4 ± 1.3 g/L and OD_{600} of 14.3 ± 0.7 at 120 h. The slower growth rate and the higher isoprenol production of the engineered strain in fed-batch condition confirms that the CRISPRi with *adhE*/*ldhA*/*fabH* gRNAs system worked and positively influenced the metabolic pathway for isoprenol biosynthesis not only at the small volume batch condition but also at the large scale fed-batch condition and led to improved production.

The use of fed-batch cultivation with continuous feeding of an additional glucose and ammonium chloride solution aimed to provide a sustained supply of carbon and nitrogen sources and contribute to the enhanced isoprenol production in the engineered strain. However, the rapid evaporation of isoprenol by agitation and air flow in the bioreactor remains a challenge in achieving higher titer. Even though our final titer was measured at 12.4 ± 1.3 g/L at 120 h, we observed a rapid decrease of the isoprenol level after 120 h and we also smelled a strong scent of isoprenol throughout the cultivation which suggests a significant evaporational loss from offgas of the bioreactor. Further optimization of the cultivation conditions such as optimization of the feeding strategy and the product recovery strategy both to improve the solvent extraction process and to capture isoprenol from the offgas could lead to higher isoprenol titers.

4 Discussion

In this work, we applied a combinatorial multiplex repression system using CRISPRi to downregulate endogenous genes in competing pathways for isoprenol production in *E. coli*. We demonstrated the success of the single and multiple CRISPRi system to improve isoprenol production in *E. coli*. The downregulation of target genes resulted in a considerable increase in isoprenol production, indicating that the CRISPRi system can modulate gene expression without adversely affecting cellular functions or knocking out essential genes. The observed increase in production may be attributed to reduced competition for precursors and altered cellular metabolism and redox balance. Furthermore, our results highlight the potential of using sgRNA arrays for multiple gene repression to modulate cellular metabolism and enhance isoprenol production in *E. coli*. The specific combinations of sgRNAs can significantly impact isoprenol titers, although careful consideration of their effects on cellular growth and fitness is necessary. This study also shows that the CRISPRi system is applicable to bench-top scale fermentation and the fed-batch condition with improvement of target bioproducts.

Interestingly, the identified downregulation gene targets that improved the isoprenol production were somewhat different from those in the previous report by Tian et al., 2019 which showcased the CRISPRi approach using an EZ-Rich defined medium and the report by Wang et al. which used an M9 medium but in conjunction with yeast extract (Wang et al., 2022). As previously reported by Goodall et al., disparities in culture media give rise to variations in gene expression (Goodall et al., 2018). This phenomenon likely explains the non-replication of genes such as *ackA*, *poxB*, and *pta* from Tian et al.'s study and genes *yggV* and *accA* from Wang et al.'s work. These differences underscore the potential impact of culture medium on genetic outcomes.

A nuanced constraint of CRISPRi-mediated downregulation lies in the inherent bias of the sgRNA library towards genes with anticipated functional relevance for titer enhancement, encompassing precursors, intermediates, and energy-related factors. However, comprehending precisely why gene downregulation amplifies the production beyond their initial functions remains a challenge. Consequently, meaningful analysis necessitates library-scale assessment of growth and production data rankings. To attain deeper insights, the synergy of CRISPRi, metabolomics, proteomics or transcriptomics, and machine learning becomes imperative. This fusion promises a more coherent rationale behind the beneficial effects of downregulation on production. Acquiring a finer grasp of altered metabolic networks

could potentially enable time-dependent gene downregulation to refine biosynthesis.

In this study, we were able to identify the genes involved in improving isoprenol production by testing a limited number of single and multiple gRNAs. As knocking-out multiple genes is not a trivial task and gene knockouts frequently result in growth retardation, the gene knockdown strategy using CRISPRi is a powerful tool to screen multiple combinations of target genes. However, screening various combinations of multiple target genes is still not a trivial task as the total number of combinations easily reaches an enormous level. Therefore, it is quite important to build an automated process to generate a library of multiplex gRNA arrays and to perform accurate and reproducible experiments in a high throughput manner. The data generated by this automated platform will make predictions using artificial intelligence feasible and can significantly reduce experimental variables and accelerate rapid and more accurate engineering.

It is interesting that there are many genes downregulated by the CRISPRi system but did not result in the isoprenol production enhancement. These genes might not be directly or indirectly associated with metabolic pathways (i.e., accumulation of precursors and cofactors to enhance isoprenol production and suppression of competing by-products). Nonetheless, it is important to note that genes which did not directly contribute to improving isoprenol production could also be the focus of future research. Further investigations would help to deepen our understanding of the metabolic pathways necessary for the production of desired bio-products.

Data availability statement

The data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

JK: Conceptualization, Data curation, Formal Analysis, Methodology, Writing–original draft, Writing–review and editing. TL: Conceptualization, Funding acquisition, Investigation, Investigation, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Small Business Innovation Research (SBIR) Program of the US Department of Energy, Office of Science through Cooperative Research and Development Agreement (CRADA) FP00003594 between the Regents of the University of California/Lawrence Berkeley National Laboratory and Technology Holding LLC. This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through Contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Acknowledgments

We thank Dr. David Carruthers for helpful discussion and writing support. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Conflict of interest

The 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2023.1296132/full#supplementary-material

References

Alper, H., Jin, Y.-S., Moxley, J. F., and Stephanopoulos, G. (2005). Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab. Eng.* 7, 155–164. doi:10.1016/j.ymben.2004.12.003

Alper, H., and Stephanopoulos, G. (2008). Uncovering the gene knockout landscape for improved lycopene production in *E. coli*. *Appl. Microbiol. Biotechnol.* 78, 801–810. doi:10.1007/s00253-008-1373-x

Baral, N. R., Yang, M., Harvey, B. G., Simmons, B. A., Mukhopadhyay, A., Lee, T. S., et al. (2021). Production cost and carbon footprint of biomass-derived dimethylcyclooctane as a high-performance jet fuel blendstock. *ACS Sustain. Chem. Eng.* 9, 11872–11882. doi:10.1021/acssuschemeng.1c03772

Beld, J., Lee, D. J., and Burkart, M. D. (2015). Fatty acid biosynthesis revisited: structure elucidation and metabolic engineering. *Mol Biosyst.* 11 (1), 38–59.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L. A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* 41, 7429–7437. doi:10.1093/nar/gkt520

Carruthers, D. N., Kim, J., Mendez-Perez, D., Monroe, E., Myllenbeck, N., Liu, Y., et al. (2023). Microbial production of high octane and high sensitivity olefinic ester biofuels. *Biotechnol. Biofuels Bioprod.* 16, 60. doi:10.1186/s13068-023-02301-7

Chae, T. U., Choi, S. Y., Kim, J. W., Ko, Y.-S., and Lee, S. Y. (2017). Recent advances in systems metabolic engineering tools and strategies. *Curr. Opin. Biotechnol.* 47, 67–82. doi:10.1016/j.copbio.2017.06.007

Cho, S., Shin, J., and Cho, B.-K. (2018). Applications of crispr/cas system to bacterial metabolic engineering. *Int. J. Mol. Sci.* 19, 1089. doi:10.3390/ijms19041089

Cho, J. S., Yang, D., Prabowo, C. P. S., Ghiffary, M. R., Han, T., Choi, K. R., et al. (2023). Targeted and high-throughput gene knockdown in diverse bacteria using synthetic sRNAs. *Nat Commun.* 14(1), 2359

Czajka, J. J., Banerjee, D., Eng, T., Menasalvas, J., Yan, C., Munoz, N. M., et al. (2022). Tuning a high performing multiplexed-CRISPRi Pseudomonas putida strain to further enhance indigoidine production. *Metab. Eng. Commun.* 15, e00206. doi:10.1016/j.mec.2022.e00206

Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6640–6645. doi:10.1073/pnas.120163297

Dhiman, R. K., Pujari, V., Kincaid, J. M., Ikeh, M. A., Parish, T., and Crick, D. C. (2019). Characterization of MenA (isoprenyl diphosphate:1,4-dihydroxy-2-naphthoate isoprenyltransferase) from *Mycobacterium tuberculosis*. *PLoS ONE* 14, e0214958. doi:10.1371/journal.pone.0214958

Foo, J. L., Jensen, H. M., Dahl, R. H., George, K., Keasling, J. D., Lee, T. S., et al. (2014). Improving microbial biogasoline production in *Escherichia coli* using tolerance engineering. *MBio* 5, e01932. doi:10.1128/mBio.01932-14

Gallagher, R. R., Li, Z., Lewis, A. O., and Isaacs, F. J. (2014). Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. *Nat. Protoc.* 9, 2301–2316. doi:10.1038/nprot.2014.082

George, K. W., Thompson, M. G., Kang, A., Baidoo, E., Wang, G., Chan, L. J. G., et al. (2015). Metabolic engineering for the high-yield production of isoprenoid-based C5 alcohols in *E. coli*. *Sci. Rep.* 5, 11128. doi:10.1038/srep11128

George, K. W., Thompson, M. G., Kim, J., Baidoo, E. E. K., Wang, G., Benites, V. T., et al. (2018). Integrated analysis of isopentenyl pyrophosphate (IPP) toxicity in isoprenoid-producing *Escherichia coli*. *Metab. Eng.* 47, 60–72. doi:10.1016/j.ymben.2018.03.004

Goodall, E. C. A., Robinson, A., Johnston, I. G., Jabbari, S., Turner, K. A., Cunningham, A. F., et al. (2018). The essential genome of *Escherichia coli* K-12. *MBio* 9, e02096-17. doi:10.1128/mBio.02096-17

Kang, A., George, K. W., Wang, G., Baidoo, E., Keasling, J. D., and Lee, T. S. (2016). Isopentenyl diphosphate (IPP)-bypass mevalonate pathways for isopentenol production. *Metab. Eng.* 34, 25–35. doi:10.1016/j.ymben.2015.12.002

Kang, A., Meadows, C. W., Canu, N., Keasling, J. D., and Lee, T. S. (2017). High-throughput enzyme screening platform for the IPP-bypass mevalonate pathway for isopentenol production. *Metab. Eng.* 41, 125–134. doi:10.1016/j.ymben.2017.03.010

Kang, A., Mendez-Perez, D., Goh, E.-B., Baidoo, E. E. K., Benites, V. T., Beller, H. R., et al. (2019). Optimization of the IPP-bypass mevalonate pathway and fed-batch fermentation for the production of isoprenol in *Escherichia coli*. *Metab. Eng.* 56, 85–96. doi:10.1016/j.ymben.2019.09.003

Keasling, J., Garcia Martin, H., Lee, T. S., Mukhopadhyay, A., Singer, S. W., and Sundstrom, E. (2021). Microbial production of advanced biofuels. *Nat. Rev. Microbiol.* 19, 701–715. doi:10.1038/s41579-021-00577-w

Kim, S. K., Han, G. H., Seong, W., Kim, H., Kim, S.-W., Lee, D.-H., et al. (2016). CRISPR interference-guided balancing of a biosynthetic mevalonate pathway increases terpenoid production. *Metab. Eng.* 38, 228–240. doi:10.1016/j.ymben.2016.08.006

Kim, S. K., Seong, W., Han, G. H., Lee, D.-H., and Lee, S.-G. (2017). CRISPR interference-guided multiplex repression of endogenous competing pathway genes for redirecting metabolic flux in *Escherichia coli*. *Microb. Cell. Fact.* 16, 188. doi:10.1186/s12934-017-0802-x

Kong, M. K., and Lee, P. C. (2011). Metabolic engineering of menaquinone-8 pathway of *Escherichia coli* as a microbial platform for vitamin K production. *Biotechnol. Bioeng.* 108, 1997–2002. doi:10.1002/bit.23142

Korz, D. J., Rinas, U., Hellmuth, K., Sanders, E. A., and Deckwer, W. D. (1995). Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. *J. Biotechnol.* 39, 59–65. doi:10.1016/0168-1656(94)00143-Z

Lai, C.-Y., and Cronan, J. E. (2003). β-Ketoacyl-Acyl carrier protein synthase III (FabH) is essential for bacterial fatty acid synthesis. *J. Biol. Chem.* 278, 51494–51503. doi:10.1074/jbc.M308638200

Liu, Z., Dong, H., Cui, Y., Cong, L., and Zhang, D. (2020). Application of different types of CRISPR/Cas-based systems in bacteria. *Microb Cell Fact* 19 (1), 172. doi:10.1186/s12934-020-01431-z

Lee, S. Y., and Kim, H. U. (2015). Systems strategies for developing industrial microbial strains. *Nat. Biotechnol.* 33, 1061–1072. doi:10.1038/nbt.3365

Rosenkoetter, K. E., Kennedy, C. R., Chirik, P. J., and Harvey, B. G. (2019). [4 + 4]-cycloaddition of isoprene for the production of high-performance bio-based jet fuel. *Green Chem.* 21, 5616–5623. doi:10.1039/C9GC02404B

Rousset, F., Cabezas-Caballero, J., Piastra-Facon, F., Fernández-Rodríguez, J., Clermont, O., Denamur, E., et al. (2021). The impact of genetic diversity on gene essentiality within the *Escherichia coli* species. *Nat Microbiol.* 6 (3), 301–12.

Tan, S. Z., and Prather, K. L. (2017). Dynamic pathway regulation: recent advances and methods of construction. *Curr. Opin. Chem. Biol.* 41, 28–35. doi:10.1016/j.cbpa.2017.10.004

Tao, S., Qian, Y., Wang, X., Cao, W., Ma, W., Chen, K., et al. (2018). Regulation of ATP levels in *Escherichia coli* using CRISPR interference for enhanced pinocembrin production. *Microb. Cell. Fact.* 17, 147. doi:10.1186/s12934-018-0995-7

Tian, T., Kang, J. W., Kang, A., and Lee, T. S. (2019). Redirecting metabolic flux via combinatorial multiplex CRISPRi-mediated repression for isopentenol production in *Escherichia coli*. *ACS Synth. Biol.* 8, 391–402. doi:10.1021/acssynbio.8b00429

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., et al. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894–898. doi:10.1038/nature08187

Wang, J., Jiang, T., Milligan, S., Zhang, J., Li, C., and Yan, Y. (2022). Improving isoprenol production via systematic CRISPRi screening in engineered *Escherichia coli*. *Green Chem.* 24, 6955–6964. doi:10.1039/D2GC02255A

Wu, J., Du, G., Chen, J., and Zhou, J. (2015). Enhancing flavonoid production by systematically tuning the central metabolic pathways based on a CRISPR interference system in *Escherichia coli*. *Sci. Rep.* 5, 13477. doi:10.1038/srep13477

Wu, J., Zhou, P., Zhang, X., and Dong, M. (2017). Efficient *de novo* synthesis of resveratrol by metabolically engineered *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* 44, 1083–1095. doi:10.1007/s10295-017-1937-9

Zada, B., Wang, C., Park, J.-B., Jeong, S.-H., Park, J.-E., Singh, H. B., et al. (2018). Metabolic engineering of *Escherichia coli* for production of mixed isoprenoid alcohols and their derivatives. *Biotechnol. Biofuels* 11, 210. doi:10.1186/s13068-018-1210-0

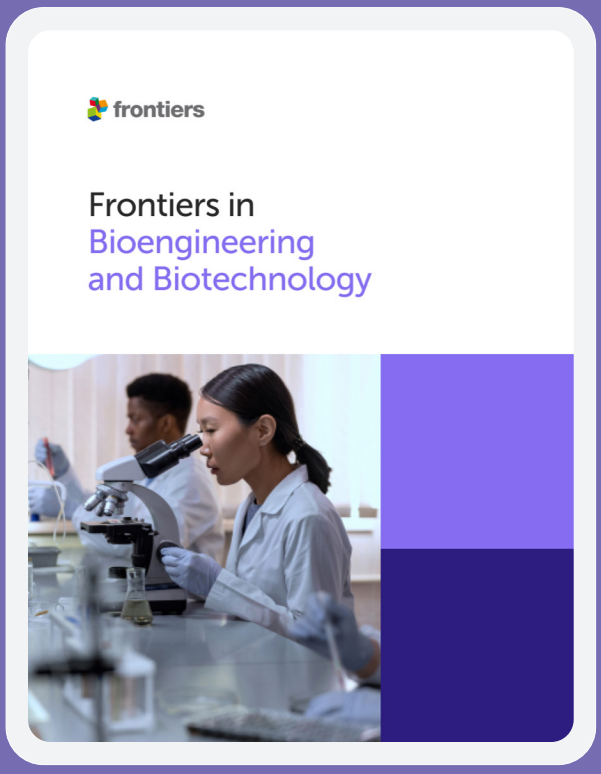
Frontiers in
Bioengineering and Biotechnology

Accelerates the development of therapies, devices, and technologies to improve our lives

A multidisciplinary journal that accelerates the development of biological therapies, devices, processes and technologies to improve our lives by bridging the gap between discoveries and their application.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact