Metabolic Engineering of Escherichia Coli BL21 Strain Using Simplified CRISPR-Cas9 and Asymmetric Homolog Arms Recombineering

Sudha Shukal
Agency for Science Technology and Research
Xiao Hui Lim
Agency for Science Technology and Research
Congqiang Zhang
Agency for Science Technology and Research
Xixian Chen (xixian_chen@sifbi.a-star.edu.sg)
Agency for Science Technology and Research https://orcid.org/0000-0002-0335-2058

Research

Keywords: CRISPR-Cas9, asymmetric homolog arms, lycopene, cell size, triacylglycerol pathway, acetyl-CoA availability

Posted Date: September 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-889623/v1

Metabolic engineering of *Escherichia coli* BL21 strain using simplified CRISPR-Cas9 and Asymmetric Homolog Arms recombineering

Sudha Shukal[#], Xiao Hui Lim[#], Congqiang Zhang, Xixian Chen[*]

*Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A*STAR), Singapore.*

Both authors contribute equally.

* To whom correspondence should be addressed.

Chen Xixian: SIFBI, A*STAR, Proteos level 4, Singapore 138673;

Email: xixian_chen@sifbi.a-star.edu.sg or xixian.chen@outlook.sg

Abstract

Background

The recent CRISPR-Cas coupled with λ recombinase mediated genome recombineering has become a common laboratory practise to modify bacterial genomes. It requires supplying a template DNA or homolog arms for precise genome editing. However, it is often overlooked the process to generate the homolog arms which is a time-consuming, costly and inefficient step.

Results

In this study, we first optimized CRISPR-Cas protocol in BL21 strain and successfully deleted 10 kb gene from the genome in one round of editing. To further simplify the

protocol, asymmetric homolog arms as PCR fragments was used. It can be obtained by one-step PCR reaction with two primers and purified with desalting columns. Unlike conventional homolog arms that are prepared through overlapping PCR, cloning to plasmid or annealing synthetic DNA fragments, our method significantly shortened the time taken and reduced the cost to prepare the homolog arms. To test the robustness of the optimized workflow, we successfully deleted 26 / 27 genes across BL21 genome. Noteworthy, gRNA design is important for CRISPR-Cas system and a general heuristic gRNA design was proposed in the study. To apply our established protocol, we targeted 16 genes and iteratively deleted 7 genes from BL21 genome. The resulting strain increased lycopene production from ~15,000 ppm to > 40,000 ppm.

Conclusions

Our work has optimized the homolog arms design for gene deletion in BL21 strains. The protocol efficiently edited BL21 strain to improve lycopene production. The same workflow is applicable to all *E. coli* strain which would be useful for genome rewiring to further increase metabolite production in microbial cell factories.

Keyword:

CRISPR-Cas9, asymmetric homolog arms, lycopene, cell size, triacylglycerol pathway, acetyl-CoA availability

Introduction

Escherichia coli BL21 strain is one of the most utilized bacterial strains for recombinant protein production and metabolic engineering [1–3]. To modify the genome of BL21 strain efficiently and easily is highly desirable to further improve the strain performance for industrial applications. With the breakthrough in genome editing technology, clustered regularly interspaced short palindromic repeats (CRISPR) and associated proteins (Cas) system has become a common genome editing tool to be applied from microbes to mammals [4]. CRISPR-Cas9, a class 2 type II system is well-characterized and the most widely applied because of its simple design [1]. It consists of a single Cas9 nuclease protein complex with a CRISPR RNA (crRNA) and trans-acting crRNA (tracrRNA) duplex or a single guide RNA (sgRNA) with fused 3' end of crRNA and 5' end of tracrRNA [5]. The complex specifically targets the protospacer DNA which is complementary to crRNA and generates a double-strand (ds) break via the recognition of a protospacer adjacent motif (PAM) sequence, 5'-NGG-3' [5]. Subsequently, the dsDNA break (DSB) can be repaired via homologous recombination (HR) or non-homologous end-joining (NHEJ). A donor DNA is required for specific deletion, insertion, or mutation of the genome sequence during HR. In *E.coli*, λ-Red recombinases are often used for efficient recombination of the chromosome and donor DNA [6,7].

Many pioneering works in bacteria optimized the CRISPR-Cas system in *E. coli* K-12 strain [6,8,9]. For example, the widely used two-plasmid system (pTarget and pCas) developed by Jiang and co-workers utilizes CRISPR-Cas9 and λ-Red recombination for scarless genome modification in *E. coli* K-12 strain MG1655 [6]. While the manuscript was in preparation, the same group published a modified pCas/pTarget system as the original

two-plasmid system failed to work in *E. coli* BL21 strain [10]. Indeed, very few reports optimized CRISPR-cas system in BL21 strain, except a large-scale validation study by Zerbini and co-workers who has shown CRISPR-Cas mediated gene knockout in BL21 Δ*ompA* strain [11]. In the study, 1-10 µg of synthetic DNA was used as homolog arms for recombineering, which is costly and restricts the homolog arm length to be 70-120 bp. The design of homolog arms is often overlooked. In the two-plasmid system, the Cas9 and λ-Red recombinases are co-expressed on pCas plasmid, whereas the donor DNA and sgRNA are carried on the pTarget plasmid (Figure 1a). Cloning of the pTarget plasmid would take at least 3 days to assemble four DNA fragments. The authors noted that cloning pTarget plasmid became complicated when multiple donor DNAs were included. They attempted to simplify the procedure with donor DNA supplied as a PCR fragment. However, the recombination efficiency was significantly decreased when the homolog arm length was shortened from 400 bp to 40 bp [6].

In this study, we have improved the CRISPR-Cas mediated deletion in BL21 strain and optimized the homolog arms design. Instead of symmetrical homolog arms, we applied asymmetric homolog arms, which can be obtained in a single PCR step. In addition, 100 ng of homolog arms was transformed to edit the genome. We have validated the optimized protocol with 27 gene targets across the BL21 genome and achieved successful gene deletion up to 3.4 kb. 18 out of 27 genes achieved ≥ 75% knockout efficiencies. We termed the workflow as CRASH (*CR*ISPR-cas9 and *as*ymmetric *h*omolog arm mediated genome modification). Lastly, we applied CRASH protocol to iteratively modify BL21 genome. By increasing cell size, regulating triacylglycerol (TAG) production, and redirecting acetyl-CoA flux to mevalonate pathway, we have successfully improved lycopene production from ~15,000 ppm to > 40,000 ppm.

Results

Optimization of CRISPR-Cas9 knock-out in E. coli B strain

We adapted the two-plasmid system developed by Jiang and co-workers to *E. coli* BL21 strain [6,11], in which Cas9 and sgRNA were overexpressed on pCas and pTarget plasmid, respectively (Fig. 1a). Curing of pTarget plasmid was modified to overexpress the toxic *sacB* gene from *Bacillus subtilis* in the presence of 5% sucrose (Fig. 1a) [11]. The gRNA-pMB1 expression cassette from pCas plasmid was removed. To establish the protocol, five gRNAs targeting *adhE* were tested first based on the distance from the PAM sequence to the nearer homolog arm sequence (PAM-to-HA distance). Four gRNAs (gRNA2, gRNA3, gRNA4 and gRNA5) were within 30 bp for the PAM-to-HA distance, while gRNA1 was >130 bp away from the nearer homolog arm sequence (Table 1). Moreover, gRNA2-5 were targeting different DNA strands at the 5' and 3' regions of *adhE* gene (Fig. 1b). All the five gRNAs were cloned into the pTarget plasmid which carried the 500 bp homolog arm sequences (Fig. 1a, method 2). The five pTarget plasmids were then transformed into *E. coli* BL21 cells overexpressing *cas9* and λ-recombinase genes and plated on agar plates with kanamycin and spectinomycin antibiotics. Both the gRNA targeting sequence and PAM sequence were deleted after homolog recombination, and the colonies formed on agar plate were expected to have genome modifications. After overnight incubation, 8 colonies from each agar plate were picked and analysed by colony-PCR. As shown in Fig.1b, only gRNA3 was effective to achieve nearly 100% knockout efficiencies. The knockout efficiencies were ranked to be gRNA3 > gRNA4 > gRNA2 > gRNA1 > gRNA5. Hardly any successful gene deletion was observed when either gRNA1 or gRNA5 was used. Although fewer colonies were formed on gRNA1 and gRNA5 agar plates (supplementary table S1), there were many "escapers" where the targeted genome region

remained intact [12]. We hypothesized that homology-directed recombination activities were insufficient, and endogenous recA-mediated SOS response rescued cells from CRISPR-cas9 induced cell death [13]. However, overexpressing recA56, the dominant negative form of recA, did not improve the knockout efficiencies when gRNA1 was used (supplementary Table S2, pCas-V6). Replacing arabinose promoter with a stronger T7 promoter to overexpress λ-recombinase genes did not yield successful knock out clones from gRNA1. In fact, it reduced the colony forming units or transformation efficiency significantly for both gRNA1 and gRNA3 (supplementary table S2, pCas-V2). Moreover, when λ-recombinase was only expressed under leaky T7 expression, similar knockout efficiencies were observed for gRNA3 as compared to induced λ-recombinase activity (supplementary table S2, pCas-V2). Similarly, when λ-recombinase genes were removed from pCas plasmids, one of 8 colonies was successfully deleted when gRNA3 was introduced into BL21 cells overexpressing cas9 only (Fig. 1b). These observations reiterated the importance of gRNA designs which may be the critical bottleneck in the CRISPR-Cas9 system. It is worth noting that *in silico* gRNA design tools predicted gRNA3 being the lowest efficiencies among the 5 gRNAs tested (supplementary table S1) [14], indicating that there were unidentified factors governing the efficiencies of gRNAs.

Next, we challenged the CRISPR-cas9 system to knockout longer length of DNA from *E. coli* BL21 genome. gRNA3 was used, the PAM-to-HA distance was kept at 29 bp, and only the homolog arm sequence proximate to the 5' or downstream of *adhE* was changed (Fig.1b). Despite a decrease in knockout efficiency was observed when the deletion length increased from 300 bp to 10 kb, >15% knockout efficiency was achieved for 10 kb deletion (Fig. 1c). Longer length (15 kb) was not tested, due to the presence of an essential gene

[15]. Deleting 10 kb DNA simultaneously knocked out ~10 chromosome genes from *E. coli* BL21 genome, which would be sufficient for subsequence genome modification.

Optimization of asymmetric donor DNA

While establishing the CRISPR-cas9 protocol, we realized that generating the pTarget plasmid with homolog-arm sequences was a time-consuming step; at least 3 days were required to clone the pTarget plasmid [10,16]. The cloning efficiency was lower, since four DNA fragments were amplified and assembled simultaneously (Fig. 1b, method 2). Reports shown that double stranded DNA (dsDNA) could be used as an alternative donor DNA, and was more efficient than single stranded DNA (ssDNA) [6,9]. Zerbini and co-workers have validated a CRISPR-Cas9 protocol in BL21 Δ*ompA* strain, which required 1-10 μg of donor dsDNA to achieve genome modification [11]. dsDNA was generated by annealing two synthetic ssDNA. This inevitably increased the cost for donor DNA synthesis and limited the length of donor DNA. Here, we explored the use of asymmetric homolog arms (aHA) as donor DNAs which can be obtained by one-step PCR (Fig. 1a). The efficiencies of aHA were tested with gRNA3 targeting *adhE* [17]. Four different homolog arm lengths were tested: 50 bp upstream of deletion site and 50 bp downstream of the deletion site (U50D50); 500 bp upstream of deletion site and 50bp downstream of the deletion site (U500D50); 50 bp upstream of deletion site and 500 bp downstream of the deletion site (U50D500); 500 bp upstream of deletion site and 500 bp downstream of the deletion site (U500D500) (Fig. 2a). ~100 ng of HA PCR products and pTarget-gRNA3 were transformed into BL21 cell overexpressing Cas9. All four HA designs resulted in similar transformation efficiency (Fig. 2a). However only U50D500 and U500D500 HA gave rise to ~100% knockout efficiency, whereas the other two HA designs only achieved

~40% knockout efficiency (Fig. 2a and 2b). Richard et al. showed that cas9 protein asymmetrically released the PAM-distal nontarget strand after double strand DNA break [17]. This may be the reason attributed to the preferred asymmetric design of HA. In addition, the study of Richard and co-workers demonstrated that 36 bp distal to the PAM is sufficient to achieve 60% knockout efficiency. Thus, we systematically tested different lengths of HA upstream (UP) of the deletion site (0, 10, 20, 30, 40, 50 or 60 bp), while keeping the length of HA downstream (DW) of the deletion site to be 500 bp. As shown in Figure 2c, transformation efficiency decreased for UP HA length of 0, 10 and 20 bp. However, knockout efficiency was maintained when the UP HA was ≥ 20 bp; the target 300 bp was successfully deleted in all the four randomly picked colonies (Fig. 2b). As a result, shorter primer can be used to generate the aHA without compromising knockout efficiencies. We termed the method CRASH (CRISPR-cas9 and Asymmetric Homolog arm directed genomic engineering). With the encouraging results, we again challenged the CRASH protocol with aHA (U50D500) to knockout longer length of DNA from *E. coli* BL21 genome. Both the knockout and transformation efficiencies decreased when the deletion length increased to 5 kb. Acceptable recombination efficiency, approximately 30%, was achieved when targeting to delete 2 kb. The length would be long enough to delete 1-2 genes from BL21 genome in one-step.

gRNA design testing for multiple gene knockout targets

To validate the CRASH protocol, we set to investigate gene deletion efficiencies across various positions along *E. coli* BL21 genome (Fig. 3b). It is well known that gRNA design plays a paramount role in ensuring the success of CRISPR-cas9 mediated genome modifications [18,19]. *In silico* design for gRNAs displays discrepancies with experimental

183 results to some extent. Thus, we also examined at least two gRNA designs for each gene
184 target (Fig. 3a). Both gRNAs would target either the positive or the negative strand of the
185 genomic DNA; each of the two gRNA will target either the 5' or 3'end of the gene. The
186 PAM-to-HA distance was mostly kept within 50 bp (Table 1). For the aHA design, the UP
187 HA length was varied between 40-45 bp, whereas the DW HA length was kept at 500 bp.
188 Out of the 27 gene targets, only *rodZ* was not successfully deleted, possibly because of
189 an essential gene (ispG) was immediately downstream of *rodZ* [20]. The rest 26 genes
190 were successfully deleted by CRASH protocol with 18 out of 26 genes achieving ≥ 75%
191 knockout efficiencies (Fig. 3c and Table 1). The data was clustered by target gene
192 direction (forward or reverse), gRNA targeting strand (positive or negative) and proximate
193 gene end (5' or 3'), as shown in Figure 3c. Unfortunately, there was no apparent pattern
194 that dictated the more efficient gRNA designs. We then analysed the mean or median
195 knockout efficiencies (supplementary table S3) and generalized the following heuristics
196 for gRNA design: for forward genes and gRNAs targeting negative strand, targeting 5' end
197 of the gene is preferred; for forward genes and gRNAs targeting positive strand, targeting
198 3' end of the gene is preferred; for reverse gene and gRNAs targeting negative strand,
199 targeting 3' end of the gene is preferred; for reverse gene and gRNAs targeting positive
200 strand, targeting 5' end of the gene is preferred (Fig. 3a and supplementary table S3).
201 Interestingly, regardless of forward or reverse genes, the preferred gRNA targeting sites
202 on the genome are similar: for gRNA targeting negative strand, proximate to UP HA is
203 preferred; for gRNA targeting positive strand, proximate to DW HA is preferred (Fig. 3a).
204 We have validated the heuristics with new gene targets and achieved successful deletions.
205 Similar design has been tested and successfully applied to other *E. coli* strain such as K12
206 strain MG1655 (results not shown). In addition, multiplex knockout in BL21 strain using

CRASH was tested by deleting *adhE* and *ldhA* simultaneously. 87.5% colonies were successfully modified (Supplementary Figure S1). It is noted that to create the pTarget plasmids with multiple gRNA cassettes is time-consuming when repeating sequences (e.g. promoter and terminator regions) and possible secondary structures are involved. To apply the CRASH deletion protocol, we systematically deleted single and combinatorial genes in *E. coli* BL21 to push-and-pull more flux towards lycopene production.

Increasing E. coli size

Lycopene was produced from the mevalonate pathway which was optimized in our previous studies [21–23] (Fig. 5a). Since lycopene is an intracellular compound, increasing cellular storage capacity may provide a favourable pull to enhance the specific yield of lycopene [24,25]. Thus, we hypothesized that by altering the cell size might increase the storage capacity. To test the hypothesis, we selected four gene targets that were shown to alter cell morphology after being deleted [26,27]. EnvC activates the peptidoglycan (PG) amidases, regulates septal PG splitting and daughter cell separation; *EnvC* null mutant results in improper cell division and forming long and chained cells [28,29]. *HolD* encodes the DNA polymerase III subunit ψ. Deleting *HolD* affects DNA replication and cell division, leading to filamentous morphologies [26]. TatA is the part of the twin arginine translocation complex and transports the PG amidase to periplasm. *TatA* mutation leads to improper location of PG amidase, thus leading to long and chained bacteria [30]. ZapB is required for proper Z ring formation for cell division [27]. *ZapB* null strain displays elongated cell shape. Using CRASH protocol, we generated the four single knockout strains and overexpressed the lycopene pathway genes in these cells. Lycopene content was measured after culturing the cells at 28 ºC for 24 h. Interestingly, not all single knockout

230 strains led to improved lycopene specific yield; only Δ*envC* and Δ*zapB* strains produced
231 17% and 30% more lycopene per cell, respectively (Fig. 4a). Moreover, *envC* mutation
232 led to much lower biomass, possibly attributed to the compromised cell envelope [29],
233 whereas deleting *zapB* had little effect on biomass. To be ascertain that these mutant
234 strains indeed altered the cell morphologies, flow cytometry was used to gauge the size
235 distribution for the wild-type and single knockout strains (Fig. 4b). When lycopene
236 pathway was not induced, the distributions for bacterial cells were generally gaussian.
237 Both *envC* and *tatA* null strains displayed overlapping size distributions with significantly
238 increased proportions of bigger cells. This is expected as both genes impacted the PG
239 amidase activity. *ZapB* null strain displayed similar size distribution as wild-type strain
240 except there was a slight increase in percentage of bigger cells. *HolD* null strain
241 unexpectedly displayed slightly reduced cell sizes, possibly because the cell was in a
242 different growth phase due to its much slower growth rate. In contrast, when lycopene
243 pathway was induced, the size distributions became asymmetric, indicating heterogenous
244 populations for all the strains tested. Interestingly, all the single knockout strain contained
245 higher percentage of bigger cells as compared to wild type. To verify the observation, we
246 observed wild-type and Δ*zapB* strains under phase contrast microscope and images were
247 analysed with imageJ software. Without lycopene production, wild-type strain remained
248 rod shape whereas Δ*zapB* strain was more heterogenous with long cells observed. In
249 contrast, irregular and elongated cells were observed for lycopene-producing cells in both
250 wild-type and Δ*zapB* strains, indicating lycopene accumulation possibly impacted cell
251 division (Fig. 4c). Size distribution analysed by imageJ also showed that Δ*zapB* strain had
252 an increased median size as compared to wild-type strain (Fig. 4d). Increased
253 heterogeneity or outliers were observed for lycopene producing cells (Fig. 4d). To avoid

plasmid instability issues due to heterogenous populations observed, we created an auxotrophic strain B2 by iteratively deleting semi-essential genes in addition to Δ*zapB* for subsequent lycopene production (Table 2) [31].

Engineering triacylglycerol biosynthesis pathway

Another way to increase lycopene accumulation is to engineer the neutral lipid pathway [25]. Previous reports have systematically optimized triacylglycerol (TAG) production in *E. coli*, which both *dgkA* and *fadE* genes were deleted, and TAG pathway genes were overexpressed (Fig. 5a) [32]. We tested single or double knockout of *dgkA* and *fadE* genes for lycopene production based on B2 strain. As shown in Figure 5b, all the three strains showed increased lycopene specific yield, especially *dgkA* null strain or B3 strain producing ~30,000 ppm lycopene. Even though *fadE* null strain was beneficial to increase lycopene content, double knockout of *fadE* and *dgkA* did not yield synergistic effect to improve the lycopene content further. Notably, we observed significant decrease in biomass when *dgkA* was deleted, possibly because of the accumulation of toxic diacylglycerol (Fig. 5b) [33]. Channelling diacylglycerol towards TAG may alleviate the toxicity. We thus overexpressed the TAG pathway genes as 4th module in B3 strain, which comprised long-chain-fatty-acid—CoA ligase (fadD), phosphatidic acid phosphatase (PAP) and wax ester synthase/diacylglycerol acyltransferase (WS/DGAT) (Fig. 5a). Two *fadD* and two *WS/DGAT* genes were screened [32]. As shown in Figure 5c, when *fadD* from *E. coli*, *PAP* from *Rhodococcus opacus* and *WS/DGAT* from *Thermomonospora curvata* where overexpressed (DTP1b), lycopene content was further increased to ~35,000 ppm, even though biomass was not increased. It was noted that when only *fadD* was overexpressed (D1 and D2), B3 strain accumulated to higher biomass while lycopene

content remained the same (Fig. 5c). This was possibly because flux towards diacylglycerol was partially diverted to fatty acid which could be re-converted back to acetyl-CoA via β-oxidation (Fig. 5a) [32]. This observation led us to hypothesize that acetyl-CoA availability may be limiting biomass accumulation [34].

Increasing acetyl-CoA availability

To increase acetyl-CoA availability, the competing flux from pyruvate and acetyl-CoA was removed [35]. We first tested lycopene production in single knockout strains based on B2 strain to remove divergent flux from pyruvate and acetyl-CoA: namely *adhE*, alcohol dehydrogenase; *ldhA*, lactate dehydrogenase; *pflB*, pyruvate-formate lyase; *poxB*, pyruvate dehydrogenase; *ackA-pta*, acetate kinase and phosphate acetyltransferase (Fig. 5a) [36]. The three modules of lycopene pathway genes were overexpressed, and the strains were cultured in rich media for 3-4 days. As shown in Figure 6a, all the single deleted strains produced higher lycopene specific yield as compared to B2 strain, partially attributed to the reduced biomass. Even though *ackA-pta* deleted strain gave rise to the highest lycopene yield and biomass, the strain was growing extremely slowly. Thus, we decided to test *adhE* and *ldhA* deletions which has the highest biomass and lycopene yield, respectively, among the remaining four single-deleted strains. Iterative gene knockout was carried out in B3 strain to create single deleted—*adhE* (B4) or *ldhA* (B5)—strains and double deleted—*adhE* and *ldhA* (B6)—strain (Table 2). All four modules were overexpressed with DTP1b as the 4th module to produce TAG. All the strains were cultured in rich media for 2 days. As expected, when *adhE* was deleted, acetyl-CoA could not be converted to ethanol, and the increased availability of acetyl-CoA improved biomass of B4 strain (Fig. 6b). Unfortunately, lycopene titer was not improved as compared to B3 strain,

leading to reduced lycopene specific yield in B4 strain. Similarly, *ldhA* null strain (B5 strain) increased pyruvate availability which activates pyruvate dehydrogenase allosterically to produce acetyl-CoA [35]. B5 strain accumulated higher biomass and lycopene titer than B3 strain, although lycopene specific yield was slightly decreased (Fig. 6b). Prolonging the culturing time to 4 days resulted in further increase in lycopene titer of B5 strain (~135 mg/L), resulting lycopene specific yield reaching ~42,000 ppm. Double deletion of *adhE* and *ldhA* did not show synergistic effect on lycopene production, indicating additional factor(s) influencing lycopene production in addition to acetyl-CoA availability (Fig. 6b).

Discussion

The field of metabolic engineering in *E. coli* has advanced significantly over the past decades. Impressive "TYR" (Titer, Yield and Rate) data have been achieved by modulating the metabolic pathway, improving the pathway enzymatic activities, and increasing co-factor availabilities [31,34]. Recently, with the discovery and development of CRISPR-Cas system, many studies have coupled the system with λRed recombineering to modify the host genome, redirect flux towards desired products and minimize regulatory inhibitions [4,37,38]. In this study, we have further simplified the CRISPR-Cas9 mediated gene deletion protocol by using asymmetric homolog arm as donor DNA. The donor DNA can be obtained by one single PCR step within 2 h. This resolves the tedious cloning steps involving multiple-piece gene assembly to carry the donor DNA on plasmid, or the use of overlapping PCR which requires optimization and gel purification (Supplementary Table S4) [16]. Moreover, our protocol only requires 100 ng donor DNA, which can be obtained in sufficient quantity within 50 μl PCR reaction. More importantly, our protocol has achieved 100% knockout efficiency for deletion length up to 3.4 kb, which is higher than

previously reported (Supplementary Table S4). Each round of gene knockout took approximately 6 days, from cloning to plasmid curing (Supplementary Figure S2). To further shorten the cloning time to half a day, we have also tested directly transforming the PCR products of pTarget plasmid (with mutated gRNA region) and donor DNA into BL21 cell, and 1 out of 8 randomly selected colonies was successfully deleted. The same method works with simultaneously deleting two genes on the genome, though the recombination efficiency was slightly decreased. It is worthwhile to note that generating the multiple gRNAs may become a bottleneck, and innovative solutions have been demonstrated such as CRISPathBrick [39]. Even though CRASH works for gene deletion, integrating DNA longer than 20 bp still requires cloning of the homolog arms into the plasmid or overlapping PCR.

Another important aspect of CRISPR-Cas system is the gRNA design. Though *in silico* prediction tools provide a preliminary guide for the choice of gRNA, more fundamental studies are required to understand the efficiencies of gRNAs [12]. Here, we propose the general heuristics to design gRNA based on our CRASH protocol, where a slightly higher efficiency was observed when gRNA is targeting the positive strand of DNA nearer to the downstream homolog arm or when gRNA is targeting the negative strand of DNA nearer to the upstream homolog arm. Even though successful deletions can be achieved with heuristic gRNA design and CRASH protocol, knockout efficiencies vary for different genes and many “escaper” colonies have been observed. This warrants further mechanistic study to understand the causes of non-edited cells which may further increase the knockout efficiency [12].

To apply CRASH protocol, we have tested 11 gene knockouts and iteratively knockout 7 genes from BL21 genome to improve lycopene production. Recently, several works have demonstrated that increasing the storage capacity of the production microorganisms is beneficial to lycopene production [24,25,34]. Here, we demonstrated two more strategies to increase the storage capacity: first is to increase cell size and second is to enhance the neutral lipid production. Heterogenous and elongated cell populations were observed when lycopene accumulated in *E. coli*, suggesting lycopene may have affected cell division. In fact, our initial size screening includes Δ*rodZ*, which reduces the size of *E. coli*. We hypothesized that rodZ null strain may have increased membrane surface to volume ratio, thus enhanced storage capacity to store lycopene. However, we failed to knockout *rodZ* after a few attempts of changing gRNA and donor DNA designs. Though *rodZ* is not an essential gene, its downstream gene *ispG* is essential for cell growth. Deleting *rodZ* may affect ispG expression. Alternative approaches such as CRISPR interference or base mutation to repress or inactivate rodZ, respectively may be tested. In addition to regulating the cell size, our results show that overexpressing neutral lipid pathway is helpful to enhance lycopene production, although it competes for the central metabolite acetyl-CoA [34]. Removing divergent flux from acetyl-CoA to organic acids is shown to increase both the biomass and lycopene production. Pathway optimization between the lipid pathway and lycopene pathway have not been performed in this study, which potentially could be useful to further boost up the lycopene production. Taken together, with 6 rounds of screening, knocking out 16 genes and combining 7 deletions in BL21 strain, we have improved lycopene content from ~15,000 ppm to > 40,000 ppm. The strain performance needs to be further validated in bioreactor.

Methods

Strain and plasmid construction

E. coli Bl21-Gold DE3 strain (Stratagene) was used in this study. The three lycopene plasmids were the same as previously described [21,23]. The gene *fadD* from *E. coli* was amplified from E. coli genome. The genes *fadD* from *Rhodococcus opacus*, *PAP* from *Rhodococcus opacus*, *WS/DGAT* from *Acinetobacter baylyi,* and *WS/DGAT* from *Thermomonospora curvata* were codon optimized and synthesized by Integrated DNA Technologies. They were cloned into the plasmid p15A-amp (L2-9) under mutant Tm1 promoter [21]. The information about plasmids and strains was summarized in Table 2.

CRISPR-Cas9 mediated gene deletion

Different pTarget plasmids with various sgRNAs were obtained by restriction free (RF) cloning methods [23]. The asymmetric homolog arm (HA) donor DNA was amplified from *E. coli* genome by using iProof PCR mix (BioRad) and column purified by Zymoclean Gel DNA Recovery Kit (Zymo Research). Generally, 100-200 ng/µl of donor DNA in 30 µl can be obtained in a 100 µl PCR reaction. For the primer design, the forward primer is a fusion of the upstream homolog arm (40-45 bp) sequence and downstream homolog arm (15-20 bp) sequence (Figure 1a and Supplementary Figure S2). The 15-20 bp downstream homolog arm is for annealing during initial cycles of PCR, and its length is chosen based on Tm ~50 ºC. The total length of forward primer is kept at 60 bp. The reverse primer is a normal PCR primer about 15-20 bp with Tm ~50 ºC. The length of downstream homolog arms can be varied based on the reverse primer chosen. For this study, the downstream homolog arm length was kept at 500 bp.

BL21 chemical competent cells were prepared using the Mix & Go! *E. coli* Transformation Kit (Zymo Research). For the construction of BL21 cells harbouring the pCas plasmid, 10 µl of cells were mixed with 50 ng/µl of pCas plasmid and heat shocked at 42 °C for 45 s. The cell was resucued in 200 µl of LB broth, at 30 °C, 300 rpm for 1 hour before spreading onto LB agar containing kanamycin (50 µg/ml) and incubated overnight at 30 °C. A single colony was picked and inoculated into 1 ml LB medium containing kanamycin (50 µg/ml) and incubated at 30 °C, 300 rpm overnight for making electrocompetent cells.

For the preparation of electrocompetent cells, OD600 0.1 of the overnight BL21 cell culture harbouring the pCas plasmid was inoculated into 10 ml of LB medium containing kanamycin (50 µg/ml) and cultured at 30 °C, 300 rpm. 20 mM arabinose was added to the culture at OD600 0.2 for the induction of λ-Red recombinase. The bacterial cells were harvested at OD600 0.6 and centrifuged at 3800 rpm for 10 minutes at 4 °C. The supernatant was discarded and the cells were re-suspended in 10 ml 10% glycerol. The washing step was repeated twice. The electrocompetent cells was then suspended in 100 µl of 10% glycerol.

For electroporation, 20 µl of cells were mixed with 100 ng/µl of pTarget plasmid and 100 ng of donor DNA in the 1 mm Gene Pulser cuvette (Bio-Rad) and electroporated at 1.8 kV. The cells were rescued in 500 µl of LB broth, at 30 °C, 300 rpm for 3 hours before spreading onto LB agar containing kanamycin (50 µg/ml) and spectinomycin (100 µg/ml) and incubated overnight at 30 °C. Colonies were screened by colony PCR using 2x PCRBIO Ultra Mix (PCR Biosystems) along with an unedited BL21 strain as control.

411 The edited colony harbouring both the pTarget and pCas plasmids was inoculated into 1
412 ml of LB medium containing kanamycin (50 μg/ml) and spectinomycin (100 μg/ml) and
413 incubated overnight at 30 °C, 300 rpm. For the curing of the pTarget plasmid, the culture
414 was streaked onto LB agar containing 5% sucrose and kanamycin (50 μg/ml) and
415 incubated overnight at 30 °C. The curing of the pTarget plasmid was confirmed by verifying
416 cell's sensitivity to spectinomycin (100 μg/ml) before proceeding on to the next round of
417 genome editing. For the curing of the pCas plasmid, the cells harbouring the pCas plasmid
418 were streaked onto LB agar and incubated at 42 °C overnight. The curing of the pCas
419 plasmid was confirmed by verifying cell's sensitivity to kanamycin (50 μg/ml). To cure both
420 plasmids at the same time, the cells can be plated on LB agar with 5% sucrose and
421 incubated at 42 °C overnight.

422 *Media and culture conditions*

423 All the cells were grown in modified autoinduction media (20g/L Peptone, 10g/L Yeast
424 extract and 10g/L NaCl), supplemented with 0.5 g/L glucose, 10 g/L glycerol, 30 mM
425 lactose, 75 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), and 0.5%
426 Tween 80. Briefly, OD_{600} 0.1 cell from overnight culture was inoculated into 1 mL fresh
427 media in 14mL BD Falcon™ tube. Cells were incubated at 28 °C for 24 h or longer as
428 indicated in the text before harvest. The media were supplemented with appropriate
429 antibiotics (100 mg/L ampicillin, 34 mg/L chloramphenicol, 50 mg/L kanamycin and 100
430 mg/L spectinomycin) to maintain corresponding plasmids.

431 *Extraction and quantification of lycopene*

432 Intracellular lycopene was extracted from cell pellets using HAE organic solvent,
433 comprising hexane : acetone : ethanol ratio to be 2 : 1 : 1 by volume. Briefly, 10–50 μL

bacterial culture was collected and centrifuged. The supernatant was discarded. 200 µL HAE buffer was then added to the cell pellets. The mixture was mixed and heated at 50 ^{o}C in thermoshaker at 1000 rpm for 30 mins, and further vortexed at room temperature for 30 min in order to completely extract the lycopene from the pellet. Subsequently, the mixture was centrifuged at 14,000 g for 10 min to pellet down the cell debris. 100 µL of the supernatant was added to 100 µL ethanol in a microplate reader and the absorbance at 472 nm was taken to calculate lycopene concentration against an external standard curve.

Microscope Image and Flow cytometry analysis

For microscopy assay, 5 µL *E. coli* cells were mounted directly on microscope slides and observed immediately under microscope. Microscopy was carried out by using a 100x Leica HCX PL FLUOTAR oil objective lens on a Leica DM6000 B microscope. Images were acquired by Leica Application Suite X software. Image analysis was carried out by ImagJ software. Flow cytometry analysis was carried out with Attune^{TM} NxT Flow Cytometer (ThermoFisher Scientific). Cell cultures were diluted 1000x in deionised water and 100 µL cells were analysed at a speed of 12.5 µL/min. The scatter signal was recorded in logarithmic scale. Threshold values for forward scatter and side scatter were set at 1000 and 300, respectively to eliminate background signals from debris. The signal was gated using forward and side scatter to exclude non-singlet cells. The cytograms was drawn using Attune NxT Flow Cytometer software version 3 and edited using Inkscape.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data supporting the findings of this study are available in the article, Supplementary Information, or upon request from the corresponding author.

Competing interests

The authors declare that they have no competing interests.

Funding

This research is supported by the Agency for Science, Technology and Research (A*STAR) under IAFPP3 - H20H6a0028, AME Young Individual Research Grants: A1984c0040 (2018) and A2084c0064 (2019), and Accelerate Technologies Gap-fund: ACCL/19-GAP044-R20H. GAP supports the research collaboration with Fermatics (www.fermatics.com), a biotech start-up and a co-venture between Hafnium Venture Capital Pte. Ltd. and A* STAR.

Authors' contributions

X.C., S.S, X.L. collected and analyzed the data. X.C., S.S, X.L. wrote the manuscript. X.C., Z.C. reviewed and revised the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

We acknowledge Dr Nicholas David Lindley, Dr Ee Lui Ang, and Dr Hazel Khoo, Singapore Institute of Food and Biotechnology Innovation for the support of the project. This research was supported by Singapore Institute of Food and Biotechnology Innovation (Previously

as Biotransformation Innovation Platform), Agency for Science, Technology and Research (A*STAR), Singapore.

Reference

1. Liu L, Zhao D, Ye L, Zhan T, Xiong B, Hu M, et al. A programmable CRISPR/Cas9-based phage defense system for Escherichia coli BL21(DE3). Microb Cell Factories. BioMed Central; 2020;19:1–9.

2. Miroux B, Walker JE. Over-production of Proteins inEscherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. J Mol Biol. 1996;260:289–98.

3. Seo J-H, Baek S-W, Lee J, Park J-B. Engineering Escherichia coli BL21 genome to improve the heptanoic acid tolerance by using CRISPR-Cas9 system. Biotechnol Bioprocess Eng. 2017;22:231–8.

4. Zhao D, Zhu X, Zhou H, Sun N, Wang T, Bi C, et al. CRISPR-based metabolic pathway engineering. Metab Eng. 2021;63:148–59.

5. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337:816–21.

6. Jiang Y, Chen B, Duan C, Sun B, Yang J, Yang S. Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. Appl Environ Microbiol. 2015;81:2506–14.

7. Ao X, Yao Y, Li T, Yang T-T, Dong X, Zheng Z-T, et al. A Multiplex Genome Editing Method for Escherichia coli Based on CRISPR-Cas12a. Front Microbiol [Internet]. Frontiers; 2018 [cited 2021 Jul 22];0. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2018.02307/full

8. Pyne ME, Moo-Young M, Chung DA, Chou CP. Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified Chromosomal Gene Replacement in Escherichia coli. Appl Environ Microbiol. American Society for Microbiology; 2015;81:5103–14.

9. Li Y, Lin Z, Huang C, Zhang Y, Wang Z, Tang Y-J, et al. Metabolic engineering of Escherichia coli using CRISPR-Cas9 meditated genome editing. Metab Eng. 2015;31:13–21.

10. Li Q, Sun B, Chen J, Zhang Y, Jiang Y, Yang S. A modified pCas/pTargetF system for CRISPR-Cas9-assisted genome editing in Escherichia coli. Acta Biochim Biophys Sin. 2021;53:620–7.

11. Zerbini F, Zanella I, Fraccascia D, König E, Irene C, Frattini LF, et al. Large scale validation of an efficient CRISPR/Cas-based multi gene editing protocol in Escherichia coli. Microb Cell Factories. BioMed Central; 2017;16:1–18.

12. Vento JM, Crook N, Beisel CL. Barriers to genome editing with CRISPR in bacteria. J Ind Microbiol Biotechnol. 2019;46:1327–41.

13. Moreb EA, Hoover B, Yaseen A, Valyasevi N, Roecker Z, Menacho-Melgar R, et al. Managing the SOS Response for Enhanced CRISPR-Cas-Based Recombineering in E. coli through Transient Inhibition of Host RecA Activity. ACS Synth Biol. 2017;6:2209–18.

14. Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res. 2019;47:W171–4.

15. Goodall ECA, Robinson A, Johnston IG, Jabbari S, Turner KA, Cunningham AF, et al. The Essential Genome of Escherichia coli K-12. mBio. 2018;9.

16. Zhang Y, Wang J, Wang Z, Zhang Y, Shi S, Nielsen J, et al. A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in Saccharomyces cerevisiae. Nat Commun. 2019;10:1053.

17. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat Biotechnol. Nature Publishing Group; 2016;34:339–44.

18. Wilson LOW, O'Brien AR, Bauer DC. The Current State and Future of CRISPR-Cas9 gRNA Design Tools. Front Pharmacol [Internet]. Frontiers; 2018 [cited 2021 Jun 17];9. Available from: https://www.frontiersin.org/articles/10.3389/fphar.2018.00749/full

19. Liu G, Zhang Y, Zhang T. Computational approaches for effective CRISPR guide RNA design and evaluation. Comput Struct Biotechnol J. 2020;18:35–44.

20. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2:2006.0008.

21. Zhang C, Chen X, Lindley ND, Too H-P. A "plug-n-play" modular metabolic system for the production of apocarotenoids. Biotechnol Bioeng. 2018;115:174–83.

22. Zhang C, Seow VY, Chen X, Too H-P. Multidimensional heuristic process for high-yield production of astaxanthin and fragrance molecules in Escherichia coli. Nat Commun. 2018;9:1–12.

23. Chen X, Shukal S, Zhang C. Integrating Enzyme and Metabolic Engineering Tools for Enhanced α-Ionone Production. J Agric Food Chem. 2019;

24. Liu Y, Low ZJ, Ma X, Liang H, Sinskey AJ, Stephanopoulos G, et al. Using biopolymer bodies for encapsulation of hydrophobic products in bacterium. Metab Eng. 2020;61:206–14.

25. Ma T, Shi B, Ye Z, Li X, Liu M, Chen Y, et al. Lipid engineering combined with systematic metabolic engineering of Saccharomyces cerevisiae for high-yield production of lycopene. Metab Eng. 2019;52:134–42.

26. Bacteria Getting into Shape: Genetic Determinants of E. coli Morphology [Internet]. mBio. [cited 2021 Jun 8]. Available from: https://journals.asm.org/doi/abs/10.1128/mBio.01977-16

27. Ebersbach G, Galli E, Møller-Jensen J, Löwe J, Gerdes K. Novel coiled-coil cell division factor ZapB stimulates Z ring assembly and cell division. Mol Microbiol. 2008;68:720–35.

28. Uehara T, Parzych KR, Dinh T, Bernhardt TG. Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis. EMBO J. 2010;29:1412–22.

29. Uehara T, Dinh T, Bernhardt TG. LytM-domain factors are required for daughter cell separation and rapid ampicillin-induced lysis in Escherichia coli. J Bacteriol. 2009;191:5094–107.

30. Stanley NR, Findlay K, Berks BC, Palmer T. Escherichia coli strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope. J Bacteriol. 2001;183:139–44.

31. Shukal S, Chen X, Zhang C. Systematic engineering for high-yield production of viridiflorol and amorphadiene in auxotrophic Escherichia coli. Metab Eng. 2019;55:170–8.

32. Xu L, Wang L, Zhou X-R, Chen W-C, Singh S, Hu Z, et al. Stepwise metabolic engineering of Escherichia coli to produce triacylglycerol rich in medium-chain fatty acids. Biotechnol Biofuels. 2018;11:177.

33. Wang L, Jiang S, Chen W-C, Zhou X-R, Huang T-X, Huang F-H, et al. The Phospholipid:Diacylglycerol Acyltransferase-Mediated Acyl-Coenzyme A-Independent Pathway Efficiently Diverts Fatty Acid Flux from Phospholipid into Triacylglycerol in Escherichia coli. Appl Environ Microbiol. 2020;86.

34. Chen X, Zhang C, Lindley ND. Metabolic Engineering Strategies for Sustainable Terpenoid Flavor and Fragrance Synthesis. J Agric Food Chem [Internet]. 2019 [cited 2020 Jan 30]; Available from: https://doi.org/10.1021/acs.jafc.9b06203

35. Ku JT, Chen AY, Lan EI. Metabolic Engineering Design Strategies for Increasing Acetyl-CoA Flux. Metabolites. Multidisciplinary Digital Publishing Institute; 2020;10:166.

36. Guo L, Diao W, Gao C, Hu G, Ding Q, Ye C, et al. Engineering Escherichia coli lifespan for enhancing chemical production. Nat Catal. Nature Publishing Group; 2020;3:307–18.

583 37. Meadows AL, Hawkins KM, Tsegaye Y, Antipov E, Kim Y, Raetz L, et al. Rewriting
584 yeast central carbon metabolism for industrial isoprenoid production. Nature.
585 2016;537:694–7.

586 38. Shanmugam S, Ngo H-H, Wu Y-R. Advanced CRISPR/Cas-based genome editing
587 tools for microbial biofuels production: A review. Renew Energy. 2020;149:1107–19.

588 39. Cress BF, Toparlak ÖD, Guleria S, Lebovich M, Stieglitz JT, Englaender JA, et al.
589 CRISPathBrick: Modular Combinatorial Assembly of Type II-A CRISPR Arrays for dCas9-
590 Mediated Multiplex Transcriptional Repression in E. coli. ACS Synth Biol. 2015;4:987–
591 1000.

592

Tables

Table 1. Target genes and gRNA designs used in the study with reference to Figure 3.

Gene	Gene direction	gRNA targeting Strand	Proximate gene end	gRNA sequence	Deletion size	PAM genomic Position	PAM-to-HA distance	knockout efficiency
adhE	Reverse	+	3'	ccttccctgactctgggttg	300	1283650	138	1/15
	Reverse	+	3'	aaatctatctacttccgccg	300	1283521	9	10/16
	Reverse	-	3'	cggcggaagtagatagattt	300	1283542	29	14/15
	Reverse	+	5'	ctcttgcctgtacactgacc	300	1283789	23	12/16
	Reverse	-	5'	ggttatcctggtcagtgtac	300	1283801	12	0/16
aroA	Forward	-	3'	gcagctggcgcggattagcc	1284	965175	11	2/9
	Forward	+	3'	gccagctgctcgaaataatc	1284	965143	43	5/10
	Forward	+	5'	agcgatgggttgtaacgtca	1284	963911	10	3/4
	Forward	-	5'	gttgtagagagttgagttca	1284	963903	1	0/4
zapB	Forward	-	5'	attagaagtgtttgagaaac	246	4026469	31	4/4
	Forward	-	3'	gcaggccctgctgggtcgca	246	4026670	14	1/4

tnaA	Forward	+	5'	cacgaatgcggaacggttca	1416	3777548	23	4/8
	Forward	+	3'	ttaaacttctttcagttttg	1416	3778922	20	1/8
aroB	Reverse	-	5'	aaccagatgcgatggtaatt	1089	3378824	41	1/4
	Reverse	-	3'	ttgttacgctgattgacaat	1089	3377793	14	2/4
aroC	Reverse	-	3'	ttattcatttttaccagcg	1086	2333650	9	1/12
	Reverse	-	5'	cccgtgcgattcgccgaaag	1086	2334687	40	1/12
sdhABCD	Forward	+	5'	gtctgtaggtccagattaac	2497	713776	30	0/4
	Forward	+	3'	gcaacaacatcgacttgata	2497	716941	22	4/4
serA	Reverse	-	3'	aggggaattagtacagcaga	1233	2888049	13	2/4
	Reverse	-	5'	gaaggctttccagcgccttt	1233	2889207	7	0/4
serB	Forward	-	5'	ccttaatgcctaacattacc	969	4542265	15	3/4
	Forward	-	3'	ggggtattctgcatcctctc	969	4543197	22	3/4
adeD	Forward	+	5'	ggctaatgtgatgaaattta	1767	3732350	20	0/4
	Forward	+	3'	cgtgacttccagcgtagtga	1767	3734071	26	3/4
speD	Reverse	-	3'	aaataaatctggcggagcct	873	137573	20	3/4
	Reverse	-	5'	caatttcttatcttctcctt	873	138404	6	3/4

metJ	Reverse	+	5'	gtctcaatttattgacgaag	318	4036957	24	2/4
	Reverse	+	3'	ggggattaacccggagacgt	318	4036635	11	0/4
tatA	Forward	-	5'	ttaatcatcatctaccacag	317	3928846	37	3/4
	Forward	-	3'	ggcggatacgaatcaggaac	317	3929079	44	4/4
hold	Forward	+	3'	gcaaatttgttgccataacg	383	4526535	24	2/12
	Forward	+	5'	ctgtaactgccagtctcgtc	383	4526187	10	1/12
rodZ	Reverse	+	3'	cagtacagatccagtatcaa	943	2506397	27	0/4
	Reverse	+	5'	caaaatgaagcacttactac	943	2507280	33	0/4
envC	Forward	-	5'	aggcgattaataccatgaca	1282	3656045	29	3/4
	Forward	-	3'	gtcaatccacagccgtggtt	1282	3657289	10	2/4
pdxH	Reverse	-	3'	tccacgcatcattttcacgc	657	1662590	46	1/4
	Reverse	-	5'	acccgcctttggtgtattca	657	1663160	44	1/4
ybaS	Forward	-	5'	aaacaaattacagcaggcag	933	480819	31	4/4
	Forward	-	3'	gcatcggtcgctaagcaact	933	481696	18	4/4
gadC	Reverse	-	3'	tcattcatcacaatatagtg	1605	1524468	33	4/4
	Reverse	-	5'	ccctaaaacggtattcctgt	1605	1525998	39	4/4

dgkA	Forward	+	5'	ggtgaatccagtggtattat	369	4164015	4	7/8
	Forward	+	3'	cgaccataacaggatgcacc	369	4164345	32	4/8
Pta	Forward	+	5'	agggatcagcataataatac	2145	2302006	4	0/10
	Forward	+	3'	ctgaatcgcagtcagcgcga	2145	2304106	24	1/10
fadE	Reverse	-	5'	gattttgagtattctcgcta	1542	247852	5	3/4
	Reverse	-	3'	gattgccatcaccgttgaag	1542	246385	25	4/4
pflB	Reverse	+	3'	gctgactaaagaacagcagc	2283	956400	35	1/4
	Reverse	+	5'	atgaaaagttagccacagcc	2283	958612	33	3/4
poxB	Reverse	+	3'	tggcgaaaacgaactggcta	1719	914040	3	4/4
	Reverse	+	5'	tatcgccaaaacactcgaat	1719	915710	43	4/4
ldhA	Reverse	-	3'	ttaaaccagttcgttcgggc	990	1417433	0	4/4
	Reverse	-	5'	gtactgttttgtgctataaa	990	1418390	10	4/4
aceAB	Forward	+	5'	ccttgtgaaagccagttcat	2436	4123387	22	1/4
	Forward	+	3'	agtacggaagaagccttcac	2436	4125768	30	0/3
ackA-pta	Forward	+	5'	accatttactgcatcgatga	3422	2300780	55	2/2
	Forward	+	3'	ctgaatcgcagtcagcgcga	3422	2304106	38	1/1

595 Table 2. Strains and plasmids used in this study

Name	Description	Reference	Remarks
E. coli BL21-Gold (DE3)	F$^-$ *ompT hsdS* (r_B^- m_B^-) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* Hte	Stratagene	Base strain for genome editing
B2	*E. coli* BL21-Gold (DE3) Δ*aroA* Δ*aroB* Δ*aroC* Δ*serC* Δ*zapB*	This study	Auxotrophic strain with increased size
B3	B2 Δ*dgkA*	This study	Strain to increase TAG production
B4	B3 Δ*adhE*	This study	Strain to increase acetyl-coA
B5	B3 Δ*ldhA*	This study	Strain to increase acetyl-coA
B6	B3 Δ*adhE* Δ*ldhA*	This study	Strain to increase acetyl-coA
pTarget	Plasmid used to express sgRNA under J23119 promoter. *SacB* gene under its native promoter is inserted between aadA and pMB1 origin of replication.	This study	Modified based on [6]

pCas-V5	Plasmid to express cas9 under its native promoter, λRed from pBAD promoter. It carries a temperature sensitive origin of replication and is kanamycin resistent. The sgRNA targeting pMB1 origin of replication is removed.	This study	Modified based on [6]
pCas-V2	Similar to pCas-v5 except λRed is controlled by T7 promoter	This study	Modified based on [6]
pCas-V6	Similar to pCas-v5 with inactive *recA56* being overexpressed with *cas9* in a polycistronic manner.	This study	Modified based on [6]
p15A-spec-Tm1-hmgS-atoB-hmgR	Plasmid for overexpression of *hmgs, atoB, thmgR* genes, controlled by mutated Tm1 promoter. It carries spectromycin resistance gene.	[21]	Module 1
p15A-cam-Tm2-mevK-pmk-pmd-idi	Plasmid for overexpression of *mevK, pmK, pmd* and *idi* genes, controlled by Tm2 promoter. It carries chloramphenicol resistance gene.	[21]	Module 2

p15A-kan-Tm1-crtEBI-ispA	Plasmid for overexpression of *crtE, crtB, crtI* and *ispA* genes, controlled by Tm1 promoter. It carries kanamycine resistance gene.	[21]	Module 3
p15A-amp-Tm1-Ec.fadD (D1)	Plasmid for overexpression of *fadD* gene from *E. coli*, controlled by Tm1 promoter. It carries ampicilin resistance gene.	This study	Module 4
p15A-amp-Tm1-Ec.fadD-a.DGT-Ro.PAP (DTP1a)	Plasmid for overexpression of *fadD* gene from *E. coli*, *WS/DGAT* from *Acinetobacter baylyi* and *PAP from Rhodococcus opacus,* controlled by Tm1 promoter. It carries ampicilin resistance gene.	This study	Module 4
p15A-amp-Tm1-Ec.fadD-t.DGT-Ro.PAP (DTP1b)	Plasmid for overexpression of *fadD* gene from *E. coli*, *WS/DGAT* from *Thermomonospora curvata* and *PAP from Rhodococcus opacus,* controlled by Tm1 promoter. It carries ampicilin resistance gene.	This study	Module 4

p15A-amp-Tm1-Ro.fadD (D2)	Plasmid for overexpression of *fadD* gene from *Rhodococcus opacus*, controlled by Tm1 promoter. It carries ampicilin resistance gene.	This study	Module 4
p15A-amp-Tm1-Ro.fadD-a.DGT-Ro.PAP (DTP2a)	Plasmid for overexpression of *fadD* gene from Rhodococcus opacus, *WS/DGAT* from *Acinetobacter baylyi* and *PAP from Rhodococcus opacus,* controlled by Tm1 promoter. It carries ampicilin resistance gene.	This study	Module 4
p15A-amp-Tm1-Ro.fadD-t.DGT-Ro.PAP (DTP2b)	Plasmid for overexpression of *fadD* gene from *Rhodococcus opacus, WS/DGAT* from *Thermomonospora curvata* and *PAP* from *Rhodococcus opacus*, controlled by Tm1 promoter. It carries ampicilin resistance gene.	This study	Module 4

596

Figure Legends

Figure 1. The modified CRISPR-Cas9 based genome deletion in *E. coli* BL21 strain. a). Schematic representation of the modified two-plasmid system adapted from Jiang et al [6]. Two different methods have been tested. Both requires the pCas plasmid being transformed into *E. coli* cell first. Subsequently, the pTarget plasmid with homology arm either as PCR fragments (method 1) or carried on pTarget plasmid (method 2) were transformed into the cell. For method 1, the homolog arm is obtained with a simple PCR step where the forward primer 1 (p1) carries the upstream (40-45 bp) homologous sequence fused with the downstream (15-20bp) homologous sequence for priming, and the reverse primer 2 (p2) is about 15-20 bp targeting 500 bp downstream of p1 priming sequence. The total length of p1 primer is 60 bp to ensure efficient synthesis and PCR. The gRNA sequence is changed with restriction-free cloning method [23]. In total, 4 primers were used for each target gene modification. For method 2, to clone the pTarget plasmid, 4 PCR fragments with 8 primers are used to assemble the plasmid. b). Knockout efficiency for CRISPR-Cas9 method 2 in BL21 cell targeting *adhE* gene. Five different gRNA designs were tested where their targeting positions are illustrated. Knockout efficiency is calculated based on the number of colonies with successful deletion over the total number of colonies tested. c). Knockout efficiency for deleting various length from BL21 genome using gRNA3 targeting *adhE* region. All the efficiencies were obtained through replicate experiments.

Figure 2. Optimizing CRASH protocol targeting *adhE* gene with gRNA3. a). Recombination efficiencies with four different homolog arm (HA) design. U50D50 HA design carries 50 bp upstream homologous sequence and 50 bp downstream homologous

sequence. U500D50 HA design carries 500 bp upstream homologous sequence and 50 bp downstream homologous sequence. U50D500 HA design carries 50 bp upstream homologous sequence and 500 bp downstream homologous sequence. U500D500 HA design carries 500 bp upstream homologous sequence and 500 bp downstream homologous sequence. All of them are obtained via PCR using the pTarget-*adhE* plasmid carrying the homolog arm as template. The template is completely removed before transforming the PCR products into BL21 cells for gene deletion. The number of colonies formed on the plates were counted by Qpix (Molecular Device) to determine the transformation efficiency. b). Gel image of colony PCR results to check successfully deleted colonies. C is the PCR products obtained from non-edited cell. The size of non-edited (WT) and edited (Deleted) is indicated on the side of the gel. c). Knockout efficiency and transformation efficiency with varied upstream HA length. The downstream HA length is kept at 500 bp. d). Knockout efficiency and transformation efficiency when various length of gene to be deleted from BL21 genome. All the efficiencies were obtained with replicate experiments.

Figure 3. Testing gRNA designs and CRASH protocol across BL21 genome. a). Illustrations of various gRNA designs. The red arrow indicates the most preferred gRNA design targeting the negative (-) strand of DNA. The orange arrow indicates the most preferred gRNA design targeting the positive (+) strand of DNA. b). Positions of the targeted genes along the BL21 genome. The reverse genes are labeled in red or anti-clockwise, and the forward genes are labeled in green or clockwise. c). The knockout efficiencies for the various genes tested. Refer to table 1 for details.

642 Figure 4. The effect of changing the cell size on lycopene production. a) Specific lycopene
643 yield and biomass of *E. coli* when genes affecting cell size were deleted. The lycopene
644 pathway comprising module 1, 2 and 3 (Table 2) are overexpressed in each of single-
645 deleted cells. All the measurements were average of triplicates with standard error bar
646 shown in the figure. b). Flow cytometry analysis on cell size distribution of single-deleted
647 genotype. c). Microscopy image of wildtype and *zapB* null strain with and without lycopene
648 production. d). Size distribution of wildtype and *zapB* null strain with and without lycopene
649 production. They were obtained by imageJ analysis.

650 Figure 5. Engineering the triacylglycerol (TAG) pathway to increase lycopene production.
651 a). Schematic representation of the TAG pathway and lycopene biosynthetic pathway. The
652 two pathways are grouped into 4 modules, which is boxed with green dash lines (Table 2).
653 The genes boxed in orange dash lines are targets to be deleted. The abbreviations are as
654 follows. *adhE*, alcohol dehydrogenase; *ldhA*, lactate dehydrogenase; *poxB*, pyruvate
655 oxidase; *pflB*, pyruvate-formate lyase; *pta*, phosphate acetyltransferase; *ackA*, acetate
656 kinase; *atoB*, Acetoacetyl-CoA thiolase; *hmgS*, HMG-CoA synthase; *thmgR*, truncated
657 HMG-CoA reductase; *mevk*, mevalonate kinase; *pmk*, phosphomevalonate kinase; *pmd*,
658 mevalonate pyrophosphate decarboxylase; *idi*, IPP isomerase; *ispA*, FPP synthase; *crtE*,
659 GGPP synthase; *crtB*, phytoene synthase; *crtI*, phytoene desaturase; FAB, fatty acid
660 biosynthesis; *fadD*, long-chain-fatty-acid—CoA ligase; *PAP*, phosphatidic acid
661 phosphatase; *WS/DGAT,* wax ester synthase/diacylglycerol acyltransferase; *dgkA*,
662 diacylglycerol kinase; *fadE*, acyl-coA dehydrogenase. b). Specific lycopene yield and
663 biomass of *E. coli* when *dgkA* and/or *fadE* were deleted. c). Specific lycopene yield and
664 biomass of *E. coli* when TAG pathway genes were overexpressed. All the measurements
665 were average of triplicates with standard error bar shown in the figure.

Figure 6. Increasing acetyl-CoA availability to improve lycopene production. a). Specific lycopene yield and biomass of *E. coli* when divergent flux from pyruvate or acetyl-coA was deleted. b). Specific lycopene yield and biomass of *E. coli* when combining both TAG overexpression and increasing acetyl-CoA availability. Refer to table 2 for strain description. All the measurements were average of triplicates with standard error bar shown in the figure.

Figures

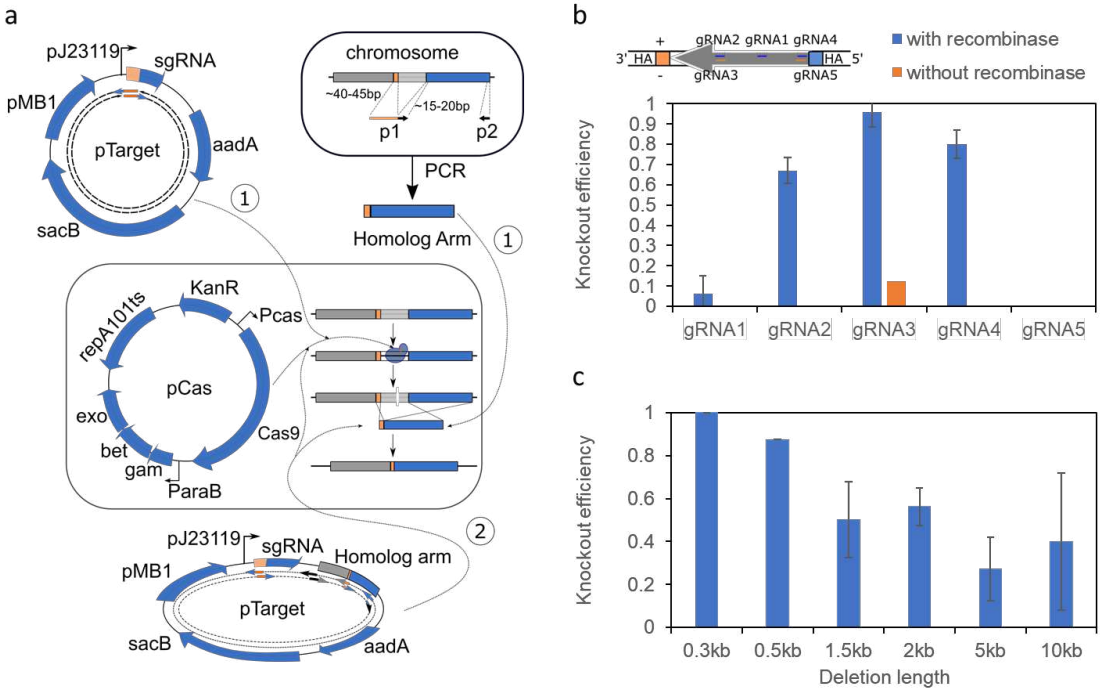


Figure 1.

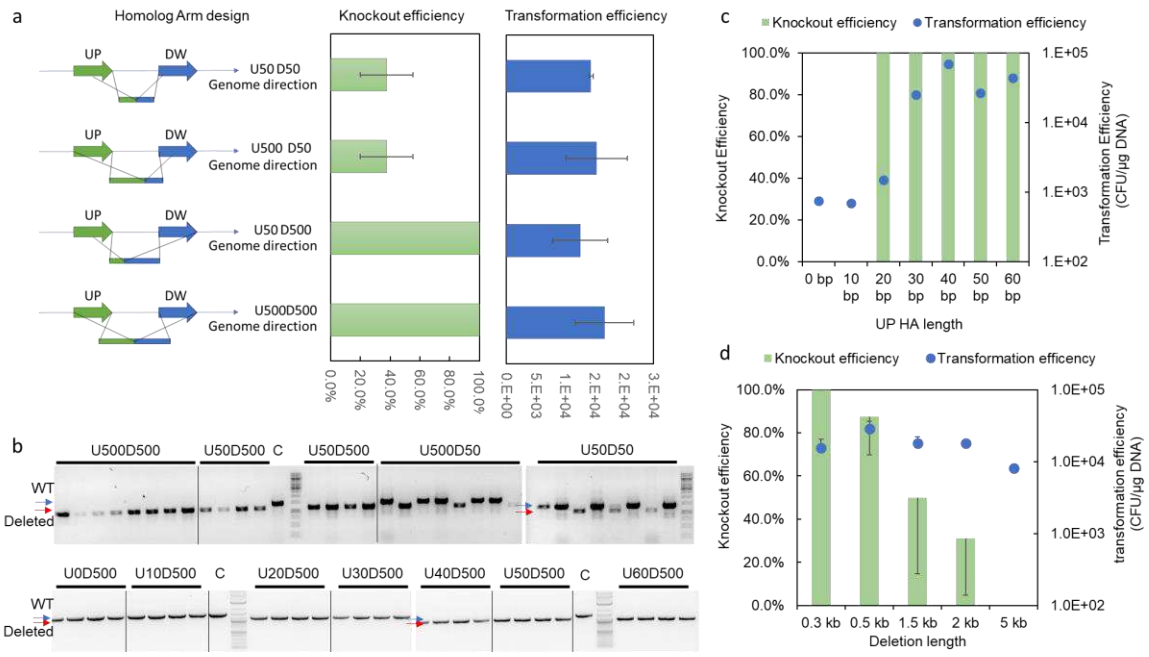

a
Homolog Arm design
Knockout efficiency
Transformation efficiency
UP
DW
U50 D50
Genome direction
U500 D50
U50 D500
U500D500
0.0%
20.0%
40.0%
60.0%
80.0%
100.0%
0.E+00
5.E+03
1.E+04
2.E+04
2.E+04
3.E+04
b
U500D500
U50D500
C
U50D500
U500D50
U50D50
WT
Deleted
U0D500
U10D500
C
U20D500
U30D500
U40D500
U50D500
C
U60D500
WT
Deleted
c
Knockout efficiency
Transformation efficiency
100.0%
80.0%
60.0%
40.0%
20.0%
0.0%
Knockout Efficiency
1.E+05
1.E+04
1.E+03
1.E+02
Transformation Efficiency
(CFU/μg DNA)
0 bp
10 bp
20 bp
30 bp
40 bp
50 bp
60 bp
UP HA length
d
Knockout efficiency
Transformation efficency
100.0%
80.0%
60.0%
40.0%
20.0%
0.0%
Knockout efficiency
1.0E+05
1.0E+04
1.0E+03
1.0E+02
transformation efficiency
(CFU/μg DNA)
0.3 kb
0.5 kb
1.5 kb
2 kb
5 kb
Deletion length

Figure 2

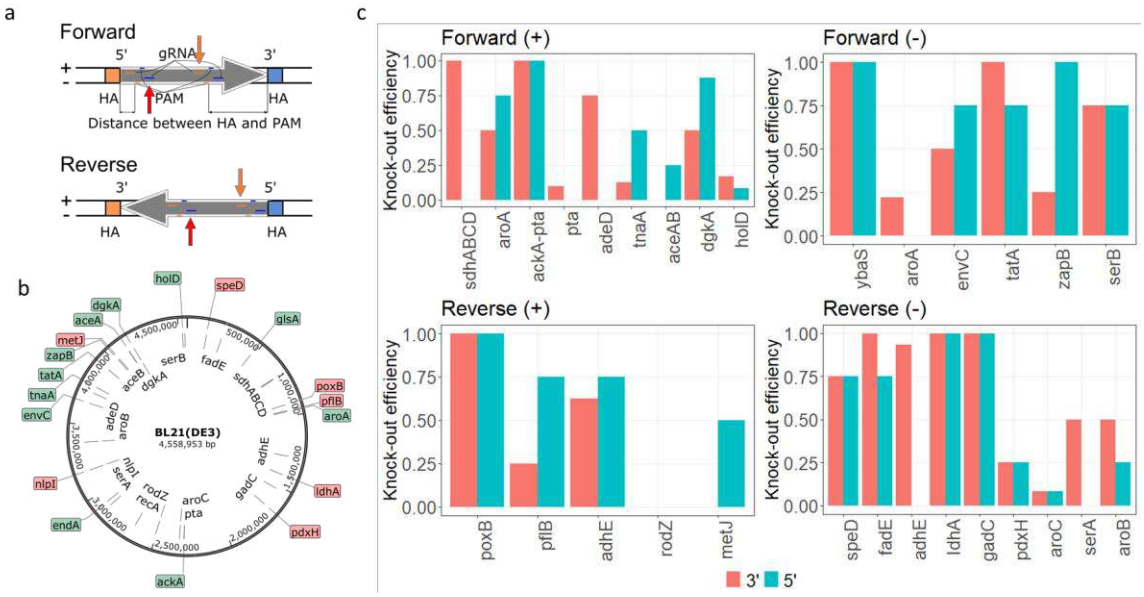

a
Forward
5'
gRNA
3'
+
-
HA
PAM
HA
Distance between HA and PAM
Reverse
3'
5'
HA
HA
b
BL21(DE3)
4,558,953 bp
c
Forward (+)
Forward (-)
Reverse (+)
Reverse (-)
Knock-out efficiency
1.00
0.75
0.50
0.25
0.00
sdhABCD
aroA
ackA-pta
pta
adeD
tnaA
aceAB
dgkA
holD
ybaS
aroA
envC
tatA
zapB
serB
poxB
pflB
adhE
rodZ
metJ
speD
fadE
adhE
ldhA
gadC
pdxH
aroC
serA
aroB
3'
5'

Figure 3.

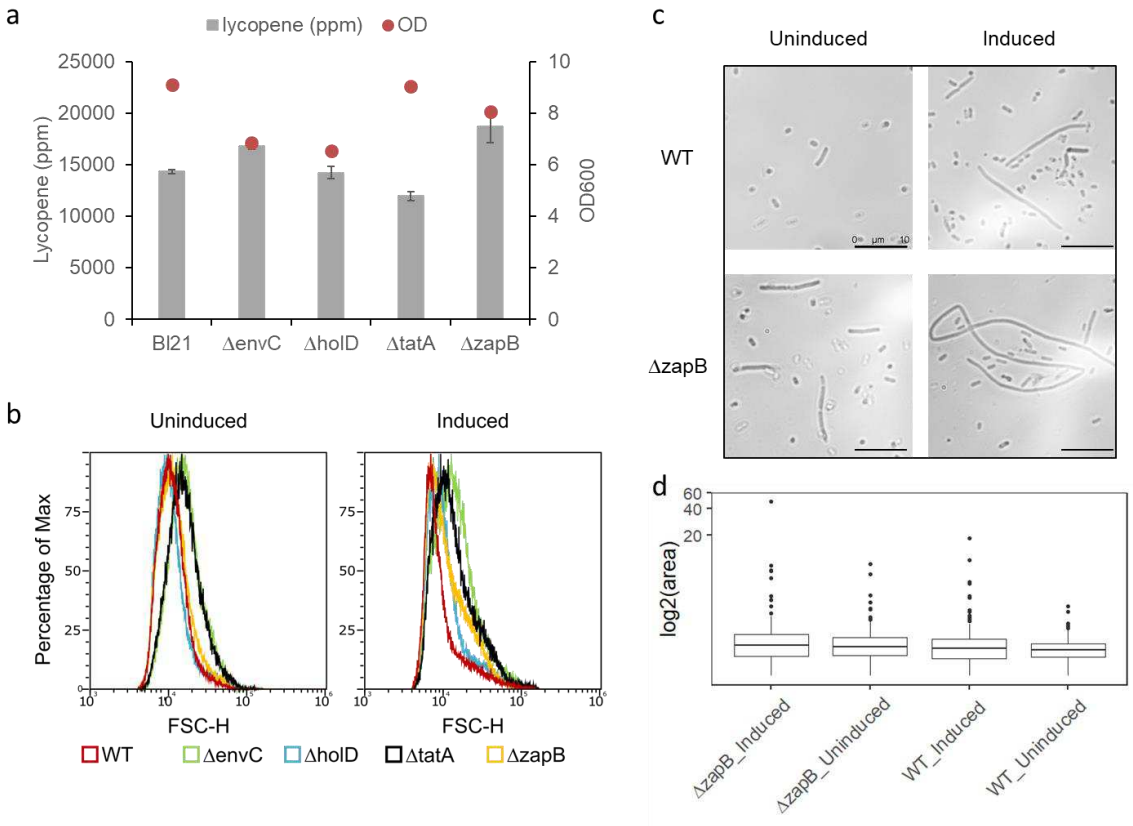

a
lycopene (ppm)
OD
Lycopene (ppm)
OD600
25000
20000
15000
10000
5000
0
10
8
6
4
2
0
BI21
ΔenvC
ΔholD
ΔtatA
ΔzapB
b
Uninduced
Induced
Percentage of Max
75
50
25
0
FSC-H
WT
ΔenvC
ΔholD
ΔtatA
ΔzapB
c
Uninduced
Induced
WT
ΔzapB
d
log2(area)
60
40
20
ΔzapB_Induced
ΔzapB_Uninduced
WT_Induced
WT_Uninduced

Figure 4.

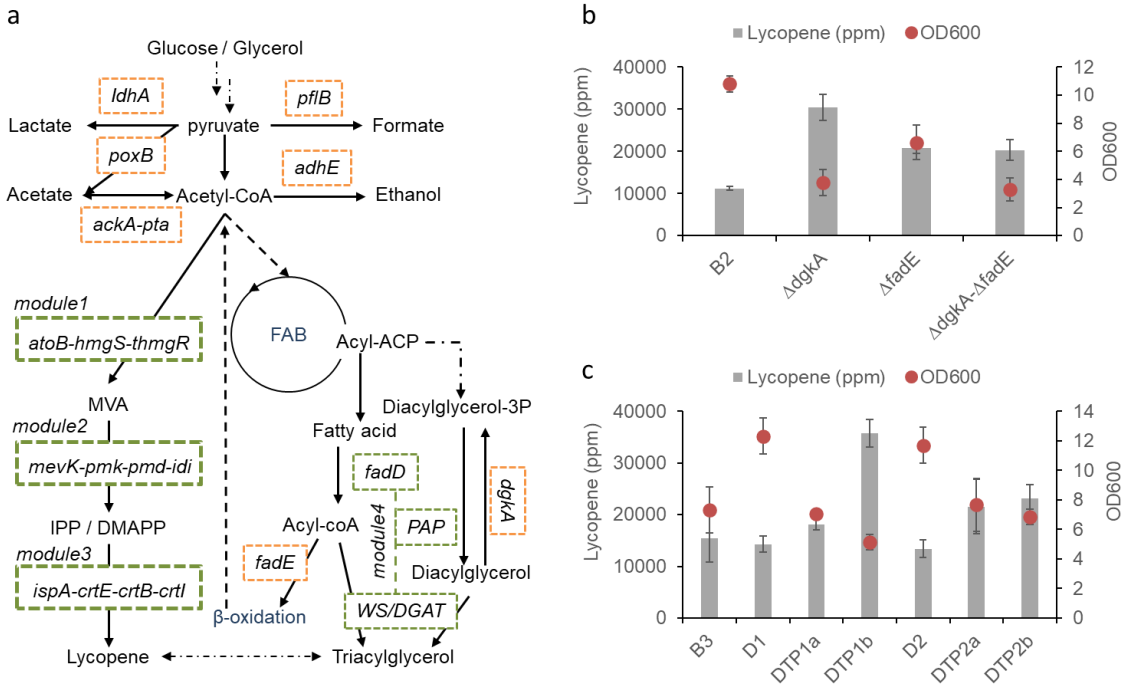
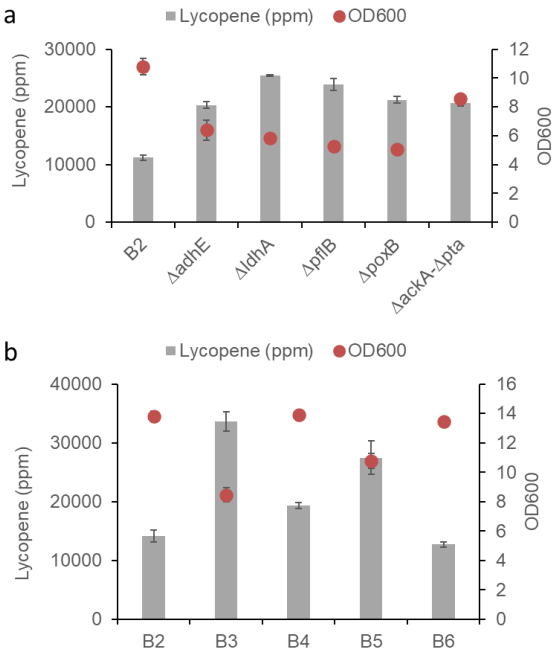

a
Glucose / Glycerol
ldhA
pflB
Lactate
pyruvate
Formate
poxB
adhE
Acetate
Acetyl-CoA
Ethanol
ackA-pta
module1
atoB-hmgS-thmgR
FAB
Acyl-ACP
MVA
Diacylglycerol-3P
module2
Fatty acid
mevK-pmk-pmd-idi
fadD
dgkA
IPP / DMAPP
Acyl-coA
PAP
module3
module4
fadE
Diacylglycerol
ispA-crtE-crtB-crtI
β-oxidation
WS/DGAT
Lycopene
Triacylglycerol
b
Lycopene (ppm)
OD600
B2
ΔdgkA
ΔfadE
ΔdgkA-ΔfadE
c
B3
D1
DTP1a
DTP1b
D2
DTP2a
DTP2b

Figure 5.

755



756

757 Figure 6.

758

759

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Crisprmanuscriptsupplementaryinformation.pdf