Title

Toward industrial production of isoprenoids in Escherichia coli: Lessons learned from CRISPR-Cas9 based optimization of a chromosomally integrated mevalonate pathway

Journal

Biotechnology and Bioengineering, 115(4)

ISSN

0006-3592

Authors

Alonso-Gutierrez, Jorge
Koma, Daisuke
Hu, Qijun
et al.

Publication Date

2018-04-01

DOI

10.1002/bit.26530

Peer reviewed

Engineering Science of Biological Systems **Biotechnology and Bioengineering**
DOI 10.1002/bit.26530

Towards industrial production of isoprenoids in *Escherichia coli*: lessons learned from CRISPR-Cas9 based optimization of a chromosomally integrated mevalonate pathway[†]

Jorge Alonso-Gutierrez[1,2,#,†], Daisuke Koma[1,2,3,#], Qijun Hu[1,2], Yuchen Yang[1,2], Leanne Jade G. Chan[1,2], Christopher J. Petzold[1,2], Paul D. Adams[1,4], Claudia E. Vickers[5], Lars K. Nielsen[5], Jay D. Keasling[1,2,6,7,8], Taek Soon Lee[1,2,*]

[1] Joint BioEnergy Institute (JBEI), 5885 Hollis Street, Emeryville, CA 94608, USA
[2] Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[3] Osaka Municipal Technical Research Institute, Osaka, Japan
[4] Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[5] Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St. Lucia, QLD 4072, Australia
[6] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2970 Horsholm, Denmark
[7] Department of Chemical & Biomolecular Engineering, University of California, Berkeley, Berkeley, CA 94720, USA
[8] Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA

[#]These authors contributed equally

[†] Current affiliation: ZymoChem Inc., 4 Anchor Dr. Suite 231, Emeryville, CA 94608, USA

[*]Corresponding author: Dr. Taek Soon Lee, Joint BioEnergy Institute, 5885 Hollis St. 4th floor, Emeryville, CA 94608, USA; Phone: +1-510-495-2470, Fax: +1-510-495-2629, E-mail: tslee@lbl.gov

Additional Supporting Information may be found in the online version of this article.

Received September 28, 2017; Revision Received December 18, 2017; Accepted December 22, 2017

Abstract

Escherichia coli has been the organism of choice for the production of different chemicals by engineering native and heterologous pathways. In the present study, we simultaneously address some of the main issues associated with *E. coli* as an industrial platform for isoprenoids, including an inability to grow on sucrose, a lack of endogenous control over toxic mevalonate (MVA) pathway intermediates, and the limited pathway engineering into the chromosome. As a proof of concept, we generated an *E. coli* DH1 strain able to produce the isoprenoid bisabolene from sucrose by integrating the *cscAKB* operon into the chromosome and by expressing a heterologous MVA pathway under stress-responsive control. Production levels dropped dramatically relative to plasmid-mediated expression when the entire pathway was integrated into the chromosome. In order to optimize the chromosomally integrated MVA pathway, we established a CRISPR-Cas9 system to rapidly and systematically replace promoter sequences. This strategy led to higher pathway expression and a 5-fold improvement in bisabolene production. More interestingly, we analyzed proteomics data sets to understand and address some of the challenges associated with metabolic engineering of the chromosomally integrated pathway. This report shows that integrating plasmid-optimized operons into the genome and making them work optimally is not a straightforward task and any poor engineering choices on the chromosome may lead to cell death rather than just resulting in low titers. Based on these results, we also propose directions for chromosomal metabolic engineering.

Key words

Chromosomal engineering, sucrose utilization, mevalonate pathway, CRISPR-Cas9, *E. coli* DH1, biofuel

This article is protected by copyright. All rights reserved

Introduction

Bio-based production of fuels and chemicals makes use of cheap, renewable carbon sources to synthesize molecules that the traditional chemical industry manufactures from fossil resources. Increasing costs and volatility of petroleum prices, in combination with substantial improvements in biotechnology, have stimulated the use of microbial production systems for fuels and chemicals (Keasling et al., 2012). Advances in synthetic biology in the last two decades (Cameron et al., 2014) have enabled the engineering of microbial hosts able to transform renewable feedstocks into valuable products ranging from drugs (Ajikumar et al., 2010; Paddon et al., 2013) and commodity chemicals (Yim et al., 2011) to biofuels (Atsumi et al., 2008; Bokinsky et al., 2011; Dellomonaco et al., 2011; Steen et al., 2010).

Industrial microbial production of high-volume, low cost chemicals and fuels, however, requires very high efficiency in order to compete with the petroleum-based counterparts (Meadows et al., 2016). The efficiency of a bioprocess is defined by the microbial host ability to transform cheap carbon sources into valuable molecules at high titers, rates and yields. Industry is therefore constantly looking for faster ways to engineer these capabilities in order to reduce production costs and maximize the profit. The following five host characteristics would be keys to bring a microbe closer to industrial application: i) deep understanding of the host metabolism, ii) availability of genetic tools, iii) capacity to grow on inexpensive carbon sources with a short doubling time, iv) an autonomous, inducer-free expression system that doesn't allow accumulation of toxic metabolites, and v) chromosomally integrated pathway operons to avoid genetic instability and the use of expensive antibiotics required to maintain plasmids.

The two most popular microbial hosts for industrial production are the model eukaryote, baker's yeast (*Saccharomyces cerevisiae*), and the model prokaryote, *Escherichia coli* (Westfall

and Gardner, 2011), due to the plethora of tools and knowledge available for them. *E. coli* is possibly the most studied organism, and its advantages include: i) rapid doubling time, ii) the most diverse set of tools to control gene expression (promoters, expression vectors, etc), iii) ease of engineering, iv) fully annotated genome, and v) detailed understanding of its metabolism. All these qualities have enabled extensive applications of engineered *E. coli* for industrial use (Keasling, 2010; Yim et al., 2011), and made it the organism of choice by many leaders in the biotechnology field (Schirmer et al., 2010; Steen et al., 2010; Westfall and Gardner, 2011). However, there are several issues to using *E. coli* as an industrial host, including i) susceptibility to phage contamination (Junker et al., 2006), ii) inefficiency in growing on certain desirable carbon sources (especially sucrose), and iii) low recombination capacity of DNA into its chromosome (Jeong et al., 2013; Mosberg et al., 2010). The first two issues have been previously addressed. Countermeasures for bacteriophage contamination have been developed (Los et al., 2004; Los et al., 2007), and engineering approaches to introduce effective sucrose utilization capabilities are now available even though most industrial strains cannot grow on sucrose (Lee et al., 2010; Bruschi et al., 2011). Feedstocks are the most expensive component of a bioprocess (Davis et al., 2013; Humbird et al., 2011), and high feedstock cost severely hampers the economics of a bioprocess and its competitiveness with petroleum-based chemicals. Sucrose from sugarcane is a high purity and low energy input feedstock, and performs as well or better than lignocellulosics for greenhouse gas reduction, making it a preferred feedstock for industrial fermentation (Renouf et al., 2008; Klein-Marcuschamer et al., 2013). Therefore, engineering sucrose utilization is an important feature for an industrial host.

Extensive research has proven the *E. coli* DH1 genotype suitable for the production of a myriad of isoprenoids as well as fatty acid-derived molecules at high titers (Alonso-Gutierrez et

al., 2013, 2015; George et al., 2015; Goh et al., 2014; Steen et al., 2010). Dynamic control of expression of the MVA pathway has also been developed in this strain to circumvent endogenous control while allowing DH1 cells to sense their own environment and avoid accumulation of toxic intermediates (Dahl et al., 2013); this translates into better conversion efficiencies, avoids the need for expensive inducers, provides higher isoprenoid yields and improved strain stability – all key features for industrial production. DH1 has also been engineered to utilize sucrose albeit with some issues such as very slow growth rates and high level of instability (Garcia, 1985; Bruschi et al., 2011)). Introduction of the non-PTS (permease) utilization mechanism from *E. coli* W encoded by the *csc* regulon, overcomes these problems. Key to this success was removal of the *cscR* regulator from the *csc* regulon and, most critically, introduction in single copy onto the chromosome to avoid severe toxicity and instability which results from plasmid-mediated expression. This approach allows K-12 MG1655, B and C strains to grow at the same rate on sucrose as they do on glucose (Bruschi et al., 2011). Engineering sucrose utilization using the *cscAKB* operon would bring DH1 even closer to industrial application.

In contrast, there is a paucity of publications addressing the metabolic engineering of chromosomally integrated production pathways, especially for the MVA pathway in *E. coli*. Although the use of plasmid-based expression systems can achieve high protein levels, this presents an inherent segregational and spatial instability and often require the use of expensive antibiotics for plasmid maintenance (Friehs, 2004). Therefore, chromosomal integration and optimization of heterologous pathways directly in the genome is an important step to generate stable and reproducible strains preferred by industry. While yeast has shown highly efficient recombination capability to easily integrate large linear pieces of DNA onto specific locations of the genome (Szostak et al., 1983), *E. coli* has a lower capacity to chromosomally integrate large

DNA fragments (Jeong et al., 2013; Mosberg et al., 2010; Sawitzke et al., 2007) and this is one of the major drawbacks. This probably explains the deficiency in publications addressing chromosomal integration of large heterologous pathways, especially in *recA*$^{-}$ *E. coli* strains used for production (e.g. DH1).

These considerations demonstrate that the ability to grow on sucrose and the expression of the heterologous pathway genes from the host chromosome under the control of self-regulated promoters are important features to implement in *E. coli* for industrial applications. Here, we report on approaches to engineer *E. coli* DH1 as a potential industrial host for the production of isoprenoid-based biofuels. We use a transferable *cscAKB* operon to confer capacity to efficiently grow on sucrose (Bruschi et al., 2011), chromosomally integrate the heterologous MVA pathway controlled by stress responsive regulation for the production of bisabolene, and develop a new engineering approach using CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and its associated Cas9 endonuclease) to add/replace T7 promoters (P_{T7}) of different strength directly onto the chromosome. Finally, we report the challenges associated with this new engineering approach on the chromosome and propose future directions for complex pathway optimization in *E. coli*.

Results

Generation of an *E. coli* DH1 strain able to grow and produce bisabolene from sucrose

To engineer bisabolene production from sucrose as a carbon source, there are two options: either express an isoprenoid pathway in naturally occurring sucrose-utilizing *E. coli* strain such as W (Archer et al., 2011) or integrate the sucrose utilization genes into a strain previously engineered for the production of isoprenoids, such is DH1. We first explored the

former option and transformed plasmids expressing the MVA pathway for the production of bisabolene into the natural sucrose utilizer *E. coli* W. However, the cells grew poorly and did not produce bisabolene at detectable levels (data not shown), suggesting there is an intrinsic problem with the expression of the MVA pathway genes in this host strain. Thus, we attempted the second option and transferred the sucrose utilizing capability from *E. coli* W into DH1. As previously described (Bruschi et al., 2011), the sucrose utilizing operon (*cscAKB*) from W strain was integrated into the DH1 chromosome at the *arsB* locus using 'knock-in/knock-out (KIKO)' plasmid vectors (Figure 1), designed to successfully integrate large pieces of DNA into different locations of the *E. coli* chromosome (Sabri et al., 2013). After the *cscAKB* operon was recombined onto the chromosome (under the control of its native constitutive promoter), a new strain (DS) was generated. DS and its derivatives were able to grow on sucrose to ODs comparable to, or sometimes even higher than those on glucose (Figure 2b).

DS strain harboring the best combination of plasmids reported for bisabolene production (JPUB_002475 + JPUB_002466, Table 1) yielded ~1 g/L of bisabolene using either sucrose or glucose (Figure 2a). Under the same conditions, the DH1 strain produced approximately the same amount of bisabolene from glucose as DS, but it only showed residual growth and production of bisabolene when grown on sucrose (Figure 2a, b), confirming the successfully transferred sucrose utilization capability in DS strain.

Inducer-free, dynamic-controlled production of bisabolene from sucrose

Strain DS was transformed with plasmid pBbA0c-P_{gadE}-MevT-MBIS (JBEI-2872, Table 1) that provides MVA pathway enzymes to synthesize farnesyl diphosphate (FPP) under the control of FPP-responsive *gadE* promoter (P_{gadE}) (Dahl et al., 2013). *RstA* promoter (P_{rstA}) is known to be up-regulated in the presence of FPP and used to control the terpene synthase

expression that consumes FPP, but its strength was too weak to efficiently express the terpene synthase and showed lower terpene production level than the strain with the terpene synthase under a stronger promoter (*trc* promoter, P_{trc}) (Dahl et al., 2013). We tried both combinations in DS strain using either glucose or sucrose as a carbon source, and surprisingly, the DS strain harboring pJBEI-2872 and pRstA-BIS, a complete dynamic system, gave a higher production (~850 mg/L; Figure 2c, d) when grown in sucrose over glucose.

Initial attempts for chromosomal integration of a self-regulated MVA pathway

We chose the MVA pathway genes for chromosomal integration from the system previously shown to have the highest bisabolene yields in *E. coli* (Alonso-Gutierrez et al., 2015). The construct containing all genes in the pathway is over 10Kb and exceeds the optimal size for successful λ-red based integration (~2Kb) (Kuhlman and Cox, 2010). Although we successfully integrated the sucrose operon (~4Kb) into the *E. coli* DH1 genome using KIKO vectors, the low recombination efficiency in *E. coli* DH1 (Hanahan, 1983) made this simple integration a difficult task even with the aid of well-established recombinases (Datsenko and Wanner, 2000; Jeong et al., 2013). We attempted to integrate the entire pathway including the terpene synthase (~12 Kb) into the DH1 chromosome using a more efficient methodology based on suicide plasmids that would recombine into the genome by a single cross-over (Haldimann and Wanner, 2001; Sabri et al., 2013). Nevertheless, we could not get any colonies with the correct integration using this method. The large size of the pathway genes and the complexity associated with the pathway (Pitera et al., 2007) seem to be the most plausible explanation for these unsuccessful attempts. Therefore, we decided to split the pathway into three operons shorter than 5Kb and integrate them in three different locations in the genome (Figure 3). With this approach we intended not

only to integrate the whole pathway but also to provide a way to achieve the optimal expression ratio among the three operons for the best production (Alonso-Gutierrez et al., 2015).

MVA pathway splitting and integration at different chromosomal locations

The MVA pathway was split into three operons as previously described (Alonso-Gutierrez et al., 2015) (Figures 3 and 4). DNA parts were collected from previously published BglBrick plasmids (Table 1) and assembled together into KIKO vectors to integrate the operons for modules: T (pKIKO-Cm-poxB-P_{gadE}-MevT), M (pKIKO-Kan-lacZ-P_{trc}-MevB) and B (pKIKO-Kan-rbsAR-LacI-P_{trc}-BS-idi-ispA) of the pathway (Figure 1, Table 1). Module T includes the MevT operon to transform acetyl-CoA into MVA (Table 1) under the control of the FPP-responsive promoter P_{gadE} , which provides dynamic control of pathway expression (Dahl et al., 2013). Modules M and B contain codon optimized genes for the optimal expression of enzymes transforming MVA into isopentenyl pyrophosphate (IPP) (MevB operon) and IPP into bisabolene, respectively, under IPTG-inducible P_{trc} promoters (Table 1). Even though P_{rstA} would be necessary to provide complete dynamic regulation in combination with P_{gadE}, we anticipated that the low transcription rate of P_{rstA} (Dahl et al., 2013) would not be enough to provide the necessary enzyme levels of the bisabolene synthase once integrated and therefore we decided to use P_{trc}.

From our previous work in engineering the MVA pathway, the ideal relative expression ratio among MVA pathway modules was determined as B > M > T (Alonso-Gutierrez et al., 2015). We used different combinations of plasmid copy number and promoter strength in this work, but the chromosomally integrated system does not allow as many levels of expression control as plasmid-based system does. It is reported that chromosomal location can markedly influence the level of expression of integrated genes (Bryant et al., 2014), which is also observed

using KIKO vectors (Sabri et al, 2013), and therefore, we placed the different portions of the pathway at distinct locations of the *E. coli* genome known to have different expression levels to achieve this ratio. Three loci (*rbsAR*, *lacZ*, and *poxB*) were selected based on the reports that expression levels decreased with distance from the genomic origin of replication (Bryant et al., 2014; Sabri et al., 2013), and we integrated three modules (B, M and T) at these locations respectively. We selected *poxB* locus to achieve additional benefit of reducing carbon loss by acetate formation as previously reported (Goh et al., 2014). To avoid potential toxicity problems by prenyl diphosphate accumulation (Dahl et al., 2013; George et al., 2014), we carefully designed the order of integration. Module M was integrated first, considering that the MevB operon would not produce any toxic intermediate by itself in the absence of module T (i.e. without MVA). We integrated module B next, and finally module T to produce MVA as shown in Figure 2.

After integrating every portion of the pathway, a new strain (named DSTMB or DK1) was generated (Table 2, Figure 4). This strain grew to an OD similar to that of the wild-type and produced ~77 and ~138 μg/L of bisabolene with glucose and sucrose as carbon source, respectively (Figure 6). These titers, however, were significantly lower than those from plasmid-based system. Therefore, we investigated why such a low production level was achieved from the integrated system, and explored novel approaches to optimize the integrated pathway.

Optimization of chromosomally integrated pathway using CRISPR-Cas9 based approach.

Plasmid-based pathway expression provides multiple copies of each gene. In contrast, when integrated onto the chromosome only one copy of each gene is responsible for the expression of the corresponding enzyme required for production. Therefore, the first hypothesis to account for the lower production level is that a single copy of pathway gene could not provide

sufficiently high enzyme concentrations, and a new strategy to provide higher enzyme expression level would be required to address this issue.

Bacteriophage T7 RNA polymerase (T7RNAP), a highly processive single subunit RNA polymerase (RNAP), and its associated promoters (P_{T7}) have been successfully used to achieve high expression levels of heterologous enzymes (Koma et al., 2012; Temme et al., 2012b; Wang et al., 2012). T7RNAP has a much higher transcriptional capability than the cell's native RNAP, allowing for high expression levels even from a single gene (Studier and Moffatt, 1986), and it has been reported that the protein levels achieved from a single chromosomal copy under P_{T7} can be comparable to those achieved from a plasmid-based expression system (Koma et al., 2012). Since P_{T7} is not responsive to the native host RNAP and slight variations in its sequence translate into different transcriptional strength, using P_{T7} provides a completely orthogonal and tunable control of expression (Temme et al., 2012a). T7RNAP was transduced into strain DSTMB genome between genes *ybhB* and *ybhC* generating strain DSTMB7 (or DK4; Table 2; Figure 5). Integration was validated both by colony PCR (primer sequences in Supplementary Table S1) and by checking the presence of plaques after infection with a T7RNAP dependent phage. Strain DK4 was confirmed to contain the *csc* operon, the whole MVA pathway and a functional T7RNAP integrated on the chromosome (Table 2 and Figures 3 and 4). The promoters driving the integrated MVA pathway in DK4 were sequentially replaced by P_{T7} using λ-red mediated integration and CRISPR-Cas9 selection of the recombinant strains (Figure 5) with the goal to achieve higher transcriptional rates, overall enzyme expression, and ultimately product titers. We chose wild type P_{T7} for modules M and B, which are known to require higher levels of expression, while we used "$P_{T7.2\_P2}$" (part number SBa_000445 (Temme et al., 2012)) for the

expression of the module T, which is known to require lower expression levels relative to the rest of the operons.

As explained in Figure 4, the CRISPR system used in this work consists of two plasmids (pSMR77 and either pMCC or pMCK depending on the resistance marker). pSMR77, which provides both the λ-red protein and the Cas9 protein, was first transformed into DK4 to generate strain DK15 (Figure 5). By keeping carbenicillin in the medium and the temperature below 30°C, this plasmid was maintained during iterative rounds of modification. Both pMCC and pMCK express the guide RNA to target specific sequences in the *E. coli* genome to be cut by Cas9. gBlock fragments (IDT (Coralville, IA)) coding for the specific crRNAs and surrounded by SMR20 (pMCC_fw) and SMR21 (pMCC_rev) were successfully cloned into *Sma*I-linearized pMCC or pMCK vectors using Gibson assembly (Table 1 and Figure 4). The presence of any of these vectors alone in combination with pSMR77 would guide the Cas9 to produce double strand breaks on two genome sites (protospacers) and therefore guarantee cell death. However, DK15 cells and their descendants containing pSM77 were co-electroporated with either pMCC or pMCK (transcribing the crRNA guiding the Cas9) and the fixing template (gBlock sequence of 500 bp with upstream and downstream homology of the cutting site) to replace the protospacers in the genome with different DNA parts (P_{T7} promoters in this case). Three fixing templates (fixT, fixM and fixB) were used in combination with their respective crRNAs (gCrT-T', gCrM-M', gCrB-B', Table 1) to remove the protospacers and to integrate a P_{T7} sequence in its place (Figure 4). At each of these locations, the originally integrated weak promoters (and its associated protospacer sequences) were substituted to stronger P_{T7} responding to T7RNAP (Figure 5).

To generate strain DK2, DK15 was first transformed with pMCC-gCrB-B' and a fixing template fixB to introduce P_{T7} in module B. DK2 was subsequently electroporated with pMCK-gCrM-M' and fixM to generate DK28, a strain with both modules M and B under P_{T7} control. However, the co-transformation of pMCC-gCrT-T' and fixT in DK28 did not generate any colony with the correct P_{T7} substitution at the module M. As an alternative, we tried to generate a triple substitution first by replacing module T promoter in DK15 using the same combination of crRNA and fixing template (pMCK-gCrT-T' and fixT). This time we got colonies after the CRISPR screening and generated strain DK16, in which P_{T7} drives only module T. This verifies that the CRISPR system for module T works and suggests that the previous failure in DK28 could be related to potential imbalance of protein expression among three operons, which might result in the accumulation of toxic intermediates. DK16 was further engineered into strain DK29 where P_{T7} drives both modules T and B. However, DK29 was unable to accept the third change (a replacement of the P_{trc} by a P_{T7} to drive module M) once again. Finally, we tried the third route by substituting P_{trc} by P_{T7} in module M of DK15. DK24 was generated in this way, but once again, substitution of module T did not work, reinforcing the hypothesis of pathway imbalance.

Strain DK16, in which a P_{T7} drives module T (fixT incorporated to DK4; Table 2, Figure 4) and its descendent, DK29 (fixB incorporated on DK16 and therefore with P_{T7} driving both modules T and B) experienced increase in bisabolene production compared to its predecessor DK4. DK16 produced ~200 µg/L and ~300 µg/L of bisabolene when cultured with glucose and sucrose, respectively, representing a ~2.5-fold increase in production compared to DK4, and DK29 produced ~435 µg/L of bisabolene using both carbon sources, representing a ~4-fold increase compared to DK4 (Figure 6). However, any of the other strains with P_{T7} driving either

module M (DK24), module B (DK2) or both M and B modules (DK28) (Table 2) experienced a ~2.5-fold decrease in production compared to DK4 (Figure 6). More interestingly, neither the combination of T and M modules changes (e.g., fixT+fixM) nor the replacement of all promoters could be achieved in any order, and we could not see any colonies from these experiments. Rather than a problem with our CRISPR system, this suggests a toxicity issue from unbalanced expression of the MVA pathway since every one of these changes could be individually achieved in different strain background.

Proteomics-aided diagnosis and promoter replacement for optimization.

In this work, to achieve better understanding on the nature of the genomic changes, to pinpoint potential bottlenecks, and to inform better engineering strategies, we integrated the heterologous pathway directly into the chromosome and modulated pathway expression by changing parts such as promoters. Proteomic analysis of the first iteration, with the same promoters (P_{trc} and P_{gadE}) as the plasmid-based system, revealed that the pathway enzyme level from the chromosome was significantly lower (on average, 10-fold lower) than what was achieved in the plasmid system (Table 3). To overcome any transcriptional limitation, we replaced those promoters with P_{T7} and the expression levels of the module T enzymes responded as expected. The expression level for the first two enzymes in the operon (AtoB and HMGS) dramatically increased in strain DK29, and comparable to the levels of those enzymes achieved from plasmid-based systems (Table 3). However, the third gene in the operon (HMGR) showed lower expression levels than that in the DK4 strain. More interestingly, the proteomic analysis revealed that the fixing template we used to fix module B did not result in a higher level for any of the enzymes in the operon but rather a decrease. This could bring some issues on the efficiency of the P_{T7}-RBS sequence we used. In addition, module M also showed unexpected

behavior. The replacement of the promoter from P_{trc} to P_{T7} did not improve the levels of the first two enzymes while the level of PMD (the third and last enzyme of the operon) was increased (Table 3). A more systematic approach, beyond the scope of this work, on various components such as integration location, promoter selection, RBS selection, and the distance between the promoter and RBS will provide better understanding of the protein expression from the chromosomal system and enable more accurate design of the system for chromosomal metabolic engineering.

Discussion

In this report, we addressed several approaches to improve *E. coli* as a host for industrial production of isoprenoid-based biofuels. Even though some *E. coli* strains such as the W strain can grow on sucrose (Archer et al., 2011), integration of the MVA pathway was not successful. The growth phenotype of the W strain suggests a stress response to the expression of the MVA pathway, which might have triggered the host's natural recombination machinery against the pathway on the plasmids, but it still need more evidence to conclude so.

Although the chromosomal integration was successful at each of the three locations in the *E. coli* DH1 genome, the overall production levels from the integrated strains were significantly lower than those from plasmid-based systems. With the rare exception of AtoB, the proteomic analysis showed very low protein expression for the integrated pathway genes (only about 10% of what is detected from plasmid-based system; Table 3). This is not surprising since there is only one copy of each gene in the chromosome while multiple copies are available in plasmid-based systems. To address this, there was an effort to multiply the copies of the integrated pathway genes using antibiotic stress (Tyo et al., 2009). However, the production level was still

not comparable to those from the plasmid-based system, and furthermore, the repeated copies of antibiotic marker in the chromosome could pose safety concerns associated with the use of antibiotic resistance markers in a bioreactor that could be spread by horizontal gene transfer into the environment (Hiltunen et al., 2017; Robinson et al., 2016).

E. coli natively corrects defective protein levels by varying the promoter sequence rather than mutating or duplicating coding genes (Amorós-Moya et al., 2010). In this work, we attempted to emulate this natural strategy by replacing the promoters in strain DSTMB instead of refactoring or adding extra copies of the pathway genes. All the promoters used in strain DSTMB for heterologous gene expression respond to *E. coli*'s native RNAP. However, these promoters are rather weak and therefore we had to substitute them with stronger promoters such as $P_{T7,}$ in the presence of T7RNAP, and further optimized the integrated pathway by engineering regulatory sequences such as promoters and RBSs using CRISPR-Cas9 approach which allows a rapid, scar-less and multiplex edition of *E. coli* genome with high efficiency (Jiang et al., 2013; Jiang et al., 2015; Li et al., 2015).

As described in the results section, we replaced promoters individually at each of the three locations with P_{T7} and improved bisabolene production from sucrose up to 4-fold when both modules T and B were overexpressed under P_{T7}. However, the simultaneous replacement of all three promoters by P_{T7} was not successful in any combination. This suggests that there could be an intrinsic issue of unbalanced protein expression, which precludes the system from further optimization.

Even though we thoroughly investigated the best relative locations for the integration of each of the three operons based on previous literature, the proteomic analysis of the strain DSTMB showed that the relative enzyme expression levels were far from the optimal ratio

achieved on plasmid-based systems (Alonso-Gutierrez et al., 2015). The *rbsAR* locus was previously reported to allow higher expression levels than *lacZ* and *poxB* loci (Sabri et al., 2013), and therefore we chose it for the expression of module B of the pathway to achieve the highest expression level among three modules (Alonso-Gutierrez et al., 2015). However, the overall levels of module B proteins turned out to be the lowest even after we replaced its promoter to P_{T7} as shown in Table 3. There are many factors known to influence the level of expression including chromosomal location, directionality and genetic context (Block et al., 2012; Bryant et al., 2014; Reyes-Lamothe et al., 2012; Sabri et al., 2013), and therefore, using chromosomal location by itself might not be enough to control gene expression. Our promoter replacement strategy showed some success in engineering gene expression in the chromosome. The proteomic analysis revealed that the expression level for the first two enzymes in module T substantially increased after the promoter replacement with P_{T7} (Table 3). However, the replacement by P_{T7} in the modules M and B did not translate as expected even though we have confirmed that there were no artifacts in the proteomic analysis and the chromosomal sequence.

In this work, we originally attempted to optimize a chromosomally integrated MVA pathway by aiming to meet a specific ratio of pathway enzymes expression previously shown in the plasmid-based systems (Alonso-Gutierrez et al., 2015). Based on our proteomics data, however, the results were not consistent to our prediction. We could not clearly explain whether this was due to a poor choice of chromosomal location and/or promoter-RBS regulatory parts, but we expect a more systematic analysis would be helpful to elucidate these results in the future.

From the cell growth and the production levels of the genome integrated strains, we noticed that the engineering for sub-optimal enzyme levels may cause severe toxicity and cell death when performed directly in the genome, while it usually leads only to low-producing

phenotypes in plasmid-based strains. The chromosomal pathway optimization strategy based on iterative cycles of part replacements and analysis may only be able to cover limited subgroups and non-viable intermediate phenotypes could preclude identification of the optimal phenotype. Therefore, a technique allowing multiple changes in the chromosome in a concurrent fashion would be more desired. It has been reported that CRISPR can edit up to three locations simultaneously (Jiang et al., 2015), but many more simultaneous changes might be necessary to optimize complex metabolic pathways such as the MVA pathway. Multiplex Automated Genome Engineering (MAGE) and its subsequent improved versions (Isaacs et al., 2011; Jeong et al., 2013; Wang et al., 2009; Wang et al., 2012) enable the simultaneous change of multiple regulatory regions that translate into production improvements. P_{T7} consists of just 23 bp and slight variations in the promoter sequences as well as RBS sequences can provide significantly wide range of the transcription strength (Salis et al., 2009; Temme et al., 2012a). However, MAGE requires an efficient screening system, and without a proper screening method the use of this technique will be limited. Recently, a new method called 'CRISPR Optimized MAGE Recombineering' (CRMAGE) (Ronda et al., 2016) combines CRISPR as a means of screening for strains, and this approach would be a good fit in our chromosomal engineering approach. The development of better multiplex editing tools compatible with an automated design-build-test-learn (DBTL) cycle would be necessary to explore such complex genetic spaces and to generate industrially relevant strains.

Finally, we envision the implementation of a dynamic control system (such as *quorum sensing* (e.g. Lux I/R system) or stress-responsive promoters (e.g. P_{gadE}/P_{rstA})) to express T7RNAP, which would translate into an inducer-free, self-regulating system for chromosomally integrated pathways driven by P_{T7}. The resulting strain would work free of expensive antibiotics

and inducers while having an improved stability; all of which are key features for an industrially relevant host.

Materials and Methods

All chemicals, solvents and medium components were purchased and used without modification from Sigma-Aldrich (St. Louis, MO), Fisher Scientific (Pittsburgh, PA), or VWR (West Chester, PA) unless otherwise noted. For targeted proteomics experiments, mass spectrometric-grade trypsin was obtained from Sigma-Aldrich and prepared according to manufacturer's instructions. DNA synthesis was ordered as gBlocks through IDT (Coralville, IA). DNA concentration was estimated using Nanodrop (Thermo, Waltham, MA).

Strain construction

E. coli DH10B (Invitrogen, Carlsbad, CA) and pir+ were used as the host for cloning while *E. coli* DH1 (ATCC33849) was used as the host for bisabolene production. All plasmids and strains used in this study are described in Tables 1 and 2. MVA pathway genes were taken from the previously reported systems (Martin et al., 2003) and include different modifications that improved isoprenoid production in plasmid-based systems (Alonso-Gutierrez et al., 2015; Dahl et al., 2013). In order to enable rapid cloning and assembly of genes, we employed the BglBrick (Anderson et al., 2010) and Gibson (Gibson et al., 2009) cloning strategies. All plasmids denoted pBb were prepared using the BglBrick standard collection of expression vectors (Lee et al., 2011). Cells containing the DE3 lysogen and T7RNAP were constructed by transfection with the DE3 lysogen using the λ_{DE3} Lysogenization Kit plus (Novagen, Darmstadt, Germany). The detail of strain construction is presented in Supplementary Method.

Growth conditions and terpene analysis

E. coli strains harboring bisabolene production pathways were grown in 8 mL of EZ-Rich defined medium (Teknova, Hollister, CA) supplemented with 1% glucose or 1% sucrose and appropriate antibiotics in glass culture tubes. Working concentrations were 30, 50 and 100 μg/mL for chloramphenicol, kanamycin and ampicillin, respectively. Starter cultures were grown overnight at 37°C, and all samples were shaken in rotary shakers at 200 rpm (Kühner, Basel, Switzerland). Production cultures were inoculated at an optical density ($OD_{600\ nm}$) of 0.1, and allowed to continue growth at 30°C, as the lower temperature improved terpene production. Each of the different strains was induced at OD_{600nm} ~0.4 with 500 μM of isopropyl β-d-1-thiogalactopyranoside (IPTG) concentrations. Samples for both terpene production and $OD_{600\ nm}$ production were taken at 24 and 48 hours while proteomic measurements were only performed at 24 h during late exponential phase. For production assays, a dodecane layer (10% by volume) was added to the culture upon induction to trap the terpene, and 10-100 μL of the organic phase was sampled and diluted 2- to 100-fold (depending on the final concentration in the culture) to be analyzed using gas-chromatography mass-spectrometry (GC/MS) as reported previously (Alonso-Gutierrez et al., 2013).

λ-Red mediated chromosomal integration of the MVA pathway

Since the pathway is much bigger (~15Kb) than the recommended DNA size for λ-Red mediated integration (~2Kb (Kuhlman and Cox, 2010)), we divided the pathway into three operons with the size of 4-5Kb and PCR-amplified them as cassettes surrounded by 500bp homology arms using KIKO vectors as templates to be independently integrated at three different locations in the DH1 genome (Sabri et al., 2013). The detail of chromosomal integration procedure is presented in Supplementary Method.

CRISPR-Cas9 selection protocol

We used CRISPR-Cas9 to individually substitute the promoter driving each of the three MVA pathway modules (i.e. T, M, and B) in the DSTMB7 (or DK4) with P_{T7}. The detail of CRISPR-Cas9 selection protocol is presented in Supplementary Method.

Targeted proteomics

Targeted proteomics was performed as previously described (Alonso-Gutierrez et al., 2015), and detailed method is presented in Supplementary Method.

Acknowledgments

We thank Dr. Sarah Richardson for helpful discussion on CRISPR-Cas9 approach and providing plasmids (pSMR00077, pMCC, and pMCK). 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. This work was also partially funded by the Queensland Government through its Smart Futures National and International Research Alliances Program and grants from Boeing Research & Technology–Australia, IOR Energy, Mackay Sugar Limited and Virgin Australia. CEV was supported by a Queensland Government Accelerate Fellowship.

Competing financial interests: JDK has financial interest in Amyris, Lygos, Demetrix, and Constructive Biology.

References

Atsumi, S, Hanai, T, and Liao, JC. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**: 86–89.

Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G. 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **330**:70–4.

Alonso-Gutierrez J, Chan R, Batth TS, Adams PD, Keasling JD, Petzold CJ, Lee TS. 2013. Metabolic engineering of *Escherichia coli* for limonene and perillyl alcohol production. *Metab. Eng.* **19**:33–41.

Alonso-Gutierrez J, Kim E, Batth TS, Cho N, Hu Q, Jade L, Chan G, Petzold CJ, Hillson NJ, Adams PD, Keasling JD, Garcia-Martin H, Lee TS. 2015. Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. *Metab. Eng.* **28**:123–133.

Amorós-Moya D, Bedhomme S, Hermann M, Bravo IG. 2010. Evolution in regulatory regions rapidly compensates the cost of nonoptimal codon usage. *Mol. Biol. Evol.* **27**:2141–51.

Archer CT, Kim JF, Jeong H, Park JH, Vickers CE, Lee SY, Nielsen LK. 2011. The genome sequence of *E. coli* W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. *BMC Genomics* **12**:9.

Block DHS, Hussein R, Liang LW, Lim HN. 2012. Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Res.* **40**:8979–92.

Bokinsky G, Peralta-Yahya PP, George A, Holmes BM, Steen EJ, Dietrich J, Lee TS, Tullman-Ercek D, Voigt CA, Simmons BA, Keasling JD. 2011. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **108**:19949–54.

Bruschi M, Boyes SJ, Sugiarto H, Nielsen LK, Vickers CE. 2011. A transferable sucrose utilization approach for non-sucrose-utilizing *Escherichia coli* strains. *Biotechnol. Adv.* **30**:1001–10.

Bryant J a, Sellars LE, Busby SJW, Lee DJ. 2014. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res.* **42**:11383–11392.

Cameron DE, Bashor CJ, Collins JJ. 2014. A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12**:381–90.

Chandran SS, Kealey JT, Reeves CD. 2011. Microbial production of isoprenoids. *Process Biochem.* **46**:1703–1710.

Cherepanov PP, Wackernagel W. 1995. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**:9–14.

Dahl RH, Zhang F, Alonso-Gutierrez J, Baidoo E, Batth TS, Redding-Johanson AM, Petzold CJ, Mukhopadhyay A, Lee TS, Adams PD, Keasling JD. 2013. Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.*:1–10.

Datsenko K a, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**:6640–5.

Davis R, Tao L, Tan ECD, Biddy MJ, Beckham GT, Scarlata C, Ross J, Lukas J, Knorr D, Schoen P. 2013. Process design and economics for the conversion of lignocellulosic biomass to hydrocarbons: Dilute-acid and enzymatic deconstruction of biomass to sugars and biological conversion of sugars to hydrocarbons process design and economics for the conversion. Technical Report, NREL/TP-5100-60223.

Dellomonaco C, Clomburg JM, Miller EN, Gonzalez R. 2011. Engineered reversal of the β-

oxidation cycle for the synthesis of fuels and chemicals. *Nature* **476**:355–9.

Edwards AN, Patterson-Fortin LM, Vakulskas CA, Mercante JW, Potrykus K, Vinella D, Camacho MI, Fields JA, Thompson SA, Georgellis D, Cashel M, Babitzke P, Romeo T. 2011. Circuitry linking the Csr and stringent response global regulatory systems. *Mol. Microbiol.* **80**:1561–1580.

Farmer WR, Liao JC. 2000. Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nat. Biotechnol.* **18**:533–537.

Friehs K. 2004. Plasmid Copy Number and Plasmid Stability:47–82.

Garcia JL. 1985. Cloning in Escherichia coli and molecular analysis of the sucrose system of the Salmonella plasmid SCR-53. *Mol. Gen. Genet*. **201**:575–577.

George KW, Alonso-Gutierrez J, Keasling JD, Lee TS. 2015. Isoprenoid Drugs, Biofuels, and Chemicals-Artemisinin, Farnesene, and Beyond. *Adv. Biochem. Eng. Biotechnol.* **148**: 355-89.

George KW, Chen A, Jain A, Batth TS, Baidoo EEK, Wang G, Adams PD, Petzold CJ, Keasling JD, Lee TS. 2014. Correlation analysis of targeted proteins and metabolites to assess and engineer microbial isopentenol production. *Biotechnol. Bioeng.* **111**: 1648-58.

Gibson DG, Young L, Chuang R, Venter JC, Iii CAH, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases *Nat. Methods* **6**:12–16.

Goh E-B, Baidoo EEK, Burd H, Lee TS, Keasling JD, Beller HR. 2014. Substantial improvements in methyl ketone production in *E. coli* and insights on the pathway from in vitro studies. *Metab. Eng.* **26C**:67–76.

Haldimann A, Wanner BL. 2001. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol*

183:6384–6393.

Hanahan D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**:557–80.

Hiltunen T, Virta M, Laine A-L. 2017. Antibiotic resistance in the wild: an eco-evolutionary perspective. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **372**:20160039.

Holtz WJ, Keasling JD. 2010. Engineering static and dynamic control of synthetic pathways. *Cell* **140**:19–23.

Humbird D, Davis R, Tao L, Kinchin C, Hsu D, Aden A, Schoen P, Lukas J, Olthof B, Worley M, Sexton D, Dudgeon D. 2011. Process design and economics for biochemical conversion of lignocellulosic biomass to ethanol process design and economics for biochemical conversion of lignocellulosic biomass to ethanol. *Rep. No. NREL/TP-5100-47764. Golden, CO Natl. Renew. Energy Lab. May 2011.*

Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM. 2011. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**:348–353.

Jeong J, Cho N, Jung D, Bang D. 2013. Genome-scale genetic engineering in *Escherichia coli*. *Biotechnol. Adv.* **31**:804–810.

Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**:233–9.

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

Junker B, Lester M, Leporati J, Schmitt J, Kovatch M, Borysewicz S, Maciejak W, Seeley A,

Hesse M, Connors N, Brix T, Creveling E, Salmon P. 2006. Sustainable reduction of bioreactor contamination in an industrial fermentation pilot plant. *J. Biosci. Bioeng.* **102**:251–268.

Keasling JD. 2010. Manufacturing molecules through metabolic engineering. *Science* **330**:1355–1358.

Keasling JD, Mendoza A, Baran PS. 2012. Synthesis: A constructive debate. *Nature* **492**:188–9.

Klein-Marcuschamer D, Turner C, Allen M, Gray P, Dietzgen RG, Gresshoff PM, Hankamer B, Heimann K, Scott PT, Stephens E, Speight R, Nielsen LK. 2013. Technoeconomic analysis of renewable aviation fuel from microalgae, *Pongamia pinnata*, and sugarcane, *Biofuels, Bioprod. Bioref.* **7**: 416-428.

Koma D, Yamanaka H, Moriyoshi K, Ohmoto T, Sakai K. 2012. A convenient method for multiple insertions of desired genes into target loci on the *Escherichia coli* chromosome. *Appl. Microbiol. Biotechnol.* **93**:815–829.

Kuhlman TE, Cox EC. 2010. Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res.* **38**:e92.

Lee JW, Choi S, Park JW, Vickers CE, Nielsen LK, Lee SY. 2010. Development of sucrose-utilizing *Escherichia coli* K-12 strain by cloning β-fructofuranosidases and its application for l-threonine production. *Appl. Microbiol. Biotechnol.* **88**:905-913.

Li Y, Lin Z, Huang C, Zhang Y, Wang Z, Tang Y jie, Chen T, Zhao X. 2015. Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 meditated genome editing. *Metab. Eng.* **31**:13–21.

Łos M, Czyz A, Sell E, Wegrzyn A, Neubauer P, Wegrzyn G. 2004. Bacteriophage

contamination: is there a simple method to reduce its deleterious effects in laboratory cultures and biotechnological factories? *J. Appl. Genet.* **45**:111–20.

Łoś M, Golec P, Łoś JM, Węglewska-Jurkiewicz A, Czyż A, Węgrzyn A, Węgrzyn G, Neubauer P. 2007. Effective inhibition of lytic development of bacteriophages lambda, P1 and T4 by starvation of their host, *Escherichia coli*. *BMC Biotechnol.* **7**:13.

Martin V, Pitera D, Withers S, Newman J, Keasling J. 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **21**:796–802.

Meadows AL, Hawkins KM, Tsegaye Y, Antipov E, Kim Y, Raetz L, Dahl RH, Tai A, Mahatdejkul-Meadows T, Xu L, Zhao L, Dasika MS, Murarka A, Lenihan J, Eng D, Leng JS, Liu C-L, Wenger JW, Jiang H, Chao L, Westfall P, Lai J, Ganesan S, Jackson P, Mans R, Platt D, Reeves CD, Saija PR, Wichmann G, Holmes VF, Benjamin K, Hill PW, Gardner TS, Tsong AE. 2016. Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature* **537**:694–7.

Mosberg J a, Lajoie MJ, Church GM. 2010. Lambda red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate. *Genetics* **186**:791–9.

Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, Polichuk DR, Teoh KH, Reed DW, Treynor T, Lenihan J, Fleck M, Bajad S, Dang G, Dengrove D, Diola D, Dorin G, Ellens KW, Fickes S, Galazzo J, Gaucher SP, Geistlinger T, Henry R, Hepp M, Horning T, Iqbal T, Jiang H, Kizer L, Lieu B, Melis D, Moss N, Regentin R, Secrest S, Tsuruta H, Vazquez R, Westblade LF, Xu L, Yu M, Zhang Y, Zhao L, Lievense J, Covello PS, Keasling JD, Reiling KK, Renninger NS, Newman JD. 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**:528–32.

Peralta-Yahya PP, Ouellet M, Chan R, Mukhopadhyay A, Keasling JD, Lee TS. 2011.

Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* **2**:483.

Peters D. 2006. Carbohydrates for fermentation. *Biotechnol. J.* **1**:806–814.

Pitera DJ, Paddon CJ, Newman JD, Keasling JD. 2007. Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab. Eng.* **9**:193–207.

Redding-Johanson AM, Batth TS, Chan R, Krupa R, Szmidt HL, Adams PD, Keasling JD, Lee TS, Mukhopadhyay A, Petzold CJ. 2011. Targeted proteomics for metabolic pathway optimization: Application to terpene production. *Metab. Eng.* **13**:194–203.

Reyes-Lamothe R, Nicolas E, Sherratt DJ. 2012. Chromosome replication and segregation in bacteria. *Annu. Rev. Genet.* **46**:121–43.

Robinson TP, Bu DP, Carrique-Mas J, Fèvre EM, Gilbert M, Grace D, Hay SI, Jiwakanon J, Kakkar M, Kariuki S, Laxminarayan R, Lubroth J, Magnusson U, Thi Ngoc P, Van Boeckel TP, Woolhouse MEJ. 2016. Antibiotic resistance is the quintessential One Health issue. *Trans. R. Soc. Trop. Med. Hyg*. **110**:377–380.

Ronda C, Pedersen LE, Sommer MOA, Nielsen AT. 2016. CRMAGE: CRISPR Optimized MAGE Recombineering. *Sci. Rep.* **6**:19452.

Sabri S, Steen J a, Bongers M, Nielsen LK, Vickers CE. 2013. Knock-in/Knock-out (KIKO) vectors for rapid integration of large DNA sequences, including whole metabolic pathways, onto the *Escherichia coli* chromosome at well-characterised loci. *Microb. Cell Fact.* **12**:60.

Salis HM, Mirsky EA, Voigt CA. 2009. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**:946–50.

Sawitzke JA, Thomason LC, Costantino N, Bubunenko M, Datta S, Court DL. 2007. Recombineering: In Vivo Genetic Engineering in *E. coli*, *S. enterica*, and Beyond. *Methods Enzymol.* **421**:171–199.

Schirmer A, Rude M a, Li X, Popova E, del Cardayre SB. 2010. Microbial biosynthesis of alkanes. *Science* **329**:559–62.

Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, Del Cardayre SB, Keasling JD. 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463**:559–562.

Studier FW, Moffattf BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**:113–130.

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. 1983. The double-strand-break repair model for recombination. *Cell* **33**:25–35.

Temme K, Hill R, Segall-Shapiro TH, Moser F, Voigt CA. 2012a. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res.* **40**:8773–81.

Temme K, Zhao D, Voigt CA. 2012b. Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc. Natl. Acad. Sci. U. S. A.* **109**:7085–90.

Tsao C-Y, Hooshangi S, Wu H-C, Valdes JJ, Bentley WE. 2010. Autonomous induction of recombinant proteins by minimally rewiring native quorum sensing regulon of *E. coli*. *Metab. Eng.* **12**:291–7.

Tsuruta H, Paddon CJ, Eng D, Lenihan JR, Horning T, Anthony LC, Regentin R, Keasling JD, Renninger NS, Newman JD. 2009. High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. *PLoS One* **4**:e4489.

Tyo KEJ, Ajikumar PK, Stephanopoulos G. 2009. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat. Biotechnol.* **27**:760–5.

Wang HH, Isaacs FJ, Carr P a, Sun ZZ, Xu G, Forest CR, Church GM. 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**:894–8.

Wang HH, Kim H, Cong L, Jeong J, Bang D, Church GM. 2012. Genome-scale promoter engineering by coselection MAGE. *Nat. Methods* **9**:591–593.

Westfall PJ, Gardner TS. 2011. Industrial fermentation of renewable diesel fuels. *Curr. Opin. Biotechnol.* **22**:344–350.

Wolfe AJ. 2005. The Acetate Switch. *Microbiol. Mol. Biol. Rev.* **69**:12–50.

Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, Estadilla J, Teisan S, Schreyer HB, Andrae S, Yang TH, Lee SY, Burk MJ, Van Dien S. 2011. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **7**:445–52.

Tables caption

Table 1. Description of *E. coli* base strains, plasmids and gene operons used in this study.

Table 2. Nomenclature of engineered DH1 strains containing editions of the integrated MVA pathway. fixB(g44) is a fixing template to substitute P_{trc} by P_{T7} in front of AgBIS at *rbsR* location. fixT(g47) is a fixing template to substitute P_{gadE} by P_{T7} in front of MevT at *poxB* location. fixM(g48) is a fixing template that substitute P_{trc} by P_{T7} promoter in front of MBI at *lacZ* location.

Table 3. Relative protein levels (chromatogram peak area) from the chromosomally integrated MVA pathway in strain DK4 and its different levels of optimization (Figure 6) after 24 hours of growth. Green and red colored cells represent protein peak areas that had been up or down-regulated, respectively, compared to the original DK4 strain. Results are the average of three biological replicates (error bars are within 1-15% of the value and not shown for clarity).

Figures captions

Figure 1. Cloning methodology of pathways of interest from BglBrick vectors into KIKO vectors for chromosomal integration. POI: Promoter of interest; GOI: Gene(s) of interest; MCS: multiple cloning site; HL: Hairpin loop; HA: Homology arm; FRT: Flippase recombinase target sites; KanR, CmR and AmpR: kanamycin, chloramphenicol and ampicillin resistance gene.

Figure 2. Production of bisabolene and growth curves from different strains grown for 96 hours under the same conditions using either 1% of glucose (solid lines and bars) or 1% of sucrose (dashed lines and bars) as the carbon source. a) Bisabolene titers and b) OD_{600nm} was measured at 24, 48 and 96h (left to right stacked bars) and compared between *E. coli* DH1 (wild type; dotted bars) versus DS strain (DH1wt with *cscAKB* operon integrated in the chromosome; white bars) either grown in glucose (solid lines) or sucrose (dashed lines). Both strains were transformed with plasmids pBbA5c-MevT(co)-T61-MBIS-T21-P_{trc}-AgBIS (JPUB_002475) and pTrc-AgBIS (JPUB_002466). c) Bisabolene titers and d) OD_{600nm} was measured at 24, 48 and 96h (left to right stacked bars) and compared between *E. coli* DS strains transformed with plasmid pBbA0c-P_{gadE}-MevT-MBIS (JBx_000909) and either pTrc-AgBIS (JPUB_002466; grey lines) or pP_{rstA}-AgBIS (JBx_001052; black lines). Strains were grown either in glucose (solid lines) or sucrose (dashed lines). Results are the average of three or more biological replicates.

Figure 3. Integration strategy of the mevalonate pathway (MVA) (for production of bisabolene), the *cscAKB* operon (for sucrose ultilization) and the DE3 prophage (for T7RNAP expression) into the *E. coli* chromosome; 1st) Module M was integrated at the *lacZ* locus; 2nd) The sucrose utilization operon was integrated at the *arsB* locus; 3rd) Module B was integrated at the *rbsAR* locus close to the origin of chromosomal replication (*oriC*) and therefore with likely higher

transcription activity. 4th) The top portion of the MVA pathway (module T) was integrated at *poxB*. Finally, the T7 RNA polymerase (T7RNAP) gene was transduced into the chromosome as part of the DE3 lysogen. Since *E. coli* is not able to produce MVA, the integration of module T as the last step avoids potential toxicity problems of intermediates that would otherwise accumulate in the absence of a complete MVA pathway. All integrations (except for the T7RNAP that was mediated by phage transduction) were done by homologous recombination between ~500 bp homology arms ("_up" and "_dw") surrounding each cassette and the chromosome. Recombinants were screened using different antibiotic markers that were subsequently removed using flippase recognition target (FRT) sites surrounding each marker to loop it out and enable reuse of the marker in subsequent engineering steps. The number of base pair next to each target loci, indicates its chromosomal location relative to the threonine (*thr*) locus.

Figure 4. Chromosomally integrated pathway optimization via CRISPR-Cas9. a) A two-plasmid system was used. pSMR77 codes for λ-Red expression under arabinose-inducible P_{BAD} promoter and the Cas9 enzyme under a native constitutive promoter from *Streptococcus pyogenes*. Plasmid pMCC or pMCK (with either chloramphenicol or kanamycin antibiotic marker, respectively) expresses crRNA (CRISPR RNA) under the native constitutive promoter from *S. pyogenes*. b) This crRNA guides the Cas9 protein to two regions of chromosomal DNA (# and #'). New targets can be changed by a two-piece Gibson assembly between a linearized pMCC/K plasmid using *Sma*I and a gBlock fragment providing the crRNA and the promoter sequence. Co-selection of pSMR77 and either pMCC or pMCK leads to cell death unless a "fixing template" (double stranded DNA with homology to chromosomal DNA) is provided to replace the target sequences in the chromosome (protospacers)). c) Two protospacer target sites

surrounding each of the promoters controlling expression of module T (CrT/T'), module M (CrM/M') and module B (CrB/B') were selected to screen against cells which did not recombine the P_{T7} (fixing templates fixT, fixM and fixB respectively). AcCoa; acetyl-CoA; MVA: mevalonate; IPP: isopentenyl pyrophosphate; BIS: bisabolene.

Figure 5. Strain nomenclature, characteristics and genealogy. Starting from a wild type strain of *E. coli* DH1 (DH1wt) a new strain called DSTMB7 was generated. DSTMB7 (also known as DK4) contains into the chromosome: i) the *cscAKB* operon to grow on sucrose (S, open triangle), ii) the MVA pathway operons (i.e. MevB (M), AgBIS-idi-ispA (B) and MevT (T)) to produce bisabolene from acetyl-CoA, and iii) the DE3 prophage to express T7RNAP. The different parts were sequentially integrated into its chromosome using different strategies: 1st) both the MVA pathway and the sucrose operons were first cloned from BglBrick parts into KIKO vectors (Figure 1), PCR amplified and chromosomally integrated using classic λ-Red recombination and antibiotic marker selection (left box). 2nd) the λ(DE3) prophage was integrated via transduction providing expression of T7RNAP under the control of P_{lac} (middle box). 3rd) Finally, MVA pathway optimization was achieved directly from the chromosome of DSTMB7 by replacing original promoters (small bended arrows) for P_{T7} (big bended arrows) using CRIPSR-Cas9 selection (right box).

Figure 6. Bisabolene production from chromosomally integrated MVA pathway strains with different level of optimization after 48 hours of growth. Red dashed line indicates the basal production level of strain DK4 before any promoter replacement via CRISPR was performed.

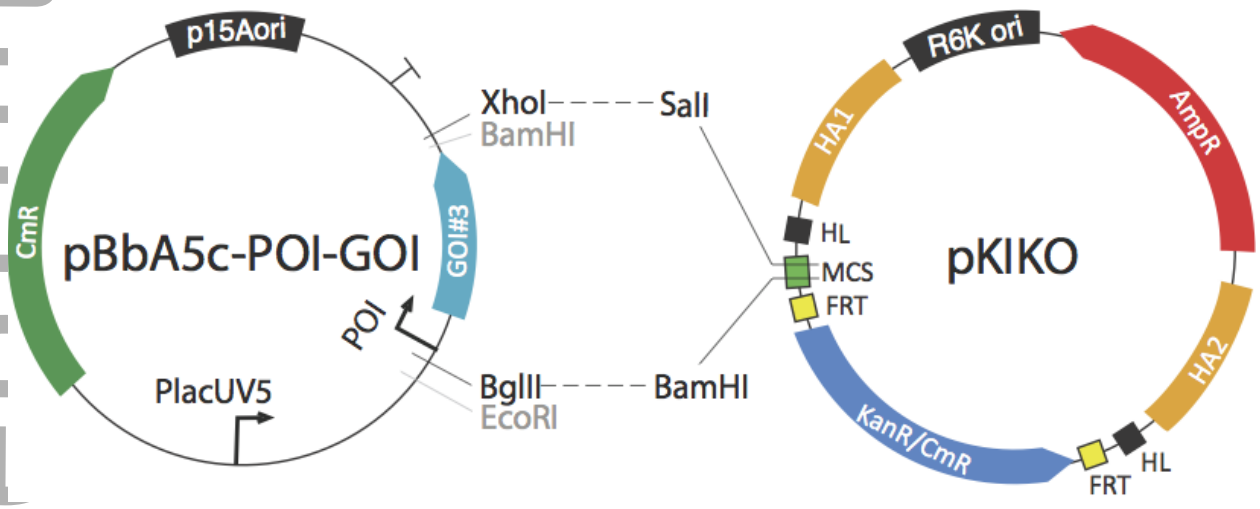

Figure 1

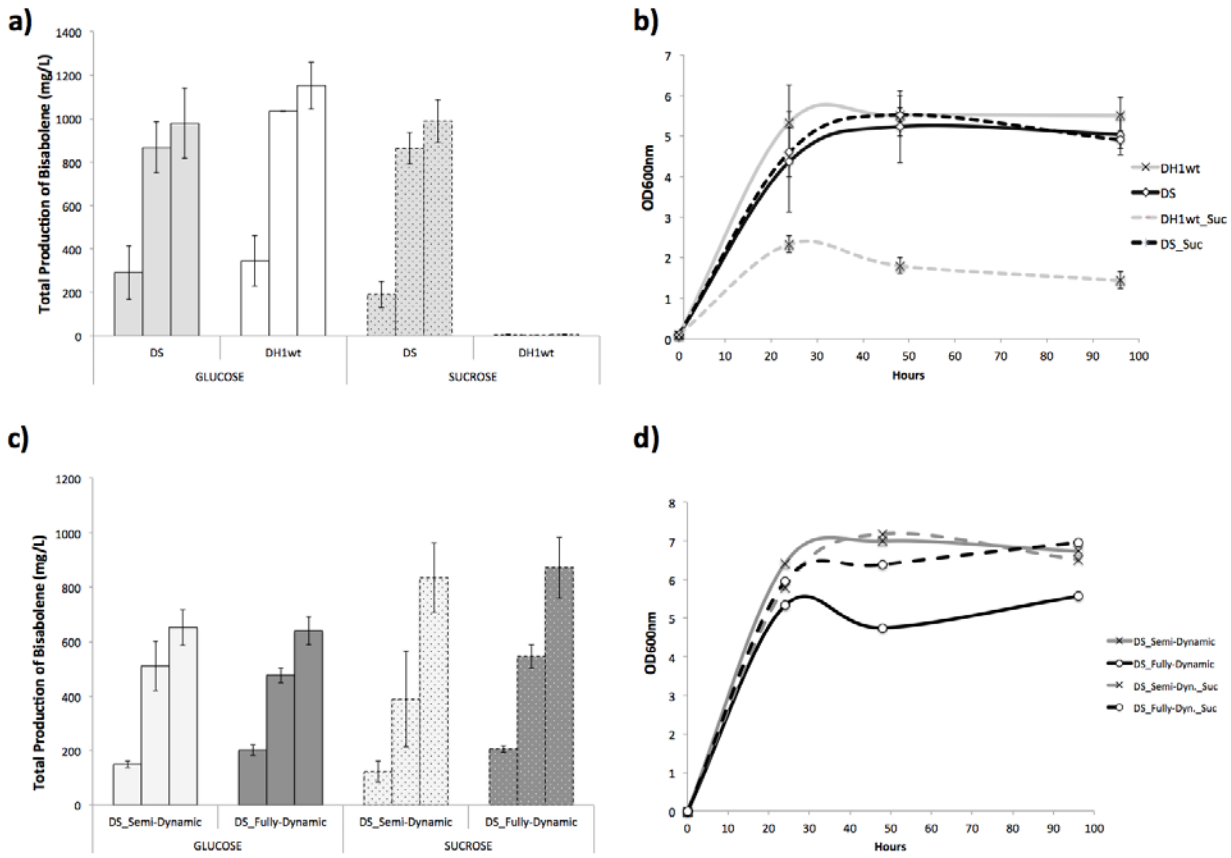

Figure 2

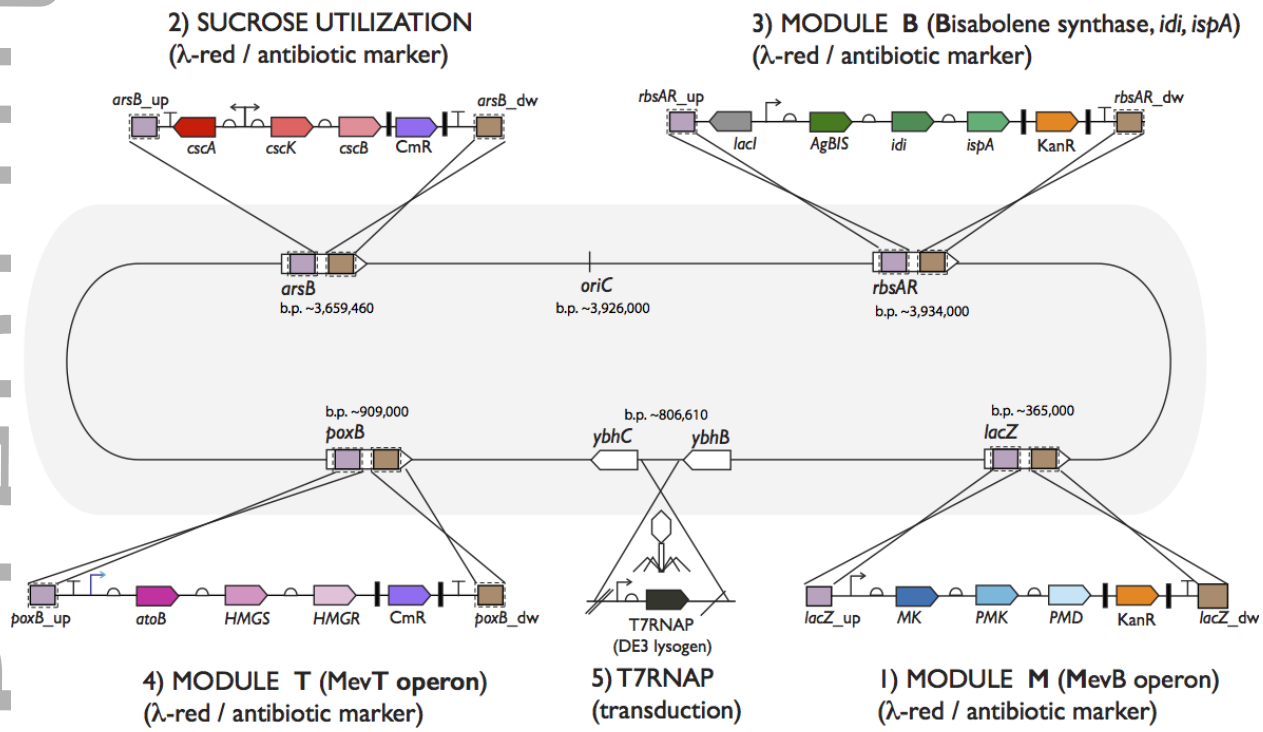

Figure 3

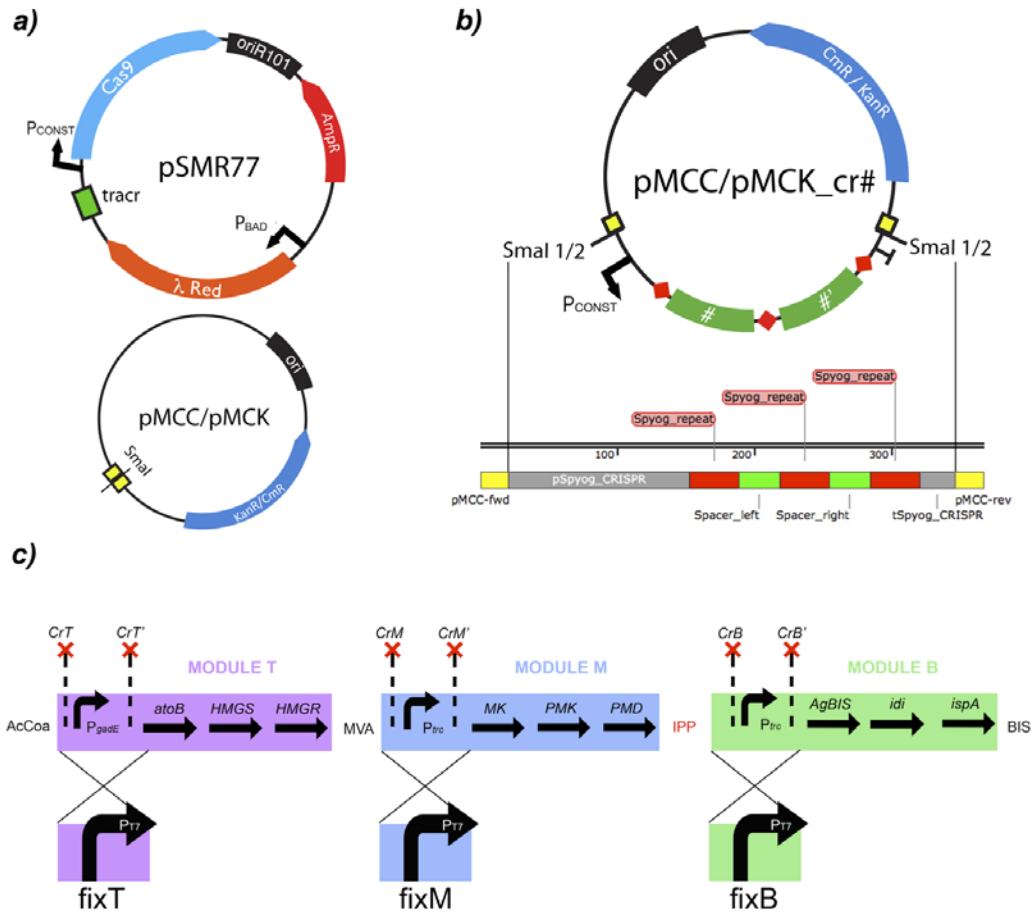

Figure 4

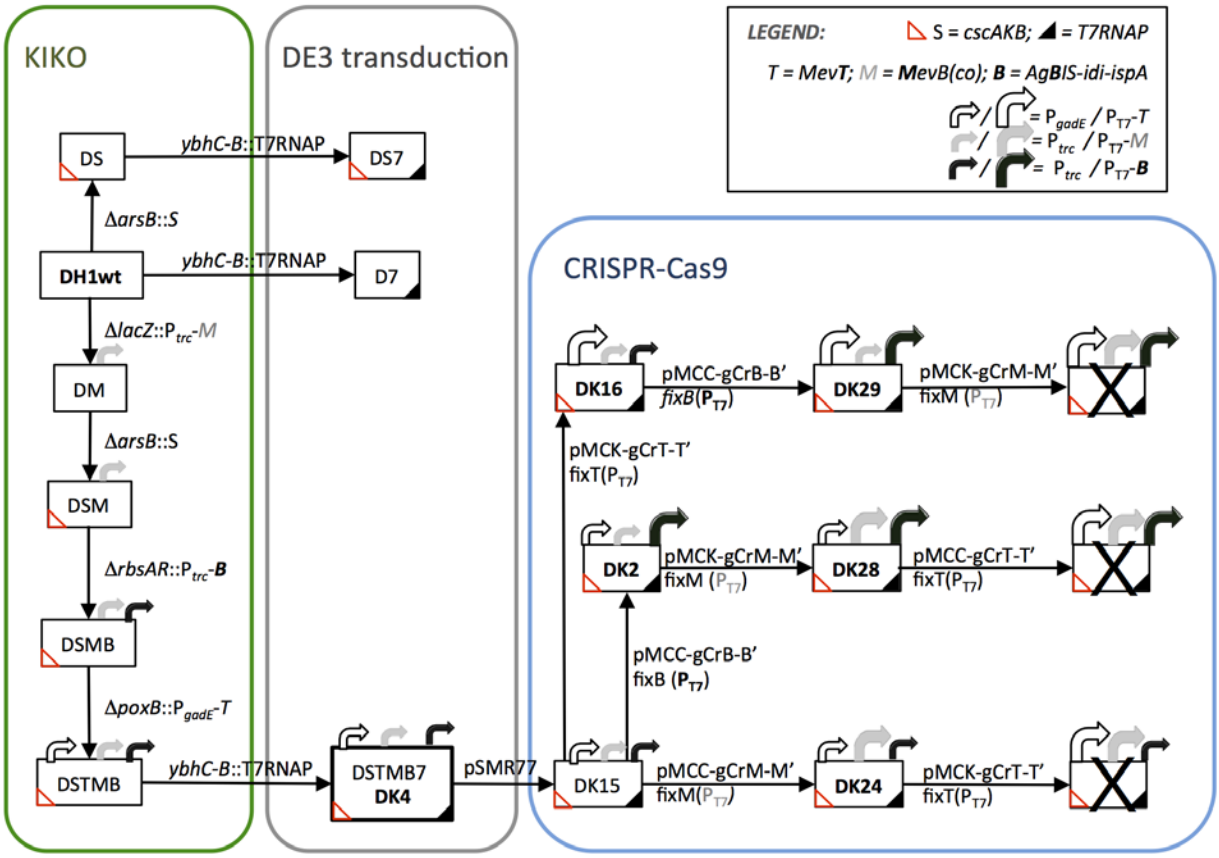

Figure 5

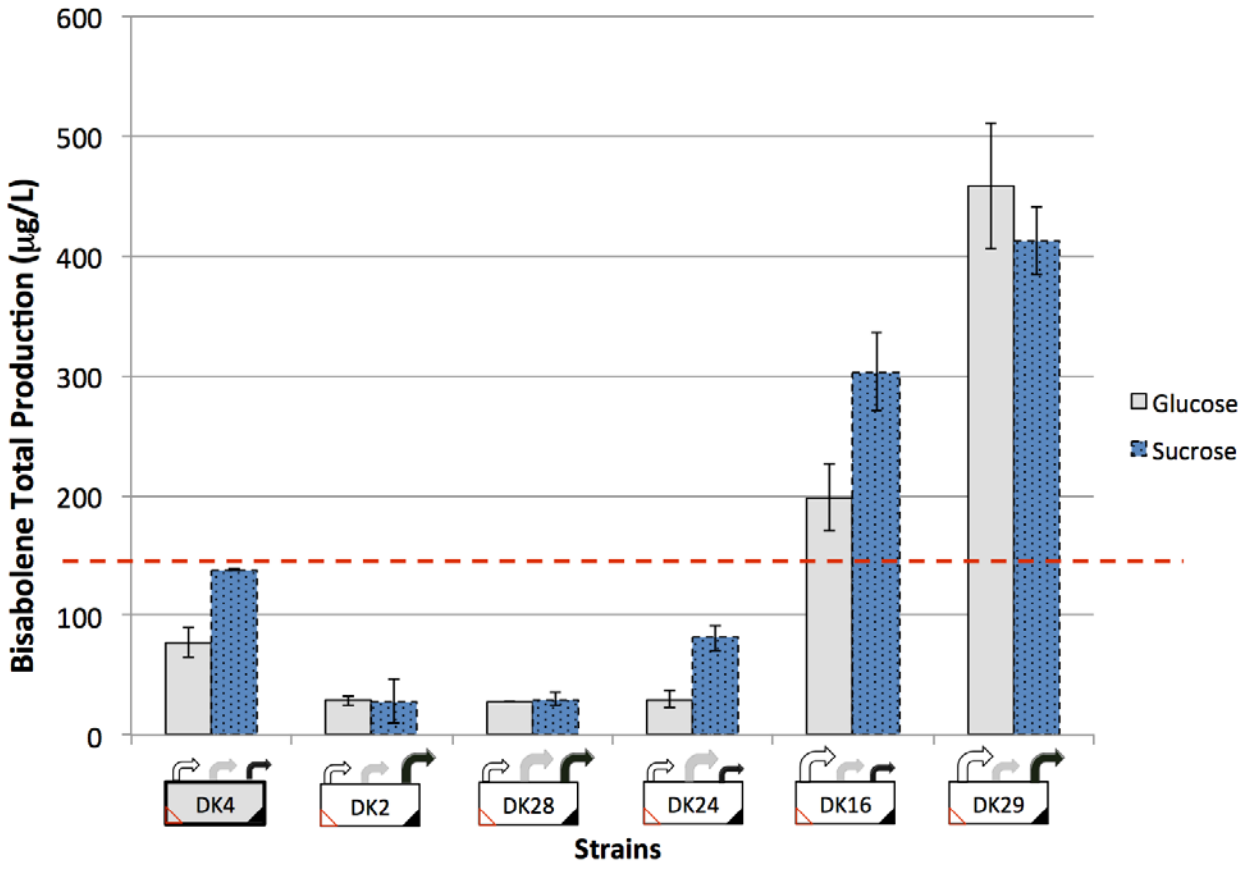

Figure 6

Table 1. Description of *E. coli* base strains, plasmids and gene operons used in this study.

E. coli strain	Description	Reference(s)
DH1	*E. coli* K-12; F- *endA1 recA1 gyrA96 thi-1 glnV44 relA1 hsdR17(rK- mK+)* λ-	(Hanahan, 1983)
DH10B	F- *endA1 recA1 galE15 galK16 nupG rpsL* Δ*lacX74* Φ80*lacZΔM15 araD139* Δ(*ara,leu*)*7697 mcrA* Δ(*mrr-hsdRMS-mcrBC*) λ-	(Casdaban & Cohen, 1980)
S-17 λpir+	TpR SmR *recA thi pro hsdR-M*+RP4: 2-Tc:Mu: Km Tn7 λpir	(Edwards et al., 2011)
PIR1	F- Δ*lac169 rpoS(Am) robA1 creC510 hsdR514 endA recA1 uidA(ΔMluI)::pir-116*	Invitrogen

Plasmid reference	Plasmid name	Description (Origin of replication, Antibiotic marker, Promotors and Operons)	Reference(s)
JPUB_002475	pBbA5c-MevT(co)[a]-T61[b] -MBIS[c]-T21[d]-Ptrc-AgBIS[e]	p15A, Cm^r, P_{lacUV5}, MevT(co), T1, MBIS, T1002, P_{trc}, AgBS	(Alonso-Gutierrez et al., 2015)
JBx_000909	pBbA0c-P_{gadE}-MevT-MBIS[f]	p15A, Cm^r, P_{gadE}, MevT, MBIS	(Dahl et al., 2013)
JBx_013443	pBbA0c-P_{gadE}-MevT	p15A, Cm^r, P_{gadE}, MevT	(Dahl et al., 2013)
JPUB_002466	pTrc-AgBIS	colE1, Amp^r, P_{trc}, AgBIS	(Alonso-Gutierrez et al., 2015)
JBx_014141	pBbE1a-T1002-Trc	colE1, Amp^r, P_{trc}, Terminator T1002	(This work)
JBx_001334	pBbB1k-MevB(CO)	pBBR1, Kan^r, MevB[i]	(This work)
JBx_001052	pRstA-AgBIS	colE1, Amp^r, P_{trc}, AgBIS	(This work)
JBx_074937	pKIKO-Kan-lacZ-Ptrc-MevB	R6K, Kan^r, *lacZ* gene homology, P_{trc}, MevB	(This work)
JBx_026115	pKIKO-Cm-arsB-cscAKB	R6K, Cm^r, *arsB* gene homology, P_{const}, *cscAKB* for sucrose uptake	(This work)
JBx_074936	pKIKO-Kan-rbsAR-LacI-P_{trc}-BIS-idi-ispA	R6K, Kan^r, *rbsAR* gene homology, P_{trc}, AgBIS, *idi*, *ispA*	(This work)
JBx_088101	pKIKO-Cm-poxB-P_{gadE}-MevT	R6K, Cm^r, *poxB* gene homology, P_{gadE}, MevT	(This work)
JBEI-2706	pKD46	OriR101/*repA101ts*, Amp^r, P_{bad}, λ Red (γ, β, exo)	(Datsenko and Wanner, 2000)
JBx_026997	pCP20	OriR101/*repA101ts*, Amp^r, FLP recombinase	(Cherepanov and Wackernagel, 1995)
JBEI-9382	pKD46-Cas9 (pSMR00077)	OriR101/*repA101ts*, Amp^r, P_{bad}, λ Red (γ, β, exo), P_{Spyog}, Cas9, tracr[j]	(This work)
JBEI-9390	pMCC	p15A, CasE1, Cm^r,	(This work)

JBEI-10488	pMCK	p15A, CasE1, Kan^r,	(This work)
JBx_063298	pMCC-gCrB-B'[k]	p15A, CasE1, Cm^r, P_{Spyog}-R^k-SpacerB-R^k-SpacerB'-R^l	(This work)
JBx_063300	pMCK-gCrT-T'[k]	p15A, CasE1, Kan^r, P_{Spyog}-R^k-SpacerT-R^k-SpacerT'-R^l	(This work)
JBx_063300	pMCC-gCrT-T'[k]	p15A, CasE1, Cm^r, P_{Spyog}-R^k-SpacerT-R^k-SpacerT'-R^l	(This work)
JBx_063302	pMCK-gCrM-M'[k]	p15A, CasE1, Kan^r, P_{Spyog}-R^k-SpacerM-R^k-SpacerM'-R^l	(This work)
JBx_074941	pMCC-gCrM-M'[k]	p15A, CasE1, Cm^r, P_{Spyog}-R^k-SpacerM-R^k-SpacerM'-R^l	(This work)

[a] MevT(co): contains genes for the conversion of acetyl CoA to mevalonate: acetoacetyl-CoA synthase from *E. coli* (AtoB), HMG-CoA synthase from *Saccharomyces cerevisiae* (HMGS), and HMG-CoA reductase from *Saccharomyces cervisiae* (HMGR) (Tsuruta et al., 2009) codon optimized for *E. coli*

[b] T61: T1006 terminator followed by P_{trc}

[c] MBIS: contains genes for the conversion of mevalonate to isoprenoid precursors (IPP and DMAPP): mevalonate kinase (MK), phosphomevalonate kinase (PMK), mevalonate diphosphate decarboxylase from *S. cerevisiae* (PMD) and isopentenyl diphosphate isomerase (*idi*) and farnesyl diphosphate synthase (*ispA*) from *E. coli* (idi) (Martin et al., 2003). These genes have been fixed for expression in *E. coli* as previously described (Redding-Johanson et al., 2011)

[d] T21: T1002 terminator followed by P_{trc}

[e] AgBIS: Bisabolene synthase coding sequence from *Abies grandis* codon-optimized for *E. coli*

[f] Plasmid that expresses all genes necessary to transform AcCoA into IPP/DMAPP, which contains *atoB*, *idi* and *ispA* from *E. coli* and HMGS, HMGR, MK, PMK and PMD from *S. cerevisiae* in its original form (i.e non codon-optimized), under the control of a stress-responsive dynamic promoter (P_{gadE}) as previously described (Dahl et al., 2013)

[g] MTSA operon contains genes for the conversion of acetyl CoA to mevalonate: acetoacetyl-CoA synthase from *E. coli* (*atoB*), HMG-CoA synthase (HMGS) and HMG-CoA reductase (HMGR) from *Staphylococcus aureus* (Baliobar et al., 2009) codon optimized for *E. coli*.

[h] T1: double terminator followed by P_{trc} promoter

[i] MevB: MevB operon contains genes for the conversion of mevalonate to isoprenoid precursor (IPP): MK, PMK, PMD from *S. cerevisiae* (Martin et al., 2003). These genes have been codon optimized for expression in *E. coli* as previously described (Redding-Johanson et al., 2011)

[j] trans-activating crRNA (TRACR)

[k] gCr stands for 'gBlock transcribing crRNA' and B-B', M-M' and T-T' stand for pairs of spacers surrounding the promoter driving modules B, M or T, respectively

[1]Clustered interspaced Repeat from *Streptococcus pyogenes* (Spyog)

Table 2. Nomenclature of engineered DH1 strains containing editions of the integrated MVA pathway. fixB(g44) is a fixing template to substitute P_{trc} by P_{T7} in front of AgBIS at *rbsR* location. fixT(g47) is a fixing template to substitute P_{gadE} by P_{T7} in front of MevT at *poxB* location. fixM(g48) is a fixing template that substitute P_{trc} by P_{T7} promoter in front of MBI at *lacZ* location.

Acronym	*Description*	*Synonym*	*Reference*
DH1	*E. coli* K-12; F-, λ-, *endA1, recA1, gyrA96, thi-1, glnV44, relA1, hsdR17(rK- mK+)*	D	Hanahan, 1983
DS	DH1 *arsB*::*escAKB*	DS	This Study
DS7	DS λ(DE3)	DS7	This Study
DM	DH1 *lacZ*::P_{trc}-MevB	DM	This Study
DSM	DS *lacZ*::P_{trc}-MevB	DSM	This Study
DSMB	DSM *rbsR*::P_{trc}-BIS-*idi-ispA*	DSMB	This Study
DSTMB	DSMB *poxB*::P_{gadE}-MevT	DK1	This Study
DSTMB7	DSMB *poxB*::P_{gadE}-MevT, λ(DE3)	DK4	This Study
DSTMB7_fixB	DSMBT7; *rbsR*::P_{T7}-BIS-*idi*-i*spA*	DK2	This Study
DSTMB7_fixT	DSMBT7; *poxB*::P_{T7}-MevT	DK16	This Study
DSTMB7_fixM	DSMBT7; *lacZ*::P_{T7}-MevB	DK24	This Study
DSTMB7_fixB_fixT	DSMBT7; *rbsR*::P_{T7}-BS-*idi-ispA, poxB*::PT7-MevT	DK29	This Study
DSTMB7_fixB_fixM	DSMBT7; *rbsR*::P_{T7}-BS-*idi-ispA, lacZ*::PT7-MevB	DK28	This Study
DSTMB7_fixT_fixM	DSMBT7; *poxB*::P_{T7}-MevT, *lacZ*::P_{T7}-MevB	No colonies	This Study
DSTMB7_fixB_fixM_fixT	DSMBT7; *poxB*::P_{T7}-MevT, *lacZ*::P_{T7}-MevB, *rbsR*::P_{T7}-BS-*ispA*	No colonies	This Study

Table 3. Relative protein levels (chromatogram peak area) from the chromosomally integrated MVA pathway in strain DK4 and its different levels of optimization (Figure 6) after 24 hours of growth. Green and red colored cells represent protein peak areas that had been up or down-regulated, respectively, compared to the original DK4 strain. Results are the average of three biological replicates (error bars are within 1-15% of the value and not shown for clarity).

		Module 1			Module 2			Module 3		
Carbon source	**Strain**	AtoB	HMGS	HMGR	MK	PMK	PMD	IDI	ISPA	BIS
SUCROSE	*Plasmid Control*	*1877*	*18423*	*3784*	*34693*	*20883*	*74977*	*195998*	*69861*	*84941*
	DK4	**3309**	**1714**	**383**	**3345**	**3064**	**12667**	**763**	**11945**	**1409**
	DK2	1789	1205	209	3216	2855	13038	788	11111	356
	DK28	1264	724	188	1299	3932	54611	619	13023	444
	DK24	884	529	302	531	1273	18411	734	8993	1004
	DK16	4922	4447	40	1525	890	5518	499	6580	636
	DK29	12701	11639	382	2953	2389	10626	912	11446	2503
GLUCOSE	*Plasmid Control*	*1229*	*13363*	*3766*	*22691*	*38889*	*149373*	*215637*	*64590*	*138784*
	DK4	**1027**	**698**	**375**	**3177**	**3445**	**11064**	**998**	**15079**	**1449**
	DK2	1120	769	416	3996	4416	14183	1111	17492	296
	DK28	1441	984	156	1531	5335	55231	1007	18000	1208
	DK24	502	413	408	555	2276	18571	2226	10981	1979
	DK16	3983	3655	208	1171	789	4526	219	6859	539
	DK29	6288	6841	349	1872	746	5529	547	8696	2526