REVISED VERSION OF AEM06785-11

Submitted for the Biotechnology section of *Applied and Environmental Microbiology*

Engineering of bacterial methyl ketone synthesis for biofuels

Ee-Been Goh† [1,3], Edward E. K. Baidoo[1,3], Jay D. Keasling [1,3,4], Harry R. Beller*† [1,2]

† H.R.B. and E.B.G. contributed equally to this study

[1]Joint BioEnergy Institute, 5885 Hollis Avenue, Emeryville, CA 94608

[2]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

[3]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

[4]Departments of Chemical & Biomolecular Engineering and of Bioengineering, University of California, Berkeley, CA 94720

Running title: Engineering bacterial methyl ketone synthesis

***Corresponding author:**

Harry R. Beller
Joint BioEnergy Institute
5885 Hollis Avenue
Emeryville, CA 94608

E-mail HRBeller@lbl.gov
Phone (510) 486-7321
Fax (510) 486-5686

ABSTRACT

We have engineered *Escherichia coli* to overproduce saturated and monounsaturated aliphatic methyl ketones in the C_{11} to C_{15} (diesel) range; this group of methyl ketones includes 2-undecanone and 2-tridecanone, which are of importance to the flavor and fragrance industry and also have favorable cetane numbers (as we report here). We describe specific improvements that resulted in a 700-fold enhancement in methyl ketone titer relative to that of a fatty acid-overproducing *E. coli* strain, including the following: (a) overproduction of β–ketoacyl-coenzyme A (CoA) thioesters achieved by modification of the β-oxidation pathway (specifically, overexpression of a heterologous acyl-CoA oxidase and native FadB, and chromosomal deletion of *fadA*) and (b) overexpression of a native thioesterase (FadM). FadM was previously associated with oleic acid degradation, not methyl ketone synthesis, but outperformed a recently identified methyl ketone synthase (*Sh*MKS2, a thioesterase from wild tomato) in β–ketoacyl-CoA-overproducing strains tested. Whole-genome transcriptional (microarray) studies led to the discovery that FadM is a valuable catalyst for enhancing methyl ketone production. The use of a two-phase system with decane enhanced methyl ketone production by 4 to 7-fold in addition to increases from genetic modifications.

INTRODUCTION

Aliphatic methyl ketones are naturally occurring compounds that were first discovered in rue (*Ruta graveolens*) more than a century ago (30) and have since been commonly found in microorganisms, plants, insects, and mammalian cells (10). These compounds have a variety of important natural and commercial roles, including acting as pheromones and natural insecticides in plants (1), or providing scents in essential oils and flavoring in cheese and other diary products (10). Biosynthesis of methyl ketones has been hypothesized to derive from a variety of different

69 biological pathways such as fatty acid β-oxidation or aerobic alkene/alkane degradation (10, 21).
70 However, studies to elucidate the genes and biochemical pathways involved in the synthesis of
71 these compounds have been quite rare until recently. One research group in particular has carried
72 out extensive biochemical and genetic studies in a wild tomato species (*Solanum habrochaites*)
73 and identified two key genes, methyl ketone synthase I (*ShMKS1*) and methyl ketone synthase II
74 (*ShMKS2*), that are essential for methyl ketone synthesis from fatty acid intermediates in this
75 plant (6, 11, 31). *Sh*MKS2, which belongs to the 4-hydroxybenzoyl-CoA thioesterase (4-HBT)
76 family, is hypothesized to hydrolyze a β-ketoacyl-acyl carrier protein (ACP) thioester
77 intermediate to generate a β-keto acid; *Sh*MKS1, an enzyme that belongs to the α/β-hydrolase
78 superfamily, apparently decarboxylates the β-keto acid released by *Sh*MKS2 to yield a methyl
79 ketone (31).
80 Despite the commercial relevance of methyl ketones and their prevalence in nature, no
81 genes other than *ShMKS1*, *ShMKS2,* and *At1g68260* (a *ShMKS2* homolog from *Arabidopsis*
82 *thaliana*)*,* have been recombinantly expressed and shown to be associated with methyl ketone
83 biosynthesis (31). Metabolic engineering of microbes to overproduce methyl ketones may merit
84 additional attention, as these compounds could be relevant to the biofuel industry as well as the
85 flavor and fragrance industry by virtue of their highly reduced, aliphatic character. Indeed, a
86 range of other fatty-acid derived compounds have already been successfully synthesized from
87 metabolically engineered microbes for use as biofuels, such as fatty acid ethyl esters (26),
88 alkanes (24), alkenes (5, 18, 22, 28), and *n*-alcohols (9).
89 In this article, we report on engineering of *E. coli* to overproduce saturated and
90 monounsaturated methyl ketones in the C_{11} to C_{15} (diesel) range for potential application to
91 biofuel production. We describe specific improvements that resulted in more than 4500-fold

92 enhancement in methyl ketone titer relative to that of a fatty acid-overproducing *E. coli* strain,
93 including re-engineering of β–oxidation, overexpression of a thioesterase native to *E. coli*
94 (FadM), and use of decane overlays. Methyl ketone titers in the best producing strains were in
95 the range of 380 mg/L.

96 **MATERIALS AND METHODS**

97 **Bacterial strains, plasmids, oligonucleotides, and reagents.** Bacterial strains and plasmids
98 used in this study are listed in Table 1. Plasmid extractions were carried out using the QIAGEN
99 (Valencia, CA) miniprep and midiprep kits. Oligonucleotide primers were designed using the
100 web–based PrimerBlast program (http://www.ncbi.nlm.nih.gov/tools/primer-
101 blast/index.cgi?LINK_LOC=BlastHomeAd) and synthesized by Bioneer (Alameda, CA). Primer
102 sequences for amplification of *E. coli* DH1 and *Micrococcus luteus* ORFs are listed in Table 2.
103 The coding sequences (CDS) corresponding to the enzymes *Sh*MKS1 (GenBank accession no.
104 AAV87156) (11) and *Sh*MKS2 (GenBank accession no. ADK38536) (31) from *S. habrochaites*,
105 and *Uc*FatB1 (GenBank accession no. Q41635) from *Umbellularia californica* (32) were
106 synthesized and codon optimized for expression in *E. coli* by GenScript (Piscataway, NJ).
107 Codon-optimized sequences are listed in Table S1.

108 **Media and bacterial growth.** *E. coli* was propagated as previously described (23). For studies
109 of heterologous gene expression in *E. coli* strains, cells were grown in 15 ml of tryptic soy broth
110 (containing 0.2% glucose) in 30-ml glass tubes with 200-rpm agitation at 37°C, unless indicated
111 otherwise, for up to 72 hours before being harvested for analysis. Frozen glycerol stocks were
112 used as inocula for the studies described here, unless noted otherwise. When required, antibiotics
113 were added to the growth medium at the following final concentrations: chloramphenicol, 25

114 μg/ml; kanamycin, 50 μg/ml. A final concentration of 0.5 mM IPTG was added to cultures after
115 6 hours when induction of genes was required.
116 **Plasmid and strain construction for heterologous expression in *E. coli*.** Cloning of *M. luteus*
117 and *E. coli* genes into expression plasmids were carried out as previously described (5). All
118 primers used to amplify target genes are listed in Table 2. PCR products and plasmid DNA were
119 digested with the appropriate restriction enzymes and purified with QIAquick gel extraction
120 and/or PCR purification kits (QIAGEN) before being ligated and transformed into *E. coli*. When
121 no appropriate restriction sites were available for generating cohesive ends for ligation, sequence
122 and ligation independent cloning (SLIC) was performed as described by Li and Elledge (17).
123 Proper clone construction was confirmed by DNA sequencing, which was performed by
124 Quintara Biosciences (Berkeley, CA). Expression of heterologous genes in constructs was
125 confirmed by extraction of proteins, tryptic digestion, and analysis of the resulting peptides by
126 electrospray ionization liquid chromatography-tandem mass spectrometry (LC/MS/MS)(QSTAR
127 Elite Hybrid Quadrupole TOF, Applied Biosystems). Mutations of genes were performed as
128 described for the QuikChange site-directed mutagenesis kit (Agilent) using primers designed
129 with nucleotide changes that corresponded to the desired amino acid substitutions. To knock out
130 *E. coli* genes, in-frame chromosomal deletion of *E. coli* genes was carried using the method of
131 Datsenko and co-workers (2, 7).
132 **Extraction of methyl ketones and related metabolites from bacterial cultures.** For most
133 samples, methyl ketones and other metabolites were extracted from cultures using a decane
134 overlay. For overlay extractions, 1 ml of decane (Sigma, ReagentPlus ≥ 99% purity) amended
135 with perdeuterated decane ($C_{10}D_{22}$) and tetracosane ($C_{24}D_{50}$) internal standards was added to
136 fifteen-ml cultures in 30-ml glass tubes following induction with IPTG. 50 μl of decane overlay

was removed at specified time points, up to 72 hrs, for direct gas chromatography-mass spectrometry (GC/MS) analysis. For low-concentration samples in which methyl ketones were not detectable using decane overlays, extractions of cell pellets were performed as previously described (5). For all extractions, culture tubes were pre-cleaned with high purity acetone before being autoclaved. All other glass and PTFE surfaces were also rigorously cleaned with high-purity acetone and an effort was made to ensure that solvent extracts contacted only glass or PTFE surfaces, whenever possible. Metabolite data described in the Results section are from 72-hr overlays unless indicated otherwise. For fatty acid analysis, 50-µl aliquots of extracts were derivatized with ethereal diazomethane to generate fatty acid methyl esters (FAME), as previously described (5).

Analysis by GC/MS. For electron ionization (EI) GC/MS analyses with a quadrupole mass spectrometer, studies were performed with a model 7890A GC (Agilent) with a DB−5 fused silica capillary column (30-m length, 0.25-mm inner diameter, 0.25-µm film thickness; J & W Scientific) coupled to an HP 5975C series mass selective detector; 1 µl injections were performed by a model 7683B autosampler. The GC oven was typically programmed from 40ºC (held for 3 min) to 300ºC at 15ºC/min and then held for 5 min; the injection port temperature was 250ºC, and the transfer line temperature was 280ºC. The carrier gas, ultra high-purity helium, flowed at a constant rate of 1 ml/min. Injections were splitless, with the split turned on after 0.5 min. For full-scan data acquisition, the MS typically scanned from 50 to 600 atomic mass units at a rate of 2.7 scans per s. For saturated methyl ketones (C_{11}, C_{13}, C_{15}), external standard quantification (*m/z* 58 areas) was performed with authentic standards. For monounsaturated ketones, no authentic standards were available, so external standard quantification relied on total ion chromatogram (TIC) areas and saturated methyl ketone standards with the appropriate chain

length. Thus, in the absence of authentic standards, unsaturated methyl ketone data should be considered as estimates.

Analysis by liquid chromatography – atmospheric pressure chemical ionization – time of flight (LC-APCI-TOF) mass spectrometry. Liquid chromatographic separation of methyl ketones was conducted at 55°C with an Inertsil ODS-3 reverse-phase column (250-mm length, 2.1-mm internal diameter, 3-µm particle size; GL Sciences, Inc., Torrance, CA) using a 1200 Series HPLC (high-performance liquid chromatography) system (Agilent Technologies, CA). The injection volume for each measurement was 2 µl. The mobile phase was composed of water (solvent A) and methanol (solvent B) (HPLC grade, Honeywell Burdick & Jackson, CA). Methyl ketones were separated with the following gradient: 60% to 98% B for 10 min, held at 98% B for 15 min, 98% to 60% B for 17 min, held at 60% B for 8 min. A flow rate of 0.19 mL/min was used throughout.

The HPLC system was coupled to an Agilent Technologies 6210 time-of-flight mass spectrometer (TOF MS) with a 1:4 post-column split. Nitrogen gas was used as both the nebulizing and drying gas to facilitate the production of gas-phase ions. The drying and nebulizing gases were set to 10 l/min and 30 psi, respectively, and a drying gas temperature of 325°C was used throughout. The vaporizer and corona were set to 350°C and 4 µA, respectively. APCI was conducted in the positive-ion mode with a capillary voltage of 3 kV. MS experiments were carried out in the full-scan mode (*m/z* 102–1000) at 0.86 spectra per s for the detection of $[M + H]^+$ ions. The instrument was tuned for a range of *m/z* 50 – 1700. Prior to LC-APCI-TOF MS analysis, the TOF MS was calibrated with the Agilent APCI TOF tuning mix. Data acquisition and processing were performed by the MassHunter software package (Agilent Technologies).

***In vitro* assay to generate pentadecenone.** His-tagged acyl-CoA oxidase (Mlut_11700) and His-tagged *E. coli* FadB were purified as previously described (5). A 1-ml acyl-CoA oxidase assay was conducted in a screw-cap glass vial containing 1.5 mM palmitoleoyl-CoA (Sigma), 400 µg of acyl-CoA oxidase, 150 µg/ml BSA, 0.1 mM FAD, and 0.1 M potassium phosphate buffer (pH 7.5). The reaction was incubated on a rotary shaker at 30ºC for 3 hr and 4 U of catalase (Sigma) was added to the mixture and incubated as before for another 30 min at 37ºC to remove the H_2O_2 generated by the acyl-CoA oxidase. 250 µl of the acyl-CoA oxidase reaction mixture was added to a 4-ml screw-cap glass vial with a polytetrafluoroethylene (PTFE)-lined septum for the 1-ml FadB assay, which also contained 400 µg/ml of BSA, 300 mM NAD, 600 µg of FadB, and 0.1 M potassium phosphate buffer (pH 7.5). Controls included assay mixtures without FadB. Reactions were incubated on the rotary shaker overnight (~18 hrs) at 37ºC. For extraction of assay products, 1 ml hexane (amended with $C_{10}D_{22}$ internal standard) was added to the assay solution, mixed well, allowed to sit for 20 min, and the solvent layer was transferred to a 10-ml conical glass vial. The extraction step was repeated and the two 1-mL aliquots of hexane were combined and then concentrated to 50 µl under a gentle stream of ultra high-purity N_2 for subsequent analysis by GC/MS.

Transcriptional studies of *E. coli* with reverse transcription-quantitative Polymerase Chain Reaction (RT-qPCR) and microarray analyses. For transcriptional studies, *E. coli* cultures were grown in 15 ml of tryptic soy broth in a 30-ml glass tube as described above, induced with IPTG after 6 hours, and harvested at 8 hours into 2 ml of ethanol solution containing 5% phenol to stop further transcription and preserve RNA integrity. Cell cultures were spun down and the pellets were immediately frozen in liquid nitrogen and stored at -80ºC until RNA extraction. Extraction and purification of RNA were carried out with the QIAGEN RNeasy Mini kit and

206 treated on-column with RNase-free DNase I (Qiagen). Concentration and integrity of RNA were
207 determined with a Thermo Scientific Nanodrop ND-1000 spectrophotometer and an Agilent
208 2100 BioAnalyzer, respectively.
209 Synthesis of cDNA for RT-qPCR analysis was carried out using 1 μg of total RNA
210 primed with 60 μM of random hexamers and reverse transcribed using a Transcriptor First
211 Strand cDNA synthesis kit (Roche, Germany). qPCR analyses were then conducted on an
212 Applied Biosystems StepOne system using 1 μl of the reverse transcription reaction and gene-
213 specific primers (Table 2) and the PerfeCTa SYBR Green FastMix (Quanta Biosciences,
214 Gaithersburg, MD). Quantitative PCR cycle parameters were as follows: initial denaturation at
215 95°C for 5 min, followed by 40 cycles of 1 s denaturation at 95°C and 30 s annealing and
216 extension at 60°C. Fluorescence measurements were taken between each cycle. At the
217 conclusion of the qPCR cycle, melting curve analysis was conducted by denaturing the PCR
218 products from 60°C to 95°C and making fluorescence measurements at 0.3°C increments. All
219 reactions were performed in triplicate. Transcripts were quantified with a standard curve
220 generated by serial dilution of pEG855 (from 10^5 to 10^{10} copies/reaction) and normalized to the
221 internal reference gene, *hcaT* (34).
222 To perform microarray analyses, 10 μg of total RNA primed with 5 μg of random
223 hexamers (Roche, Germany) were reverse transcribed using the SuperScript™ Indirect cDNA
224 labeling kit (Invitrogen). Alexa Fluor 555 dyes (Invitrogen) were then incorporated into amino-
225 allyl-dUTP-labeled cDNA, the fluorescently labeled cDNA was purified with the QiaQuick PCR
226 purification kit (Qiagen) and dried under vacuum (Vacufuge Speed Vac, Eppendorf). Labeled
227 cDNA was hybridized to the four-plex NimbleGen *E. coli* K-12 (Roche) Expression microarray
228 chip (catalog no. A6697-00-01), which contains duplicates of 8 different 60-mer probes for each

of the 4,254 genes in the *E. coli* K-12 genome, at 42ºC for 20 - 24 hours as recommended by the manufacturer. After hybridization, microarray chips were scanned with a GenePix 4000B scanner and data were extracted using NimbleScan software. Array normalization was performed using the Robust Multiarray Average (RMA) technique as described by Irizarry *et al* (13). The normalized expression values generated in RMA pair files were imported into Excel and statistical analyses were performed with the Significance Analysis of Microarray (SAM) add-on (29).

Cetane number determination. Cetane number (CN) determinations of selected methyl ketones (Sigma) were performed by the Southwest Research Institute (San Antonio, TX) according to ASTM (American Society for Testing and Materials) method D613, with no modifications.

Microarray data accession number. Microarray data have been deposited in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo) under accession number GPL14649.

RESULTS

Detection of methyl ketones in *E. coli* fatty acid-overproducing strains. Previous studies of alkene biosynthesis in *Micrococcus luteus* (5) in which *M. luteus* condensing enzymes [e.g., FabH (β–ketoacyl-ACP synthase III) and FabF (β–ketoacyl-ACP synthase II)] were heterologously expressed in a fatty acid-overproducing strain of *E. coli* DH1 resulted in unexpected GC/MS detection of methyl ketones. Authentic standards were used to confirm that these compounds were 2-undecanone (C_{11}), 2-tridecanone (C_{13}; the predominant methyl ketone), and 2-pentadecanone (C_{15}). Furthermore, we observed that overexpression of the *M. luteus fabH* (*MlfabH;* Mlut_09310) resulted in an increase in methyl ketone concentration relative to the fatty acid-overproducing control strain, particularly on an OD-normalized basis (Figure 1; Figure S1).

Enhancement of methyl ketone generation by overproduction of β–ketoacyl-CoAs. Several factors led us to hypothesize that increasing the production of β–ketoacyl-CoAs would lead to better production of methyl ketones: (a) the long-held hypothesis that, in fungi, methyl ketones arise from incomplete β–oxidation of fatty acids and decarboxylation of β–keto acids (10), (b) methyl ketones were observed at higher concentration in fatty acid-overproducing DH1 strains than in wild-type DH1 (data not shown), and (c) the carbon-chain lengths of the observed methyl ketones were consistent with decarboxylation of prominent fatty acids in DH1 (i.e., C_{12}, C_{14}, and C_{16}). To test this hypothesis and increase levels of β-ketoacyl-CoAs, we constructed a modified, truncated fatty acid β-oxidation pathway in DH1 (Figure 2).

The native fatty acid β–oxidation pathway in *E. coli* strain DH1 begins with the conversion of free fatty acids into acyl-CoAs by an acyl-CoA synthetase (FadD). The acyl-CoA is then oxidized to a 2,3-enoyl-CoA by a FAD-dependent acyl-CoA dehydrogenase (FadE). Next, FadB catalyzes a hydratase reaction to form a β–hydroxyacyl-CoA, which is then oxidized

to a β–ketoacyl-CoA (also catalyzed by the bifunctional FadB). The cycle is completed by CoA-mediated thiolytic cleavage of a β–ketoacyl-CoA to acetyl-CoA and a shorter (n-2) acyl-CoA, a reaction catalyzed by FadA. Our strategy to increase levels of β-ketoacyl-CoAs involved the following steps: (a) overexpression of a heterologous acyl-CoA oxidase used in lieu of FadE, (b) overexpression of the native FadB, and (c) deletion of *fadA* from the chromosome to truncate the β–oxidation cycle at β–ketoacyl-CoA. We chose to replace FadE with an acyl-CoA oxidase because the latter enzyme is a highly soluble protein (FadE is membrane associated) and has much higher specific activity than FadE (3, 4). Based upon reports of a high-activity acyl-CoA oxidase from *Arthrobacter ureafaciens* (3), we selected an apparent homolog (Mlut_11700; 63% protein sequence identity) from the related actinobacterium, *M. luteus*. Both Mlut_11700 and *E. coli fadB* were cloned into the low-copy pKS1 vector downstream of the *'tesA* (thioesterase) gene (Table 1). The chromosomal deletion of *fadA* in *E. coli* DH1 was performed as described in the Materials and Methods section.

GC/MS analyses of extracts of β-ketoacyl-CoA-overproducing strains indicated dramatic increases in methyl ketone production relative to fatty acid-overproducing strains (e.g., a ~75-fold increase for strain EGS560 versus strain EGS084) (Table 3, Figure 1). Concentration trends were similar on an OD-normalized basis (compare Figure 1 and Figure S1). 2-Tridecanone was the predominant methyl ketone observed in β-ketoacyl-CoA-overproducing strains as it was in fatty acid-overproducing strains (Figure 1).

Identification of candidate *E. coli* thioesterase genes involved in methyl ketone production. We demonstrated that overproduction of β-ketoacyl-CoAs increased methyl ketone production, however it was unclear whether native *E. coli* proteins were facilitating conversion of the β-ketoacyl-CoAs to methyl ketones (e.g., by hydrolysis of the CoA thioester bond to generate a

free β-keto acid and/or decarboxylation of the β-keto acid; Figure 2). Further investigation of the enhancement of methyl ketone production in the presence of *MlfabH* suggested that indeed native *E. coli* proteins were facilitating conversion of β-ketoacyl-CoAs to methyl ketones. More specifically, when we mutated the conserved, well-characterized catalytic triad residues (C123S–H275A–N306A) of *Ml*FabH (strain EGS735, Table 1), which should have rendered FabH enzymatically inactive (8), enhancement of methyl ketones was comparable to that observed in the strain expressing wild-type *Ml*FabH (EGS212) (within 10%). This suggested that *MlfabH* expression had an epigenetic rather than catalytic effect, potentially upregulating native genes whose products facilitated methyl ketone production.

To explore the possibility that native *E. coli* DH1 proteins that could facilitate methyl ketone synthesis were being upregulated in the presence of *MlfabH*, we performed whole-genome transcriptional (microarray) analysis of strains EGS212 (*MlfabH*; Table 1) and EGS084 (control; empty vector). Using the Significance Analysis of Microarray (SAM) software package, we were able to narrow down the number of significantly upregulated genes to 55 that had a false discovery rate (FDR) of 0.6% or less (Figure S2 and Table S2). Of these significantly upregulated genes, only 7 were annotated to be associated with metabolism, and two thioesterases (*paaI* and *fadM*) were the most upregulated genes in this group (Table 4). RT-qPCR analyses confirmed that *fadM* was upregulated approximately 2-fold in strain EGS212 compared to strain EGS084.

Overexpression of the *E. coli fadM* thioesterase enhances methyl ketone production. The two thioesterase genes observed to be upregulated in the presence of *Ml*FabH were overexpressed in a fatty acid-overproducing host (*fadM* in strain EGS860 and *paaI* in strain EGS790; Table 1) and the effect on methyl ketone production was assessed. Overexpression of

313 *paaI* slightly decreased methyl ketone production (~ 30%; data not shown) but overexpression of
314 *fadM* resulted in approximately a 2-fold increase in 2-tridecanone (relative to the empty-vector
315 control, strain EGS084) (Figure 1). Furthermore, overexpression of *fadM* in a β-ketoacyl-CoA-
316 overproducing strain (strain EGS895; Table 1) resulted in a 9-fold increase in methyl ketone
317 production (relative to the empty-vector control, strain EGS560) (Table 3, Figure 1).
318 **A broader range of methyl ketones (including monounsaturates) is produced in β-ketoacyl-**
319 **CoA-overproducing strains expressing FadM**. In addition to producing higher concentrations
320 of 2-undecanone, 2-tridecanone, and 2-pentadecanone relative to fatty acid-overproducing strains
321 and/or strains without *fadM* overexpression (Figure 1), strain EGS895 also produced a wider
322 range of detectable methyl ketones. This included 2-nonanone (C_9) and 2-heptadecanone (C_{17}) at
323 low relative concentration (< 1% of 2-tridecanone levels) and prominent peaks that are identified
324 as monounsaturated methyl ketones. A representative GC/MS chromatogram of a diluted decane
325 overlay of strain EGS895 is presented in Figure 3A. Peaks A and B (Figure 3A) are identified as
326 tridecenone ($C_{13}H_{24}O$) and pentadecenone ($C_{15}H_{28}O$), respectively, based upon electron-
327 ionization GC/MS spectra (Figure 3B and C), LC-APCI-TOF MS analysis, and comparison to a
328 pentadecenone standard synthesized *in vitro*. Although authentic standards are not commercially
329 available for tridecenone and pentadecenone, the TOF-determined accurate masses of the
330 molecular ions representing peaks A and B agreed extremely well (within 0.5 ppm relative error)
331 with the calculated masses for $C_{13}H_{24}O$ and $C_{15}H_{28}O$. Furthermore, the base peak at *m/z* 43 in
332 both EI spectra (Figure 3B,C) is consistent with the [CH_3-CO^+] fragment characteristic of methyl
333 ketones. Finally, an *in vitro* assay containing the CoA thioester of palmitoleic acid [(*Z*)-9-
334 hexadecenoic acid), acyl-CoA oxidase (from *M. luteus*), *E. coli* DH1 FadB, and appropriate co-
335 factors resulted in the formation of a compound with an identical GC/MS retention time and

mass spectrum as Peak B; this compound was not observed in an assay lacking FadB. Notably, an analogous assay using tetradecanoyl-CoA rather than palmitoleoyl-CoA resulted in the formation of 2-tridecanone. This strongly suggests that Peak B is (*Z*)-8-pentadecen-2-one (15:1 methyl ketone), which was derived from palmitoleic acid (16:1 fatty acid). By analogy to Peak B, it is logical to conclude that Peak A is (*Z*)-8-tridecen-2-one derived from myristoleic acid (14:1 fatty acid). However, the mass spectral fragmentation patterns of Peaks A and B differ somewhat in the region between *m/z* 50 and 120, so the position of the double bond in the tridecenone is less certain.

A summary of the quantitative relationships between methyl ketones (both saturated and unsaturated) and their presumed fatty acid precursors is presented in Table 5. Among the trends apparent from Table 5 is that ratios of fatty acid precursors to the daughter methyl ketones are much greater in fatty acid-overproducing strains (EGS084 and EGS860) than in β-ketoacyl-CoA-overproducing strains (EGS560 and EGS895), suggesting that overall conversion of fatty acids to methyl ketones is far more efficient in the β-ketoacyl-CoA-overproducing strains. In addition, ratios of fatty acid precursors to the daughter methyl ketones are typically lower in strains with overexpressed FadM (EGS860 and EGS895) than in those without (EGS084 and EGS560, respectively), further suggesting that FadM improves the conversion of fatty acids to methyl ketones.

Further characterization of the best methyl ketone-producing strain (EGS895). The relative distribution of methyl ketones produced by strain EGS895 (the best producing strain in this study) is as follows (expressed as percent of total methyl ketones): 2-undecanone (15%), 2-tridecenone (16%), 2-tridecanone (36%), 2-pentadecenone (26%), 2-pentadecanone (6%). The total concentration of methyl ketones produced by strain EGS895 was 380 $\pm$ 38 mg/L for freshly

359 transformed cells (pEG855) and 110 ± 32 mg/L in cells grown from frozen glycerol stocks. A
360 times series of methyl ketone production by strain EGS895 over 72 hr (Figure S3) indicates that
361 production begins in post-exponential phase and that the production rate decreases between 48
362 and 72 hr.
363 **Strategies to modify methyl ketone composition.** Degree of unsaturation and chain length are
364 important factors that mediate key properties of diesel fuels (e.g., low-temperature properties,
365 represented here by melting point, and CN). Three modifications to the genotype or cultivation
366 of strain EGS895 were examined to determine their impact on overall methyl ketone
367 composition and production.
368 The first strategy involved changing the cultivation temperature of EGS895 to increase
369 degree of unsaturation and thereby decrease melting point. We found that indeed the ratios of the
370 dominant unsaturated methyl ketones (C_{13} and C_{15}) to their saturated analogs increased
371 considerably when strain EGS895 was cultivated at lower temperature. To illustrate, at 37ºC, the
372 ratio of tridecenone/tridecanone was 0.45, but at 15ºC it increased to 0.93. Similarly, at 37ºC,
373 the ratio of pentadecenone/pentadecanone was 4, but at 15ºC it increased to 8.5.
374 The second strategy was to replace the native 'TesA acyl-ACP thioesterase with *Uc*FatB1
375 (strain EGS975, Table 1), a plant-derived thioesterase that has a stronger preference toward $C_{12:0}$
376 acyl-ACP than does 'TesA (32). Based on the substrate preference of *U*cFatB1, we anticipated
377 an increase in the proportion of undecanone (derived from C_{12} fatty acid) and a corresponding
378 decrease in melting point. As expected, the ratio of undecanone to tridecanone increased from
379 0.1 in strain EGS895 to 0.4 in strain EGS975, but unexpectedly the pentadecanone to
380 tridecanone ratio increased from 0.24 in strain EGS895 to 0.82 in strain EGS975.

Although both strategies achieved the intended objective of altering methyl ketone composition, they also resulted in lower total methyl ketone production (from 2- to 5-fold lower) than strain EGS895 cultivated at 37ºC. Finally, an attempt was made to increase methyl ketone production by increasing the flux of free fatty acids into the β-oxidation pathway. To accomplish this, *E. coli* FadD (fatty acyl-CoA synthetase; see Figure 2) was overexpressed in strain EGS895. However, this modification also resulted in a 2-fold decrease rather than an increase in methyl ketone production.

Methyl ketone production in strains containing *fadM* compared to production in strains containing known methyl ketone synthases. To date, the only proteins that have been experimentally verified as methyl ketone synthases are *Sh*MKS1 and *Sh*MKS2 from *S. habrochaites* and homologous proteins in other plants (6, 31). *Sh*MKS2 has been described as a "hot-dog"-fold-family thioesterase that hydrolyzes β–ketoacyl-ACPs (intermediates of fatty acid biosynthesis) and *Sh*MKS1 is a decarboxylase that acts on β–keto acids (such as those produced by *Sh*MKS2) (31). Since FadM, like *Sh*MKS2, is a thioesterase belonging to the "hot-dog" fold protein family (in this case hydrolyzing long-chain acyl-CoAs) (20), we were curious about the relative effects of overexpression of these proteins on methyl ketone production. Comparisons were made of methyl ketone (2-tridecanone) production in wild-type, fatty acid-overproducing, and β–ketoacyl-CoA-overproducing DH1 strains overexpressing *fadM*, *ShMKS2*, or *ShMKS1+ShMKS2* (Figure 4). Proteomics analyses confirmed ample expression of *Sh*MKS1 and *Sh*MKS2 in these studies. In all strains tested, constructs overexpressing *Sh*MKS2 or *Sh*MKS1+*Sh*MKS2 never produced a 2-tridecanone concentration exceeding 5% that of strain EGS895 (a β–ketoacyl-CoA-overproducing, FadM-overexpressing strain). Two aspects of the data in Figure 4 were unexpected: (a) the best methyl ketone production in a strain containing

404 *Sh*MKS2 was in the wild-type host (strain EGS1140) rather than in a fatty acid- or β–ketoacyl-
405 CoA-overproducing host, and (b) overexpression of *Sh*MKS1 with *Sh*MKS2 never improved
406 methyl ketone production, and in some cases it detracted considerably from methyl ketone
407 production. Regarding the latter point, overexpression of *Sh*MKS1 also detracted from methyl
408 ketone production in strains overexpressing FadM. To illustrate, in β–ketoacyl-CoA-
409 overproducing DH1 strains overexpressing *fadM* plus ShMKS1 (with or without its own P_{trc}
410 promoter; strains EGS1115 and EGS1120, respectively), 2-tridecanone concentrations were
411 approximately 5-fold lower than in strain EGS895, which did not contain ShMKS1 (data not
412 shown). The reason that ShMKS1 decreased methyl ketone production is unknown.
413 **Effect of decane overlay on production.** In strains with very low methyl ketone production
414 (primarily wild-type *E. coli* DH1), an exhaustive extraction of the cell pellet (using methods
415 described previously (5)) was necessary. However, decane overlays were usable for all other
416 strains. Methyl ketone production was considerably higher when fatty acid- or β-ketoacyl-CoA-
417 overproducing strains were incubated with a decane overlay than when they were sacrificed and
418 the cell pellet exhaustively extracted. To illustrate, for the best producing strain (EGS895; Table
419 1), the methyl ketone concentration was more than 4-fold greater in the overlay than in the pellet
420 extract at 39 hrs (Table 3). This may be explained by one or more of several factors, including
421 the following: (a) removal of the methyl ketone products provides a thermodynamic driving
422 force for production, (b) the overlay efficiently sequesters methyl ketones that might otherwise
423 be volatilized during cultivation, and (c) removal of methyl ketones (or other metabolites) from
424 the medium may alleviate potentially inhibitory or toxic effects from their accumulation. A
425 comparison between the results of overlay and pellet extractions supports both points (a) and (b).
426 First, the ratio of $C_{14:0}$ fatty acid to $C_{13:0}$ methyl ketone for strain EGS895 was 30-fold lower in

overlays than in pellet extractions; this lower ratio in overlays could be explained by more efficient flux of fatty acids to methyl ketones in the presence of the overlay. Second, the ratio of $C_{11:0}$ methyl ketone to $C_{15:0}$ methyl ketone is 2-fold higher for the overlay than for the pellet. Since the $C_{11:0}$ methyl ketone is more volatile than the $C_{15:0}$ methyl ketone, the higher ratio in overlays supports the notion that the decane overlay facilitates capture of volatile compounds that would be lost without an overlay. Regarding the final explanation (toxicity mitigation), this seems unlikely because OD values for strain EGS895 were similar in the presence and absence of an overlay, suggesting that methyl ketones are not particularly toxic (at least, not at these concentrations).

Cetane number determination of selected methyl ketones. Cetane number (CN) is a key index indicating overall diesel fuel quality, much as octane number is a widely used indicator of gasoline fuel quality. More specifically, CN is a measure of ignition delay during compression ignition; a higher CN indicates a shorter ignition delay period and is more favorable than a lower CN (up to a CN of 55 to 60). In the U.S., diesel fuel must have a minimum CN of 40, in accordance with ASTM Standard D975. The CN for 2-undecanone (Sigma) was 56.6 and for a 50/50 (wt/wt) mixture of 2-undecanone and 2-tridecanone was 58.4.

DISCUSSION

We have engineered a small number of modifications into *E. coli* DH1 that resulted in a 700-fold increase in methyl ketone concentration relative to a fatty acid-overproducing strain. Accounting for the use of decane overlays, the overall increase was more than 4500-fold (Table 3). The modifications included overproduction of β-ketoacyl-CoAs (by overexpression of an acyl-CoA oxidase from *M. luteus* and native FadB, as well as chromosomal deletion of *fadA*) and

overexpression of the native thioesterase, FadM. In all host strains tested (wild-type, fatty acid-overproducing, β-ketoacyl-CoA-overproducing DH1), overexpression of the methyl ketone synthase *Sh*MKS2 never produced methyl ketones at concentrations that were more than 5% of those observed for the best-producing FadM-overexpressing strain.

To some extent, the difference in behavior of the two thioesterases, FadM and *Sh*MKS2, can be explained by their known substrates. FadM has relatively high activity on acyl-CoA substrates between C_{12} and C_{18} (particularly 3,5-*cis*-tetradecadienoyl-CoA) (20), whereas *Sh*MKS2 appears to be well suited to β-ketoacyl-ACPs (31). It follows that a thioesterase that hydrolyzes CoA thioesters (FadM) would be more amenable to acting on β-oxidation intermediates whereas a thioesterase that hydrolyzes ACP thioesters would be more effective at hydrolyzing fatty acid biosynthetic intermediates (β-ketoacyl-ACPs in particular). That said, a limited amount of information is available on the substrate ranges of these two thioesterases (particularly *Sh*MKS2), so the extent to which each favors CoA versus ACP thioesters is unknown (25). Although FadM apparently hydrolyzes β-ketoacyl-CoAs sufficiently to markedly increase methyl ketone yields, it is reported to have considerably (at least 10-fold) higher activity on C_{16} acyl-CoA than on C_{16} β-ketoacyl-CoA (20).

The best methyl ketone producer studied here (strain EGS895) did not have an added decarboxylase to convert free β-keto acids to methyl ketones. Either a native enzyme catalyzed this reaction, or it occurred abiotically, as β-keto acids are well known to be inherently unstable and prone to spontaneous decarboxylation (16). Spontaneous decarboxylation would not be surprising, as we observed substantial methyl ketone yields from *in vitro* reaction mixtures that produced β-ketoacyl-CoAs from acyl-CoAs; these reaction mixtures lacked both decarboxylases and thioesterases (the only enzymes they contained were acyl-CoA oxidase and FadB). For

unknown reasons, overexpression of the *Sh*MKS1 decarboxylase, which is reported to play a role in methyl ketone synthesis in *S. habrochaites*, markedly decreased methyl ketone synthesis in this study (including strains EGS1115 and EGS1120, which were simply *ShMKS1*-amended versions of EGS895).

As is the case for other fatty acid-derived biofuels, such as fatty acid ethyl esters, saturated, medium-chain methyl ketones addressed in this article have favorable cetane numbers (CN). A less favorable property of the saturated methyl ketones addressed in this article is relatively high melting point (e.g., 30.5ºC for 2-tridecanone; (12)), which is related to cold-temperature diesel fuel properties such as cloud point. This disadvantage could be significantly mitigated by the prominent monounsaturated methyl ketones observed in the best producing strains (monounsaturated methyl ketones account for ~40% of total methyl ketones in strain EGS895). Melting point depression caused by monounsaturation in fatty acid methyl esters illustrates this point. For example, for C_{16} and C_{18} fatty acid methyl esters, the *cis*-Δ^9 monounsaturated homologs have melting points approximately 60ºC lower than those of their saturated counterparts [the melting point of methyl palmitoleate (16:1) is -33.9ºC whereas that of methyl palmitate (16:0) is 30ºC; the melting point of methyl oleate (18:1) is -19.5ºC whereas that of methyl stearate (18:0) is 39ºC](15). However, unsaturation can also be expected to decrease CN (e.g., a decrease of ~30 in CN applies to C_{16} fatty acid methyl esters; (15)). In addition to degree of unsaturation, chain length will also affect fuel properties (increasing chain length increases CN and melting point). The ensemble of saturated and unsaturated methyl ketones generated by strain EGS895 (and related strains) may have sufficiently favorable collective fuel properties to be appropriate for blending with petroleum-based diesel. Nonetheless, future efforts will be directed at enhancing methyl ketone production (e.g., by enhancing intracellular malonyl-

496 CoA levels; (33)) and modulating the methyl ketone composition to optimize diesel fuel
497 properties.

498 **ACKNOWLEDGMENTS**

499 We thank Tanveer Batth and Christopher Petzold (Functional Genomics Department,
500 Technology Division, JBEI) for mass spectrometric analysis of protein samples, Kenneth
501 Childress (Southwest Research Institute) for cetane number analysis, and Taek Soon Lee (JBEI)
502 for helpful comments on the manuscript.

503 J.D.K. has a financial interest in LS9, Amyris, and Lygos.

504 This work conducted by the Joint BioEnergy Institute was supported by the Office of
505 Science, Office of Biological and Environmental Research, of the U.S. Department of Energy
506 under Contract No. DE-AC02-05CH11231.

507

REFERENCES

1. **Antonious, G. F., D. L. Dahlman, and L. M. Hawkins.** 2003. Insecticidal and acaricidal performance of methyl ketones in wild tomato leaves. Bull. Environ. Contam. Toxicol. **71:**400-7.

2. **Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori.** 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. **2:**2006.0008.

3. **Bakke, M., C. Setoyama, R. Miura, and N. Kajiyama.** 2007. *N*-ethylmaleimide-resistant acyl-coenzyme A oxidase from *Arthrobacter ureafaciens* NBRC 12140: molecular cloning, gene expression and characterization of the recombinant enzyme. Biochim. Biophys. Acta **1774:**65-71.

4. **Baltazar, M. F., F. M. Dickinson, and C. Ratledge.** 1999. Oxidation of medium-chain acyl-CoA esters by extracts of *Aspergillus niger*: enzymology and characterization of intermediates by HPLC. Microbiology **145:**271-8.

5. **Beller, H. R., E. B. Goh, and J. D. Keasling.** 2010. Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. Appl. Environ. Microbiol. **76:**1212-23.

6. **Ben-Israel, I., G. Yu, M. B. Austin, N. Bhuiyan, M. Auldridge, T. Nguyen, I. Schauvinhold, J. P. Noel, E. Pichersky, and E. Fridman.** 2009. Multiple biochemical and morphological factors underlie the production of methylketones in tomato trichomes. Plant Physiol. **151:**1952-64.

7. **Datsenko, K. A., and B. L. Wanner.** 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.

8. **Davies, C., R. J. Heath, S. W. White, and C. O. Rock.** 2000. The 1.8 Å crystal structure and active-site architecture of β-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*. Structure **8:**185-95.

9. **Dellomonaco, C., J. M. Clomburg, E. N. Miller, and R. Gonzalez.** 2011. Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals. Nature **476:**355-9.

10. **Forney, F. W., and A. J. Markovetz.** 1971. The biology of methyl ketones. J. Lipid Res. **12:**383-95.

11. **Fridman, E., J. Wang, Y. Iijima, J. E. Froehlich, D. R. Gang, J. Ohlrogge, and E. Pichersky.** 2005. Metabolic, genomic, and biochemical analyses of glandular trichomes from the wild tomato species *Lycopersicon hirsutum* identify a key enzyme in the biosynthesis of methylketones. Plant Cell **17:**1252-67.

12. **Haynes, W. M.** 2010-2011. CRC handbook of chemistry and physics. 91st Edition.

13. **Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed.** 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics **4:**249-64.

14. **Kirchner, O., and A. Tauch.** 2003. Tools for genetic engineering in the amino acid-producing bacterium *Corynebacterium glutamicum*. J. Biotechnol. **104:**287-99.

15. **Knothe, G.** 2008. "Designer" biodiesel: optimizing fatty ester composition to improve fuel properties. Energy & Fuels **22:**1358-1364.

16. **Kornberg, A., S. Ochoa, and A. H. Mehler.** 1947. Spectrophotometric studies on the decarboxylation of β-keto acids. Fed. Proc. **6:**268.

17. **Li, M. Z., and S. J. Elledge.** 2007. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. Nat. Methods **4:**251-6.

18. **Mendez-Perez, D., M. B. Begemann, and B. F. Pfleger.** 2011. Modular synthase-encoding gene involved in α-olefin biosynthesis in *Synechococcus* sp. strain PCC 7002. Appl. Environ. Microbiol. **77:**4264-7.

19. **Meselson, M., and R. Yuan.** 1968. DNA restriction enzyme from *E. coli*. Nature **217:**1110-4.

20. **Nie, L., Y. Ren, and H. Schulz.** 2008. Identification and characterization of *Escherichia coli* thioesterase III that functions in fatty acid β-oxidation. Biochemistry **47:**7744-51.

21. **Patel, R. N., C. T. Hou, A. I. Laskin, A. Felix, and P. Derelanko.** 1980. Microbial oxidation of gaseous hydrocarbons: production of methylketones from corresponding *n*-alkanes by methane-utilizing bacteria. Appl. Environ. Microbiol. **39:**727-33.

22. **Rude, M. A., T. S. Baron, S. Brubaker, M. Alibhai, S. B. Del Cardayre, and A. Schirmer.** 2011. Terminal olefin (1-alkene) biosynthesis by a novel P450 fatty acid decarboxylase from *Jeotgalicoccus* species. Appl. Environ. Microbiol. **77:**1718-27.

23. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning : a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

24. **Schirmer, A., M. A. Rude, X. Li, E. Popova, and S. B. del Cardayre.** 2010. Microbial biosynthesis of alkanes. Science **329:**559-62.

25. **Spencer, A. K., A. D. Greenspan, and J. E. Cronan, Jr.** 1978. Thioesterases I and II of *Escherichia coli*. Hydrolysis of native acyl-acyl carrier protein thioesters. J. Biol. Chem. **253:**5922-6.

26. **Steen, E. J., Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S. B. Del Cardayre, and J. D. Keasling.** 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature **463:**559-62.

27. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189:**113-30.

28. **Sukovich, D. J., J. L. Seffernick, J. E. Richman, J. A. Gralnick, and L. P. Wackett.** 2010. Widespread head-to-head hydrocarbon biosynthesis in bacteria and role of OleA. Appl. Environ. Microbiol. **76:**3850-62.

29. **Tusher, V. G., R. Tibshirani, and G. Chu.** 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. U S A **98:**5116-21.

30. **Williams, C. G.** 1858. On the constitution of the essential oil of rue. Philosophical Transactions of the Royal Society of London **148:**199-204.

31. **Yu, G., T. T. Nguyen, Y. Guo, I. Schauvinhold, M. E. Auldridge, N. Bhuiyan, I. Ben-Israel, Y. Iijima, E. Fridman, J. P. Noel, and E. Pichersky.** 2010. Enzymatic functions of wild tomato methylketone synthases 1 and 2. Plant Physiol. **154:**67-77.

32. **Yuan, L., T. A. Voelker, and D. J. Hawkins.** 1995. Modification of the substrate specificity of an acyl-acyl carrier protein thioesterase by protein engineering. Proc. Natl. Acad. Sci. U S A **92:**10639-43.

33. **Zha, W., S. B. Rubin-Pitel, Z. Shao, and H. Zhao.** 2009. Improving cellular malonyl-CoA level in *Escherichia coli* via metabolic engineering. Metab. Eng. **11:**192-8.

34. **Zhou, K., L. Zhou, Q. Lim, R. Zou, G. Stephanopoulos, and H. P. Too.** 2011. Novel reference genes for quantifying transcriptional responses of *Escherichia coli* to protein overexpression by quantitative PCR. BMC Mol. Biol. **12:**18.

599 TABLE 1. Bacterial strains, plasmids and primers used in this study

Strain or plasmid	Relevant Characteristics	Source or reference
E. coli strains		
BL21 (DE3)	F^- *ompT gal dcm lon hsdSB*($r_B^- m_B^-$) λ(DE3)	(27)
DH1	*endA1 recA1 gyrA96 thi-1 glnV44 relA1 hsdR*17($r_K^- m_K^+$) λ^-	(19)
LT-ΔfadE	DH1 *ΔfadE* with pKS1	(26)
EGS084	LT-ΔfadE with pEC-XK99E	(5)
EGS212	LT-ΔfadE with pEG205	(5)
EGS514	BL21(DE3) with pEG513	This study
EGS517	BL21(DE3) with pEG516	This study
EGS522	DH1; *ΔfadE;* Δ*fadA*	This study
EGS560	EGS522 with pEG530 & pEC-XK99E	This study
EGS700	EGS522 with pEG530 & pEG205	This study
EGS735	LT-ΔfadE with pEG705	This study
EGS790	LT-ΔfadE with pEG775	This study
EGS860	LT-ΔfadE with pEG855	This study
EGS895	EGS522 with pEG530 & pEG855	This study
EGS975	EGS522 with pEG955 & pEG855	This study
EGS1015	EGS522 with pEG530 & pEG990	This study
EGS1080	EGS522 with pEG530 & pEG1065	This study
EGS1085	EGS522 with pEG530 & pEG1070	This study
EGS1090	EGS522 with pEG530 & pEG1075	This study
EGS1115	EGS522 with pEG530 & pEG1101	This study
EGS1120	EGS522 with pEG530 & pEG1106	This study
EGS1135	DH1 with pEG1065	This study
EGS1140	DH1 with pEG1075	This study
EGS1150	LT-ΔfadE with pEG1065	This study
EGS1155	LT-ΔfadE with pEG1075	This study
M. luteus strains		
ATCC 4698	Wild type	ATCC
Plasmids		
pEC-XK99E	Km^r; *E. coli - C. glutamicum* shuttle expression vectors based on the medium copy number plasmid pGA1 and containing the *trc* promoter	(14)
pKS1	Cm^r; p15a derivative containing '*tesA* under the *lacUV5* promoter	(26)
pKS104	Amp^r, ColE1 derivative with *fadD* (M335I), *atfA* under the *lacUV5* promoter	(26)
pSKB3	Km^r; A derivative of the expression vector pET-28a with the thrombin protease site replaced by a TEV protease site.	Burley[a]
pEG205	Km^r; ~1-kb fragment of Mlut_09310 (*MlfabH*)	(5)

	cloned into pEC-XK99E at EcoRI and XbaI sites.	
pEG513	Km^r; ~2.2-kb fragment of *fadB* (EcDH1_4315) cloned into pSKB3 at NdeI and SalI sites.	This study
pEG516	Km^r; ~2.1-kb fragment of Mlut_11700 cloned into pSKB3 at NdeI and SalI sites.	This study
pEG530	Cm^r; ~2.1-kb fragment of Mlut_11700 and ~2.2-kb fragment of *fadB* (EcDH1_4315) cloned downstream of the '*tesA* gene in pKS1 by SLIC.	This study
pEG705	Km^r; QuikChange mutagenesis of *MlfabH* in pEG205 to the following residues*:* C123S, H275A, and N306A.	This study
pEG775	Km^r; ~0.4-kb fragment of *paaI* (EcDH1_2249) cloned into pEC-XK99E at EcoRI and XbaI sites.	This study
pEG855	Km^r; ~0.4-kb fragment of *fadM* (EcDH1_3166) cloned into pEC-XK99E at EcoRI and XbaI sites.	This study
pEG955	Cm^r; ~2.2-kb fragment of L-*mbp-UCfatB1*, ~2.1-kb fragment of Mlut_11700 and 2.2-kb fragment of *fadB* cloned into pKS1 (digested with MfeI and SalI to remove *'tesA*) by SLIC.	This study
pEG990	Km^r; ~1.7-kb of *fadD* M335I allele from pKS104 cloned downstream of *fadM* in pEG855 by SLIC.	This study
pEG1065	Km^r; ~0.8-kb fragment of *ShMKS1 and* ~0.6-kb fragment of *ShMKS2* cloned into pEC-XK99E at BamHI and SalI sites by SLIC.	This study
pEG1070	Km^r; ~0.8-kb fragment of *ShMKS1* cloned into pEC-XK99E at BamHI and XbaI sites.	This study
pEG1075	Km^r; ~0.6-kb fragment of *ShMKS2* cloned into pEC-XK99E at BamHI and XbaI sites.	This study
pEG1101	Km^r; ~ 0.9-kb fragment of p_{trc}-*ShMKS1* cloned downstream of *fadM* in pEG855 by SLIC.	This study
pEG1106	Km^r; ~ 0.8-kb fragment of *ShMKS1* cloned downstream of *fadM* in pEG855 by SLIC.	This study
pEG1145	Km^r; ~1.2-kb fragment of *hcaT* (EcDH1_1132) into pEC-XK99E at EcoRI and XbaI sites.	This study

a Stephen K. Burley.

601 TABLE 2. Primers used in this study

Target genes	Primer name	Primer Sequence[a,b] (5'→3')
Primers used for target gene amplification		
fadB	DH1_fadB_SLIC_F1	GCGAAGCAGTTGCAGCCTTTAGTAAATCATGACTCATAAGAGCTCGGTACGACCAGATCACCTTGCGG
	DH1_fadB_SLIC_R1	TGGACGGTCATGACGATGCTCCTGTTCGTGAGTGGGGGCGTTCGAACGGCCCATCGGGGT
	DH1_fadB_F1	CTGCCATATGCTTTACAAAGGCGACACCCTGT
	DH1_fadB_R1	TACAGAATTCGAACGGCCCATCGGGGTG
fadM	DH1_fadM_F1	CGCTGAATTCACAACGTAAGGTTATTGCGCTATGC
	DH1_fadM_R1	ATGTTCTAGACTTGAGCATCCGGCACCACAAAAC
hcaT	DH1_hcaT_F1	TACTGAATTCCCTGACGGGAGGGACTCATGGT
	DH1_hcaT_R1	GCTATCTAGAGGAGCAGATCCGCAAAATGCTCG
l-mbp	L-mbp_SLIC_F1	TGTGGAATTGTGAGCGGATAACAATTGCACCAACAAGGACCATAGCATATGAAAATCGAAGAAGGTAAACTGGT
	L-mbp_SLIC_R1	AAGGCGCTTGCCAGGCTCGTCGTTGCCATCCCGAGGTTGTTGTTATTGTTATTGTTG
paaI	DH1_paaI_F1	AGTGGAATTCGGGCGCTTCTGGAGAGCGGTTA
	DH1_paaI_R1	TTATTCTAGAGGCTTCACGCATCAGGCTTCTCC
ptrc	Ptrc_SLIC_F1	GTTTTGTGGTGCCGGATGCTCAAGTCTAGATATCATCGACTGCACGGTGC
	Ptrc_SLIC_R1	TTCCATGTTTCCTCCTGCGCAGGGAATTCCATGGTCTGTTTCCTGTGTGA
ShMKS1	MKS1_SLIC_F1(MKS2)	CGTCCAGCATCATCTGTAATCTAGACCTGCGCAGGAGGAAACATGGAA
	MKS1_SLIC_F2 (fadM)	TTTTGTGGTGCCGGATGCTCAAGTCTAGACCTGCGCAGGAGGAAACATGGAA
	MKS1_SLIC_F3 (ptrc)	TCACACAGGAAACAGACCATGGAATTCCCTGCGCAGGAGGAAACATGGAA
	MKS1_SLIC_R1	GCCAAGCTTGCATGCCTGCAGGTCGACTCATTTGTATTTATTAGCGATGG
ShMKS2	MKS2_SLIC_F1	TCACACAGGAAACAGACCATGGGATCCCCTGCGCAGGAGGAAACATGTCAC
	MKS2_SLIC_R1	TTCCATGTTTCCTCCTGCGCAGGTCTAGATTACAGATGATGCTGGACG

Mlut_09310	Mlut_09310_C123S_F1	TCTCCGCCGCG**A**GCGCCGGCTAC
	Mlut_09310_C123S_R1	GTAGCCGGCGC**T**CGCGGCGGAGA
	Mlut_09310_H275A_F1	CCGCGTTCATCCCG**GC**CCAGGCCAACATGC
	Mlut_09310_H275A_R1	GCATGTTGGCCTGG**GC**CGGGATGAACGCGG
	Mlut_09310_N306A_F1	GCGGACGCCGGC**GC**CACGTCGGCCGC
	Mlut_09310_N306A_R1	GCGGCCGACGTG**GC**GCCGGCGTCCGC
Mlut_11700	Mlut_11700_SLIC F1	GTCATTGTCGATGCAATTCGCACCCCGATG GGCCGTTCGAACGCCCCCACTCACGAACAG G
	Mlut_11700_SLIC R1	TGCCTCTAGCACGCGTCTCACTATAGGGCG AATTGGAGCTCCACCGCGAGGTGACGGGG
	Mlut_11700_F2	GATTCATATGACCGTCCACGAGAAGCTCGC
	Mlut_11700_R2	GATTGAATTCACCGCGAGGTGACGGGGG
UcfatB1	UcfatB1_SLIC_F1	CAACAATAACAATAACAACAACCTCGGGAT GGCAACGACGAGCCTGGCAAGCGCCTT
	UcfatB1_SLIC_R1	ATCCGCAAGGTGATCTGGTCGTACGAGCTC TCACACACGCGGTTCAGCCGGAAT
Primers used for real-time PCR		
fadM	fadM_qPCR_ F1	CCGCTACCTTGAATTTCTCG
	fadM_qPCR_ R1	ACGACGAAGGCGATGTTATG
hcaT	hcaT_qPCR_ F1	GCTGATGCTGGTGATGATTG
	hcaT_qPCR_R1	AGTCGCACTTTGCCGTAATC

[a] Underlined sequences indicate restriction sites or homology regions used for cloning purposes.

[b] Bold sequences indicate nucleotide changes from wild-type gene to generate site-directed mutations.

TABLE 3. Fold improvements in total methyl ketone production[a] resulting from genetic modifications and the presence of a decane overlay

Strains		Overlay		
		EGS895[b]	EGS560[c]	EGS084[d]
Overlay	EGS084	700	76	
	EGS560	9.0		
	EGS895			
Pellet[e]	EGS084	4600	500	6.6
	EGS560	61	6.6	
	EGS895	4.7		

[a] Ratios of total methyl ketone concentrations at 39 hrs. Individual and total methyl ketone concentrations in these strains are presented in Table S3.

[b] Strain EGS895 - β-Ketoacyl-CoA-overproducing, FadM-overexpressing (full description in Table 1)

[c] Strain EGS560 - β-Ketoacyl-CoA-overproducing control without FadM (full description in Table 1)

[d] Strain EGS084 – Fatty acid-overproducing control without FadM (full description in Table 1)

[e] Cell pellet extracted after incubation, no decane overlay used.

TABLE 4. List of metabolic genes that were significantly upregulated during heterologous expression of *Ml*FabH[a].

Gene ID	Gene Name	Fold Change	Function
b1396	***paaI***	**3.4**	**predicted thioesterase[b]**
b0443	***fadM***	**2.3**	**long-chain acyl-CoA thioesterase III[b]**
b0459	*maa*	2.1	maltose *O*-acetyltransferase
b4040	*ubiA*	2.0	*p*-hydroxybenzoate octaprenyltransferase
b3769	*ilvM*	2.0	acetolactate synthase II, small subunit
b4039	*ubiC*	1.9	chorismate pyruvate-lyase
b1400	*paaY*	1.7	predicted hexapeptide repeat acetyltransferase

a Based upon whole-genome microarray analysis of strain EGS212 and control strain EGS084.

b The two thioesterase genes used for further characterization are indicated in bold.

TABLE 5. Molar ratios of precursor fatty acids to their daughter methyl ketones in fatty acid- and β-ketoacyl-CoA-overproducing strains of *E. coli* DH1 with and without *fadM* overexpression.

Strains		C_{12} fatty acid / C_{11} methyl ketone[a]	C_{14} fatty acid / C_{13} methyl ketone	$C_{14:1}$ fatty acid / $C_{13:1}$ methyl ketone[b]	C_{16} fatty acid / C_{15} methyl ketone	$C_{16:1}$ fatty acid / $C_{15:1}$ methyl ketone
low ↓ Methyl ketone yield ↓ high	EGS084	35	112	NA[c]	220	NA
	EGS860	33	30	6.2	41	68
	EGS560	0.50	0.40	0.13	0.97	0[d]
	EGS895	0.078	0.041	0.018	0.17	0.0052

[a] Fatty acids were determined as methyl esters.

[b] "X:Y" notation represents "# carbon atoms : # C=C double bonds"

[c] Not Applicable; unsaturated methyl ketone was not detected.

[d] Fatty acid (16:1) not detected

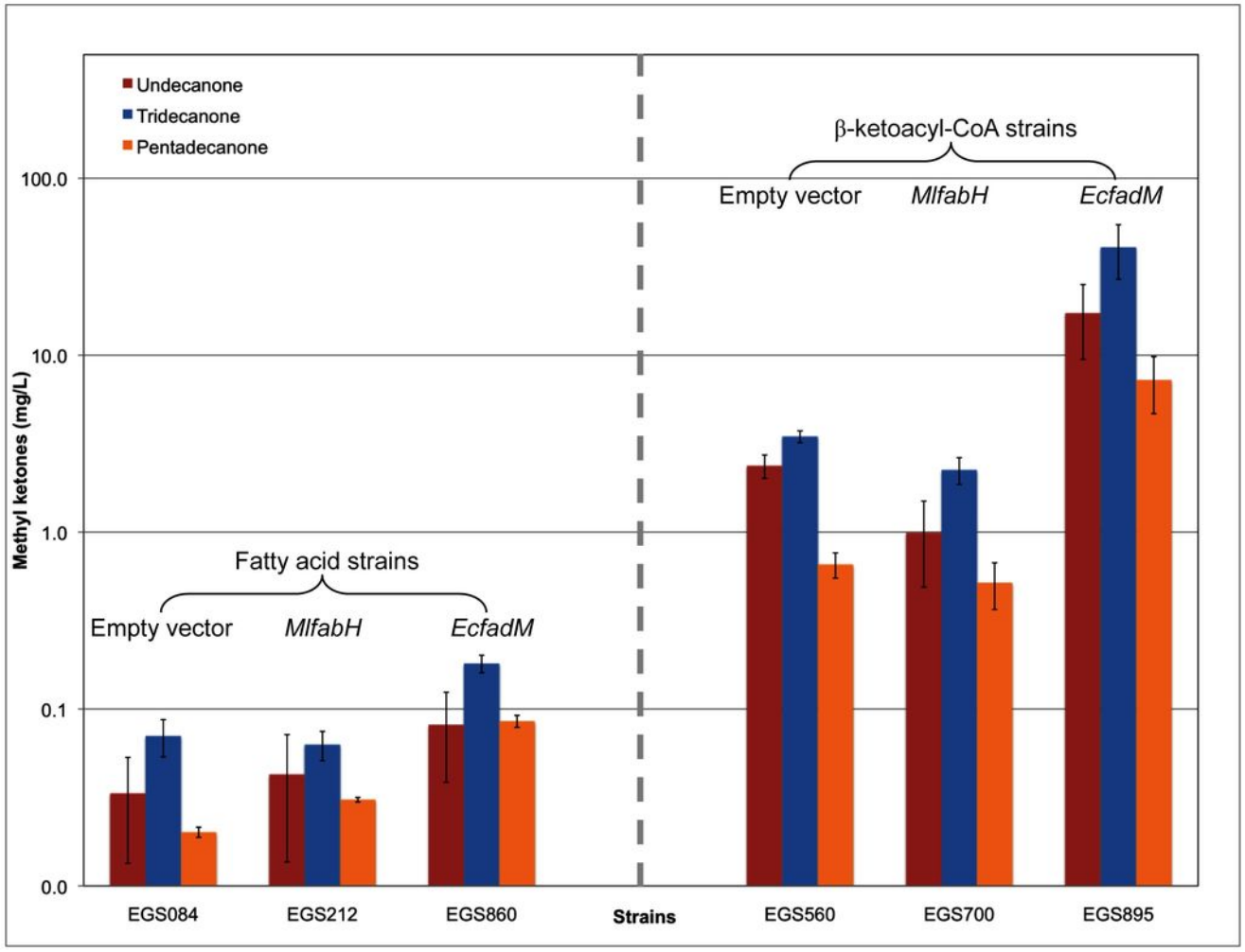
FIGURE LEGENDS

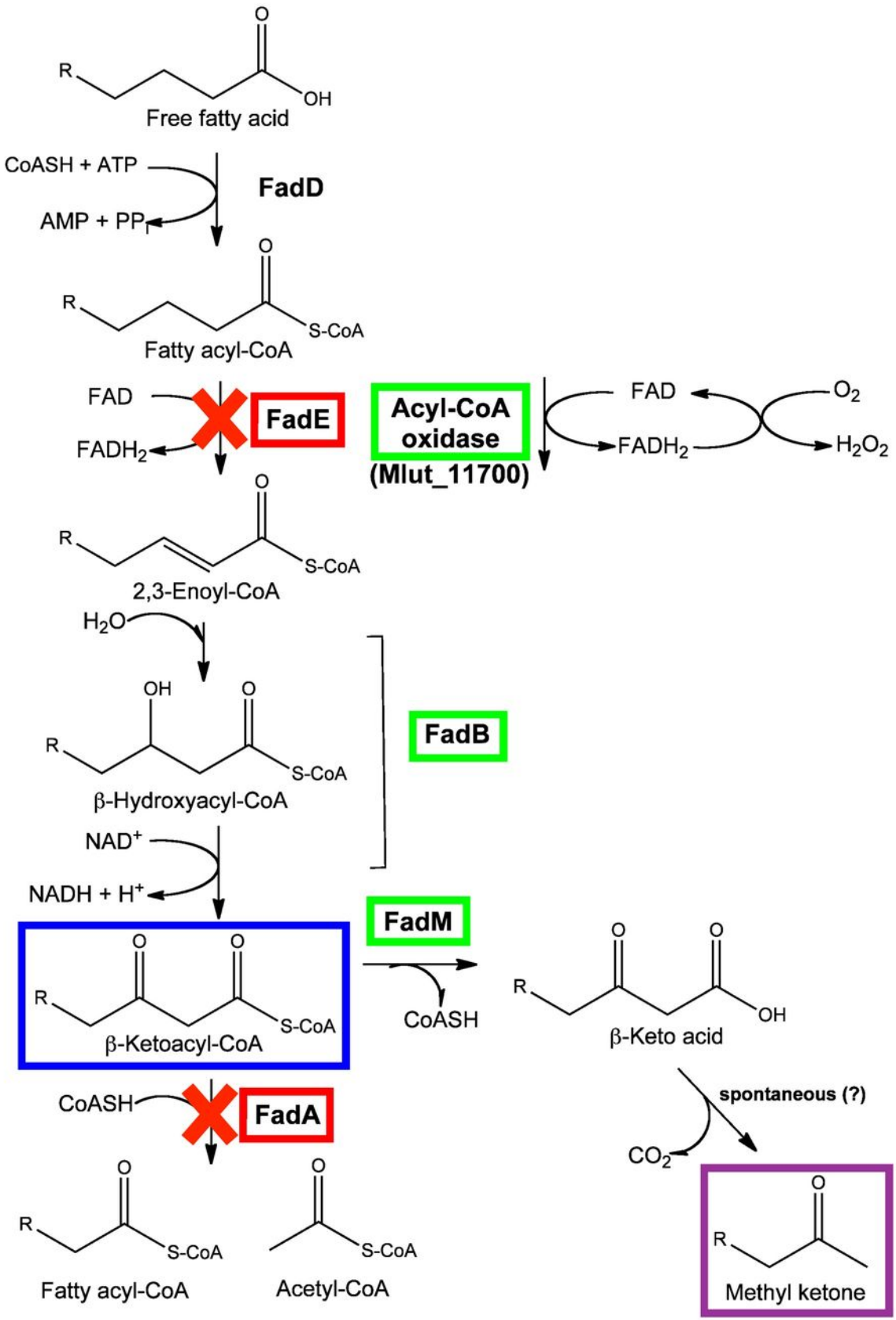
FIG. 1. Methyl ketone production in fatty acid- and β-ketoacyl-CoA-overproducing strains. Bar heights represent the averages of at least three biological replicates and error bars represent one standard deviation.

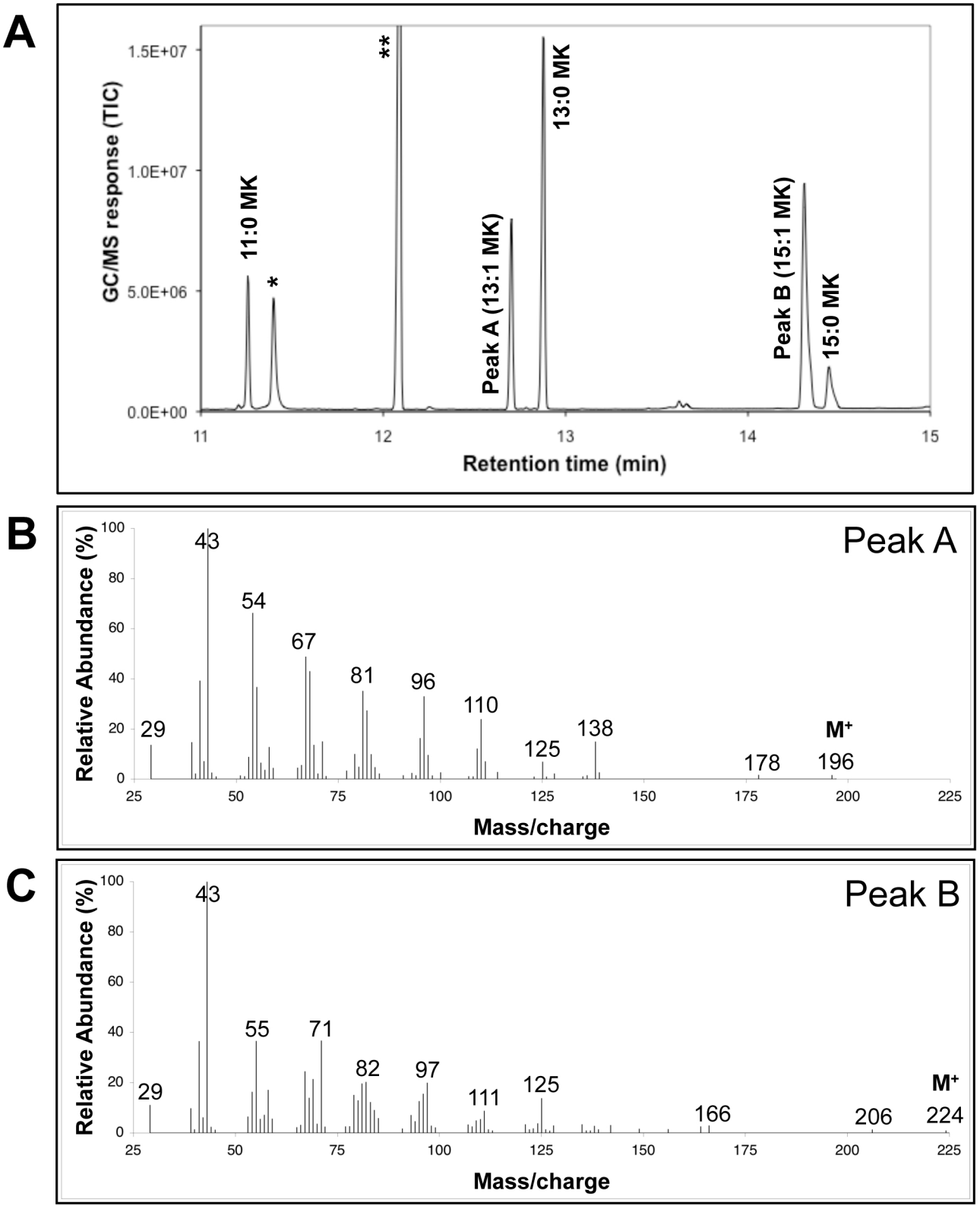
FIG.2. Summary of engineered pathway to convert fatty acids to methyl ketones in *E. coli* DH1. Green boxes indicate overexpressed genes and red boxes indicate chromosomal deletions. The blue box indicates the putative substrate for FadM (producing free β-keto acids) and the purple box indicates the final methyl ketone product (putatively generated by spontaneous decarboxylation of β-keto acids). The 'TesA thioesterase used for fatty acid overproduction is not depicted in this figure.

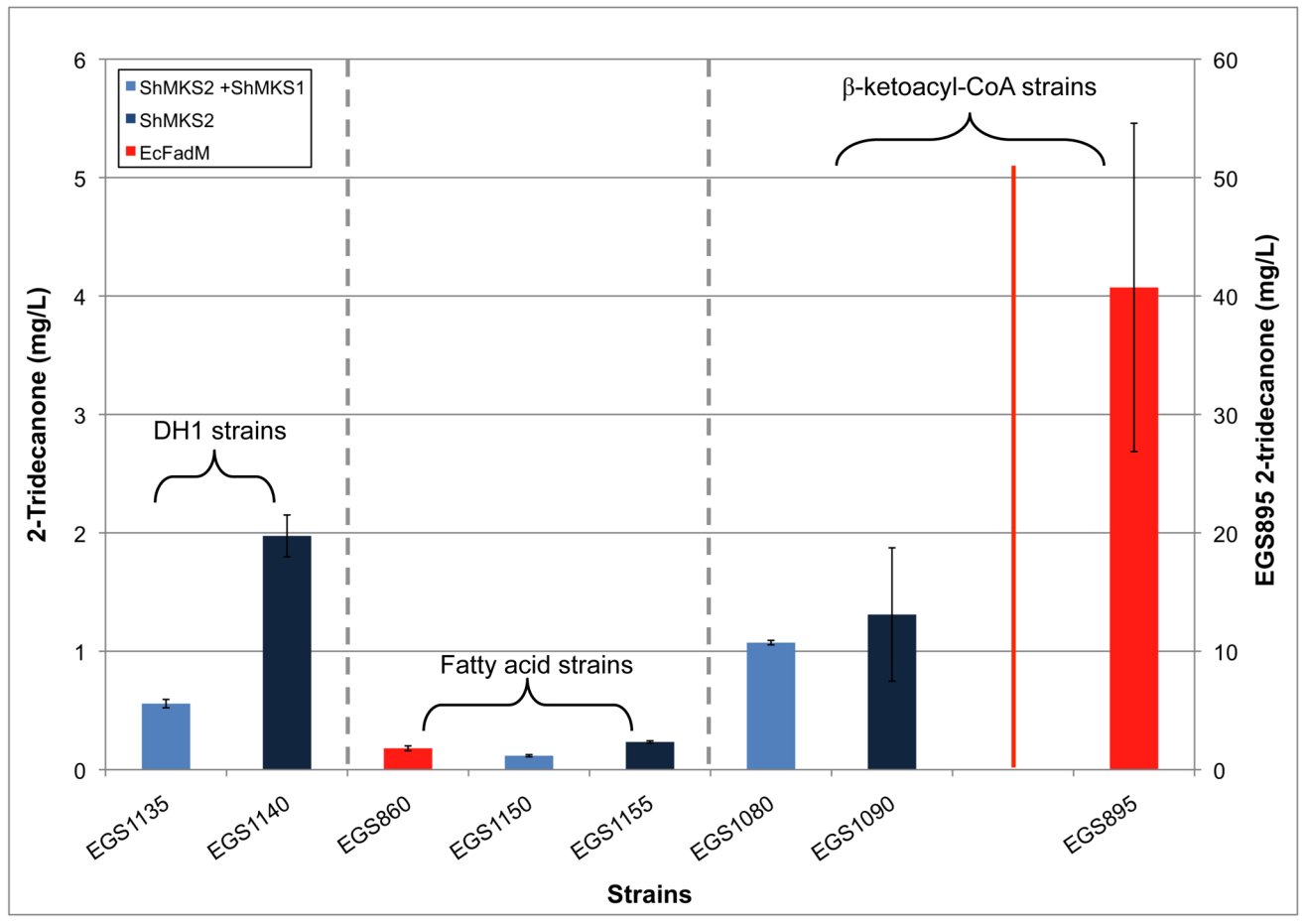
FIG.3. GC/MS chromatogram of methyl ketone mixture generated by the best producing strain (strain EGS895) and mass spectra of prominent monounsaturated methyl ketones. (A) GC/MS total ion chromatogram (TIC) of diluted decane overlay featuring region with C_{11} to C_{15} saturated and monounsaturated methyl ketones (MK). X:Y notation is described in Table 5. *, component of growth medium. **, hydrocarbon contaminant in decane. (B) 70-eV electron ionization mass spectrum of Peak A, which was identified as tridecenone (see text). (B) 70-eV electron ionization mass spectrum of Peak B, which was identified as pentadecenone (see text).

FIG. 4. 2-Tridecanone concentration in DH1 wild-type, fatty acid-overproducing, or β-ketoacyl-CoA-overproducing strains expressing various methyl ketone synthases. Note that the scale for 2-tridecanone concentration in strain EGS895 is on the right-hand *y*-axis. Bar heights represent averages and error bars represent one standard deviation.


Undecanone
Tridecanone
Pentadecanone
β-ketoacyl-CoA strains
Empty vector
MlfabH
EcfadM
Fatty acid strains
Empty vector
MlfabH
EcfadM
Methyl ketones (mg/L)
100.0
10.0
1.0
0.1
0.0
EGS084
EGS212
EGS860
Strains
EGS560
EGS700
EGS895


Free fatty acid
CoASH + ATP
AMP + PPi
FadD
Fatty acyl-CoA
FAD
FADH2
FadE
Acyl-CoA oxidase
(Mlut_11700)
O2
H2O2
2,3-Enoyl-CoA
H2O
β-Hydroxyacyl-CoA
FadB
NAD+
NADH + H+
β-Ketoacyl-CoA
FadM
CoASH
β-Keto acid
spontaneous (?)
CO2
CoASH
FadA
Fatty acyl-CoA
Acetyl-CoA
Methyl ketone


A
GC/MS response (TIC)
1.5E+07
1.0E+07
5.0E+06
0.0E+00
11:0 MK
*
**
Peak A (13:1 MK)
13:0 MK
Peak B (15:1 MK)
15:0 MK
11
12
13
14
15
Retention time (min)
B
Peak A
Relative Abundance (%)
43
54
67
81
96
110
125
138
29
178
196
M+
Mass/charge
C
Peak B
Relative Abundance (%)
43
55
71
82
97
111
125
29
166
206
224
M+
Mass/charge


ShMKS2 +ShMKS1
ShMKS2
EcFadM
DH1 strains
Fatty acid strains
β-ketoacyl-CoA strains
2-Tridecanone (mg/L)
EGS895 2-tridecanone (mg/L)
0
1
2
3
4
5
6
0
10
20
30
40
50
60
EGS1135
EGS1140
EGS860
EGS1150
EGS1155
EGS1080
EGS1090
EGS895
Strains

DISCLAIMER