ARTICLES
DOI: 10.1038/s41564-017-0028-z
nature microbiology

A metabolic pathway for catabolizing levulinic acid in bacteria

Jacqueline M. Rand[1], Tippapha Pisithkul[2], Ryan L. Clark[1], Joshua M. Thiede[1], Christopher R. Mehrer[1], Daniel E. Agnew[1], Candace E. Campbell[1], Andrew L. Markley[1], Morgan N. Price[3], Jayashree Ray[3], Kelly M. Wetmore[3], Yumi Suh[3], Adam P. Arkin[3,4], Adam M. Deutschbauer[3], Daniel Amador-Noguez[2,5] and Brian F. Pfleger[1*]

Microorganisms can catabolize a wide range of organic compounds and therefore have the potential to perform many industrially relevant bioconversions. One barrier to realizing the potential of biorefining strategies lies in our incomplete knowledge of metabolic pathways, including those that can be used to assimilate naturally abundant or easily generated feedstocks. For instance, levulinic acid (LA) is a carbon source that is readily obtainable as a dehydration product of lignocellulosic biomass and can serve as the sole carbon source for some bacteria. Yet, the genetics and structure of LA catabolism have remained unknown. Here, we report the identification and characterization of a seven-gene operon that enables LA catabolism in *Pseudomonas putida* KT2440. When the pathway was reconstituted with purified proteins, we observed the formation of four acyl-CoA intermediates, including a unique 4-phosphovaleryl-CoA and the previously observed 3-hydroxyvaleryl-CoA product. Using adaptive evolution, we obtained a mutant of *Escherichia coli* LS5218 with functional deletions of *fadE* and *atoC* that was capable of robust growth on LA when it expressed the five enzymes from the *P. putida* operon. This discovery will enable more efficient use of biomass hydrolysates and metabolic engineering to develop bioconversions using LA as a feedstock.

Levulinic acid (LA) is a five-carbon γ-keto acid that can be readily obtained from biomass through non-enzymatic, acid hydrolysis of a wide range of feedstocks[1,2]. LA was named one of the US Department of Energy's 'top 12 value-added chemicals from biomass'[3] because it can be used as a renewable feedstock for generating a variety of molecules, such as fuel additives[4–6], flavours, fragrances[7,8] and polymers[9,10], through chemical catalysis. In addition, microbes can use LA as a sole carbon source and have been shown to convert LA into polyhydroxyalkanoates (PHAs)[11–13], short-chain organic acids[14–16] and trehalose[14]. All of these bioconversion studies were conducted with natural bacterial isolates because the enzymes comprising the LA assimilation pathway were unknown[14]. This knowledge gap limited metabolic engineering and the potential of creating LA-based bioconversions.

Although the enzymes involved with LA assimilation were unknown at the time of these bioconversion demonstrations, other studies identified putative intermediates and suggested pathways for LA catabolism. In a study where crude cell lysates of *Cupriavidus necator* were fed LA, the concentration of LA and free CoA decreased over time, while acetyl-CoA and propionyl-CoA concentrations increased, suggesting that LA is catabolized via CoA thioesters like other short-chain organic acids[17]. In a second study, cultures of *Pseudomonas putida* KT2440 expressing a heterologous TesB thioesterase were fed LA. Here, 4-hydroxyvalerate (4HV) and 3-hydroxyvalerate (3HV) transiently accumulated extracellularly, before ultimately disappearing[18]. This observation strongly suggested that 4HV and 3HV (or their CoA thioesters) were pathway intermediates. Finally, a metabolomic study of rat livers suggested that LA is catabolized to acetyl-CoA and propionyl-CoA via a unique phosphorylated acyl-CoA[19,20]. In summary, these observations suggested a relatively direct route from LA to β-oxidation intermediates, but the enzymes comprising such a pathway remained unknown. Here, we have investigated the genetic and biochemical factors that allow *P. putida* KT2440 to catabolize LA, and have demonstrated that the pathway could be reconstituted in vitro and in *Escherichia coli*.

Results

Identification of genes involved in LA metabolism. *P. putida* KT2440 can metabolize LA as a sole carbon source and demonstrates diauxic growth in the presence of glucose and LA (Supplementary Fig. 1a). Therefore, we initiated a genetic study to identify the genes involved in LA catabolism. We constructed a mutant library with a Tn5 mini transposase[21] and screened for *P. putida* mutants lacking the ability to grow on LA as the sole carbon source. Thirteen of 7,000 colonies screened demonstrated LA growth deficiencies. The location of each transposon insertion was determined by sequencing PCR products created with a primer nested in the transposon paired with a degenerate random primer. Table 1 shows the ten unique isolates from these 13 hits and the putative function of the disrupted genes. Two mutants had disruptions in genes involved in propionate metabolism, supporting the hypothesis that LA is catabolized to the central metabolites, acetyl-CoA and propionyl-CoA.

[1]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA. [2]Graduate Program in Cellular and Molecular Biology, University of Wisconsin-Madison, Madison, WI 53706, USA. [3]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [4]Department of Bioengineering, University of California, Berkeley, CA 94720, USA. [5]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI 53706, USA. *e-mail: pfleger@engr.wisc.edu

Table 1 | *P. putida* LA transposon insertion sites

Locus	Insertion point*	Gene name	Description/homology
PP_0364	442685	*bioH*	Pimeloyl-ACP methyl ester esterase
PP_0988	1128706	*gcvP*-1	Glycine dehydrogenase
PP_2332	2660666	N/A	ATP-dependent zinc protease family
PP_2336	2666405	*acnA*-II	Aconitate hydratase
PP_2337	2666944	*prpF*	Aconitate isomerase
PP_2791	3181098	N/A	Phosphotransferase family
PP_2793	3182533	N/A	Acyl-CoA dehydrogenase family protein
PP_2794	3183601	N/A	Short-chain dehydrogenase/ reductase family
PP_3741	4271628	*mrdA*-I	Transpeptidase
PP_4217	4765953	*fpvA*	TonB-dependent outer membrane ferripyoverdine receptor

*Insertion point based on location from *P. putida* KT2440 origin. N/A, not available.

Three transposon mutants had disruptions in a putative operon that had not been previously characterized (disrupting genes PP_2791, PP_2793 and PP_2794). Other mutants had disruptions in genes with no obvious connection to LA catabolism (*bioH*, *gcvP*, a hypothetical zinc protease, *mrdA* and *fpvA*). To confirm that we had screened a sufficient number of clones, we performed random bar code transposon-site sequencing (RB-TnSeq) for cultures enriched by growth on LA and 4HV relative to growth on glucose. RB-TnSeq is an efficient method for determining gene essentiality under different conditions with high genomic coverage[22]. This analysis identified additional genes involved in LA metabolism, including an acetoacetyl-CoA transferase important for growth on LA, genes functioning in β-oxidation and propionyl-CoA metabolism and 14 transcriptional regulators potentially involved in regulating LA metabolism. The RB-TnSeq data set also revealed that 3-hydroxybutyryl-CoA dehydrogenase (FadB) and β-ketothiolase (FadA) were also necessary for growth on LA and 4HV, supporting our hypothesis that LA metabolism terminates through β-oxidation. For a more complete summary and analysis of the fitness data, see Supplementary Table 1 and the Supplementary Note.

Operon characterization and induction. Given the propensity of bacteria to cluster related genes into operons, we examined the putative seven-gene operon PP_2791-PP_2797, which contains three of our transposon hits (PP_2791, PP_2793 and PP_2794). We analysed the sequence homology of the seven genes in the operon using the basic local alignment search tool (BLAST[23]) and assigned predicted functions (listed in Table 2). We were unable to find any published studies about these genes beyond the automated sequence annotations. Therefore, we investigated the expression and function of these genes involved in LA catabolism. First, we isolated RNA from wild-type *P. putida* grown in minimal medium with LA as the carbon source and demonstrated that we could locate all seven genes by PCR amplification of cDNA created with a reverse primer specific to PP_2797 (Fig. 1a–c). The transcription start site (TSS) of the operon was isolated by 5′ RACE[24] (Supplementary Fig. 1b) and implicated a different start codon for PP_2791, 72 bp downstream of the one originally reported[25]. A σ^{54} promoter sequence located upstream of PP_2791 was identified by comparing the sequence upstream of the new TSS with published σ^{54} promoter consensus sequences[26] (Supplementary Fig. 1c). The data presented below suggest that the proteins encoded by this operon are important in LA catabolism and we propose that the polycistronic genes be designated as *lvaABCDEFG*.

Upstream of *lvaABCDEFG*, we identified a gene oriented divergently from the operon (PP_2790) and predicted to encode a transcription factor with a σ^{54} interaction domain and homology to the propionate metabolism activator, *prpR*. The genomic organization strongly suggested that the gene encoded a regulator for the *lva* operon. Consequently, we deleted PP_2790 and evaluated the growth of *P. putida* strains on both LA and a likely intermediate, 4HV. The ΔPP_2790 mutant was unable to grow on LA and 4HV, suggesting that it acts as an activator. Expression of PP_2790 on a plasmid restored growth of the deletion strain on LA and 4HV. To identify compounds that activate *lvaABCDEFG* expression, we built a transcriptional reporter system that linked sfGFP to the σ^{54} promoter sequence located upstream of *lvaA*. The reporter cassette was cloned onto a broad host range vector (Fig. 1d) and the resulting construct was transformed into wild-type *P. putida*. We tested a variety of short- and medium-chain-length acids by adding them to rich medium and evaluating the corresponding sfGFP expression levels. We observed strong sfGFP fluorescence only when LA or 4HV was added to the system (Fig. 1d). For these reasons, we suggest that PP_2790 encodes a transcriptional regulator responsive to the LA pathway and should be designated *lvaR*.

Genetic and biochemical studies of the *lvaABCDEFG* operon. To confirm the involvement of the *lva* operon in LA catabolism, we created a deletion mutant of each *lva* gene predicted to encode an enzymatic protein and a corresponding complementation plasmid using the P_{araBAD} promoter. We tested the ability of the resulting strains to grow on LA and 4HV (Table 2 and Supplementary Fig. 2). In addition, we purified the five enzymes from cultures of *E. coli* BL21 (DE3), reconstituted the enzymatic reactions in vitro, and used liquid chromatography/mass spectrometry (LC–MS) to identify reaction products. We used selective ion scanning to monitor the masses for likely intermediates based on previous studies[17,18,20] and developed the following hypothesized pathway (Fig. 1e and Supplementary Fig. 3). First, LA is activated as a coenzyme A-thioester, levulinyl-CoA (LA-CoA). Second, LA-CoA is reduced to 4-hydroxyvaleryl-CoA (4HV-CoA). Third, 4HV-CoA is phosphorylated at the γ-position to yield 4-phosphovaleryl-CoA (4PV-CoA). Fourth, 4PV-CoA is dephosphorylated to yield a pentenoyl-CoA species (probably 3-pentenoyl-CoA). Finally, pentenoyl-CoA is hydrated to yield 3-hydroxyvaleryl-CoA (3HV-CoA), which can be further oxidized via β-oxidation to yield acetyl-CoA and propionyl-CoA or incorporated into PHA polymers. The remainder of this manuscript will provide evidence supporting our hypothesized metabolic pathway for converting LA to 3-hydroxyvaleryl-CoA (3HV-CoA) and assignment of enzymes to each reaction.

LvaE. The presence of an enzyme (encoded by *lvaE*) with homology to an acyl-CoA synthetase (including a putative CoA binding region and an AMP binding site) suggested that the degradation pathway acts on CoA thioesters and begins with the activation of acids to acyl-CoAs. The Δ*lvaE* strain grew on LA but not on 4HV, indicating that LA may also be activated by other CoA-synthetases in *P. putida*. We quantified the activity of purified LvaE (6×-His N-terminal fusion) on a variety of organic acid substrates using the EnzChek Pyrophosphate Assay Kit, which detects pyrophosphate released in the first half of the reaction, which creates an acyl-AMP intermediate (Fig. 2a). LvaE demonstrated activity on C_4–C_6 carboxylic acids, including LA and 4HV (Fig. 2b), but showed minimal activity on other organic acids (pyruvate, acetate, propionate, octanoate).

Table 2 | *P. putida* LA operon knockout and complementation

		Growth on LA		Growth on 4HV	
Genotype	**Predicted function**	**EV**	**Complement**	**EV**	**Complement**
WT		++	N/A	++	N/A
Δ*lvaR*	σ^{54}-dependent sensory box protein	–	++	–	++
Δ*lvaA*	Phosphotransferase family	–	++	–	++
Δ*lvaB*	Hypothetical protein	–	++	–	++
Δ*lvaC*	Acyl-CoA dehydrogenase family protein	–	++	+	++
Δ*lvaD*	Short-chain dehydrogenase reductase family	–	++	++	++
Δ*lvaE*	Acyl-CoA synthetase	++	++	–	+

EV, empty vector plasmid; N/A, not applicable; –, no growth; +, visible growth; ++, robust growth.

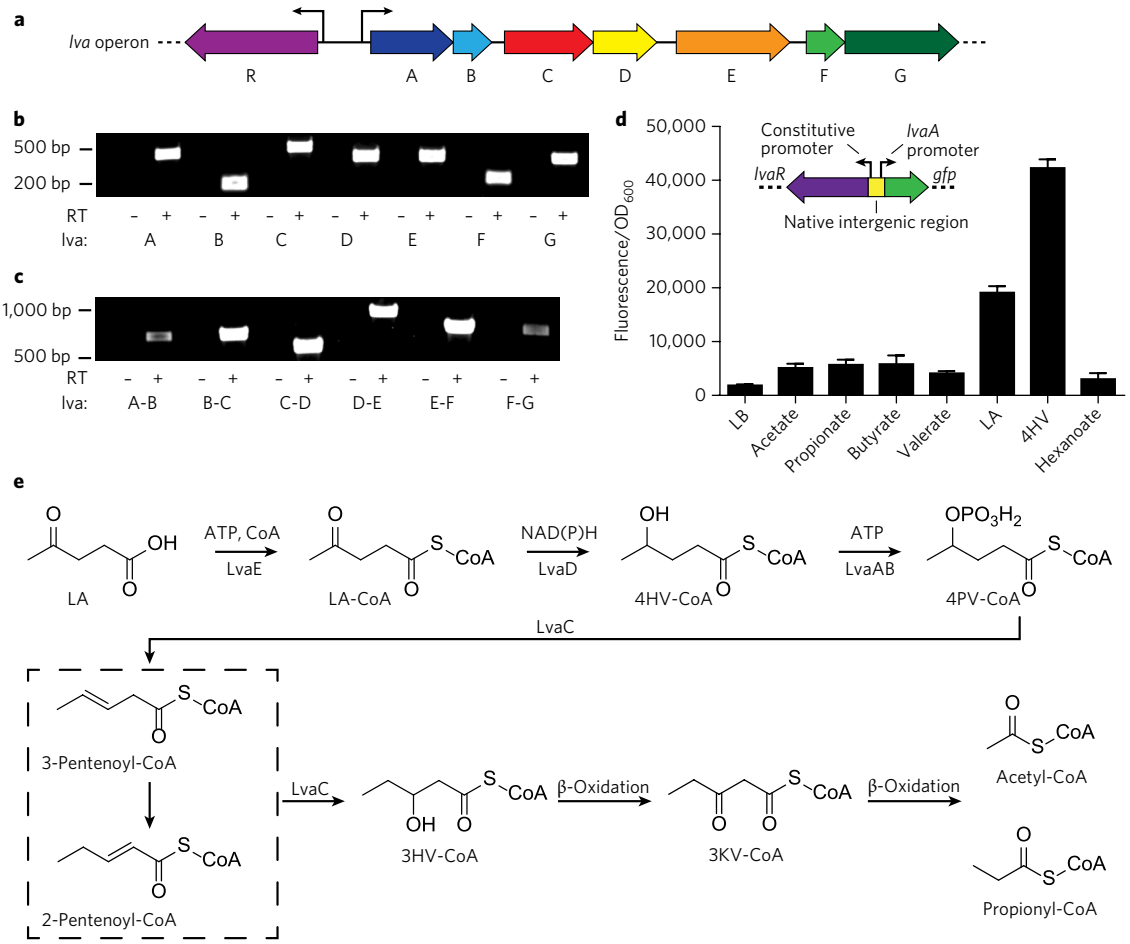

Fig. 1 | Genetic characterization and proposed catabolic activity of the *P. putida lva* operon. a, Organization of the *lvaRABCDEFG* (9,323 bp) operon. **b**, Reverse transcriptase (RT) PCR demonstrates that each gene is expressed in cells grown on LA. Samples were compared with the negative control (– RT), where reverse transcriptase was omitted from the reaction ($n=1$). **c**, RT–PCR of cDNA created with primer JMR237 demonstrates that the operon is polycistronic. Note that a product spanning each intergenic region was observed ($n=1$). **d**, *lva* operon induction assay. GFP fluorescence was measured from LB cultures supplemented with various organic acids (20 mM) ($n=3$, biological). Error bars represent s.d. Inset: schematic of transcriptional GFP fusion used to test induction of the *lva* operon. *lvaR* was cloned onto a plasmid containing its native constitutive promoter and the native promoter region for *lvaA*. Fluorescent protein sfGFP was cloned in place of *lvaA*. **e**, Proposed pathway for LA metabolism. LA, levulinic acid; 4HV, 4-hydroxyvalerate; 3HV, 3-hydroxyvalerate; LA-CoA, levulinyl-CoA; 4HV-CoA, 4-hydroxyvaleryl-CoA; CoA, coenzyme-A; ATP, adenosine triphosphate; 4PV-CoA, 4-phosphovaleryl-CoA; 3KV-CoA, 3-ketovaleryl-CoA; NAD(P)H, nicotinamide adenine dinucleotide (phosphate) reduced; GFP, green fluorescent protein.

Using LC–MS to detect reaction products (Fig. 2c), we demonstrated that LvaE was necessary and sufficient to catalyse the ligation of CoA to LA, generating LA-CoA. None of the other enzymes from the operon catalysed this or any other reaction using LA as substrate (Supplementary Fig. 4), confirming that the pathway proceeds via acyl-CoA intermediates.

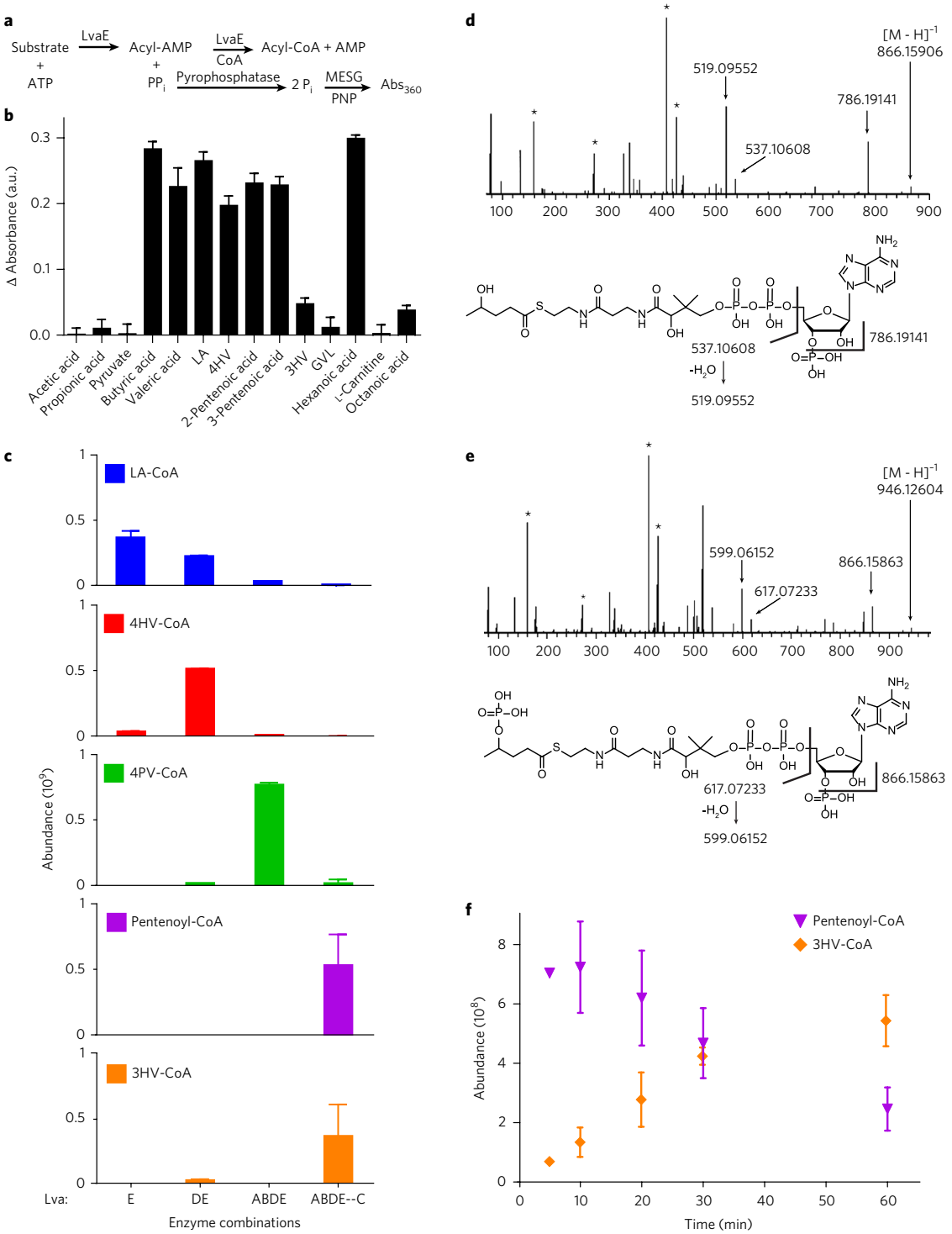

Fig. 2 | Enzymatic activity and pathway characterization for the *lva* operon. a, CoA-ligase activity assay schematic. Using the EnzChek Pyrophosphatase Assay Kit, the amount of pyrophosphate released during the CoA ligase reaction was measured as an increase of absorbance at 360 nm. **b**, Activity of LvaE towards short- and medium-chain acids ($n=3$, technical). Baseline subtraction was performed on all samples with a control reaction containing no substrate, indicated by Δ absorbance. **c**, CoA species abundance in LC–MS analysis of in vitro enzyme combinations following 30 min incubation. Reactions contained LA, CoA, ATP and NAD(P)H with varying enzyme combinations ($n=3$, technical). ABDE--C indicates that the LvaABDE reaction was performed first, metabolites were separated from LvaABDE and the resulting solution was supplemented with LvaC. The reaction confirms that LvaC is capable of converting 4PV-CoA to 3HV-CoA. **d**, MS/MS spectra for 4HV-CoA. Assignment of selected fragments from 4HV-CoA are given below. **e**, MS/MS spectra for 4PV-CoA. Assignments of selected fragments from 4PV-CoA are given below. The masses between the selected fragments of 4PV-CoA and 4HV-CoA differ by the mass of PO_3H^- (79.967), indicating that 4PV-CoA contains a phosphate group not found in 4HV-COA. Bold values indicate the mass of the parent ion. Peaks identified with * are fragments resulting from CoA. See Supplementary Fig. 6 and Supplementary Table 2 for additional fragmentation information. **f**, Abundance of pentenoyl-CoA and 3HV-CoA over a 60 min timecourse for a mixture of LvaABCDE, LA, CoA, ATP and NAD(P)H ($n=3$, technical). AMP, adenosine monophosphate; PP_i, pyrophosphate; P_i, phosphate; MESG, 2-amino-6-mercapto-7-methylpurine ribonucleoside; PNP, purine nucleoside phosphorylase; Abs, absorbance; a.u., arbitrary units. Error bars represent s.d.

LvaD. The second step in our proposed pathway is the reduction of LA-CoA to 4HV-CoA, which we predicted to be catalysed by *lvaD*. LvaD is annotated as an oxidoreductase containing an NADH binding domain and was found to be required for growth on LA but not necessary for growth on 4HV (Table 2). We purified LvaD in a similar manner to LvaE but used an N-terminal maltose binding protein (MBP) tag to increase the solubility of the enzyme[27]. The in vitro reaction containing LvaD and LvaE verified that LvaD was involved in the production of 4HV-CoA (Fig. 2c). Furthermore, LvaDE was the only enzyme combination capable of generating 4HV-CoA *in vitro* (Supplementary Fig. 4). LvaD can catalyse the reduction of LA-CoA with either NADH or NADPH (Supplementary Fig. 4).

LvaAB. We hypothesized that the third intermediate would be 4-phospho-valeryl-CoA (4PV-CoA) based on its observation in LA degradation in rat livers[19,20]. The first enzyme encoded in the operon—LvaA—has putative homology, including an ATP binding site, that associated it with the kinase superfamily and phosphotransferase family of enzymes. The second protein in the operon (LvaB) has no listed function and is predicted to be only 12 kDa in size. Orthologous sequence alignments of *lvaB* reveal that in all other organisms the gene encoding this hypothetical protein is located immediately downstream of an *lvaA* orthologue. Therefore, a pulldown experiment was used to determine if the two proteins interact[28,29]. LvaA was N-terminally tagged with MBP and cloned into a pET expression vector. LvaB was cloned directly downstream of LvaA as it is found in *P. putida's* genome. The recombinant proteins were expressed in *E. coli* BL21(DE3) and purified using the MBP tag. An SDS–PAGE gel of the eluent contained two bands at 85 kDa and 12 kDa, closely matching the predicted sizes of MBP-LvaA and untagged LvaB, respectively (Supplementary Fig. 5a,b). We performed liquid chromatography tandem mass spectrometry (LC–MS/MS) on a trypsin digest of the 12 kDa band and identified the protein sequence to be LvaB (Supplementary Fig. 5c).

Growth studies of deletion mutants revealed that *lvaA* and *lvaB* are both essential genes for growth on either LA or 4-HV. This supports the hypothesis that they catalyse a reaction after the conversion of LA-CoA to 4HV-CoA. To confirm that the association between LvaA and LvaB was important for enzymatic activity, we tested the following enzymatic combinations: (1) LvaA, LvaD and LvaE, (2) LvaB, LvaD and LvaE, and (3) LvaAB, LvaD and LvaE. We observed a decrease of 4HV-CoA and an increase of the predicted 4PV-CoA intermediate, only when all four of the enzymes were present (Fig. 2c and Supplementary Fig. 4).

To verify the identity of 4PV-CoA, we performed MS/MS (Fig. 2d,e, Supplementary Fig. 6 and Supplementary Table 2). We compared the MS/MS spectra of 4HV-CoA and 4PV-CoA and detected major ion fragments at *m/z* 786.191, 537.106 and 519.095 (4HV-CoA) and 866.158, 617.072 and 599.061 (4PV-CoA). For each compound, these fragments can be assigned to the cleavage of a P–O bond an O–C bond and the dehydration of O–C cleaved product, respectively. Both compounds are fragmenting at the same bonds, but the resulting *m/z* values for the daughter ions differ by 79.967. This mass corresponds to the *m/z* of PO_3H^-, supporting the existence of the phosphorylated 4HV-CoA species, 4PV-CoA.

LvaC. The final step in the hypothesized pathway is the formation of 3HV-CoA. Given that the combination of LvaABDE was responsible for generating 4PV-CoA and no 3HV-CoA was detected in these reactions, we postulated that LvaC was responsible for the final conversion steps. LvaC has homology to the dehydrogenase family of enzymes and 30% amino-acid sequence identity to the *E. coli* acyl-CoA dehydrogenase protein. The Δ*lvaC* strain was unable to grow on LA, but grew weakly on 4HV. LvaC was purified as an MBP fusion and the resulting protein pellet displayed a yellow hue. This is often indicative of a co-purified flavin and an absorbance scan of the protein revealed absorbance maxima that were consistent with a flavin cofactor (Supplementary Fig. 5d,e). When the LvaC sample was treated with trichloroacetic acid and centrifuged[30], a white protein pellet and a yellow hued supernatant were observed. This indicated that the cofactor was not covalently bound to LvaC.

When LvaC was added to the in vitro reaction mixture, the concentrations of intermediates (LA-CoA, 4HV-CoA and 4PV-CoA) were reduced, while the abundance of 3HV-CoA and a pentenoyl-CoA species increased (Fig. 2c). This species is probably 2-pentenoyl-CoA and/or 3-pentenoyl-CoA, which could not be resolved with our methods. Both compounds eluted at the same retention time with the same molecular mass. To test if LvaC was solely responsible for the conversion of 4PV-CoA to 3HV-CoA, we ran a two-step reaction. First, we performed the LvaABDE reaction with LA, CoA, ATP, NAD(P)H and separated the CoA products from the enzymes. To the enzyme-free mixture, we added LvaC without additional cofactors. After 30 min, we observed signals for both pentenoyl-CoA and 3HV-CoA. This indicated that the putative oxidoreductase, LvaC, is responsible for both the removal of the phosphate group to produce the enoyl-CoA and the hydration of the double bond at the 3-position. To reconstitute the whole pathway, we set up a time-course reaction with all five Lva enzymes and LA as the starting substrate. Over time, we observed a rapid increase in pentenoyl-CoA, followed by a slow disappearance that mirrored the increase in the 3HV-CoA signal (Fig. 2f). This suggested that the hydration reaction may be the limiting step in the overall pathway.

LvaFG. Based on homology alignments, *lvaG* is predicted to encode a protein with 95% amino acid sequence identity to a *Pseudomonas aeruginosa* cation acetate symporter, and LvaF shares 33% amino-acid sequence identity with the *E. coli* inner membrane protein Yhjb[23]. Sequence alignments of *lvaF* orthologues indicate that *lvaF* and *lvaG* are found with the same spatial relationship to each other in many organisms. These proteins are probably involved in organic acid transport but are unlikely to be involved in the catabolism of LA given that they were not necessary for the enzymatic conversion of LA to 3HV-CoA in vitro.

Conferring growth on LA to *E. coli*. To demonstrate the ability of the *lvaABCDE* operon to enable LA catabolism, we augmented the metabolism of *E. coli* LS5218 by transforming a plasmid linking LvaABCDE expression to an anhydrotetracycline inducible promoter (pJMR5). *E. coli* LS5218 carries two known mutations, *fadR* and *atoC*(Con), that enhance its ability to metabolize organic acids. The disruption in *fadR* is particularly useful because it deregulates expression of β-oxidation enzymes[31], including FadB and FadA, which are needed to convert 3HV-CoA to acetyl-CoA and propionyl-CoA. While *E. coli* contains the necessary genes for metabolism of propionyl-CoA[32], elevated propionyl-CoA concentrations are known to be inhibitory[33]. Therefore, we fed 20 mM LA (~0.2% by weight) to minimize the impact of propionate toxicity on growth (Fig. 3a). Initial cultures of *E. coli* LS5218 pJMR5 failed to grow on LA as a sole carbon source, so we performed adaptive evolution in effort to obtain mutants that could. The first three rounds were conducted in media containing both LA and acetate as available carbon to stimulate growth and allow cells to adapt to the presence of LA. In these experiments, we observed an increase in final cell density when both carbon sources were present relative to parallel cultures that were fed only acetate. Subsequent rounds of evolution were conducted with LA as the sole carbon source. After 14 rounds of subculturing on LA, we isolated two mutant strains, M141 and M142, capable of robust LA catabolism.

We purified the LvaABCDE expression plasmid from each mutant and discovered a mutation in the ribosome binding

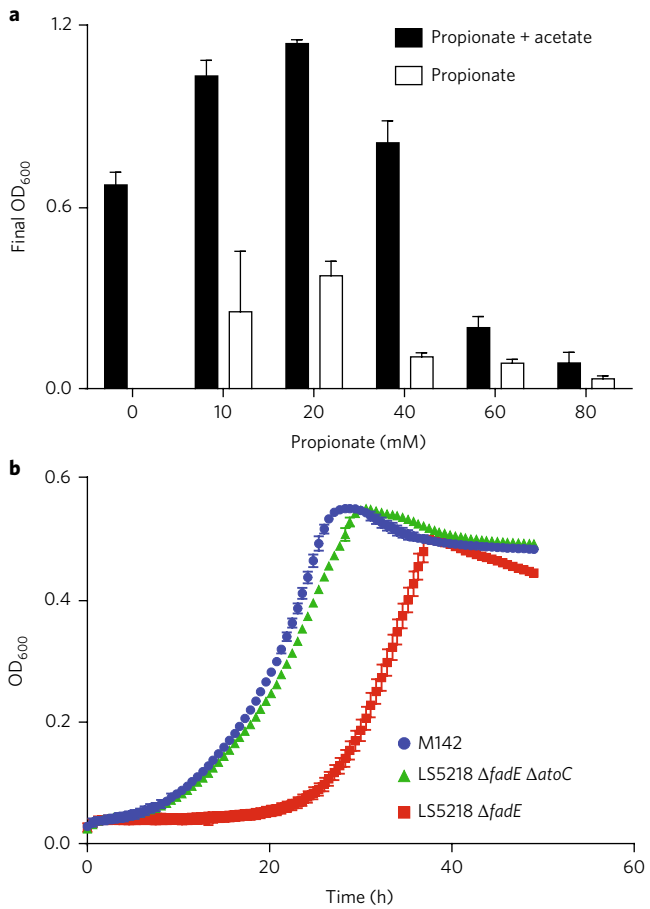

Fig. 3 | *E. coli* growth on propionate and LA. a, *E. coli* utilization of propionate ($n = 3$, biological). A growth study was performed to evaluate *E. coli* growth on various concentrations of propionate, with and without acetate as a secondary carbon source. The maximum allowable concentration that stimulated growth was 20 mM propionate, both in the presence and absence of acetate. Using this information, the LA concentration was limited to 20 mM for growth and induction studies. **b**, Growth curve of *E. coli* strains on LA. All strains harbour *lvaABCDE* on pJMR32 ($n = 3$, biological). Error bars represent s.d.

sequence (RBS) linked to *lvaA* (Supplementary Table 3). The new RBS sequence was predicted to increase the translation initiation rate relative to the original sequence[25]. Therefore, we retransformed the isolated plasmid, designated p2, back into unevolved LS5218. The resulting strain failed to grow on LA, indicating that genomic mutations were also necessary. We therefore submitted M141 and M142 for whole-genome sequencing and identified two conserved mutations (Supplementary Table 3) relative to the LS5218 genome (GCA_002007165.1). The common mutations were a point mutation in *fadE* and insertion of transposons into *atoC* that resulted in premature stop codons and functional deletion of both genes. To validate these mutations, we generated knockouts of *fadE* and/or *atoC* using CRISPR–Cas9 mediated genome engineering[34–36] and transformed each strain with pJMR32, a redesigned pJMR5 with a stronger *lvaA* RBS. Wild-type LS5218 and LS5218 Δ*atoC* were unable to grow on LA, whereas LS5218 Δ*fadE* and LS5218 Δ*atoC* Δ*fadE* grew robustly on LA as the sole carbon source (Supplementary Table 4). In liquid media, LS5218 Δ*fadE* demonstrated a significant lag relative to LS5218 Δ*atoC* Δ*fadE* and M142 (Fig. 3b). This indicated that in *E. coli* LS5218, a *fadE* deletion is necessary, but an *atoC* deletion is beneficial for growth on LA.

Discussion

The work described herein identified an operon that was essential for assimilating LA into the β-oxidation pathway of *P. putida*. Through an integrated genetic and in vitro biochemistry study, we demonstrated that genes *lvaABCDE* were upregulated in the presence of LA and were sufficient for the conversion of LA to 3HV-CoA, an intermediate of native β-oxidation. Removing any enzyme from the reaction mixture abolished 3HV-CoA production, indicating all five enzymes were necessary for this pathway. The biochemical assays confirmed the presence of 4PV-CoA, an intermediate previously observed in the metabolism of LA in rat livers. In summary, the pathway consumed at least two ATP and one reducing equivalent to produce 3HV-CoA (Fig. 1e). β-Oxidation of 3HV-CoA to acetyl-CoA and propionyl-CoA would recover the reducing equivalent. Given the energy demands of the pathway, growth on LA must be performed aerobically or in the presence of an alternative electron acceptor to enable ATP synthesis via respiration.

Like many catabolic pathways, expression of the *lva* operon is regulated by the presence of the pathway substrates. Using a transcriptional reporter assay, we demonstrated that the *lva* operon is upregulated by a transcriptional activator encoded by the divergent *lvaR* gene. Additionally, we suspect that the *lva* operon is also regulated by Crc, a global carbon catabolite repressor. Crc is an mRNA binding protein that prevents protein translation when bound to a specific mRNA sequence in *P. putida*, AAnAAnAA[37,38]. This sequence pattern is found immediately upstream of *lvaE* (Supplementary Fig. 1d), which encodes an acyl-CoA synthetase that initiates the pathway. The presence of the Crc target sequence suggests that the operon is also subject to *P. putida*'s carbon catabolite repression system, which may explain the diauxic growth curves observed for mixtures of glucose and LA.

The *lva* operon is highly conserved among the various *Pseudomonas* species (Supplementary Table 5). Gene clusters composed of the main enzymatic proteins can also be found in a variety of alpha-, beta- and gamma-proteobacteria, as graphically represented in Fig. 4. The alpha-proteobacteria species (*Azospirillum*, *Bradyrhizobium*, *Rhodopseudomonas* and *Sphingobium*) are primarily isolated from soil environments, similar to *P. putida*. The beta-proteobacteria species (*Azoarcus* and *Limnobacter*) and the gamma-proteobacteria species (*Acinetobacter* and *Marinobacter*) are isolated from both soil and ocean environments. Supplementary Table 5 lists all species that were found to contain individual homologues to LvaABCD positioned throughout the genome. Supplementary Table 6 lists all species that contain LvaACD homologues. Further investigation into the use of LA by these species could help determine whether the spatial relationship of the *lva* operon genes is important.

Although LvaB was shown to be essential for LA catabolism, its exact role remains unclear. LvaB is a small protein (~100 aa) that is unlikely to contain enzymatic activity by itself. Furthermore, LvaB co-purifies with LvaA, is essential for the phosphorylation of 4HV-CoA, and its orthologues are consistently found adjacent to orthologues of *lvaA* in the genomes of other organisms. Similar examples where small proteins provide critical support to enzymatic function have been indentified[39,40]. For example, nonribosomal peptide synthetase gene clusters often contain a small protein that belongs to the MbtH-like protein family, a family of proteins that are known to bind adenylation domains and enable catalytic activity. MbtH-like proteins form the necessary complexes required for domain activation but are not predicted to interact directly with the catalytic site[41,42]. Although LvaB does not share significant sequence homology with known MbtH-like proteins, we speculate that it could be playing a similar role with LvaA, where the presence of LvaB is required to form an active LvaAB complex. Without a crystal structure, the specific interaction between LvaA and LvaB and its role in catalysis will be difficult to unlock.

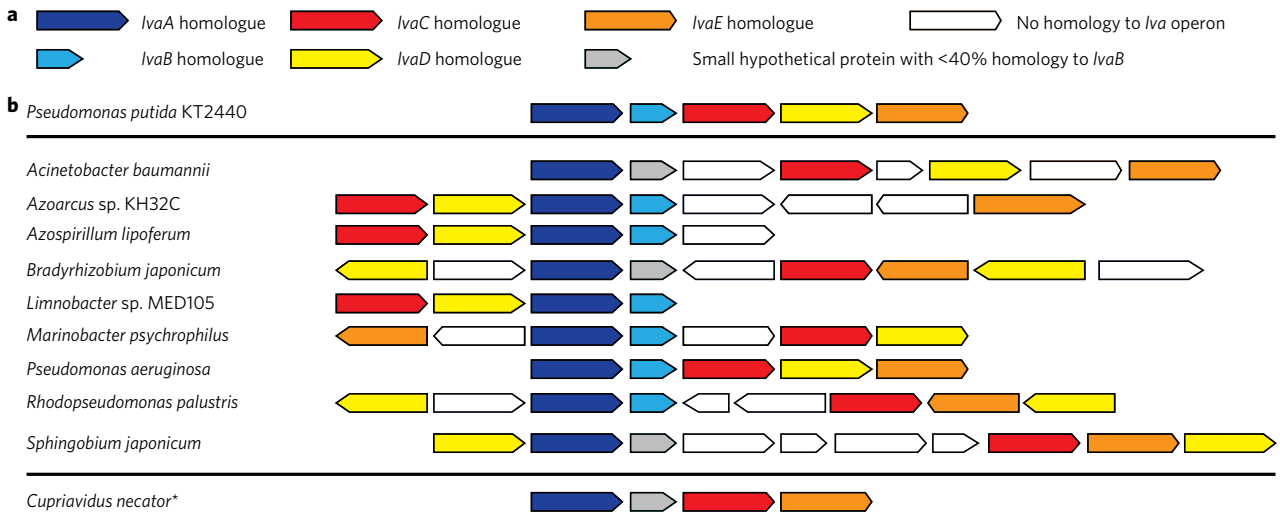

Fig. 4 | Predicted LA catabolism gene clusters in other genomes. a, Representation of *lva* operon enzymatic genes. **b**, Comparison of *lva* operon from *P. putida* KT2440 with homologous, predicted LA degradation gene clusters found in other organisms. **Cupriavidus necator* had less than 30% homology to LvaA, below the homology cutoff set for species isolation listed in Supplementary Tables 5 and 6.

Interestingly, the isomerization of 4HV-CoA to 3HV-CoA in *P. putida* proceeds through a phosphorylated intermediate, 4PV-CoA, a compound also observed in a study of LA metabolism in rat livers[20]. This study suggested the 3HV-CoA was generated via a pathway comprised of complex phosphorylated intermediates. We did not detect MS peaks corresponding to any of these compounds in our in vitro reaction mixtures. Instead, based on changes we observed in total ion abundance over time, we propose that 4PV-CoA is dephosphorylated to an enoyl-CoA and subsequently rehydrated to 3HV-CoA. We suspect that the phosphorylation of 4HV-CoA by LvaAB generates a better leaving group and makes the subsequent dehydration more thermodynamically favourable. However, the mechanism for these last steps remains unclear. Previous groups studying the non-mevalonate pathway have identified phosphate elimination steps for the formation of a double bond that is reminiscent of the intermediates we observed[43,44], but these reactions do not include a rehydration step. The timecourse measurements collected for the full reaction indicate that the formation of the pentenoyl-CoA happens quickly, but the transition from the pentenoyl-CoA to the 3HV-CoA is a much slower reaction (Fig. 2f). Our tests indicate that LvaC is capable of converting 4PV-CoA to 3HV-CoA, but those reactions still contain a higher abundance of pentenoyl-CoA compared to 3HV-CoA. A more detailed mechanistic study of the final steps may clarify the specific role of LvaC.

We selected *E. coli* strain LS5218 for our complementation studies because it constitutively expresses enzymes involved in β-oxidation and the catabolism of organic acids. Interestingly, this choice of host led to the requirement of additional mutations in *fadE* and *atoC* to permit robust growth on LA. These deletions probably prevent detrimental side reactions catalysed by FadE and AtoDA (activated by AtoC(con)[45,46]) that would compete with the desired catabolic flux to central metabolism. FadE is an acyl-CoA dehydrogenase that catalyses the formation of a *trans*-2-enoyl-CoA from an acyl-CoA[47]. We hypothesize that FadE, which is upregulated in FadR deletions mutants, may act on LA-CoA, adding a double bond between the 2- and 3-positions of the γ-ketovaleryl-CoA species and sequestering the molecule from further metabolism. Unfortunately, FadE is an inner membrane protein[48,49] that has not been purified or characterized in vitro. Although a deletion of *atoC* was not a necessary mutation, it did confer a growth benefit. We suspect that this mutation was isolated during the directed evolution process because we were screening for mutants with rapid initial growth. Constitutive activation of the *ato* regulon by the *atoC*(Con) mutation in LS5218 causes overexpression of an acetoacetyl-CoA transferase (encoded by *atoDA*), an acetyl-CoA acetyltransferase (encoded by *atoB*) and a short-chain fatty acid transporter (encoded by *atoE*)[46,50]. We suspect that 3-ketovaleryl-CoA, the product of FadB acting on 3HV-CoA, was diverted away from the desired FadA reaction by increased AtoDA activity that released 3-ketovalerate. This sequestration of LA as 3-ketovalerate would reduce overall carbon flow to central metabolites and stunt growth until cells adapt to consume 3-ketovalerate. Reducing expression of AtoDA through the deletion of *atoC* would prevent the shunt pathway and allow direct flux of LA to central metabolites. Further investigation of the competing metabolic pathways will be critical to developing LA-based bioconversions.

Methods

Chemicals, strains and media. All chemicals were obtained from Sigma-Aldrich or Fisher Scientific. 4-Hydroxyvalerate was made through the saponification of γ-valerolactone (GVL)[18]. The pH of 2 M GVL was increased to pH 12 with 10 M NaOH and incubated for 1 h. For use in bacterial growth conditions, 4HV stocks were adjusted to pH 8 with 5 M HCl.

The bacterial strains and plasmids used in this study are summarized in Supplementary Table 7. Plasmid sequences are listed in Supplementary File 'Plasmids and Jupyter notebook'. *E. coli* strains were grown at 37 °C and *P. putida* strains at 30 °C, unless otherwise noted.

Plasmid construction was completed using Phusion High Fidelity DNA Polymerase (NEB) for the PCR reactions and Gibson assembly[51]. *P. putida* genomic DNA sequences were retrieved from the NCBI database, with the following designations: PP_2791, *lvaA*; PP_2792, *lvaB*; PP_2793, *lvaC*; PP_2794, *lvaD*; PP_2795, *lvaE*; PP_2790 and *lvaR*. A 2 µl volume of the Gibson reaction mixture was transformed into chemically competent *E. coli* DH5α cells and plated on appropriate medium. Minimal media were prepared as follows: M9 minimal medium was made according to ref. [52] and MOPS minimal medium according to ref. [53]. Kanamycin was used at a final concentration of 50 µg ml^{-1}. Ampicillin was used at a final concentration of 100 µg ml^{-1}. Anhydrotetracycline (aTc) was used at a final concentration of 200 µg ml^{-1}. 5-Fluorouracil was used at a final concentration of 20 µg ml^{-1}.

Transposon library and screening. The transposon library was created following a protocol adapted from ref. [21]. Suicide vector delivery was achieved through biparental mating. Overnights of *P. putida* KT2440 and *E. coli* CC118λ pir with pBAM1 were grown with appropriate antibiotics. From overnight cultures, 1 ml of cells was pelleted by centrifugation, washed with 10 mM $MgSO_4$, and resuspended in 1 ml of 10 mM $MgSO_4$. Cells were mixed in a 1:1 ratio to a final volume of 1 ml 10 mM $MgSO_4$, with the final concentration of each strain at an optical density at 600 nm (OD_{600}) of 0.03 (3×10^7 cells). The mixture was

concentrated down to 30 µl and plated on 0.22 µm filter paper. The filter paper was incubated for 16 h on lysogeny broth (LB) agar plates at 30 °C. After incubation, the filter paper was removed from the plate and transferred into a 1.5 ml microfuge tube with 1 ml of 10 mM $MgSO_4$. The cells were resuspended by vortexing and plated onto kanamycin selective M9 citrate plates to isolate *P. putida* cells with transposon insertions. The *P. putida* transposon library was screened by replica-plating colonies from the M9 citrate plates onto LB, M9 glucose and M9 LA plates supplemented with kanamycin. Positive hits were identified as colonies that exhibited growth on LB and glucose plates but not on LA plates.

***P. putida* barcoded transposon library preparation, enrichment and analysis.** We generated a DNA-barcoded transposon mutant library of *P. putida* KT2440 using previously described methods and resources[54]. Briefly, we conjugated wild-type *P. putida* KT2440 with an *E. coli* strain (WM3064) carrying the transposon vector library pKMW3[54]. pKMW3 is a mariner class transposon vector library containing a kanamycin resistance marker and millions of random 20mer DNA barcodes. Conjugations were performed at a 1:1 donor:recipient ratio on LB + diaminopimelic acid (DAP) plates for 6 h and finally plated on LB plates supplemented with 100 µg ml^{-1} kanamycin. *E. coli* conjugation strain WM3064 is auxotrophic for DAP and does not grow on media that are not supplemented with this compound. We combined thousands of kanamycin-resistant *P. putida* colonies into a single tube, made multiple aliquots and stored these samples at −80 °C for future use. We also extracted genomic DNA and mapped the transposon insertion locations and their associated DNA barcodes via a TnSeq-like Illumina sequencing protocol, as previously described[54]. We named the final, sequenced-mapped transposon mutant library Putida_ML5.

An aliquot of the *P. putida* RB-TnSeq library (Putida_ML5) was grown for 5 h in a shake flask containing 25 ml of LB medium with 50 µg ml^{-1} kanamycin sulfate to late log phase (30 °C, 250 r.p.m.). The volume of cells corresponding to an OD_{600} of 1 was pelleted, decanted and frozen at −20 °C for barcode sequencing as the time zero inoculum control. Cells were washed with three volumes of minimal medium with no carbon source and then resuspended in 2× minimal medium with no carbon source for a new OD_{600} measurement. These cells were diluted into 2× minimal medium to an OD_{600} of 0.04. This culture was then diluted in half with 2× solutions of each carbon source of interest to a final volume of 10 ml in a culture tube for 4HV and 1.2 ml total volume in the well of a 24-well microplate for LA. The carbon sources tested were 40 mM 4HV (pH adjusted to 7 with NaOH), 40 mM LA (pH adjusted to 7 with NaOH), 20 mM potassium acetate and 40 mM glucose, each with two replicates. The 4HV and acetate experiments were performed one day and the LA experiments were performed on a different day, each day with its own 40 mM glucose control. The culture tubes were placed in a shaker incubator (30 °C, 250 r.p.m.) until they achieved an OD_{600} of ~3 for 40 mM glucose (~20 h), ~0.25 for 20 mM potassium acetate (~44 h) or ~0.3–0.5 for 40 mM 4HV (~68 h). For LA, the samples were grown in a 24-well microplate in a Multitron shaker set to 30 °C and 700 r.p.m. We monitored the OD of the microplate in a Tecan M1000 microplate reader. A 1 ml sample from each culture tube was pelleted and frozen at −20 °C for barcode sequencing.

We performed DNA barcode sequencing (BarSeq) as previously described[54], with a slight variation in our common P1 oligo design. In this study, we used a mixture of P1 oligos (Supplementary Table 10) with variable length N space regions (2–5 nt) to 'phase' our BarSeq PCR products for sequencing on the Illumina HiSeq4000.

Both the TnSeq data and the BarSeq data were processed using analysis scripts as described previously[54]. Briefly, the fitness of a strain is the normalized $\log_2$ ratio of barcode reads in the experimental sample to barcode reads in the time zero sample. The fitness of a gene is the weighted average of the strain fitness for insertions in the central 10–90% of the gene. The gene fitness values are normalized so that the typical gene has a fitness of zero. The primary statistic *t*-value is of the form of fitness divided by the estimated variance across different mutants of the same gene. All experiments described herein pass the quality metrics described previously unless noted otherwise[54].

The fitness values reported in Supplementary Table 1 are the average of two replicates. Fitness scores for LA and 4HV relative to glucose were calculated using the following equation:

$$\text{Fitness}\left(\frac{\text{LA}}{\text{Glucose}}\right) = \text{Fitness(LA)} - \text{Fitness(Glucose)}$$

Annotations in Supplementary Table 1 and discussed in the text were adapted from www.MicrobesOnline.org[55].

RNA extraction. Wild-type *P. putida* KT2440 cells were grown in MOPS minimal medium supplemented with 20 mM LA to an OD_{600} 0.8. Then, 10 OD ml were collected by centrifugation at 5,000*g* for 10 min at 4 °C in a Beckman Coulter Allegra X-15R. The supernatant was decanted and the pellet frozen at −80 °C for 24 h. The RNA extraction protocol was adapted from ref. [56]. The frozen pellet was thawed, resuspended in 1.5 ml Trizol, and transferred to a 2.0 ml microfuge tube. The suspension was incubated for 5 min at 95 °C and then for 5 min on ice. After incubation, 300 µl chloroform was added and the tube shaken vigorously for 15 s. The Trizol–chloroform mixture was incubated at room temperature for 15 min and then centrifuged for 15 min at 12,000*g* and 4 °C. The upper phase was transferred to a fresh tube and an equal volume of isopropanol was added. This mixture was incubated for 10 min at room temperature and then centrifuged for 10 min at 12,000*g* and 4 °C. The supernatant was discarded and the pellet resuspended in 1 ml of 75% ethanol. This was centrifuged for 5 min at 8,000*g* and 4 °C. The supernatant was discarded, the pellet air-dried for 3 min and then resuspended in 100 µl RNase-free water and stored at −80 °C.

Transcription start site (TSS) isolation. The TSS for genes *lvaR* and *lvaA* were isolated using an adapted 5′ Race protocol from ref. [24]. The RNA isolated from *P. putida* KT2440 was treated with a TURBO DNA-free Kit from Ambion to remove any contaminating DNA. A Promega GoScript RT PCR kit was used to generate cDNA using 1 µl of a 10 µM gene-specific oligo (JMR2 for *lvaR* and JMR287 for *lvaA*) instead of the random oligo mixture. Following inactivation of the reverse transcriptase, the cDNA was purified using a Qiagen PCR Purification kit. Tailing of the cDNA was achieved using the terminal deoxynucleotidyl transferase (TdT) enzyme from Thermo Scientific. The final reaction mixture contained 1× reaction buffer, 1 pmol cDNA fragments, 60 pmol dGTP or dCTP and 30 U TdT. The reaction was incubated at 37 °C for 15 min and then quenched by heating to 70 °C for 10 min and the tailed cDNA fragments cleaned up using a Qiagen PCR Purification kit. The tailed cDNA was amplified using GoTaq Green Master Mix with an annealing temperature of 55 °C and an extension time of 30 s. Primer GG318 was used for dGTP tailing and ALM244 was used for dCTP tailing. The reverse primer for *lvaR* was JMR150 and for *lvaA* was JMR296. The resulting PCR product was submitted for sequencing.

Polycistronic verification. Using the DNAse-treated RNA isolated from LA-grown *P. putida* KT2440, cDNA for the operon was generated with the Promega GoScript RT PCR kit using 1 µl of a 10 µM gene-specific oligo (JMR237). The cDNA was then used as the template for PCR reactions using GoTaq Green Master Mix with an annealing temperature of 55 °C and an extension time of 0:30 s. Primers used for each gene are provided in Supplementary Table 8.

***P. putida* knockouts.** The genetic knockout of *lvaD* was performed following the protocol from ref. [57]. Knockouts of the remaining genes in *P. putida* were performed following the protocol from ref. [58]. Knockout constructs were designed with 500 bp of homology up- and downstream of the deletion site. This region was cloned into the pJOE vector backbone. This suicide vector was transformed into *P. putida* KT2440 Δ*upp* (*P. putida* KTU) through electroporation, and colonies that successfully integrated the plasmid into the chromosome were selected on LB_{kan} plates. A colony was then grown in LB medium overnight to cure the counter-selection cassette. Various dilutions of the overnight culture were plated on $LB_{5\text{-FU}}$ plates to isolate colonies that had successfully excised the plasmid insertion. Colonies were then screened by colony PCR to isolate deletion strains.

Transcriptional reporter assay. *P. putida* KT2440 was transformed with pJMR74 through electroporation. pJMR74 is a broad host range plasmid containing a kan resistance marker and the predicted regulator for the *lva* operon, *lvaR*. Expressed divergent of *lvaR* is sfGFP cloned under the native promoter for *lvaA*. *P. putida* KT2440 containing empty vector pBAD35 was used as the no fluorescence control. Overnights of *P. putida* + pJMR74 or pBAD35 were inoculated at an OD_{600} of 0.05 in LB + kan[50] + 20 mM of the appropriate carboxylic acid (acetate, propionate, butyrate, valerate, LA, 4HV or hexanoate). Final timepoints were taken at 24 h in a Tecan Infinite M1000, with OD absorbance measured at 600 nm and fluorescence measured with an excitation of 485 nm and emission of 510 nm. Standard deviation error propagation was performed for the normalization of fluorescence and OD measurements.

Protein production and purification. Vectors were constructed using the pET28b backbone and individually cloned genes from the *P. putida* genome. The plasmid containing *lvaAB* was constructed using the pET28b backbone and the *lvaAB* genes cloned as an operon directly out of *P. putida*'s genome. *E. coli* BL21 (DE3) strains with sequenced verified plasmids were grown at 37 °C in LB. Cultures were induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at an OD_{600} of 0.4. The cultures were then chilled on ice for 10 min before incubation at 16 °C for 18 h in a New Brunswick Incubator I-26. The cultures were then centrifuged for 20 min at 5,000*g* in a Beckman Coulter Avanti J-E centrifuge. The supernatant was decanted and the cells resuspended in 30 ml of LB before another centrifugation at 5,000*g* for 20 min. The supernatant was removed and pellets stored at −80 °C for at least 24 h.

Purification of His6- (*lvaE*) and MBP-tagged proteins (*lvaABCD*). Frozen cell pellets were thawed on ice and resuspended in His6-lysis buffer (50 mM Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, 2 mM dithiothreitol (DTT), pH 8.0) supplemented with 2 µl of benzonase or MBP-lysis buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.4) supplemented with 2 µl of

benzonase. Cell suspensions were sonicated three times using the following program: 1.5 s pulse, 1.5 s pause, 40% duty, for a total of 30 s. Between sonication cycles, the solution was stored on ice for 5 min. Lysed cells were centrifuged at 25,000*g* at 4 °C for 30 min and the supernatant filtered through a 0.45 µm filter.

For the purification of His_6-tagged proteins, a GE Äkta Start System with a 1 ml HisTrap HP column and a constant flow rate of 1 ml min^{-1} was used. A 5 column volume (CV) of wash buffer (50 mM Na_2HPO_4, 300 mM NaCl, 40 mM imidazole, 2 mM DTT, pH 8.0) was used to equilibrate the column. The sample was loaded and washed with 15 CV wash buffer. The protein was eluted with 5 CV elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 250 mM imidazole, 2 mM DTT, pH 7.8). Fractions (1 ml) of eluted protein were collected. A GE PD-10 desalting column was used to buffer-exchange the protein into the desalting buffer (100 mM Tris, 4.1 M glycerol, 2 mM DTT). An Amicon Ultra 4 ml centrifugal filter with a 10 kDa cutoff size was used to concentrate the protein. Each protein was stored at −80 °C until use.

For purification of MBP-tagged proteins, a GE Äkta Start System with a 1 ml MBPTrap HP column and a constant flow rate of 1 ml min^{-1} was used. Wash buffer (5 CV) was used to equilibrate the column. The sample was loaded and washed with 15 CV wash buffer. The protein was eluted with 5 CV elution buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 10 mM maltose, pH 7.4). Fractions (1 ml) of eluted protein were collected. A GE PD-10 desalting column was used to buffer-exchange the protein into the desalting buffer (100 mM Tris, 4.1 M glycerol, 2 mM DTT, pH 8.2). An Amicon Ultra 4 ml centrifugal filter with a 10 kDa cutoff size was used to concentrate the protein. The protein was stored at −80 °C until use.

LvaAB pulldown experiment. All proteins for the pulldown experiment were purified on a 1 ml MBPTrap HP column, as previously described, regardless of the protein tag. LvaA was tagged with an N-terminal MBP tag. LvaAB was designed with LvaA tagged with an N-terminal MBP tag and LvaB untagged. Both proteins were expressed from the same construct as they appear in a native operon. For controls, LvaA contained an N-terminal His tag and was expressed with the native LvaB, and the last control was an N-terminal MBP tagged LvaA containing a frameshift stop codon expressed with native LvaB. The purified proteins were analysed on a 15% SDS–PAGE gel to determine the major protein products.

CoA ligase assay. A CoA ligase activity assay was performed with the EnzChek Pyrophosphate Assay Kit. The maximum allowable concentration of LvaE that did not affect assay sensitivity was determined to be 0.2 µM through an enzymatic dilution test. The final reaction volume of 100 µl contained 0.1 mM ATP, 0.1 mM CoA, 0.2 µM LvaE, 0.2 mM MESG, 1 U purine nucleoside phosphorylase, 0.01 U pyrophosphatase, 50 mM Tris-HCL, 1 mM $MgCl_2$ and 0.1 mM substrate (sodium acetate, sodium propionate, butyric acid, valeric acid, LA, hexanoic acid, octanoic acid, 4HV, 2-pentenoic acid, 3-pentenoic acid, 3HV, γ-valerolactone, pyruvate, L-carnitine). All substrate stocks were adjusted to pH 7 before use. Reactions were incubated at 25 °C for 30 min before measuring the absorbance at 360 nm in a Tecan M1000. A control reaction that did not contain a substrate was used for performing an absorbance baseline subtraction.

Enzyme assays and metabolite purification. All in vitro enzyme assays were performed in a 30 °C water bath at a pH of 7.5 and contained 50 mM Tris-HCl, 1 mM $MgCl_2$ and 2 mM DTT. Final reaction concentrations included the following components, depending on the enzymes added: 0.5 mM LA, 0.55 mM CoA, 0.55 mM ATP (1.05 mM ATP when *lvaAB* were present), 0 mM NAD(P)H (0.55 mM NAD(P)H when *lvaD* was present). Final protein concentrations were LvaA (0.2 µM), LvaB (0.8 µM), LvaAB (0.4 µM), LvaC (0.4 µM), LvaD (0.2 µM) and LvaE (0.2 µM) (Supplementary Fig. 5). The in vitro enzyme assays were incubated for 30 min, excluding the timecourse, which was incubated for various intervals up to 60 min. Reaction metabolites were purified following a modified protocol from ref. [19]. Reactions were quenched by adding methanol/water 1:1 containing 5% acetic acid in a 1:1 volume ratio (extraction buffer). Quenched reactions were run on a 1 ml ion exchange column prepacked with 100 mg 2-2(pyridyl)ethyl silica gel from Sigma. The column had been preconditioned with 1 ml methanol followed by 1 ml of extraction buffer. Metabolites loaded on the column were washed with 750 µl extraction buffer before being eluted with 1 ml of 4:1 methanol/250 mM ammonium formate, pH 6.3 and 1 ml methanol. Samples were dried using a Thermo Scientific Savant SC250EXP Speedvac Concentrator and stored at −80 °C until LC–MS analysis. Samples for LC–MS analysis were resuspended in 100 µl 50 mM ammonium formate.

LC–MS and LC–MS/MS. Samples were analysed using a high-performancce LC (HPLC)–MS/MS system consisting of a VanquishTM UHPLC system (Thermo Scientific) coupled by electrospray ionization (ESI; negative polarity) to a hybrid quadrupole high-resolution mass spectrometer (Q Exactive orbitrap, Thermo Scientific) operated in full scan mode for detection of targeted compounds based on their accurate masses. Properties of Full MS–SIM included a resolution of 140,000, AGC target of 1E6, maximum IT of 40 ms and scan range from 70 to 1,000 *m/z*. LC separation was achieved using an ACQUITY UPLC BEH C18 (2.1 × 100 mm column, 1.7 µm particle size; part no. 186002352; serial no. 02623521115711, Waters). Solvent A was 97:3 water:methanol with 10 mM tributylamine (TBA) adjusted to pH 8.1–8.2 with 9 mM acetic acid. Solvent B was 100% methanol. Total run time was 25 min with the following gradient: 0 min, 5% B; 2.5 min, 5% B; 5 min, 20% B; 7.5 min, 20% B; 13 min, 55% B; 15.5 min, 95% B; 18.5 min, 95% B; 19 min, 5% B; 25 min, 5% B. Flow rate was 200 µl min^{-1}. The autosampler and column temperatures were 4 °C and 25 °C, respectively. Fragmentation of CoA, 4HV-CoA and phosphorylated 4HV-CoA was achieved using the parameters indicated in Supplementary Table 9.

Enzymatic 'in gel' digestion and indentification of Lva proteins. 'In gel' digestion and MS analysis were carried out at the Mass Spectrometry Facility (Biotechnology Center, University of Wisconsin–Madison). Digestion was performed as outlined on https://www.biotech.wisc.edu/services/massspec/protocols/ingelprotocol. In short, Coomassie Blue R-250 stained gel pieces were destained twice for 5 min in MeOH/H_2O/NH_4HCO_3 (50%:50%:100 mM), dehydrated for 5 min in ACN/ H_2O/NH_4HCO_3 (50%:50%:25 mM) then once more for 1 min in 100% ACN, dried in a Speed-Vac for 2 min, reduced in 25 mM DTT (DTT in 25 mM NH_4HCO_3) for 30 min at 56 °C, alkylated with 55 mM IAA (iodoacetamide in 25 mM NH_4HCO_3) in darkness at room temperature for 30 min, washed twice in H_2O for 30 s, equilibrated in 25 mM NH_4HCO_3 for 1 min, dehydrated for 5 min in ACN/H_2O/NH_4HCO_3 (50%:50%:25 mM) then once more for 30 s in 100% ACN, dried again and rehydrated with 20 µl of trypsin solution (10 ng $µl^{-1}$ trypsin gold (Promega) in 25 mM NH_4HCO_3/0.01% proteaseMAX wt/vol (Promega)). An additional 30 µl of digestion solution (25 mM NH_4HCO_3/0.01% ProteaseMAX wt/vol, Promega) was added to facilitate complete rehydration and excess overlay needed for peptide extraction. The digestion was conducted for 3 h at 42 °C. Peptides generated from digestion were transferred to a new tube and acidified with 2.5% trifluoroacetic acid (TFA) to 0.3% final. Degraded proteaseMAX was removed via centrifugation (max speed, 10 min) and the peptides solid-phase-extracted (ZipTip C18 pipette tips, Millipore).

Peptides were analysed by nanoLC–MS/MS using the Agilent 1100 nanoflow system connected to a new-generation hybrid linear ion trap–orbitrap mass spectrometer (LTQ-Orbitrap Elite, Thermo Fisher Scientific) equipped with an EASY-Spray electrospray source. Chromatography of peptides before MS analysis was accomplished using a capillary emitter column (PepMap C18, 3 µM, 100 Å, 150 × 0.075 mm, Thermo Fisher Scientific) onto which 2 µl of extracted peptides was automatically loaded. A NanoHPLC system delivered solvents—A: 0.1% (vol/vol) formic acid; B: 99.9% (vol/vol) acetonitrile, 0.1% (vol/vol) formic acid at 0.50 µl min^{-1} to load the peptides (over a 30 min period) and 0.3 µl min^{-1} to elute peptides directly into the nano-electrospray with gradual gradient from 3% (vol/vol) B to 30% (vol/vol) B over 77 min and concluded with 5 min fast gradient from 30% (vol/vol) B to 50% (vol/vol) B, at which time a 5 min flashout from 50 to 95% (vol/vol) B took place. As peptides eluted from the HPLC column/electrospray source, survey MS scans were acquired in the Orbitrap with a resolution of 120,000 followed by MS2 fragmentation of the 20 most intense peptides detected in the MS1 scan from 300 to 2,000 *m/z*. Redundancy was limited by dynamic exclusion.

Raw MS/MS data were converted to the mgf file format using MSConvert (ProteoWizard: Open Source Software for Rapid Proteomics Tools Development). The resulting mgf files were used to search against a *P. putida* amino-acid sequence database containing a list of common contaminants (5,388 total entries) using an in-house Mascot search engine 2.2.07 (Matrix Science) with variable methionine oxidation with asparagine and glutamine deamidation plus fixed cysteine carbamidomethylation. Peptide mass tolerance was set at 15 ppm and fragment mass at 0.6 Da.

Identification of organisms with potential homologous LA catabolism pathways. Possible LvaABCD homologues were identified by performing a BLAST search of each protein sequence against the NCBI non-redundant protein sequence database using the BioPython library (python code provided in Supplementary File 'Plasmids and Jupyter notebook')[23,59]. From the search results, the organism name was extracted from the sequence title and added to a set for each protein. The list of organisms containing the full set of LvaABCD enzymes was found by determining the intersection of the four sets of organism names from the BLAST results from each protein. A similar list was found for those organisms containing only LvaACD homologues. These lists were then used to query the original search results and find lists of proteins that have homology to proteins in the Lva pathway.

Genome sequencing and analysis. DNA was isolated from *E. coli* strains using the Wizard Genomic DNA Purification Kit (Promega) and sequenced by the University of Wisconsin Biotechnology Center. A paired-end library (2 × 250) was run on an Illumina Hi-Seq. Sequencing reads (as FASTQ files) of *E. coli* mutants were mapped to the sequenced reference genome *E. coli* LS5218 (GCA_002007165.1) using Bowtie2 with the 'fast-local' setting[60]. The output sequence alignment map (SAM) file was converted to a binary alignment map (BAM) file and sorted using SAMtools[61]. Variants were then called using Naïve Variant Caller (Galaxy open source bioinformatics tool)[62].

Directed evolution of *E. coli* LS5218. Subculturing experiments were performed with a volume of 5 ml in glass test tubes (20 × 150mm, Fisher Scientific) with 250 r.p.m. agitation in an I26 shaker (New Brunswick Scientific). The starting medium contained 20 mM LA and 40 mM acetate or 40 mM acetate only for negative

control. Cultures were grown for 72 h and OD measurements taken with a Spectronic 20 (Milton Roy Company), then culture were diluted 1:100 into fresh media. Once the OD_{600} in the LA and acetate cultures exceeded the OD_{600} of the acetate only cultures, further growth medium was 20 mM LA only. These cultures were incubated until turbidity was observed visually, then diluted 1:100 into fresh medium. This occurred for a total of 14 dilutions steps in LA medium, spanning two weeks.

Plasmids were prepped (QIAprep Miniprep Kits, Qiagen) and sequenced (Functional Biosciences) to find mutations. Plasmids were cured from mutated strains through serial culturing in rich medium (LB broth) and patch-plated on LB and LB_{kan50}.

Genome engineering with CRISPR–Cas9. CRISPR–Cas recombineering was performed following an adapted protocol from refs [63,64].

Data availability. All data from the *P. putida* transposon sequencing experiments are available through the fitness browser at http://fit.genomics.lbl.gov/cgi-bin/exps.cgi?orgId=Putida&expGroup=carbon%20source. All plasmid sequences are provided in the Supplementary Files. *P. putida* genomic DNA sequences were retrieved from the NCBI database, with the following designations: PP_2791, *lvaA*; PP_2792, *lvaB*; PP_2793, *lvaC*; PP_2794, *lvaD*; PP_2795, *lvaE*; PP_2790, *lvaR*. The *E. coli* LS5218 reference genome can be retrieved from the NCBI database under accession no. GCA_002007165.1. The python code used to identify potential *lva* homologues is provided in the Supplementary Files. Any additional data that support the findings of this study are available from the corresponding author upon request.

Received: 24 April 2017; Accepted: 16 August 2017;
Published online: 25 September 2017

References

1. Pileidis, F. D. & Titirici, M. M. Levulinic acid biorefineries: new challenges for efficient utilization of biomass. *ChemSusChem* **9**, 562–582 (2016).
2. Bozell, J. J. et al. Production of levulinic acid and use as a platform chemical for derived products. *Resour. Conserv. Recycl.* **28**, 227–239 (2000).
3. Werpy, T. & Petersen, G. *Top Value Added Chemicals from Biomass* (US Department of Energy, 2004); http://www.pnl.gov/main/publications/external/technical_reports/PNNL-14808.pdf.
4. Lange, J. P. et al. Valeric biofuels: a platform of cellulosic transportation fuels. *Angew. Chem. Int. Ed.* **49**, 4479–4483 (2010).
5. Alonso, D. M., Wettstein, S. G. & Dumesic, J. A. Gamma-valerolactone, a sustainable platform molecule derived from lignocellulosic biomass. *Green Chem.* **15**, 584–595 (2013).
6. Joshi, H., Moser, B. R., Toler, J., Smith, W. F. & Walker, T. Ethyl levulinate: a potential bio-based diluent for biodiesel which improves cold flow properties. *Biomass Bioenergy* **35**, 3262–3266 (2011).
7. Zhang, Z., Dong, K. & Zhao, Z. Efficient conversion of furfuryl alcohol into alkyl levulinates catalyzed by an organic-inorganic hybrid solid acid catalyst. *ChemSusChem* **4**, 112–118 (2011).
8. Demolis, A., Essayem, N. & Rataboul, F. Synthesis and applications of alkyl levulinates. *ACS Sustain. Chem. Eng.* **2**, 1338–1352 (2014).
9. Moore, J. A. & Tannahill, T. Homo- and co-polycarbonates and blends derived from diphenolic acid. *High Perform. Polym.* **13**, 305–316 (2001).
10. Guo, Y., Li, K., Yu, X. & Clark, J. H. Mesoporous H3PW12O40-silica composite: efficient and reusable solid acid catalyst for the synthesis of diphenolic acid from levulinic acid. *Appl. Catal. B* **81**, 182–191 (2008).
11. Chung, S. H., Choi, G. G., Kim, H. W. & Rhee, Y. H. Effect of levulinic acid on the production of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) by *Ralstonia eutropha* KHB-8862. *Society* **39**, 79–82 (2001).
12. Berezina, N. & Yada, B. Improvement of the poly(3-hydroxybutyrate-co-3-hydroxyvalerate) (PHBV) production by dual feeding with levulinic acid and sodium propionate in *Cupriavidus necator*. *Nat. Biotechnol.* **33**, 231–236 (2016).
13. Jang, J. H. & Rogers, P. L. Effect of levulinic acid on cell growth and poly-beta-hydroxyalkanoate production by *Alcaligenes* sp SH-69. *J. Chem. Inf. Model.* **18**, 219–224 (1996).
14. Habe, H. et al. Bacterial production of short-chain organic acids and trehalose from levulinic acid: a potential cellulose-derived building block as a feedstock for microbial production. *Bioresour. Technol.* **177**, 381–386 (2015).
15. Martin, C. H., Wu, D., Prather, K. L. J. & Jones Prather, K. L. Integrated bioprocessing for the pH-dependent production of 4-valerolactone from levulinate in *Pseudomonas putida* KT2440. *Appl. Environ. Microbiol.* **76**, 417–424 (2010).
16. Yeon, Y. J., Park, H. Y. & Yoo, Y. J. Enzymatic reduction of levulinic acid by engineering the substrate specificity of 3-hydroxybutyrate dehydrogenase. *Bioresour. Technol.* **134**, 377–380 (2013).
17. Jaremko, M. & Yu, J. The initial metabolic conversion of levulinic acid in *Cupriavidus necator*. *J. Biotechnol.* **155**, 293–298 (2011).
18. Martin, C. H. & Prather, K. L. J. High-titer production of monomeric hydroxyvalerates from levulinic acid in *Pseudomonas putida*. *J. Biotechnol.* **139**, 61–67 (2009).
19. Zhang, G. F. et al. Catabolism of 4-hydroxyacids and 4-hydroxynonenal via 4-hydroxy-4-phosphoacyl-CoAs. *J. Biol. Chem.* **284**, 33521–33534 (2009).
20. Harris, S. R. et al. Metabolism of levulinate in perfused rat livers and live rats: conversion to the drug of abuse 4-hydroxypentanoate. *J. Biol. Chem.* **286**, 5895–5904 (2011).
21. Martínez-García, E., Calles, B., Arévalo-Rodríguez, M. & de Lorenzo, V. pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiol.* **11**, 38 (2011).
22. Wetmore, K. M. M. et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio* **6**, e00306-15 (2015).
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
24. Schramm, G., Bruchhaus, I. & Roeder, T. A simple and reliable 5′-RACE approach. *Nucleic Acids Res.* **28**, E96 (2000).
25. Espah Borujeni, A., Channarasappa, A. S. S. & Salis, H. M. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* **42**, 2646–2659 (2014).
26. Barrios, H., Valderrama, B. & Morett, E. Compilation and analysis of σ^{54}-dependent promoter sequences. *Nucleic Acids Res.* **27**, 4305–4313 (1999).
27. Fox, J. D., Routzahn, K. M., Bucher, M. H. & Waugh, D. S. Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. *FEBS Lett.* **537**, 53–57 (2003).
28. Striebel, F. et al. Bacterial ubiquitin-like modifier Pup is deamidated and conjugated to substrates by distinct but homologous enzymes. *Nat. Struct. Mol. Biol.* **16**, 647–651 (2009).
29. Yamamoto, S. & Kutsukake, K. FliT acts as an anti-$FlhD_2C_2$ factor in the transcriptional control of the flagellar regulon in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* **188**, 6703–6708 (2006).
30. Dijkman, W. P. & Fraaije, M. W. Discovery and characterization of a 5-hydroxymethylfurfural oxidase from *Methylovorus* sp. strain MP688. *Appl. Environ. Microbiol.* **80**, 1082–1090 (2014).
31. Simons, R. W., Egan, P. A., Chute, H. T. & Nunn, W. D. Regulation of fatty acid degradation in *Escherichia coli*: isolation and characterization of strains bearing insertion and temperature-sensitive mutations in gene *fadR*. **142**, 621–632 (1980).
32. Brock, M., Maerker, C., Schutz, A., Volker, U. & Buckel, W. Oxidation of propionate to pyruvate in *Escherichia coli*. Involvement of methylcitrate dehydratase and aconitase. *Eur. J. Biochem.* **269**, 6184–6194 (2002).
33. Man, W. J., Li, Y., O'Connor, C. D. & Wilton, D. C. The binding of propionyl-CoA and carboxymethyl-CoA to *Escherichia coli* citrate synthase. *Biochim. Biophys. Acta* **1250**, 69–75 (1995).
34. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
35. Jiang, Y. et al. Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).
36. Li, Y. et al. Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 meditated genome editing. *Metab. Eng.* **31**, 13–21 (2015).
37. Moreno, R., Marzi, S., Romby, P. & Rojo, F. The Crc global regulator binds to an unpaired A-rich motif at the *Pseudomonas putida* alkS mRNA coding sequence and inhibits translation initiation. *Nucleic Acids Res.* **37**, 7678–7690 (2009).
38. Moreno, R., Martínez-Gomariz, M., Yuste, L., Gil, C. & Rojo, F. The *Pseudomonas putida* Crc global regulator controls the hierarchical assimilation of amino acids in a complete medium: evidence from proteomic and genomic analyses. *Proteomics* **9**, 2910–2928 (2009).
39. Storz, G., Wolf, Y. I. & Ramamurthi, K. S. Small proteins can no longer be ignored. *Annu. Rev. Biochem.* **83**, 753–777 (2014).
40. Su, M., Ling, Y., Yu, J., Wu, J. & Xiao, J. Small proteins: untapped area of potential biological importance. *Front. Genet.* **4**, 286 (2013).
41. Felnagle, E. A. et al. MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry* **49**, 8815–8817 (2010).
42. Baltz, R. H. Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol.* **38**, 1747–1760 (2011).
43. Gräwert, T. et al. Structure of active IspH enzyme from *Escherichia coli* provides mechanistic insights into substrate reduction. *Angew. Chem. Int. Ed.* **48**, 5756–5759 (2009).
44. Hecht, S. et al. Studies on the nonmevalonate pathway to terpenes: the role of the GcpE (IspG) protein. *Proc. Natl Acad. Sci. USA* **98**, 14837–14842 (2001).
45. Jenkins, L. S. & Nunn, W. D. Genetic and molecular characterization of the genes involved in short-chain fatty acid degradation in *Escherichia coli*: the *ato* system. *J. Bacteriol.* **169**, 42–52 (1987).

46. Matta, M. K., Lioliou, E. E., Panagiotidis, C. H., Kyriakidis, D. A. & Panagiotidis, C. A. Interactions of the antizyme AtoC with regulatory elements of the *Escherichia coli atoDAEB* operon. *J. Bacteriol.* **189**, 6324–6332 (2007).
47. Campbell, J. W. & Cronan, J. E. J. The enigmatic *Escherichia coli fadE* gene is *yafH*. *J. Bacteriol.* **184**, 3759–3764 (2002).
48. Daley, D. O. et al. Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* **308**, 1321–1323 (2005).
49. Díaz-Mejía, J. J., Babu, M. & Emili, A. Computational and experimental approaches to chart the *Escherichia coli* cell-envelope-associated proteome and interactome. *FEMS Microbiol. Rev.* **33**, 66–97 (2009).
50. Jenkins, L. S. & Nunn, W. D. Regulation of the *ato* operon by the *atoC* gene in *Escherichia coli*. *J. Bacteriol.* **169**, 2096–2102 (1987).
51. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. & Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
52. Sambrook, J., Russell, D. W. & David, W. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 2001).
53. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture medium for enterobacteria. *J. Bacteriol.* **119**, 736–747 (1974).
54. Wetmore, K. M. et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio* **6**, e00306–15 (2015).
55. Dehal, P. S. et al. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res.* **38**, 396–400 (2009).
56. Pinto, F. L., Thapper, A., Sontheim, W. & Lindblad, P. Analysis of current and alternative phenol based RNA extraction methodologies for cyanobacteria. *BMC Mol. Biol.* **10**, 79 (2009).
57. Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G. & Pühler, A. Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* **145**, 69–73 (1994).
58. Graf, N. & Altenbuchner, J. Development of a method for markerless gene deletion in *Pseudomonas putida*. *Appl. Environ. Microbiol.* **77**, 5549–5552 (2011).
59. Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
60. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
61. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
62. Goto, H. et al. Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. *Genome Biol.* **12**, R59 (2011).
63. Li, Y. et al. Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 meditated genome editing. *Metab. Eng.* **31**, 13–21 (2015).
64. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).

Acknowledgements

Work in the Pfleger laboratory was funded by the National Science Foundation (CBET-114678) and the William F. Vilas Trust. Work in the Deutschbauer and Arkin laboratories was funded by ENIGMA, a Scientific Focus Area Program, supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research and Genomics: GTLFoundational Science through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. Work in the Amador-Noguez laboratory was funded by the HHMI International Student Research Fellowship. R.L.C. was supported by the NIH NHGRI Genomic Sciences Training Program (T32 HG002760). A.L.M. was supported by an NSF SEES fellowship (GEO-1215871). J.M.R. was supported by an NSF Graduate Research Fellowship (DGE-1256259).
The authors thank J. Escalante for providing plasmid pK18*mobsacB* and J. Altenbuchner for providing strain *P. putida* KTU and plasmid pJOE6261.2. The authors acknowledge the Mass Spectrometry/Proteomics Facility at the UW–Madison Biotechnology Center for performing the in-gel digest and providing the LC–MS/MS results, and the UW–Madison Biotechnology Center DNA Sequencing Facility for providing genomic sequencing services. The authors also thank G. Gordon for help with the analysis of the *E. coli* genomic sequencing single nucleotide polymorphisms (SNPs).

Author contributions

J.M.R., D.E.A. and B.F.P. conceived the study. J.M.R. designed and performed the experiments and analysed the data, with the following exceptions. T.P. and D.A.-N. designed the LC–MS/MS experiments and T.P. performed the LC–MS and LC–MS/MS experiments. D.E.A. and J.M.T. performed the transposon library screen. C.E.C. assisted with the promoter and CoA ligase assay. A.L.M. proposed, and helped design and perform, the pulldown experiment. Y.S. and J.R. prepared the RB-TnSeq mutant library of *P. putida* KT2440 (Putida_ML5). K.M.W., R.L.C., J.R. and A.M.D. performed the fitness assays with the Putida_ML5 library. M.N.P. performed the data analysis to determine fitness values. R.L.C. prepared the supplementary analysis of the Putida_ML5 fitness experiments. A.M.D. and A.P.A. managed the Bar-Seq experiments. C.R.M helped design and analyse the prevalence of the *lva* operon in other organisms. J.M.R. and B.F.P. wrote the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at doi:10.1038/s41564-017-0028-z.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to B.F.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Pfleger

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined. | we performed triplicate biological experiments.

2. Data exclusions

Describe any data exclusions. | N/a

3. Replication

Describe whether the experimental findings were reliably reproduced. | each growth study was performed at least twice. Enzymatic studies were performed at least twice.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups. | n/a

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis. | n/a

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ A statement indicating how many times each experiment was replicated
☒	☐ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☒	☐ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☒	☐ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this | We used excel or instrument software to collect and analyze most experiments.

study. | The code used to collect data for the bioinformatics analysis is supplied as a Jupyter notebook

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company. | n/a - plasmids will be made available through Addgene. mutant strains will be maintained in the Pfleger lab for distribution if needed.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species). | n/a

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used. | n/a

b. Describe the method of cell line authentication used. | n/a

c. Report whether the cell lines were tested for mycoplasma contamination. | n/a

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. | n/a

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study. | n/a

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants. | n/a