bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Massively parallel fitness profiling reveals multiple novel enzymes in *Pseudomonas putida* lysine metabolism

Mitchell G. Thompson[1,2,3], Jacquelyn M. Blake-Hedges[1,2,4], Pablo Cruz-Morales[1,2], Jesus F. Barajas[2,5], Samuel C. Curran[1,2,6], Nicholas C. Harris[3], Veronica T. Benites[1,2], Jennifer W. Gin[1,2], Christopher B. Eiben[1,2,7], William A. Sharpless[1,2,4], Rohith N. Krishna[1,2,5], Edward E. K. Baidoo[1,2], Christopher J. Petzold[1,2], Adam P. Arkin[7,8], Adam M. Deutschbauer[2,3], Jay D. Keasling[1,2,7,9,10]

[1]Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, CA 94608, USA.

[2]Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

[3]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

[4]Department of Chemistry, University of California, Berkeley, CA 94720, USA

[5]Department of Energy Agile BioFoundry, Emeryville, CA 94608, USA

[6]Department of Comparative Biochemistry, University of California, Berkeley, CA 94720, USA

[7]Joint Program in Bioengineering, University of California, Berkeley/San Francisco, CA 94720, USA

[8]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

[9]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 94720, USA

[10]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Denmark

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

The lysine metabolism of *Pseudomonas putida* can produce multiple important commodity chemicals and is implicated in rhizosphere colonization. However, despite intensive study, the biochemical and genetic links between lysine metabolism and central metabolism remain unresolved in *P. putida*. Here, we leverage Random Barcode Transposon Sequencing (RB-TnSeq), a genome-wide assay measuring the fitness of thousands of genes in parallel, to identify multiple novel enzymes in both L- and D-lysine metabolism. We first describe three pathway enzymes that catabolize 2-aminoadipate (2AA) to 2-ketoglutarate (2KG) connecting D-lysine to the TCA cycle. One of these enzymes, PP_5260, contains a DUF1338 domain, a family without a previously described biological function. We demonstrate PP_5260 converts 2-oxoadipate (2OA) to 2-hydroxyglutarate (2HG), a novel biochemical reaction. We expand on recent work showing that the glutarate hydroxylase, CsiD, can co-utilize both 2OA and 2KG as a co-substrate in the hydroxylation of glutarate. Finally we demonstrate that the cellular abundance of D- and L-lysine pathway proteins are highly sensitive to pathway specific substrates. This work demonstrates the utility of RB-TnSeq for discovering novel metabolic pathways in even well-studied bacteria.

Importance

- *P. putida* is an attractive host for metabolic engineering as its lysine metabolism can be utilized for the production of multiple important commodity chemicals.
- We demonstrate the first biochemical evidence of a bacterial 2OA catabolic pathway to central metabolites.

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

- DUF1338 proteins are widely dispersed across many kingdoms of life. Here we demonstrate the first biochemical evidence of function for a member of this protein family.

Introduction

Pseudomonas putida is a ubiquitous saprophytic soil bacterium and is a model organism for bioremediation (1). Interest in utilizing *P. putida* KT2440 as a chassis organism for metabolic engineering has recently surged due to the existence of well-established genetic tools, its innate tolerance to organic solvents, and its robust metabolism of aromatic compounds that resemble lignin hydrolysis products (2, 3). As lignin valorization remains essential for the economic feasibility of cellulosic bioproducts, a nuanced and predictable understanding of *P. putida* metabolism is highly desirable (4).

Although its aromatic metabolism has garnered much attention, the lysine metabolism of *P. putida* has also been rigorously studied for over fifty years (5). This metabolism is likely related to the organism's native environment, as plant root exudates contain lysine. Not surprisingly, lysine degradation is highly upregulated in *P. putida* colonizing the maize rhizosphere (6, 7). Additionally, a deeper understanding of lysine metabolism has biotechnological value, as it has previously been used to produce glutarate, 5-aminovalerate, as well as valerolactam in *P. putida* and in the other bacteria (8–11).

Our current understanding of lysine catbolism however, remains incomplete. In particular, the connection between lysine metabolism and central metabolism in *P. putida* is unclear and has not been fully biochemically and genetically characterized. It is known that *P. putida* can utilize lysine as both a sole carbon and nitrogen source. It employs a bifurcating pathway where the L- and D- isomers are separately metabolized (Figure S1a) (12). Growth on

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

L-lysine is dependent on the presence of a functional D-lysine pathway; however, D-lysine metabolism occurs independently of the L-lysine pathway (5, 12). The basis for this dependence remains unresolved. The L-lysine pathway was initially assumed to proceed from glutarate to acetyl-CoA via the canonical ketogenic pathway involving a glutaryl-CoA intermediate. Recently, though, an additional glucogenic pathway was discovered (Figure S1a) (10). The final steps of the D-lysine pathway are less clear. Conflicting biochemical reports suggest several potential pathways linking 2AA to central metabolism (5, 12–14). Given this confusion and the ecological and industrial importance of lysine metabolism, we sought to identify the missing steps in D-lysine metabolism using high-throughput genetics.

Random barcode transposon sequencing (RB-TnSeq) is a genome-wide approach that measures the importance of each gene to growth (or fitness) in a massively parallel assay (15) . RB-TnSeq has recently been used to identify phenotypes for thousands of previously uncharacterized genes (15, 16), including the levulinic acid degradation pathway in *P. putida* KT2440 (17). In this study, we applied RB-TnSeq to uncover multiple novel genes and regulators implicated in L- and D-lysine metabolism in *P. putida*. We first describe a three enzyme route connecting 2AA to 2KG (Figure S1B). Within this pathway, we discover that D-lysine metabolism connects to central metabolism through a 2HG intermediate which is directly produced from 2OA in a reaction catalyzed by a DUF1338-containing protein. This protein family, widely distributed across many domains of life, previously had no known function despite its pervasiveness. Subsequently, we describe how the glutarate hydroxylase, CsiD, can function as potential bridge between the D-lysine and L-lysine pathways, by utilizing 2OA as a co-substrate in the hydroxylation of glutarate. Finally, we show that the expression of all newly

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

discovered enzymes changes significantly in response to specific metabolites within the two catabolic pathways.

Results

Identification of lysine catabolism genes via RB-TnSeq

To identify genes defective in lysine catabolism in *P. putida* KT2440, an RB-TnSeq library of this bacterium (17) was grown on minimal media supplemented with either D-lysine or L-lysine as the sole carbon source. To evaluate whether D-lysine metabolism was required for the metabolism of other downstream metabolites of L-lysine, the library was also grown on 5-aminovaleric acid (5AVA). As a control, we also grew the library on glucose. Fitness was calculated as the $\log_2$ ratio of strain and gene abundance at the end of selective growth relative to initial abundance (15). Fitness profiling revealed 39 genes with significant fitness values of less than -2 for 5AVA, D-lysine, or L-lysine, and no less than -0.5 fitness for glucose (Figure 1a, Supplemental Table 1). Within this set, 10 of the 12 known lysine degradation genes were identified, with the exception of the two enzymes in the ketogenic route of glutarate degradation (PP_0158 and PP_0159, which both had significant fitness values ($t < -4$) but whose magnitude was greater than -2). Instead, we identified the recently-characterized genes involved in the glucogenic pathway (PP_2909 and PP_2910) (10). The fitness data corroborated previous work showing that a functional D-lysine pathway is required for L-lysine catabolism (5, 12). None of the known L-lysine catabolic genes showed fitness defects for growth on D-lysine, but transposon insertions in all previously-identified D-lysine genes showed negative fitness scores when grown on L-lysine (Figure 1b). No known D-lysine catabolic enzymes showed fitness defects when grown on 5AVA, suggesting that the D-lysine dependence of L-lysine catabolism may only occur for early catabolic steps (Figure 1c).

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In addition to catabolic enzymes, lysine transporters and multiple transcriptional regulators were identified in our screen (Figure 1a). Some of the transcriptional regulators were located near known catabolic or transport enzymes (PP_0384, PP_3592, and PP_3603), while others were not clustered with any obviously related genes (PP_1109, PP_2868, PP_3649, and PP_4482). Two of these regulators showed fitness defects in this screen: known global regulator *cbrA* (PP_4695), a histidine kinase sensor (defects on both L- and D-lysine), and the alternative sigma factor *rpoX* (PP_2088) only when grown on D-lysine. Conversely, there were 15 genes that when disrupted displayed fitness advantages greater than 2 on 5AVA, D-lysine, or L-lysine and less than 0.5 fitness when grown on glucose. This positive fitness value indicates that these mutations have a competitive advantage compared to other strains in the library when grown on these carbon sources. Most striking amongst these genes were the sigma factor *rpoS* and the LPS export system (PP_1778/9), which both displayed fitness benefits on all three non-glucose carbon sources when disrupted (Figure S2). Only one gene (PP_0787, a quinolinate phosphoribosyltransferase) showed fitness defects on all three non-glucose carbon sources (Figure 1c). However, disruption of PP_0787 also showed a significant fitness defect when grown on levulinic acid, suggesting it is unlikely to be uniquely important to lysine metabolism (17). Only 3 genes shared fitness defects between 5AVA and L-lysine (PP_0213, PP_0214, and PP_2910), all of which have been previously implicated in 5AVA metabolism (Figure 1c).

PP_4108 is a D-2-aminoadipic acid aminotransferase

In humans and other animals, L-lysine degradation proceeds through a 2AA intermediate, which a transaminase converts to 2OA (10, 12, 18). Yet, no such transaminase has been identified in *P. putida*. We were able to identify a candidate aminotransferase in our fitness data, PP_4108, in which gene inactivation showed a significant growth defect on D-lysine (-5.9) and a

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

relatively minor defect on L-lysine (-1.2). To corroborate our RB-TnSeq fitness data, we constructed a deletion mutant of PP_4108 that failed to grow on a mixture of D- and L-2-aminoadipate (Figure 2a). The mutant showed severe growth defects on D-lysine and a slightly increased lag time on L-lysine (Figure S3).

To further validate this hypothesis, the ΔPP_4108 strain was subjected to metabolomics analysis to monitor the accumulation of 2AA, its expected substrate, when grown in the presence of D-lysine. Intracellular 2AA was quantified after 12 hours of growth on minimal media supplemented with 10 mM each of glucose and D-lysine. The PP_4108 deletion strain showed a 6.3-fold increase (p = 0.00016) in 2AA peak area compared to WT (Figure 2b). To analyze its function, PP_4108 was expressed and purified from *E. coli* for biochemical characterization. After incubation with DL-2AA, 2KG, and pyridoxal phosphate (PLP) for 16 hours, the reaction mixture was subjected to LC-TOF analysis. The expected product, 2OA, was detected in the enzymatic reaction but not in a boiled enzyme control, confirming that PP_4108 transaminates 2AA to form 2OA (Figure 2c). As many transaminases have broad substrate specificity (19), we also probed the substrate range of PP_4108 using a colorimetric assay for glutamate, a stoichiometric product of the transamination reaction (Figure 2d). The enzyme was most active on D-2AA, and only showed 2.8% relative activity (p = 0.0057) on its enantiomer, L-2AA. No activity was observed on either lysine isomer; however, the enzyme had slight activity towards 4-aminobutyrate (GABA) (2.8% relative activity, p = 0.0057) and moderate activity on 5AVA (30.5% relative activity, p = 0.0139). These results suggest *P. putida* KT2440 metabolizes D-lysine to D-2AA, which is then converted to 2OA by the transaminase PP_4108.

<u>PP_5260 is a novel DUF1338 family enzyme that catalyzes the conversion of 2OA to 2HG</u>

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The literature remains divided on how *P. putida* incorporates 2OA into central metabolism. Early work proposed that 2OA is converted to 2KG via a 2HG intermediate (13, 20), while later results suggested a direct conversion of 2OA to glutarate (12). Based on these hypotheses, a decarboxylation of 2OA was assumed the likely next step in D-lysine catabolism. Our fitness data on either lysine isomer revealed no obvious decarboxylases or enzymes likely to contain a thiamine pyrophosphate (TPP) cofactor (commonly employed by decarboxylases). However, a gene near other D-lysine catabolic genes in the *P. putida* genome, PP_5260, showed a significant fitness defect. To further investigate the activity of this gene, we constructed a *P. putida* strain containing a targeted deletion of PP_5260. This ΔPP_5260 strain was unable to grow on either isomer of lysine (Figure 3a), validating its importance in lysine metabolism.

PP_5260 belongs to the DUF1338 protein family (http://pfam.xfam.org/family/PF07063). Although several unpublished crystal structures of DUF1338 domain containing proteins have been deposited into the PDB, their biological function remains elusive. However, these structures combined with protein sequence alignments suggest that a putative metal binding site is conserved throughout the DUF1338 family. As we hypothesized that PP_5260 might serve as the missing decarboxylase in D-lysine metabolism, we purified the enzyme for biochemical analysis. Enzymatic activity on 2OA was probed and analyzed via LC-MS. After incubation of 2OA with PP_5260, we observed a 13.1-fold (p=0.00034) reduction in the abundance of 2OA, whereas no 2OA was consumed in a boiled enzyme control (Figure 3b). Moreover, the addition of the metal-chelating reagent EDTA eliminated enzymatic activity. From these results, we concluded that PP_5260 catalyzes the next unconfirmed step of D-lysine catabolism, converting 2OA to an unknown intermediate in a metal-dependent reaction. To probe the potential decarboxylase activity of PP_5260, we used an enzyme-coupled assay to spectrophotometrically measure CO_2

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

evolution via NADH oxidation (21). 2OA incubation with increasing PP_5260 concentrations increased the NADH oxidation rate compared to substrate alone (Figure S4a), or boiled enzyme control (Figure S4b).

Based on these data, we hypothesized that PP_5260 is a decarboxylase which converts 2OA to glutarate semialdehyde. A glutarate semialdehyde intermediate would be consistent with previous hypotheses (12) and could be further metabolized by L-lysine pathway enzymes (Figure S1). However, no glutarate semialdehyde or its potential degradation products (water-derived geminal diol, methanol-derived hemiacetal) could be detected in our LC-TOF data from PP_5260 *in vitro* reactions (data not shown). As early biochemical work suggested 2HG as a potential intermediate in pipecolate metabolism (20), we also searched for masses consistent with 2HG. We were able to identify a metabolite in our PP_5260 LC-TOF data with the same retention time and mass-to-charge ratio as a DL-2HG standard (Figure 3c). Enzymatic *in vitro* reactions showed a 3-log increase in the peak area of 2HG compared to boiled controls (p = 3.42E-6) (Figure 3d). The isotopic distribution of [M-H] peaks in the mass spectra of these compounds was also identical (Figure S4c). These results confirmed that PP_5260 catalyzes the transformation of 2OA to 2HG, the first biochemical function identified for a DUF1338-containing protein.

<u>DUF1338 proteins are a widely distributed enzyme family with multiple putative functions</u>

DUF1338 proteins are widely distributed across the tree of life, with homologs of PP_5260 found in plants, fungi, and bacteria, though they were not found in animals or archaea (Figure 4a). Previous work has shown that DUF1338 proteins are ubiquitous in green plants and also found in green algae (22). Homologs are widely distributed amongst bacteria, with the Firmicutes being a notable exception. PP_5260 homologs within Streptophyta, Actinobacteria,

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cyanobacteria, and Bacteroidetes formed monophyletic clades, while homologs from other taxonomic groups were not monophyletic (Figure 4a). DUF1338 homologs are found in bacteria of biotechnological (*Corynebacterium glutamicum*), environmental (*Nostoc puncitforme*), and medical importance (*Yersinia pestis*, *Mycobacterium tuberculosis*, *Burkholderia pseudomallei*).

Publically available fitness data show that both *Pseudomonas fluorescens* FW300-N2C3 and *Sinorhizobium meliloti* PP_5260 homologs have L-lysine specific defects when interrupted (16). Genomic contexts within other bacteria suggest that many DUF1338-containing enzymes may be involved in lysine or other amino acid metabolism (Figure 4b). Within the Actinobacteria DUF1338 homologs are often found adjacent to sarcosine oxidases, aldehyde dehydrogenases, and transaminases implying an additional catabolic amino acid function. In both the oleaginous bacterium *Rhodococcus opacus* B4 and *M. tuberculosis* DUF1338 homologs are found next to predicted L-lysine aminotransferases potentially suggesting that an ancestral homolog functioned in lysine catabolism. Interestingly the *R. opacus* B4 genome has three DUF1338 homologs, only one of which contains genes predicted to be specific to lysine catabolism. The other two gene neighborhoods are similar in their functional content, mainly differing by containing an oxidoreductase or glycolate dehydrogenase, either potentially performing the same biochemical function. In Alphaproteobacteria, Betaproteobacteria, and Cyanobacteria the presence of aldehyde dehydrogenases, oxidoreductases, glycolate dehydrogenases, an aminotransferase implies a metabolic function similar to PP_5260.

However there is evidence to suggest roles for DUF1338-containing proteins outside of lysine metabolism. Neither *E. coli* nor *Klebsiella michiganensis* M5al PP_5260 homologs showed any fitness defect when grown on L-lysine as a sole carbon source (16). In fact, outside of *P. putida*, relatively few well characterized Gammaproteobaceteria showed obvious amino

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

acid metabolism connections based on genomic context (Figure 4b). The *E. coli* homolog, *ydcJ*, which is found next to a glucans biosynthesis protein has no significant phenotypes in 166 fitness experiments (16). These data suggest a divergent function within these homologs.

PP_4493 putatively oxidizes 2-hydroxyglutarate to 2-oxoglutarate and connects D-lysine and central metabolism

In glucogenic glutarate metabolism, PP_2910 oxidizes 2HG to 2KG, but this gene showed no fitness defect on D-lysine in our RB-TnSeq data (Figure S1a). However, a putative FAD-dependent and 4Fe-4S cluster-containing glycolate dehydrogenase, PP_4493, did show fitness defects on both D-lysine and L-lysine. Glycolate dehydrogenases oxidize the alcohol group of an alpha-hydroxyacid, glycolate, to the corresponding carbonyl group to form glyoxylate (Figure 5a). Therefore, we hypothesized that PP_4493 could potentially oxidize a similar 2-hydroxyacid, 2HG, to the corresponding alpha-ketoacid, 2KG. Moreover, many PP_5260 homologs were located next to or near putatively annotated glycolate dehydrogenases in other bacteria, underscoring their potential metabolic link (Figure 4b). To confirm these hypotheses, we again constructed a deletion strain, *P. putida* ΔPP_4493, which could not grow on D-lysine as a sole carbon source (Figure 5b), and showed attenuated growth on L-lysine (Figure S5). Furthermore, the mutant accumulated a 33.2-fold (p = 0.00023) higher abundance of intracellular 2HG than the wild type when grown on glucose and D-lysine (Figure 5c). These data and the conserved function and genomic context of glycolate dehydrogenases strongly suggest that PP_4493 catalyzes the last step of 2AA metabolism: the oxidation of 2HG to 2KG (Figure S1b).

CsiD is highly specific for glutarate hydroxylation but promiscuous in 2-oxoacid selectivity

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In addition to identifying the final steps of D-lysine catabolism, our RB-TnSeq data provided new insights into L-lysine metabolism, from fitness data of *P. putida* grown on 5AVA. Not surprisingly, disruption of the two enzymes from the ketogenic glutarate pathway, glutaryl-CoA ligase (PP_0159) and glutaryl-CoA dehydrogenase (PP_0158), resulted in mild fitness defects when grown on 5AVA (Figure 6a). Two additional genes showed much larger fitness defects when disrupted: PP_2910, a probable L-2HG oxidase, and PP_2909 (CsiD), a non-heme Fe(II) oxidase. We postulated potential functions for these enzymes based on biochemically-characterized homologs. Non-heme Fe(II) oxidases typically activate molecular oxygen and utilize a 2-oxoacid cosubstrate to perform hydroxylation reactions (23). The decarboxylation of 2-oxoacids is required to achieve a reactive Fe(IV)-oxo species for hydroxylation and results in the production of the corresponding diacid (usually a succinate derived from aKG). Given this, we hypothesized a cyclic reaction cascade wherein PP_2909 would hydroxylate glutarate to form 2HG using 2KG as a cosubstrate. PP_2910 would then subsequently oxidize 2HG to 2KG, regenerating the 2KG consumed in the initial reaction. Although seemingly futile, these reactions would result in the net incorporation of succinate into central metabolism (Figure S1). To test this, we purified PP_2909 and confirmed that it hydroxylates glutarate in a 2KG-dependent manner (Figure S6a). HPLC analysis demonstrated that as glutarate was consumed, equimolar quantities of succinate and 2HG were produced (Figure S6b). Additionally, a PP_2909 deletion mutant showed increased lag time when growing on either L-lysine or 5AVA. By deleting the glutaryl-CoA ligase PP_0158 (disrupting the ketogenic glutarate pathway), we were able to completely prevent growth on 5AVA or L-lysine (Figure S6c).

Because non-heme Fe(II) oxidases can be promiscuous with respect to the 2-oxoacid cosubstrate (23, 24) and since we had identified an alternative alpha ketoacid, 2OA, as an

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

intermediate in D-lysine catabolism, we next attempted to determine if this metabolite could be shunted between the two pathways. First, we evaluated the hydroxyl acceptor substrate specificity of CsiD family proteins by purifying two homologs from *Escherichia coli* and a halophilic bacterium, *Halobacillus* sp. BAB-2008 (Figure 6b). We probed the activity of the homologs against a panel of 3-6 carbon fatty acids and diacids in the presence of 2KG and found that only glutarate served as a hydroxylation substrate (Figure 6c). These results are consistent with the work recently reported by Zhang *et al (10)*, which claimed that PP_2909 acts only on a glutarate substrate. Our result suggests that the specificity of CsiD homologs is conserved across phyla. Although extremely specific for the hydroxylation substrate, all three CsiD homologs could utilize both 2OA and 2KG, but not oxaloacetate, as a cosubstrate for 2HG formation (Figure 6d). Homology modelling of the *E. coli* CsiD demonstrated that 2OA and 2KG can fit inside the active site (Figure 76). This result is particularly interesting as it provides a possible link between D-lysine and L-lysine catabolism. CsiD could convert the D-lysine catabolic product 2OA to glutarate where it could then be shunted into either the ketogenic or glucogenic glutarate pathway. Growth defects observed in a ΔPP_2909ΔPP_0158 double mutant grown on D-lysine also support this hypothesis (Figure S8).

Expression of lysine metabolic proteins is responsive to pathway metabolites

Multiple studies have demonstrated that the expression of lysine catabolic genes is upregulated in the presence of pathway metabolites (10, 14, 25). To investigate the regulation of the newly-discovered lysine catabolic enzymes from this study, wild-type *P. putida* KT2440 was grown in minimal media on glucose or a single lysine metabolite (e.g. D-lysine, L-lysine, 5AVA, 2AA, or glutarate) as a sole carbon source for 36 hours. We then quantified the relative abundance of D- and L-lysine catabolic proteins via targeted proteomics (Figure 7). For each

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

protein, all pairwise statistical comparisons of different carbon sources can be found in Supplemental Table 2. All five D-lysine pathway proteins measured (AmaA (PP_5257), AmaB (PP_5258), PP_4108, YdcJ (PP_5260), and YdiJ (PP_4493)) were upregulated when grown on L-lysine, D-lysine or 2AA compared to the glucose control. Neither 5AVA or glutarate significantly induced expression of any D-lysine proteins measured. Of all the targeted proteins, the three identified in this study that directly degrade 2AA were most strongly induced by 2AA. Somewhat surprisingly, we also found that the two enzymes involved in 2AA formation, AmaA and AmaB, were also more highly expressed in the presence of 2AA.

The initial two enzymes from L-lysine metabolism, DavA and DavB, were most highly expressed in the presence of L-lysine, but also significantly with D-lysine. As previously observed, DavT and DavD were most strongly upregulated on 5AVA, moderately upregulated on L-lysine, and to a lesser extent D-lysine. The induction of LhgO and CsiD was highest when grown on glutarate, although these proteins were also moderately upregulated by 5AVA and L-lysine. By comparison, PP_0159 (GcdG) expression in the presence of glutarate was stimulated to a lesser extent than LhgO and CsiD expression; in addition, GcdG was slightly upregulated on 5AVA and L-lysine. Taken together these results confirm that despite being found in distant loci within the genome, lysine catabolic genes are induced specifically in response to pathway metabolites.

Discussion

Though the final steps of of D-lysine metabolism have been debated in *P. putida* and other bacteria for some time, a complete biochemical and genetic understanding of this pathway has remained elusive. A 2OA degradation pathway has been extensively characterized in mammals, though, because of its implications in human disease (26). In this pathway, L-lysine

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is metabolized to 2OA and eventually converted to acetyl-CoA via a glutaryl-CoA intermediate (26). However, this pathway has not been observed in bacteria. Previous work suggested that 2OA is either converted via decarboxylation to glutarate or through several enzymatic steps to 2HG, yet none of these studies was able to conclusively demonstrate a genetic and biochemical basis for these hypotheses (12, 20, 27). Here we biochemically characterized plausible routes from 2OA directly to 2HG and 2KG that, unlike the mammalian pathway, do not include CoA intermediates. The first route, catalyzed by the DUF1338-containing metalloenzyme PP_5260, involves the direct conversion of 2OA to 2HG . This transformation seemingly involves two separate reactions: a decarboxylation and a hydroxylation. Hydroxymandalate synthase has been shown to catalyze a similar enzymatic reaction, via an intramolecular oxidative decarboxylation similar to 2KG dependent Fe(II) oxidases (28). Though PP_5260 and hydroxymandalate synthase share little sequence homology, this enzyme may give us insight into the molecular mechanism of DUF1338 proteins. We have given PP_5260 the tentative title of 2-hydroxyglutarate synthase (*hglS*) until further mechanistic studies (currently underway in our group) are completed and a more accurate enzyme name can be assigned.

In bacteria, homologs of PP_5260 appear widely distributed with their genomic contexts suggesting functions both within and beyond lysine metabolism. Genomic contexts in other bacteria, particularly *Actinobacteria*, suggest these homologs may be involved in other amino acid catabolic pathways. Unfortunately, there is scant evidence for homologous function in model organisms. For example, although DUF1338 proteins are present in many *Ascomycota*, there is no homolog in *Saccharomyces cerevisiae*. Interestingly, the *E. coli* homolog of PP_5260 is located next to a potential glucans biosynthesis gene: Glucans biosynthesis protein D (29). Another DUF1338-containing protein from rice has been characterized and was implicated in

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

starch granule formation (22). These results suggest that DUF1338 proteins could play a role in sugar polymerization.

Recently Zhang *et al*. thoroughly characterized a glucogenic glutarate catabolism route involving the Fe(II) dependent oxidase CsiD (10). Our RB-TnSeq screening convergently uncovered this pathway, and our biochemical and physiological results fully corroborate their findings. While both works show multiple CsiD homologs from divergent bacteria are highly specific towards glutarate as a hydroxyl acceptor, all three homologs we tested showed promiscuous activity toward 2-oxoacid cosubstrates. The ability of the *P. putida* CsiD to utilize 2OA as a cosubstrate is particularly interesting as it may directly connect L- and D-lysine metabolism. The promiscuity of CsiD may explain earlier studies which reported glutarate formation from D-lysine (12). Further studies involving labelled substrates may help elucidate the potential link between the two pathways. While CsiD plays a clear role in L-lysine metabolism in *P. putida*, its role in other organisms remains a mystery. In *E. coli*, RpoS controls the expression of CsiD, but *rpoS* mutants showed fitness benefits on all three lysine metabolites tested in our RB-TnSeq data (30). Furthermore, despite the high substrate specificity for glutarate displayed by all CsiD homologs tested, we are unaware of any reports that identify glutarate as a natural metabolite in *E. coli*. The implication that *E. coli* may produce glutarate naturally is intriguing given the multiple efforts to produce it heterologously in the host (31, 32). Uncovering other physiological roles of CsiD homologs may shed further light on the importance of glutarate as a metabolite in bacteria.

Given the large number of aminotransferases contained within the genome of *P. putida* KT2440 it is unsurprising that the enzyme catalyzing deamination of the 2AA had remained undiscovered, however PP_4108 appears to be critical for D-lysine and 2AA metabolism. The

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

preference of PP_4018 for D-2AA is surprising as work done in the 1970s had shown that *P. putida* strain P2 could not metabolize D-2AA, but could metabolize L-2AA (20). However, no genome sequence exists for *P. putida* P2, and thus it is difficult to say how closely related this strain is to KT2440, or how similar their lysine metabolisms may be. While PP_4108 is critical for D-lysine metabolism, it is not required for L-lysine metabolism, which sets it apart from other enzymes in the D-lysine catabolic pathway, which have all been essential for L-lysine catabolism. It may be reasonable to expect that other aminotransferases or deaminases may be able to complement the function of PP_4108 that are uniquely expressed when *P. putida* KT2440 is grown on L-lysine.

In both *E. coli* and *P. putida* KT2440 LhgO homologs are able to oxidize L-2HG to 2KG using molecular oxygen (10, 33). However, fitness profiling revealed PP_2910 to be dispensable for D-lysine metabolism. Our results demonstrated that a putative glycolate dehydrogenase, PP_4493, is required for D-lysine metabolism, and mutants accumulate 2HG when grown on glucose and D-lysine. Work in the late 1960s discovered a membrane bound enzyme in *P. putida* P2 capable of oxidizing D-2HG to 2KG (27). Active fractions in this work displayed absorbance spectra consistent with non-heme iron, as well flavin cofactor, and its activity could be disrupted with respiration poisons. PP_4493 is a 4Fe-4S iron-sulfur cluster protein, with a conserved flavin binding domain which would be consistent with the properties of the previously reported enzyme. Recent work with a 4Fe-FS D-lactate dehydrogenase in *P. putida* KT2440, suggests that quinone may be the terminal electron acceptor, though other electron acceptors are possible (34). There may be a selective advantage to utilizing quinones as electron acceptors rather than oxygen in order to conserve more energy during catabolism. Whether the enzyme characterized from early biochemical studies is a homolog of PP_4493 remains to be determined, but the well-

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

known difficulty of working with iron-sulfur proteins will make further biochemical characterization challenging (35).

Work presented here and previous reports have shown that expression of both lysine catabolism pathways are highly responsive to their respective metabolites. While this metabolism appears highly coordinated, the genes themselves are dispersed across the genome, with both PP_4018, and PP_4493 found outside of operons, and relatively distant from other lysine catabolic genes. At least two global regulators appeared to be important to lysine metabolism based on our Rb-TnSeq data, *cbrA* (PP_4695) and *rpoX* (PP_2088). The two-component system CbrAB has been implicated in catabolite repression and C/N balance in *P. aeruginosa*, with mutants unable to grown on multiple amino acids (36). Further work in *P. putida* KT2440 showed the CbrAB system behaved similarly to that in *P. aeruginosa* (37). It would be unsurprising if this regulator controls the expression of various genes within lysine catabolism; more work into uncovering the regulon is warranted. RpoX on the other hand has been implicated in osmotic tolerance in *P. aeruginosa* (38, 39). This is interesting as lysine metabolism, and specifically pipecolate metabolism, has been associated with osmotic tolerance across multiple bacteria (40). As *rpoX* only showed fitness defects on D-lysine, the metabolism that produces pipecolate, these results suggest that the D-lysine pathway of *P. putida* may be involved in adaptation to saline or other osmotically stressful environments.

An interesting question remains as to why *P. putida* maintains separate metabolic pathways for D- and L-lysine, and why L-lysine metabolism seems dependent on the presence of an intact D-lysine metabolism. Previously it has been proposed that the D-lysine may provide a way of resolving a C/N imbalance that may occur when L-lysine is metabolized; however we believe this is unlikely as both pathways contain one deamination and one transamination

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reaction (12). Our fitness results indicate that D-lysine metabolism is dispensable for growing on 5AVA. This would suggest that only the initial two steps of L-lysine metabolism, the oxidation of lysine to 5-aminopentanamide by DavB and its subsequent deamination to 5AVA by DavA are dependent on D-Lysine catabolism. We propose that the adjacent AsnC family regulator PP_0384 likely responds to L-lysine as many proteins within this family respond to amino acids including lysine (41, 42) and expression of these two enzymes is most responsive to L-lysine. To our knowledge there has been no rigorous characterization of the regulation of the *davAB* operon, nor of the biochemical activities of these two enzymes *in vitro*. Future studies to uncover the mechanistic regulation at the transcriptional and post-translational levels at these two steps may uncover the necessity of D-lysine dependency of the L-lysine catabolic pathway. Overall our work highlights the utility of global fitness profiling to discover novel, complex, metabolic pathways in even well-characterized bacteria.

Materials and Methods

Media, chemicals, and culture conditions

Routine bacterial cultures were grown in in Luria-Bertani (LB) Miller medium (BD Biosciences, USA). *E. coli* was grown at 37 °C, while *P. putida* was grown at 30 °C unless otherwise noted. When indicated, *P. putida* was grown on modified MOPS minimal medium (43). Cultures were supplemented with kanamycin (50 mg/L, Sigma Aldrich, USA), gentamicin (30 mg/L, Fisher Scientific, USA), or carbenicillin (100mg/L, Sigma Aldrich, USA), when indicated. D-2-aminoadipate was purchased from Takara Bioscience (USA), all other compounds were purchased through Sigma Aldrich.

Strains and plasmids

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

All bacterial strains and plasmids used in this work are listed in Supplemental Table 3. All strains and plasmids created in this work are available through the public instance of the JBEI registry. (https://public-registry.jbei.org/folders/391). All plasmids were designed using Device Editor and Vector Editor software, while all primers used for the construction of plasmids were designed using j5 software (44–46). Synthetic DNA coding for the *Halobacillus* sp. BAB-2008 *csiD* homolog was purchased from Integrated DNA Technologies (IDT, Coralville, IA). Plasmids were assembled via Gibson Assembly using standard protocols (47), or Golden Gate Assembly using standard protocols (48). Plasmids were routinely isolated using the Qiaprep Spin Miniprep kit (Qiagen, USA), and all primers were purchased from Integrated DNA Technologies (IDT, Coralville, IA).

Random barcode TnSeq experiments

P. putida RB-TnSeq library JBEI-1 was created by diluting a 1 mL aliquot of the previously described *P. putida* RB-TnSeq library (17) in 500mL of LB media supplemented with kanamycin which was then grown to an OD600 of 0.5 and frozen as 1 mL aliquots after adding glycerol to a final concentration of 20% v/v. Libraries were stored at −80 °C until used. A 1 mL aliquot of P. putida RB-TnSeq library JBEI-1 was thawed on ice and diluted in 25 mL of LB supplemented with kanamycin. The culture was grown until it reached an OD600 of 0.5, at which point 3 1-mL aliquots were taken, pelleted, decanted, and then stored at −80 °C to use as a time zero control. The library was then washed once in MOPS minimal medium without any carbon source, and then diluted 1:50 into 10 mL MOPS minimal medium supplemented with either 10 mM glucose, 5AVA, D-Lysine, or L-Lysine. Cells were grown in 50 mL culture tubes for 48 hours at 30 °C shaking at 200 rpm. After growth 2 ml aliquots from the culture tubes were pelleted, decanted and frozen at −80 °C for barcode sequencing. We performed DNA barcode

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sequencing (BarSeq) as previously described (15). Briefly, the fitness of a strain is the normalized log2 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. Statistic *t*-values of >|4| were considered significant. All experiments described herein pass the quality metrics described previously unless noted otherwise. All fitness data in this work is publically available at http://fit.genomics.lbl.gov.

Construction of deletion mutants

Deletion mutants in *P. putida* were constructed by homologous recombination and *sacB* counterselection using the allelic exchange vector pMQ30 (49). Briefly, homology fragments ranging from 1 kbp to 2 kbp up- and downstream of the target gene, including the start and stop codons respectively, were cloned into pMQ30. An exception to these design parameters was plasmid pMQ-PP_5260 which maintained an additional 21 nt at the 5' end of the gene in addition to the stop codon. Plasmids were then transformed via electroporation into *E. coli* S17 and then mated into *P. putida* via conjugation. Transconjugants were selected for on LB Agar plates supplemented with gentamicin 30 mg/mL, and chloramphenicol 30 mg/mL. Transconjugants were then grown overnight on LB media also supplemented with 30 mg/mL gentamicin, and 30 mg/mL chloramphenicol, and then plated on LB Agar with no NaCl supplemented with 10% w/v sucrose. Putative deletions were restreaked on LB Agar with no NaCl supplemented with 10% w/v sucrose, and then were screened via PCR with primers flanking the target gene to confirm gene deletion.

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Expression and purification of proteins

A 5 mL overnight culture of *E. coli* BL21 (DE3) containing the expression plasmid was used to inoculate a 500 mL culture of LB. Cells were grown at 37 ºC to an OD of 0.6 then induced with Isopropyl β-D-1-thiogalactopyranoside to a final concentration of 1 mM. The temperature was lowered to 30 ºC and cells were allowed to express for 18 hours before being harvested via centrifugation. Cell pellets were stored at -80 ºC until purification. For purification, cell pellets were resuspended in lysis buffer (50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole, 8% glycerol, pH 7.5) and sonicated to lyse cells. Insolubles were pelleted via centrifugation (30 minutes at 40,000xg). The supernatant was applied to a fritted column containing Ni-NTA resin (Qiagen, USA) which had been pre-equilibrated with several column volumes of lysis buffer. The resin was washed with lysis buffer containing 50 mM imidazole, then the protein was eluted using a step-wise gradient of lysis buffer containing increasing imidazole concentrations (100 mM, 200 mM, and 400 mM). Fractions were collected and analyzed via SDS-PAGE. Purified protein was dialyzed overnight at 4 ˚C against 50 mM HEPES pH 7.5, 5% glycerol.

For PP_5260, further purification of the protein was conducted by anion exchange chromatography using HiTrap Q FF (GE Healthcare) according to the manufacturer's instructions. Purified protein was dialyzed against minimal buffer, which consisted of 25 mM HEPES pH 7.5, 50 mM NaCl. Purified proteins were concentrated using Spin-X UF 20 (10 kDa MWCO) spin concentrators (Corning, Inc.). Concentrated protein was stored at 4 ºC until *in vitro* analysis.

CsiD *in vitro* assays

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The activity of purified CsiD homologs was analyzed in 100 µL reaction mixtures containing 50 mM HEPES (pH 7), 5 mM glutarate, 5 mM α-ketoglutarate, 25 µM $FeSO_4$, 0.1 mM ascorbate, and 0.5 mM DTT. For negative control reactions, each respective reaction component was omitted. To initiate reactions, CsiD was added to a final concentration of 7 µM. For the no enzyme control, CsiD was denatured at 98 ºC for 10 minutes prior to addition to the reaction mix. Reactions were allowed to proceed at 22 ºC for 3 hours. Products were analyzed via LC-MS method 3 after quenching via the addition of acetonitrile and methanol for a final ACN:H_2O:MeOH ratio of 6:3:1 To analyze products from substrate range as well 2-oxoacid specificity experiments, reactions were measured via LC-MS method 1.

Transamination assays

To determine product formation via PP_4108, assays were conducted in 50 mM HEPES (pH 7.5), with 5 mM 2-oxoglutarate, 0.1 mM PLP, and 5 mM of substrate, and 10 uM of purified enzyme or boiled enzyme control in 100uL volumes. Reactions were incubated at 30 ºC for 16 hours and quenched via the addition of ACN and MeOH for a final ACN:H_2O:MeOH ratio of 6:3:1 for LC-MS method 3. To determine substrate specificity reactions were set up at 75 µL scale and carried out at 30ºC for 16 hours before freezing. For analysis, reactions were diluted 15-fold in water and assessed by a colorimetric assay for glutamate (Sigma MAK004) in 96-well format via a SpectraMax M4 plate reader (Molecular Devices, USA).

PP_5260 *in vitro* assays

The activity of PP_5260 was assessed in 50 mM HEPES, with 5 mM 2OA as substrate and 10 uM purified enzyme or boiled enzyme control. Reactions were incubated for 16 hours at 30 ºC. To test the necessity of metal cofactors EDTA was added to a final concentration of 50uM. Reactions and quenched via the addition of ACN and methanol MeOH for a final

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ACN:H_2O:MeOH ratio of 6:3:1 for LC-MS analysis method 3, or with an equal volume of ice-cold methanol for HPLC analysis and LC-MS method 2.

Enzyme coupled decarboxylation assays were carried out as previously described (21). Reaction mixtures contained 100 mM Tris-HCl (pH 7), 10 mM $MgCl_2$, 0.4 mM NADH, 4 mM phosphoenol pyruvate (PEP), 100U/mL pig heart malate dehydrogenase(Roche), 2U/mL microbial PEP carboxylase (Sigma), and 10 mM 2OA. Reactions were initiated by the addition of purified PP_5260 or boiled enzyme controls, and absorbance at 340 nm was measured via a SpectraMax M4 plate reader (Molecular Devices, USA).

HPLC analysis

HPLC analysis was performed on an Agilent Technologies 1200 series liquid chromatography instrument coupled to a refractive index detector (35°C, Agilent Technologies, Santa Clara, CA). Samples were injected onto an Aminex HPX-87H Ion Exclusion Column (300 x 7.8 mm, 60°C, Bio-Rad, Hercules, CA) and eluted isocratically with 4 mM H_2SO_4 at 600 uL/min for 20 minutes. Compounds were quantified via comparison to a calibration curve prepared with authentic standards and normalized to injection volume.

Proteomics analysis

Protein lysis and precipitation were achieved by using a chloroform-methanol extraction as previously described (50). Thawed pellets were loosened from 14 mL falcon tubes and transferred to PCR 8-well tube strip, followed by the addition of 80 µL of methanol, 20 µL of chloroform, and 60 µL of water, with vortexing. The samples were centrifuged at 20,817 x g for 1 minute for phase separation. The methanol and water (top) layer was removed, then 100 µL of methanol was added and the sample was vortexed briefly. The samples were centrifuged at 20,817 x g for 1 minute to isolate the protein pellet. The protein pellet was air-dried for 10

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

minutes and resuspended in 100 mM ammonium bicarbonate with 20% methanol. The protein concentration was measured using the DC Protein Assay Kit (Bio-Rad, Hercules, CA) with bovine serum albumin for the standard curve. A total of 100 µg of protein from each sample was digested with trypsin for targeted proteomic analysis. The protein was reduced by adding tris 2-(carboxyethyl) phosphine (TCEP) at a final concentration of 5 mM, alkylated by adding iodoacetamide at a final concentration of 10 mM, and digested overnight at 37 ºC with trypsin at a ratio of 1:50 (w/w) trypsin:total protein. As previously described (51), peptides were analyzed using an Agilent 1290 liquid chromatography system coupled to an Agilent 6460QQQ mass spectrometer (Agilent Technologies, Santa Clara, CA). Peptide samples (10 µg) were separated on an Ascentis Express Peptide ES-C18 column (2.7 µm particle size, 160 Å pore size, 50 mm length x 2.1 mm i.d., 60 ºC; Sigma-Aldrich, St. Louis, MO) by using a chromatographic gradient (400 µL/min flow rate) with an initial condition of 95% Buffer A (99.9% water, 0.1% formic acid) and 5% Buffer B (99.9% acetonitrile, 0.1% formic acid) then increasing linearly to 65% Buffer A/35% Buffer B over 5.5 minutes. Buffer B was then increased to 80% over 0.3 minutes and held at 80% for two minutes followed by ramping back down to 5% Buffer B over 0.5 minutes where it was held for 1.5 minutes to re-equilibrate the column for the next sample. The peptides were ionized by an Agilent Jet Stream ESI source operating in positive-ion mode with the following source parameters: Gas Temperature: 250 ºC, Gas Flow: 13 L/min, Nebulizer Pressure: 35 psi, Sheath Gas Temperature: 250 ºC, Sheath Gas Flow: 11 L/min, Nozzle Voltage: 0 V, Chamber Voltage: 3,500 V. The data were acquired using Agilent MassHunter, version B.08.02, processed using Skyline (52) version 4.1, and peak quantification was refined with mProphet (53) in Skyline. Data are available at Panorama Public via this link: https://panoramaweb.org/massive_fitness_profiling_Pseudomonas_putida.url. All pairwise

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

combinations of spectral counts from carbon sources for each protein were compared via Student's t-test followed by a Bonferroni correction.

Detection of metabolites via LC-MS

Sampling of intracellular metabolites was conducted as described previously (54). Multiple methods were used to detect compounds in this work. Method (1) HILIC-HRMS analysis was performed using an Agilent Technologies 6510 Accurate-Mass Q-TOF LC-MS instrument using positive mode and an Atlantis HILIC Silica 5 µm column (150 x 4.6 mm) with a linear of 95 to 50% acetonitrile (v/v) over 8 min in water with 40 mM Ammonium formate, pH 4.5, at a flow rate of 1 mL min−1 . Method (2) HILIC-HRMS analysis was performed using an Agilent Technologies 6510 Accurate-Mass Q-TOF LC-MS instrument using negative mode and an Atlantis HILIC Silica 5 µm column (150 x 4.6 mm) with an isocratic mobile phase (80% acetonitrile (v/v) with 40 mM Ammonium formate, pH 4.5) for 20 min at a flow rate of 1 mL min−1 . Method (3) is described in George et al (54). Briefly, samples were separated via a SeQuantZIC-pHILIC guard column (20-mm length, 2.1-mm internal diameter, and 5-µm particle size; from EMD Millipore, Billerica, MA, USA), then with a short SeQuantZIC-pHILIC column (50-mm length, 2.1-mm internal diameter, and 5-µm particle size) followed by a long SeQuantZIC-pHILIC column (150-mm length, 2.1-mm internal diameter, and 5-µm particle size) using an Agilent Technologies 1200 Series Rapid Resolution HPLC system (Agilent Technologies, Santa Clara, CA, USA). The mobile phase was composed of 10 mM ammonium carbonate and 118.4 mM ammonium hydroxide in acetonitrile/water (60.2:39.8, v/v). Metabolites were eluted isocratically via a flow rate of 0.18 mL/min from 0 to 5.4 min, which was increased to 0.27 mL/min from 5.4 to 5.7 min, and held at this flow rate for an additional 5.4

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

min. The HPLC system was coupled to an Agilent Technologies 6210 TOF-MS system in negative mode.

Homology modelling

The CsiD homology model was generated using the structure prediction software I-TASSER (https://zhanglab.ccmb.med.umich.edu/) (55–57) The internal C-score scoring function in I-TASSER was utilized to select the top scoring homology model. The CsiD model was energy minimized using CHIMERA (58). All ligand coordinates were generated using eLBOW from the Phenix suite (59). Ligands were docked using the Swissdock web server to predict the molecular interactions (60, 61) and subsequently energy minimized using CHIMERA. All figures related to the CsiD homology model and docking were visualized and generated by PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.).

Phylogenomic analyses

Amino acid sequences of CsiD homologs were downloaded from the pFAM database and aligned with MAFFT-linsi (62). Phylogenetic trees of CsiD alignments were constructed with FastTree 2, and trees were visualized on iTOL (63, 64).

Representative DUF1338 sequences were obtained from pFAM (https://pfam.xfam.org/family/PF07063#tabview=tab3). All genomes analyzed were downloaded from the NCBI FTP site and annotated using RAST (65). Amino acid sequences of DUF1338 proteins from these genomes were retrieved using BlastP with a bit score cutoff of 150 and an E-value of 0.000001. All sequences alignments were performed using Muscle v3.8 (66) and the alignments were manually curated using Jalview V2 (67).

For the phylogenetic reconstructions, the best amino acid substitution model was selected using ModelFinder implemented on IQ-tree (68) the phylogenies were obtained using IQ-tree v

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1.6.7 (69), with 10,000 bootstrap replicates. The final trees were visualized and annotated using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). Genome neighborhoods of DUF1338 were obtained using CORASON-BGC (70) and manually colored and annotated.

Statistical analyses and data presentation

All numerical data were analyzed using custom Python scripts. All graphs were visualized using either Seaborn or Matplotlib (71, 72). Calculation of 95% confidence intervals, standard deviations, and T-test statistics were conducted via the Scipy library (73). Bonferroni corrections were calculated using the MNE python library (74).

Acknowledgements

This manuscript is dedicated to the memory of Cornell Professor Dr. Eugene Madsen. The authors would like to thank Morgan Price, Dr. Jamie Meadows, Dr. Robert Haushalter, Dr. Bo Pang, Dr. Nick Weathersby, Mary Thompson, and Catharine Adams for their helpful discussions in preparing this manuscript. We would also like to thank the UC Berkeley SMART program for providing support for R.K to conduct summer research. This work was part of the DOE Joint BioEnergy Institute (https://www.jbei.org) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, and protein purification and homology modelling components were part of the Agile BioFoundry (http://agilebiofoundry.org) supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Contributions

Conceptualization, M.G.T., and J.M.B.; Methodology, M.G.T., J.M.B, J.F.B., P.C.M., S.C.C., N.C.H, C.B.E, E.E.K.B, C.J.P., and A.M.D.; Investigation, M.G.T., J.M.B, W.N.S., R.A.K, J.F.B., V.T.B, P.C.M., J.W.G, C.J.P, N.C.H., E.E.K.B.; Writing – Original Draft, M.G.T.; Writing – Review and Editing, All authors.; Resources and supervision, A.P.A., A.M.D., and J.D.K.

Competing Interests

J.D.K. has financial interests in Amyris, Lygos, Constructive Biology, Demetrix, Napigen and Maple Bio.

References

1. Nelson KE, Weinel C, Paulsen IT, Dodson RJ, Hilbert H, Martins dos Santos VAP, Fouts DE, Gill SR, Pop M, Holmes M, Brinkac L, Beanan M, DeBoy RT, Daugherty S, Kolonay J, Madupu R, Nelson W, White O, Peterson J, Khouri H, Fraser CM. 2002. Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. Environ Microbiol 4:799–808.

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2. Nikel PI, Chavarría M, Danchin A, de Lorenzo V. 2016. From dirt to industrial applications: Pseudomonas putida as a Synthetic Biology chassis for hosting harsh biochemical reactions. Curr Opin Chem Biol 34:20–29.

3. Loeschcke A, Thies S. 2015. Pseudomonas putida-a versatile host for the production of natural products. Appl Microbiol Biotechnol 99:6197–6214.

4. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, Davison BH, Dixon RA, Gilna P, Keller M, Langan P, Naskar AK, Saddler JN, Tschaplinski TJ, Tuskan GA, Wyman CE. 2014. Lignin valorization: improving lignin processing in the biorefinery. Science 344:1246843.

5. Chang YF, Adams E. 1971. Induction of separate catabolic pathways for L- and D-lysine in Pseudomonas putida. Biochem Biophys Res Commun 45:570–577.

6. Moe LA. 2013. Amino acids in the rhizosphere: from plants to microbes. Am J Bot 100:1692–1705.

7. Ramos-González MI, Campos MJ, Ramos JL. 2005. Analysis of Pseudomonas putida KT2440 gene expression in the maize rhizosphere: in vivo [corrected] expression technology capture and identification of root-activated promoters. J Bacteriol 187:4033–4041.

8. Chae TU, Ko Y-S, Hwang K-S, Lee SY. 2017. Metabolic engineering of Escherichia coli for the production of four-, five- and six-carbon lactams. Metab Eng 41:82–91.

9. Zhang J, Barajas JF, Burdu M, Wang G, Baidoo EE, Keasling JD. 2017. Application of an Acyl-CoA Ligase from Streptomyces aizunensis for Lactam Biosynthesis. ACS Synth Biol 6:884–890.

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

10. Zhang M, Gao C, Guo X, Guo S, Kang Z, Xiao D, Yan J, Tao F, Zhang W, Dong W, Liu P, Yang C, Ma C, Xu P. 2018. Increased glutarate production by blocking the glutaryl-CoA dehydrogenation pathway and a catabolic pathway involving L-2-hydroxyglutarate. Nat Commun 9:2114.

11. Kim HT, Khang TU, Baritugo K-A, Hyun SM, Kang KH, Jung SH, Song BK, Park K, Oh M-K, Kim GB, Kim HU, Lee SY, Park SJ, Joo JC. 2018. Metabolic engineering of Corynebacterium glutamicum for the production of glutaric acid, a C5 dicarboxylic acid platform chemical. Metab Eng.

12. Revelles O, Espinosa-Urgel M, Fuhrer T, Sauer U, Ramos JL. 2005. Multiple and interconnected pathways for L-lysine catabolism in Pseudomonas putida KT2440. J Bacteriol 187:7500–7510.

13. Perfetti R, Campbell RJ, Titus J, Hartline RA. 1972. Catabolism of pipecolate to glutamate in Pseudomonas putida. J Biol Chem 247:4089–4095.

14. Revelles O, Wittich R-M, Ramos JL. 2007. Identification of the initial steps in D-lysine catabolism in Pseudomonas putida. J Bacteriol 189:2787–2792.

15. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, Deutschbauer A. 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. MBio 6:e00306-15.

16. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, Kuehl JV, Melnyk RA, Lamson JS, Suh Y, Carlson HK, Esquivel Z, Sadeeshkumar H, Chakraborty R, Zane GM, Rubin BE, Wall JD, Visel A, Bristow J, Blow MJ, Deutschbauer AM. 2018. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature 557:503–509.

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

17. Rand JM, Pisithkul T, Clark RL, Thiede JM, Mehrer CR, Agnew DE, Campbell CE, Markley AL, Price MN, Ray J, Wetmore KM, Suh Y, Arkin AP, Deutschbauer AM, Amador-Noguez D, Pfleger BF. 2017. A metabolic pathway for catabolizing levulinic acid in bacteria. Nat Microbiol 2:1624–1634.

18. Okuno E, Tsujimoto M, Nakamura M, Kido R. 1993. 2-Aminoadipate-2-oxoglutarate aminotransferase isoenzymes in human liver: a plausible physiological role in lysine and tryptophan metabolism. Enzyme Protein 47:136–148.

19. Wilding M, Peat TS, Newman J, Scott C. 2016. A β-Alanine Catabolism Pathway Containing a Highly Promiscuous ω-Transaminase in the 12-Aminododecanate-Degrading Pseudomonas sp. Strain AAC. Appl Environ Microbiol 82:3846–3856.

20. Kopchick JJ, Hartline RA. 1979. alpha-Hydroxyglutarate as an intermediate in the catabolism of alpha-aminoadipate by Pseudomonas putida. J Biol Chem 254:3259–3263.

21. Witkowski A, Joshi AK, Smith S. 2002. Mechanism of the β-Ketoacyl Synthase Reaction Catalyzed by the Animal Fatty Acid Synthase†. Biochemistry 41:10877–10887.

22. Zhang L, Ren Y, Lu B, Yang C, Feng Z, Liu Z, Chen J, Ma W, Wang Y, Yu X, Wang Y, Zhang W, Wang Y, Liu S, Wu F, Zhang X, Guo X, Bao Y, Jiang L, Wan J. 2016. FLOURY ENDOSPERM7 encodes a regulator of starch synthesis and amyloplast development essential for peripheral endosperm development in rice. J Exp Bot 67:633–647.

23. Martinez S, Hausinger RP. 2015. Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases. J Biol Chem 290:20702–20711.

24. Martinez S, Hausinger RP. 2016. Biochemical and Spectroscopic Characterization of the

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Non-Heme Fe(II)- and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme from Pseudomonas syringae pv. phaseolicola PK2. Biochemistry 55:5989–5999.

25. Revelles O, Espinosa-Urgel M, Molin S, Ramos JL. 2004. The davDT operon of Pseudomonas putida, involved in lysine catabolism, is induced in response to the pathway intermediate delta-aminovaleric acid. J Bacteriol 186:3439–3446.

26. Danhauser K, Sauer SW, Haack TB, Wieland T, Staufner C, Graf E, Zschocke J, Strom TM, Traub T, Okun JG, Meitinger T, Hoffmann GF, Prokisch H, Kölker S. 2012. DHTKD1 mutations cause 2-aminoadipic and 2-oxoadipic aciduria. Am J Hum Genet 91:1082–1087.

27. REITZ M, RODWELL V. 1969. a-Hydroxyglutarate Oxidoreductase of Pseudomonas putida. J Bacteriol.

28. Wójcik A, Broclawik E, Siegbahn PEM, Borowski T. 2012. Mechanism of benzylic hydroxylation by 4-hydroxymandelate synthase. A computational study. Biochemistry 51:9570–9580.

29. Lequette Y, Odberg-Ferragut C, Bohin J-P, Lacroix J-M. 2004. Identification of mdoD, an mdoG paralog which encodes a twin-arginine-dependent periplasmic protein that controls osmoregulated periplasmic glucan backbone structures. J Bacteriol 186:3695–3702.

30. Vijayakumar SRV, Kirchhof MG, Patten CL, Schellhorn HE. 2004. RpoS-regulated genes of Escherichia coli identified by random lacZ fusion mutagenesis. J Bacteriol 186:8499–8507.

31. Adkins J, Jordan J, Nielsen DR. 2013. Engineering Escherichia coli for renewable production of the 5-carbon polyamide building-blocks 5-aminovalerate and glutarate.

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Biotechnol Bioeng 110:1726–1734.

32. Zhao M, Li G, Deng Y. 2018. Engineering Escherichia coli for Glutarate Production as the C5 Platform Backbone. Appl Environ Microbiol 84.

33. Kalliri E, Mulrooney SB, Hausinger RP. 2008. Identification of Escherichia coli YgaF as an L-2-hydroxyglutarate oxidase. J Bacteriol 190:3793–3798.

34. Jiang T, Guo X, Yan J, Zhang Y, Wang Y, Zhang M, Sheng B, Ma C, Xu P, Gao C. 2017. A Bacterial Multidomain NAD-Independent d-Lactate Dehydrogenase Utilizes Flavin Adenine Dinucleotide and Fe-S Clusters as Cofactors and Quinone as an Electron Acceptor for d-Lactate Oxidization. J Bacteriol 199.

35. Raulfs EC, O'Carroll IP, Dos Santos PC, Unciuleac M-C, Dean DR. 2008. In vivo iron-sulfur cluster formation. Proc Natl Acad Sci USA 105:8591–8596.

36. Nishijyo T, Haas D, Itoh Y. 2001. The CbrA-CbrB two-component regulatory system controls the utilization of multiple carbon and nitrogen sources in Pseudomonas aeruginosa. Mol Microbiol 40:917–931.

37. Valentini M, García-Mauriño SM, Pérez-Martínez I, Santero E, Canosa I, Lapouge K. 2014. Hierarchical management of carbon sources is regulated similarly by the CbrA/B systems in Pseudomonas aeruginosa and Pseudomonas putida. Microbiology (Reading, Engl) 160:2243–2252.

38. Bouffartigues E, Gicquel G, Bazire A, Bains M, Maillot O, Vieillard J, Feuilloley MGJ, Orange N, Hancock REW, Dufour A, Chevalier S. 2012. Transcription of the oprF gene of Pseudomonas aeruginosa is dependent mainly on the SigX sigma factor and is sucrose induced. J Bacteriol 194:4301–4311.

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

39. Blanka A, Schulz S, Eckweiler D, Franke R, Bielecka A, Nicolai T, Casilag F, Düvel J, Abraham W-R, Kaever V, Häussler S. 2014. Identification of the alternative sigma factor SigX regulon and its implications for Pseudomonas aeruginosa pathogenicity. J Bacteriol 196:345–356.

40. Neshich IAP, Kiyota E, Arruda P. 2013. Genome-wide analysis of lysine catabolism in bacteria reveals new connections with osmotic stress resistance. ISME J 7:2400–2410.

41. Brinkman AB, Bell SD, Lebbink RJ, de Vos WM, van der Oost J. 2002. The Sulfolobus solfataricus Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. J Biol Chem 277:29537–29549.

42. Thaw P, Sedelnikova SE, Muranova T, Wiese S, Ayora S, Alonso JC, Brinkman AB, Akerboom J, van der Oost J, Rafferty JB. 2006. Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. Nucleic Acids Res 34:1439–1449.

43. LaBauve AE, Wargo MJ. 2012. Growth and laboratory maintenance of Pseudomonas aeruginosa. Curr Protoc Microbiol Chapter 6:Unit 6E.1.

44. Ham TS, Dmytriv Z, Plahar H, Chen J, Hillson NJ, Keasling JD. 2012. Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. Nucleic Acids Res 40:e141.

45. Chen J, Densmore D, Ham TS, Keasling JD, Hillson NJ. 2012. DeviceEditor visual biological CAD canvas. J Biol Eng 6:1.

46. Hillson NJ, Rosengarten RD, Keasling JD. 2012. j5 DNA assembly design automation software. ACS Synth Biol 1:14–21.

47. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009.

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345.

48. Engler C, Kandzia R, Marillonnet S. 2008. A one pot, one step, precision cloning method with high throughput capability. PLoS ONE 3:e3647.

49. Shanks RMQ, Kadouri DE, MacEachran DP, O'Toole GA. 2009. New yeast recombineering tools for bacteria. Plasmid 62:88–97.

50. González Fernández-Niño SM, Smith-Moritz AM, Chan LJG, Adams PD, Heazlewood JL, Petzold CJ. 2015. Standard flow liquid chromatography for shotgun proteomics in bioenergy research. Front Bioeng Biotechnol 3:44.

51. Batth TS, Singh P, Ramakrishnan VR, Sousa MML, Chan LJG, Tran HM, Luning EG, Pan EHY, Vuu KM, Keasling JD, Adams PD, Petzold CJ. 2014. A targeted proteomics toolkit for high-throughput absolute quantification of Escherichia coli proteins. Metab Eng 26:48–56.

52. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ. 2010. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26:966–968.

53. Reiter L, Rinner O, Picotti P, Hüttenhain R, Beck M, Brusniak M-Y, Hengartner MO, Aebersold R. 2011. mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nat Methods 8:430–435.

54. George KW, Thompson MG, Kim J, Baidoo EEK, Wang G, Benites VT, Petzold CJ, Chan LJG, Yilmaz S, Turhanen P, Adams PD, Keasling JD, Lee TS. 2018. Integrated analysis of isopentenyl pyrophosphate (IPP) toxicity in isoprenoid-producing Escherichia coli. Metab

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Eng 47:60–72.

55. Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40.

56. Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738.

57. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure and function prediction. Nat Methods 12:7–8.

58. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.

59. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66:213–221.

60. Grosdidier A, Zoete V, Michielin O. 2011. SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res 39:W270-7.

61. Grosdidier A, Zoete V, Michielin O. 2011. Fast docking using the CHARMM force field with EADock DSS. J Comput Chem 32:2149–2159.

62. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780.

63. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

trees for large alignments. PLoS ONE 5:e9490.

64. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242-5.

65. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75.

66. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

67. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191.

68. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589.

69. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274.

70. Cruz-Morales P, Ramos-Aboites HE, Licona-Cassani C, Selem-Mójica N, Mejía-Ponce PM, Souza-Saldívar V, Barona-Gómez F. 2017. Actinobacteria phylogenomics, selective isolation from an iron oligotrophic environment and siderophore functional characterization, unveil new desferrioxamine traits. FEMS Microbiol Ecol 93.

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

71. mwaskom/seaborn: v0.9.0 (July 2018) | Zenodo.

72. Matplotlib: A 2D Graphics Environment - IEEE Journals & Magazine.

73. Jones E, Oliphant T, Peterson P, Others. SciPy: Open source scientific tools for Python.

74. Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS. 2014. MNE software for processing MEG and EEG data. Neuroimage 86:446–460.

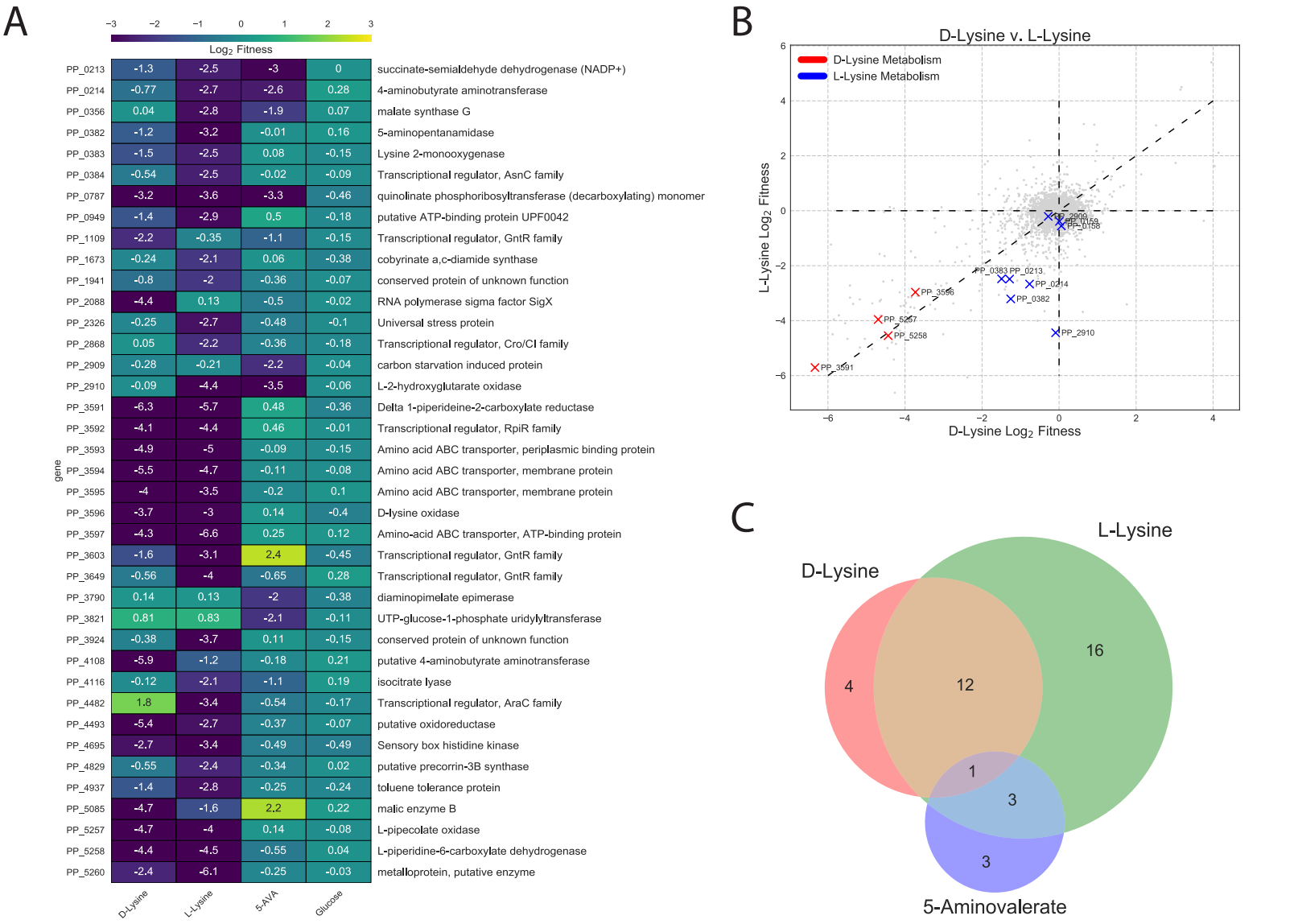


Figure 1: Results of RB-TnSeq screen. A) Genes that showed greater than -2 log2 fitness on either D-lysine, L-lysine, or 5AVA but showed no less than -0.5 log2 fitness defect when grown on glucose. B) Plot of genome wide fitness values of libraries grown on either L-lysine or D-lysine. Genes encoding for enzymes known to be involved in D-lysine metabolism are shown in red, while those known to be involved in L-lysine metabolism are shown in blue. C) Venn diagram of genes with significant fitness defects when grown on either D-lysine, L-lysine, or 5AVA.

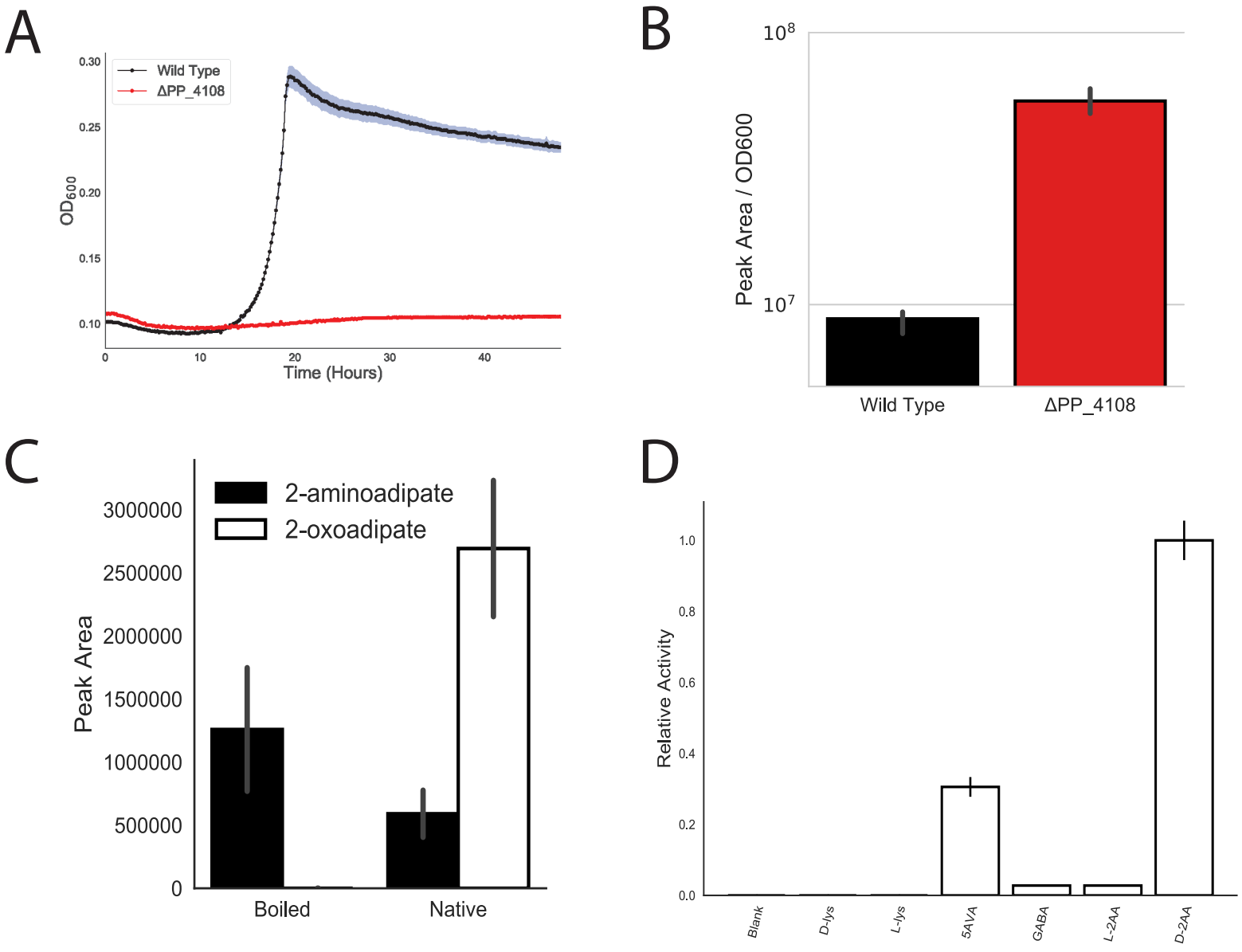


Figure 2: Identification of PP_4108 as a 2AA aminotransferase. A) Growth of wild-type KT2440 and PP_4108 mutant on 2AA as a sole carbon source. Shaded area represents 95% CI, n=3. B) In vivo accumulation of 2AA in wild-type KT2440 and a PP_4108 mutant after 12 hours of grown on minimal media supplemented with 10 mM glucose and 10 mM D-lysine. Bars represent 1og10 transformed spectral counts, error bars show 95% CI, n=3. C) In vitro transamination reactions of PP_4108. Bars represent spectral counts of either 2OA (black) or 2AA (white) in either boiled or native protein reactions. Error bars show 95% CI, n=3. D) In vitro transaminations of PP_4108 incubated with different possible amino donors. Bars represent nmols of glutamate formed after incubation for 16 hours. Error bars show standard deviation, n=2.

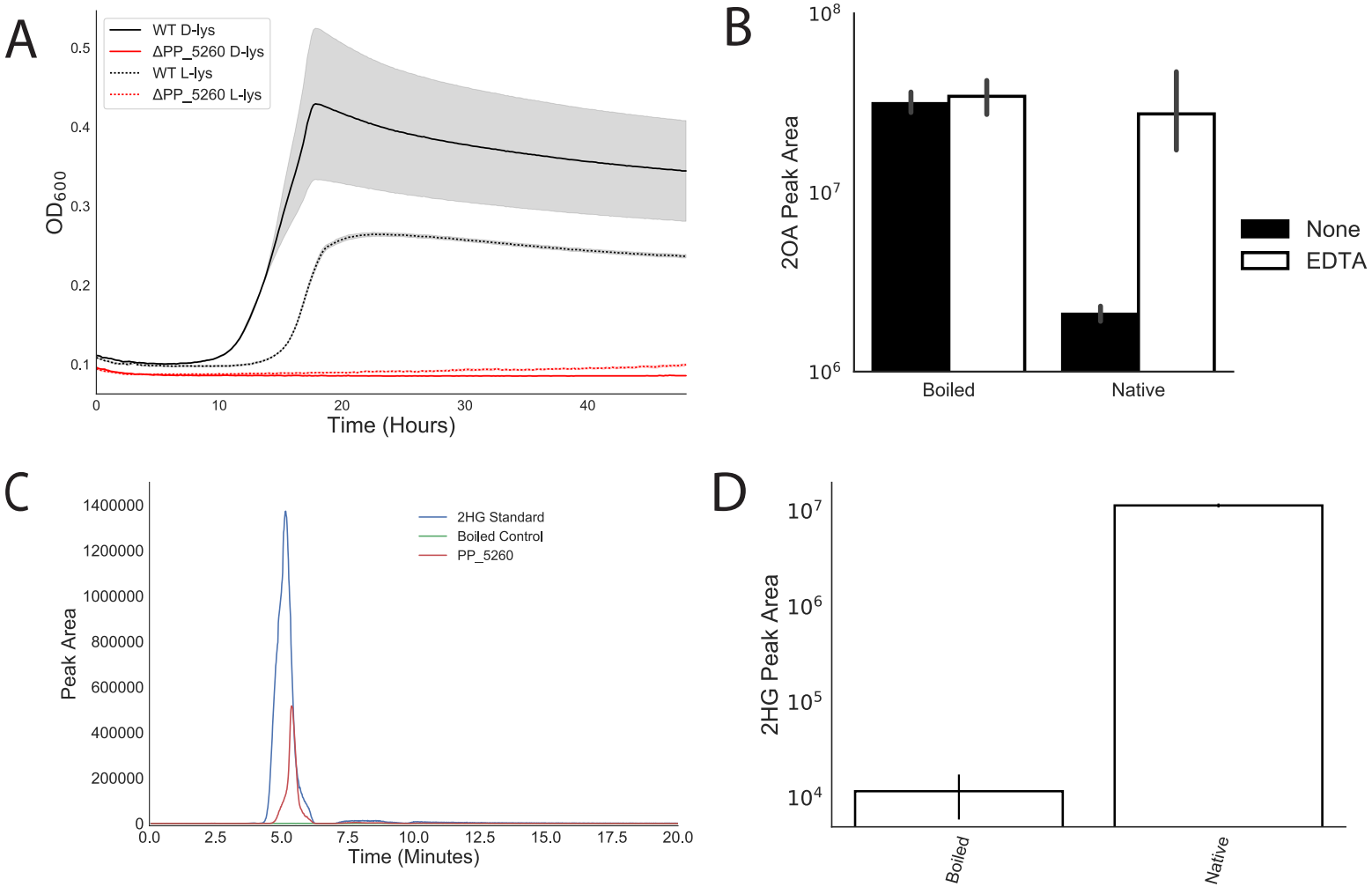
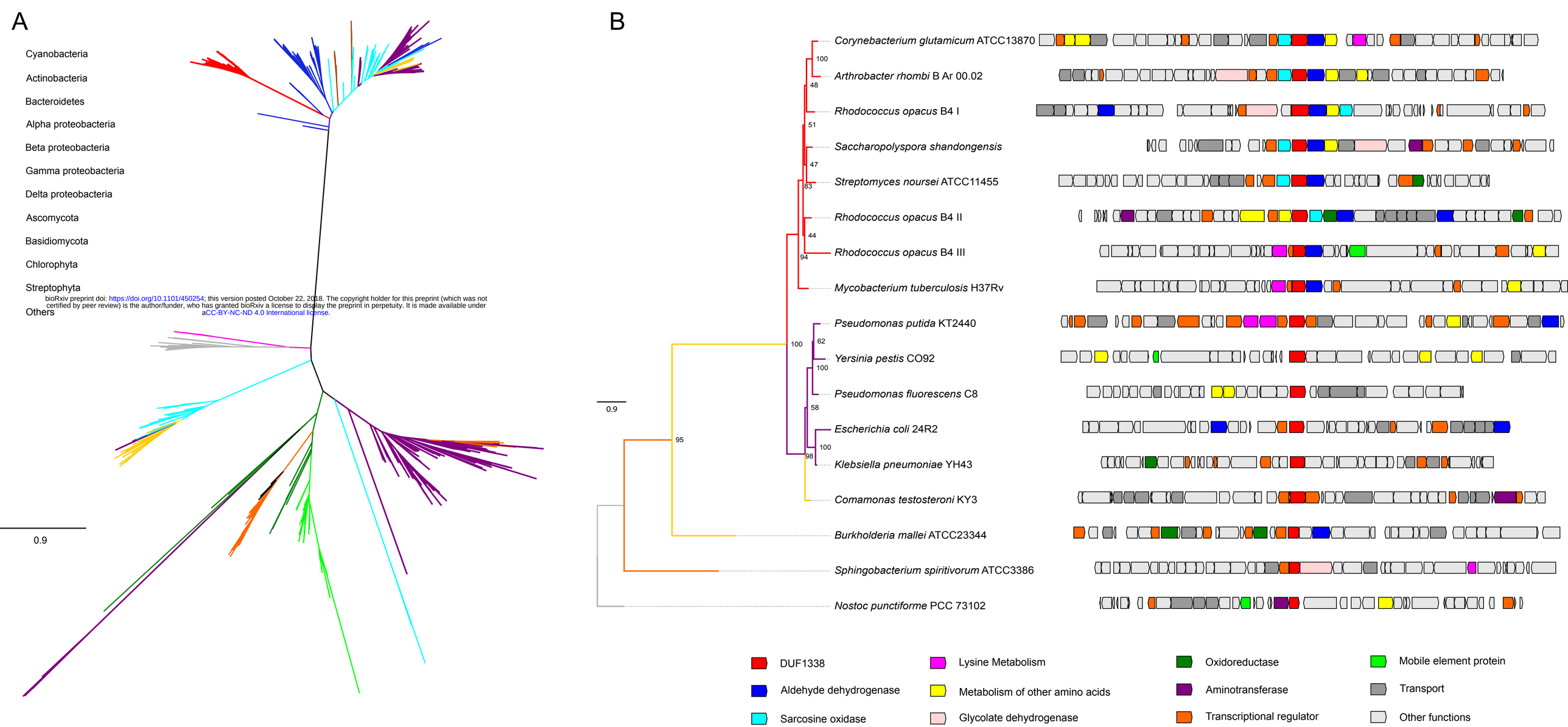


Figure 3: Identification of ydcJ (PP_5260) as a 2OA decarboxylase/oxidase. A) Growth of wild-type (black) and PP_5260 mutant (red) on D-lysine (line) or L-lysine (dashed line) as a sole carbon source. Shaded area represents 95% CI, n=3. B) In vitro reactions of boiled or native PP_5260 incubated with 2OA with 50 uM EDTA (white), or without EDTA (black). Bars represent 1og10 transformed spectral counts, errors bars represent 95% CI, n=3. C) LC-TOF analysis of a 2HG standard, and products of PP_5260 incubated with 2OA, and boiled control. D) Peak area of 2HG produced from native PP_5260 in vitro reactions compared to boiled control. Error bars represent 95% CI, n=3.



bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4: Phylogenomics of the DUF1338 enzyme family. A. Phylogentic relationships among DUF1338 homologs and their distribution among major phyla. Branches in the tree are colored by phylum. DUF1338 is found in most bacterial phyla as well as in plant and fungi. Non-monophyletic clades suggest pervasive horizontal gene transfer events in the family. B. Phylogenomics of selected DUF1338 homologs in bacteria. The phylogeny in the left shows the phylogenetic relationships between selected homologs, the branches have been colored according to their adscription to a given phylum and the support values are shown at the nodes. The boxes in the right represent the gene neighborhood for each homolog. The genes have been colored to represent their annotated functions.

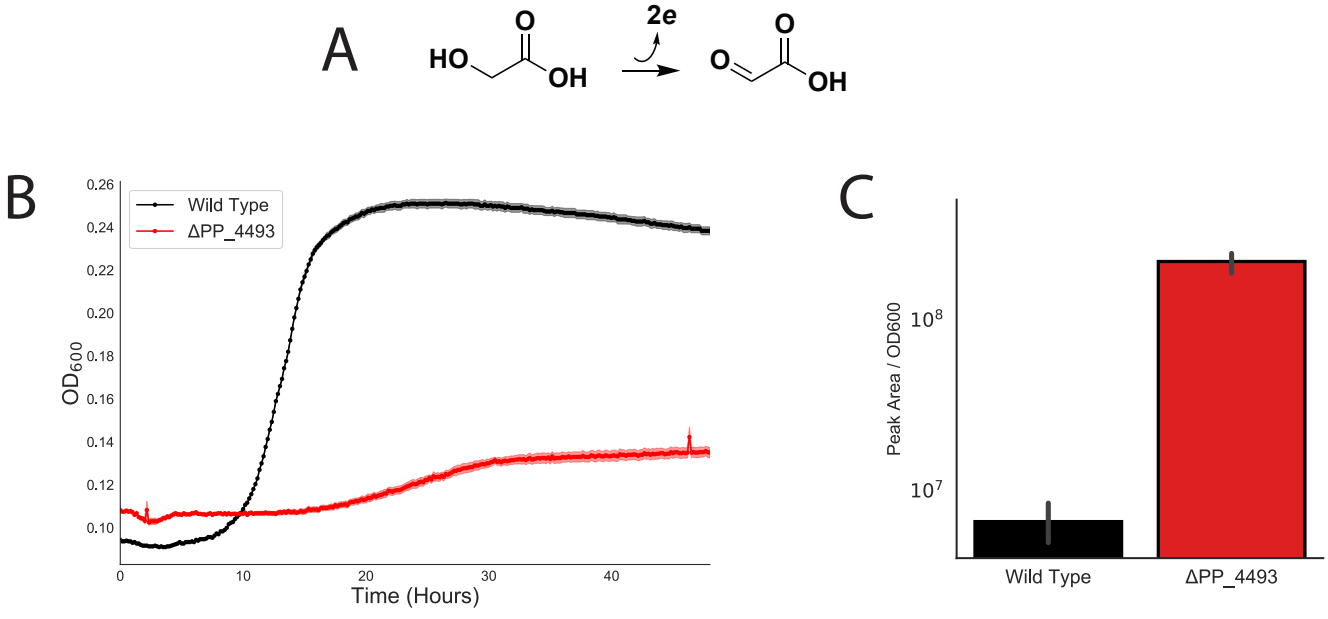


Figure 5: Identification of PP_4493 as a putative 2HG dehydrogenase. A) Enzymatic conversion of glycolate to glyoxylate B) Growth of wild-type KT2440 and PP_4108 mutant on D-lysine as a sole carbon source. Shaded area represents 95% CI, n=3. C) In vivo accumulation of 2HG in wild-type KT2440 and a PP_4108 mutant after 12 hours

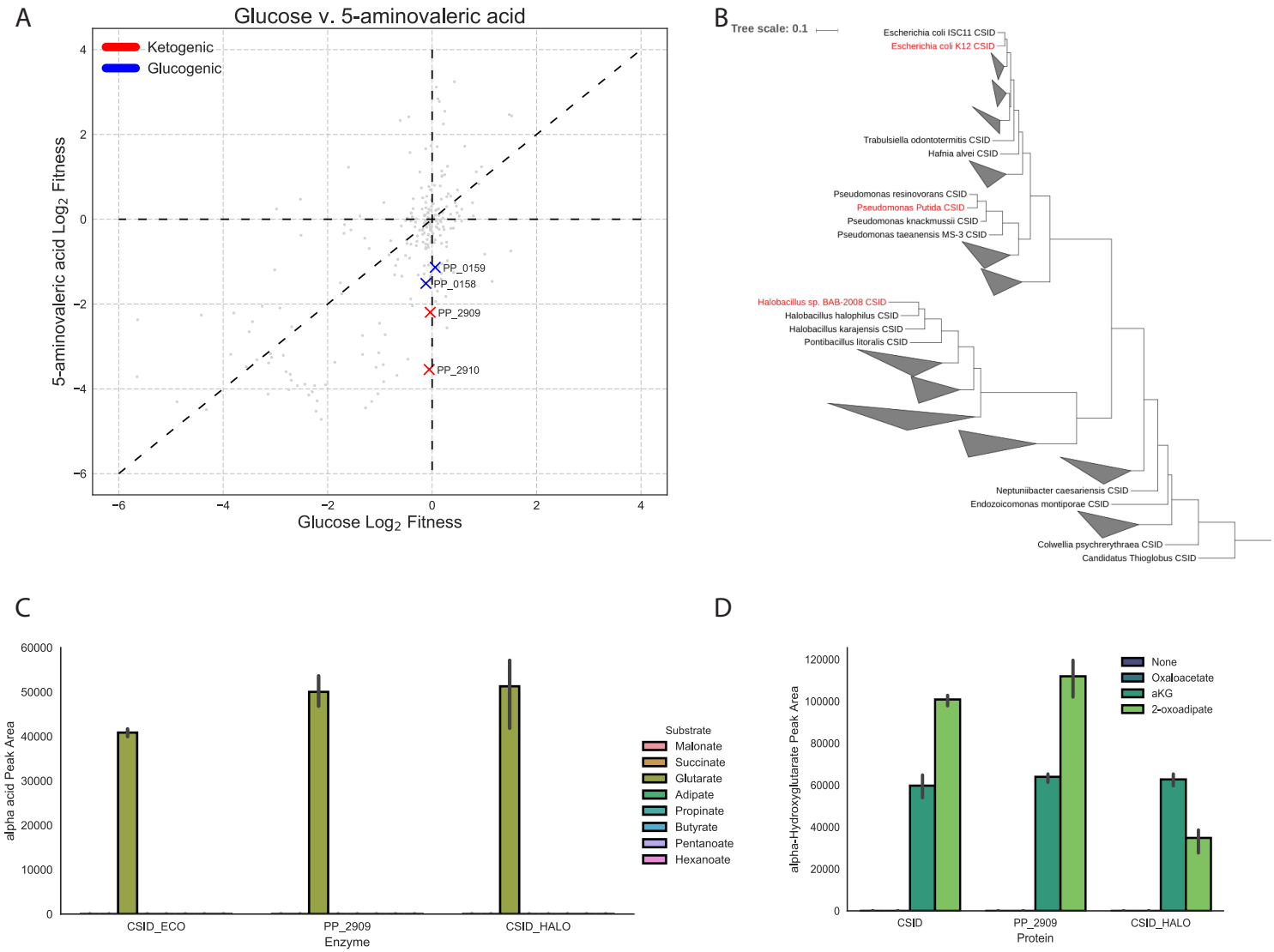


Figure 6: Novel roles of CsiD in P. putida lysine metabolism. A) Plot of genome wide fitness values of libraries grown on either 5AVA or Glucose. Ketogenic glutarate degradation genes are shown in red,, while those known to be involved glucogenic metabolism are shown in blue. B) Phylogenetic tree of bacterial CsiD homologs. Homologs used in in vitro assays are highlighted in red. C) in vitro reactions of CsiD with different substrates using 2KG as a 2-oxoacid. Bars show peak area of 2-hydroxyacid, error

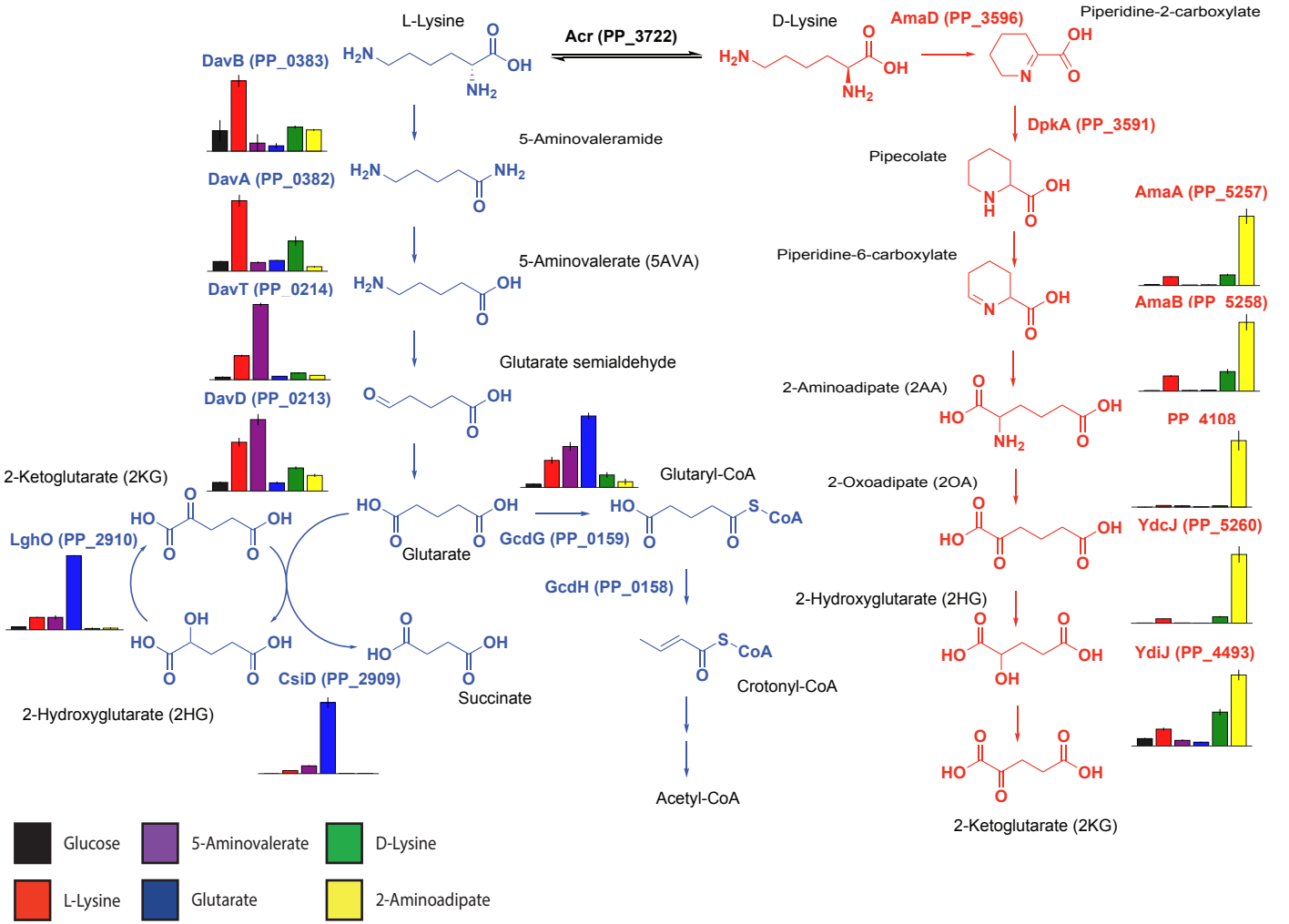


Figure 7: Expression of Lysine degradation pathways in response to different Lysine metabolites. Relative abundance of select lysine degradation enzymes expressed in wild-type KT2440 in response to different carbon sources. Bars show spectral counts of proteins after 36 hours of grown on 10 mM glucose (black), 5AVA (purple), D-lysine (green), L-lysine (red), glutarate (blue), or 2AA (yellow). Error bars show 95% CI, n=3.

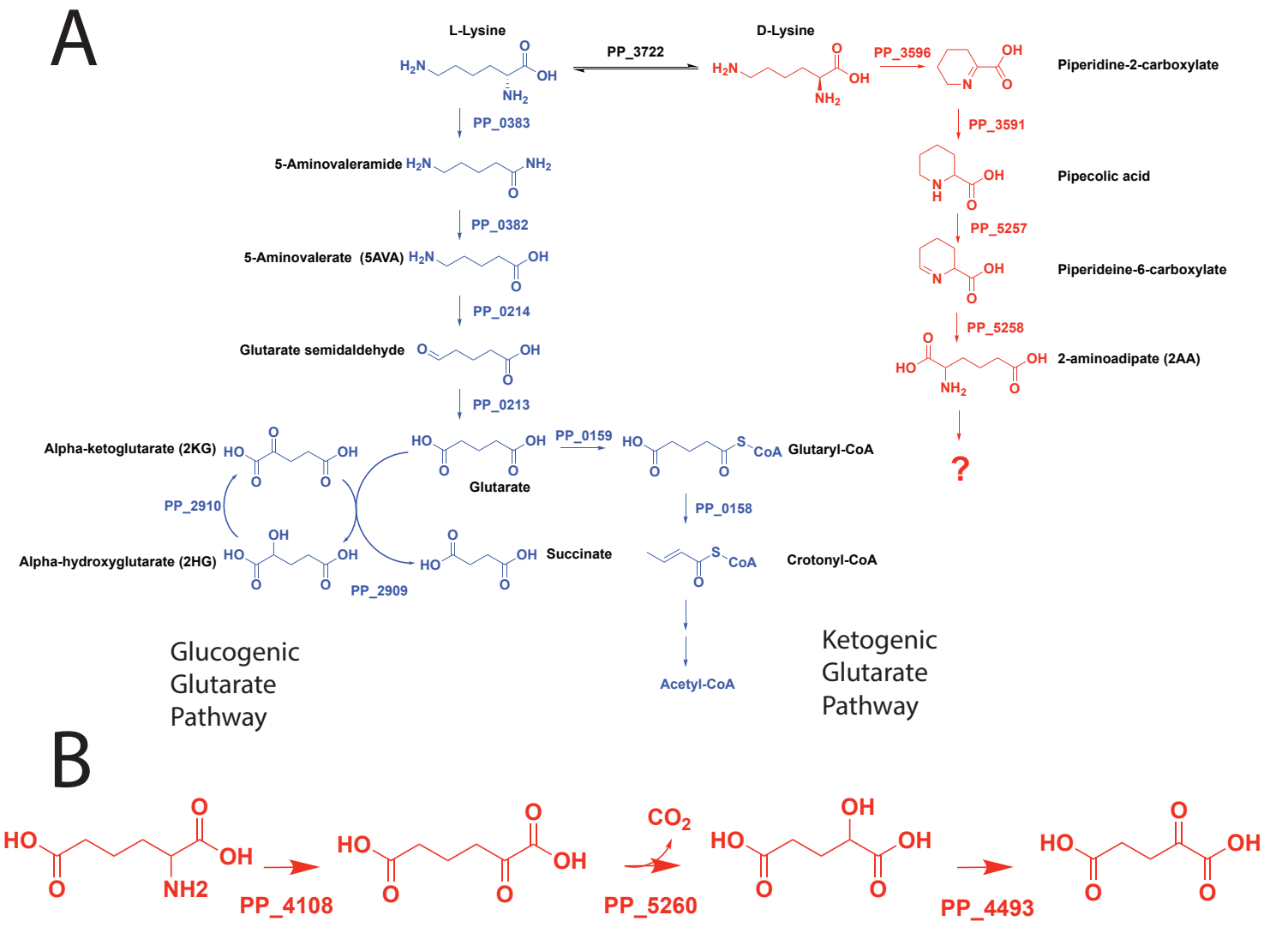


Figure S1: Known metabolic pathways of lysine catabolism in P. putida KT2440. A) L-lysine metabolic pathway is shown in blue, while the known steps of D-lysine metabolism is shown in red. B) Proposed route of 2AA metabolism in P. putida.

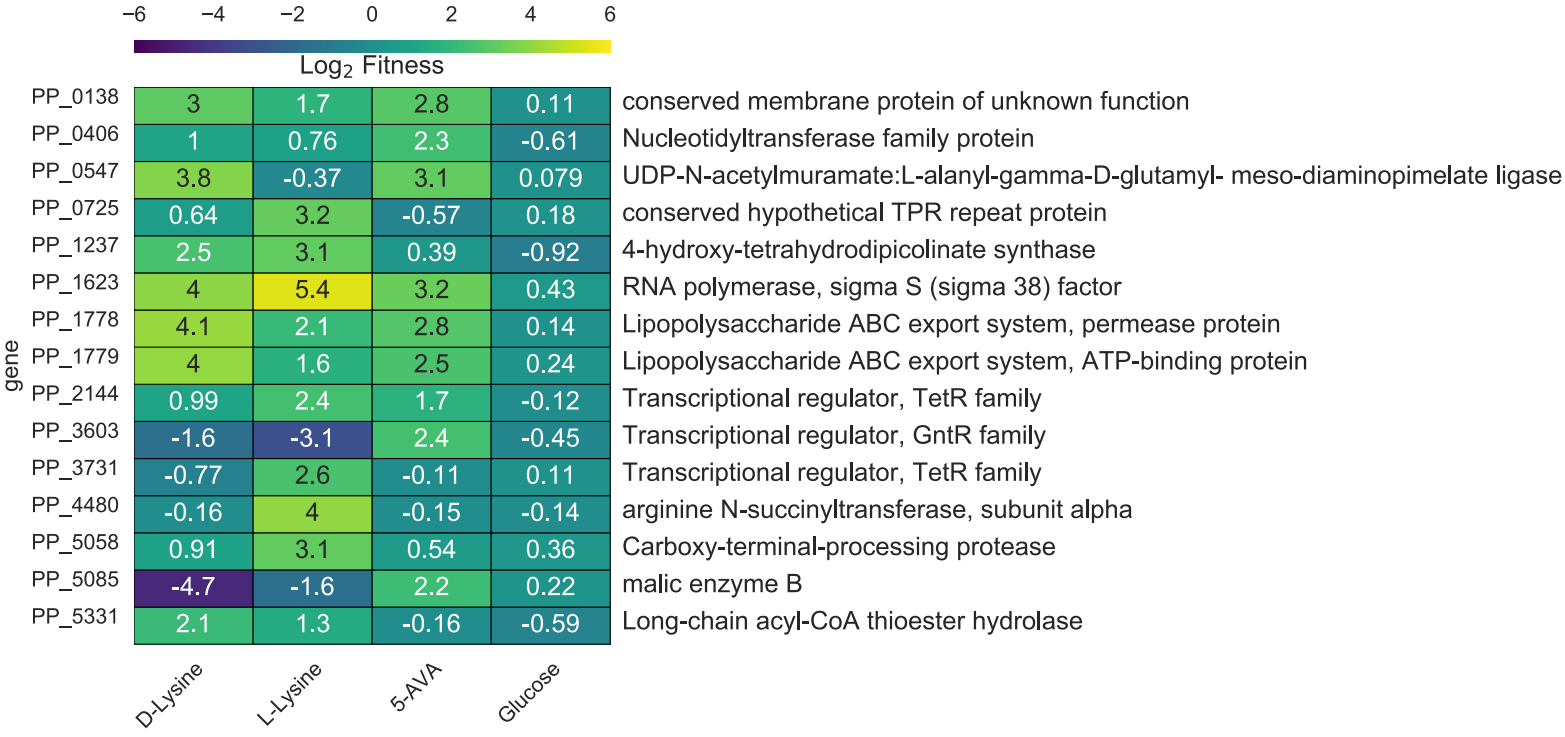


Figure S2: Results of RB-TnSeq screen. A) Genes that showed great than 2 log2 fitness on either D-lysine, L-lysine, or 5-AVA but showed no less than 0.5 log2 fitness defect when grown on glucose.

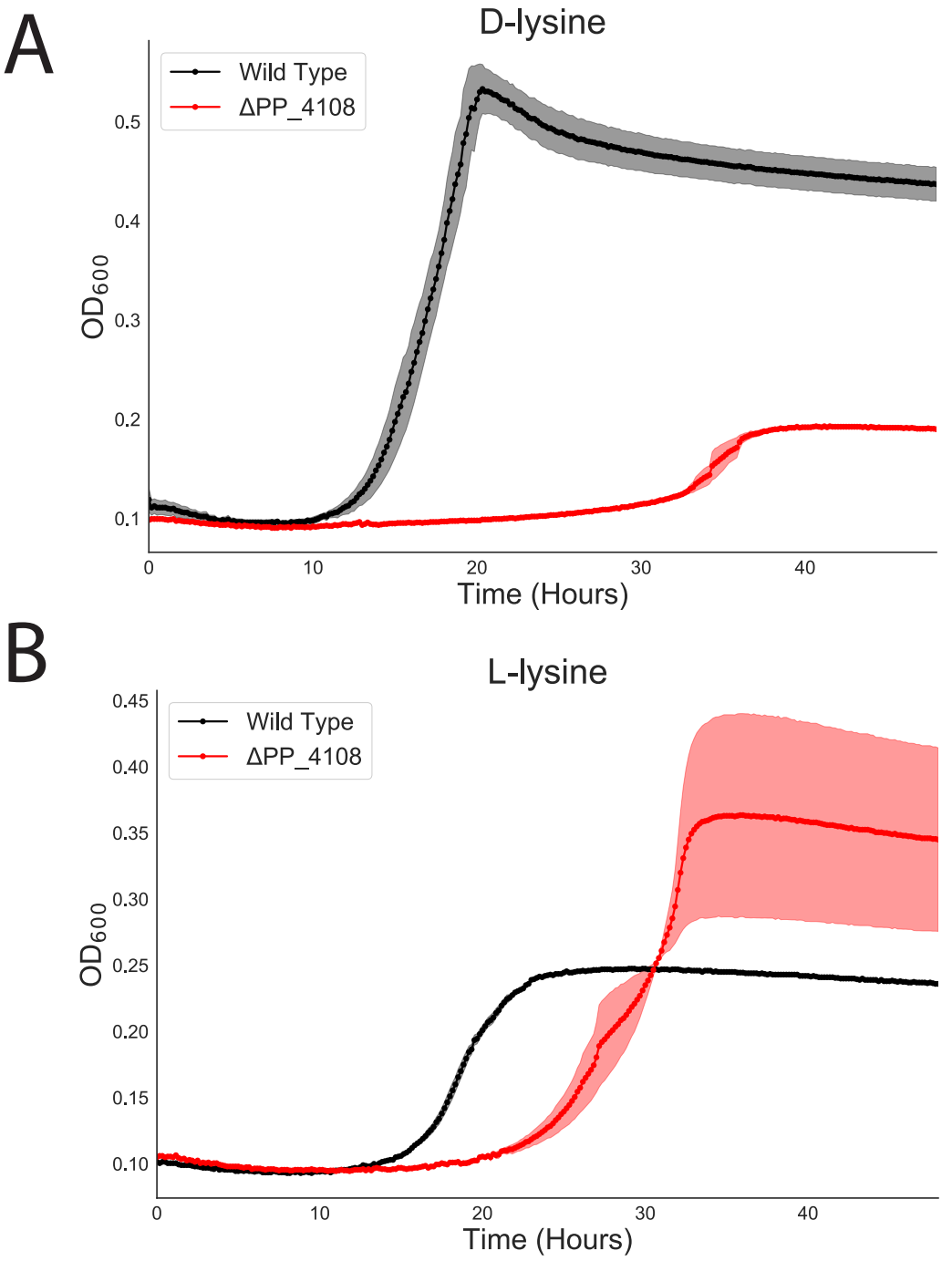
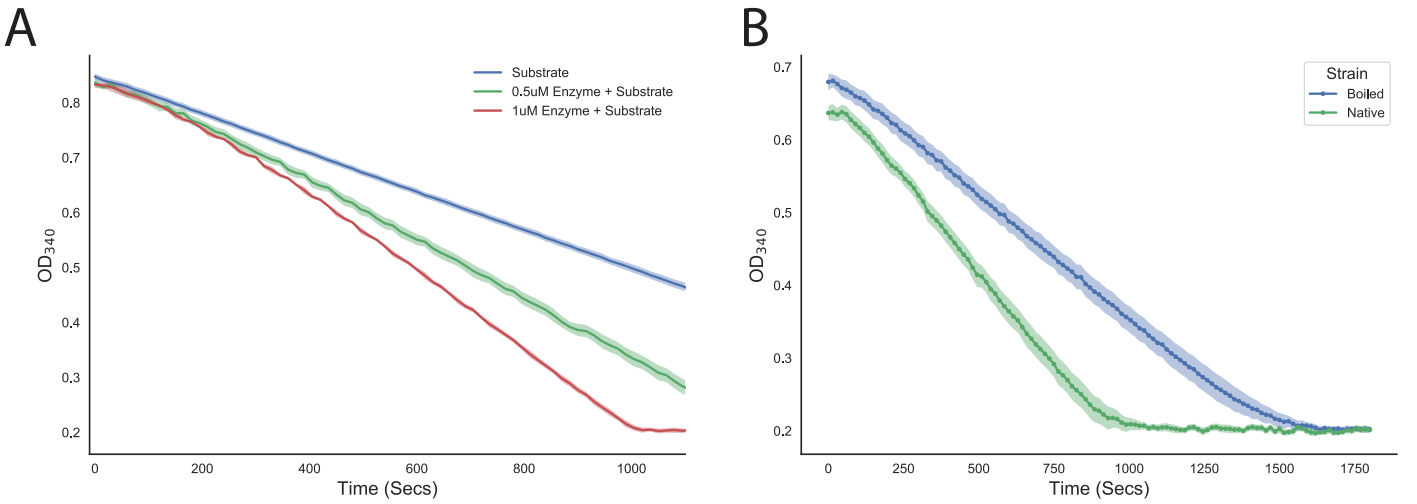


Figure S3: Growth of PP_4108 mutants on lysine. A) Growth of wild-type KT2440 and PP_4108 mutant on D-lysine as a sole carbon source. Shaded area represents 95% CI, n=3. B) Growth of wild-type KT2440 and PP_4108 mutant on L-lysine as a sole carbon source. Shaded area represents 95% CI, n=3.



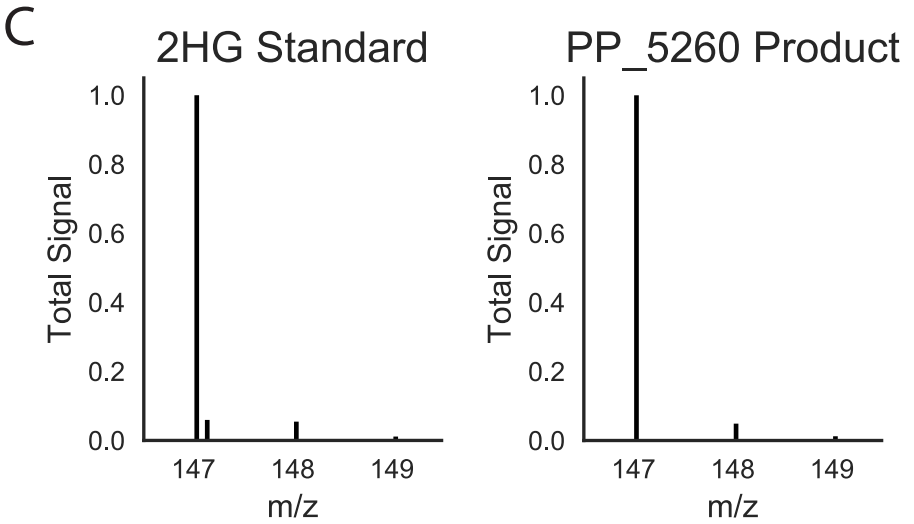


Figure S4: PP_5260 is a decarboxylase. A) Enzyme coupled decarboxylation assay of 2OA incubated with PP_5260. Lines show oxidation of NADH as measured by decrease in OD340 when 2OA alone (blue line), 2OA with 0.5uM PP_5260 (green line), or 2OA with 1.0uM PP_5260 are added. Shaded area represents 95% CI, n=3. B) Enzyme coupled decarboxylation assay of 2-OA incubated with PP_5260. Lines show oxidation of NADH as measured by decrease in OD340 with boiled PP_5260 (blue line) or native PP_5260 (green line) are added. Shaded area represents 95% CI, n=3. C) Mass spectra of 2HG standard and product of PP_5260 in vitro reaction.

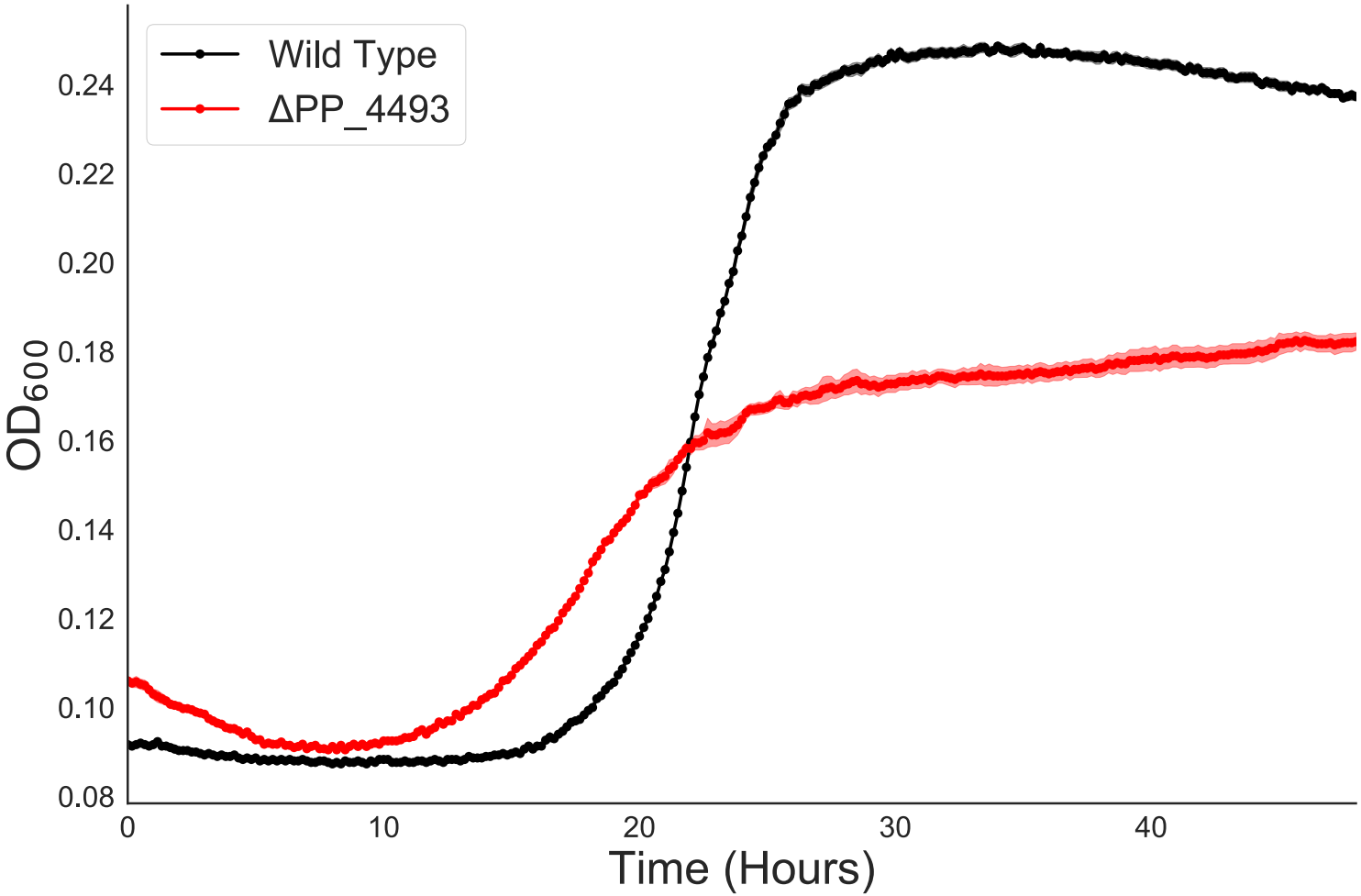


Figure S5: Growth of PP_4493 mutant on L-lysine. Growth of wild-type KT2440 and PP_4493 mutant on L-lysine as a sole carbon source. Shaded area represents 95% CI, n=3.

bioRxiv preprint doi: https://doi.org/10.1101/450254; this version posted October 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

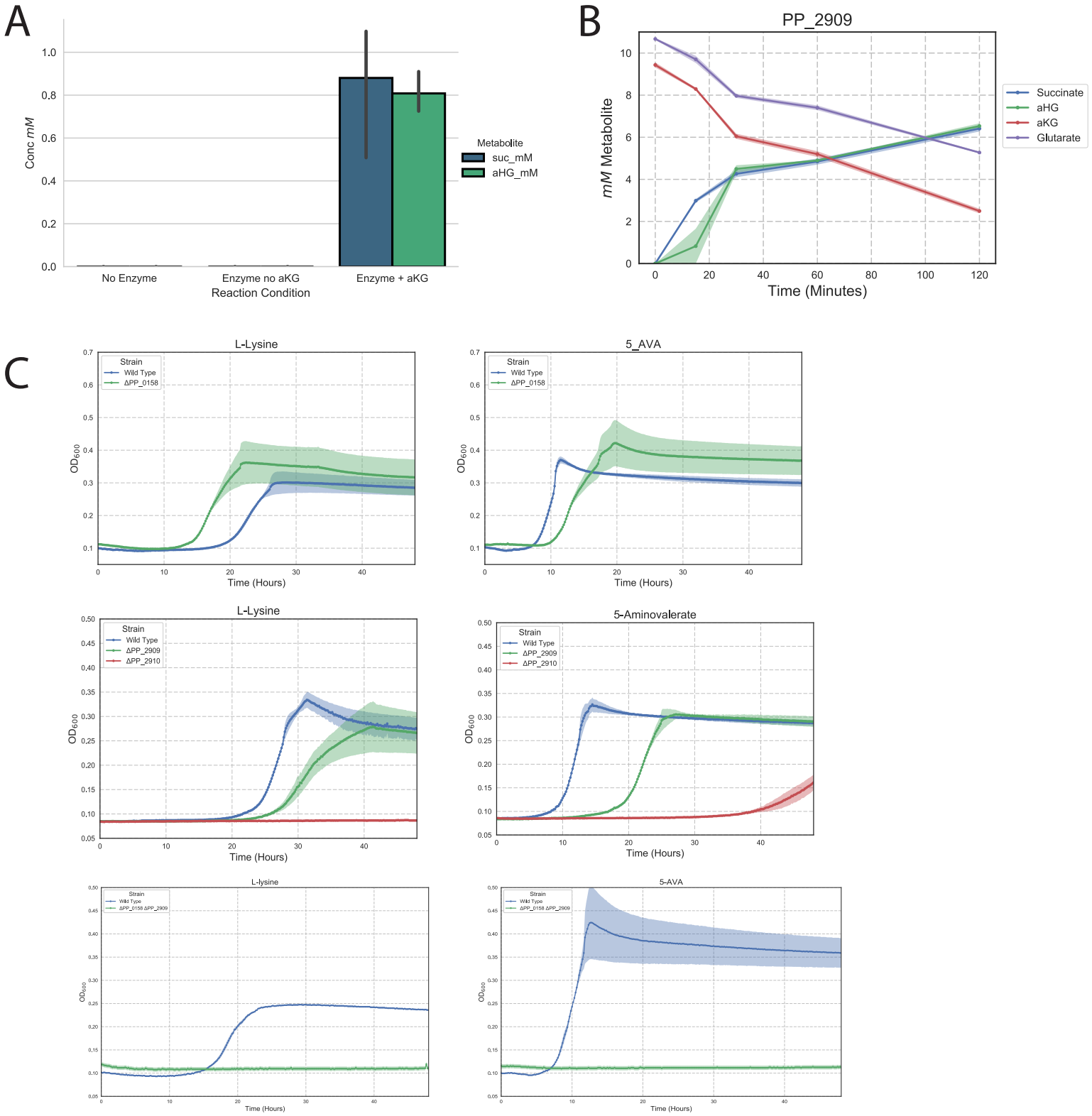


Figure S6: Characterization of PP_2909 (CsiD) in P. putida. A) In vitro reactions of PP_2909. Bars show mM of either succinate (blue), or 2-HG (green) formed by boiled enzyme control, no 2-KG control, or native enzyme with 2-KG added. Errors bars show 95% CI, n=3. B) Time course in vitro reaction of PP_2909. Plot shows 2-HG, 2-KG, succinate, and glutarate overtime. Shaded region shows 95% CI, n=3. C) Growth curves of wild-type KT2440, PP_0158, PP_2909, PP_2910, or PP_2909/PP_0158 double mutants grown on either L-lysine (left column), or 5-AVA (right column). Shaded region shows 95% CI, n=3. D) In vitro reactions of three homologs of CsiD from E. coli, P. putida, and Halobacillus sp. BAB-2008 with C3-C6 fatty acids and diacids. Bars show peak are of the corresponding 2-hydroxyacid. Error bars show 95%CI, n=3.

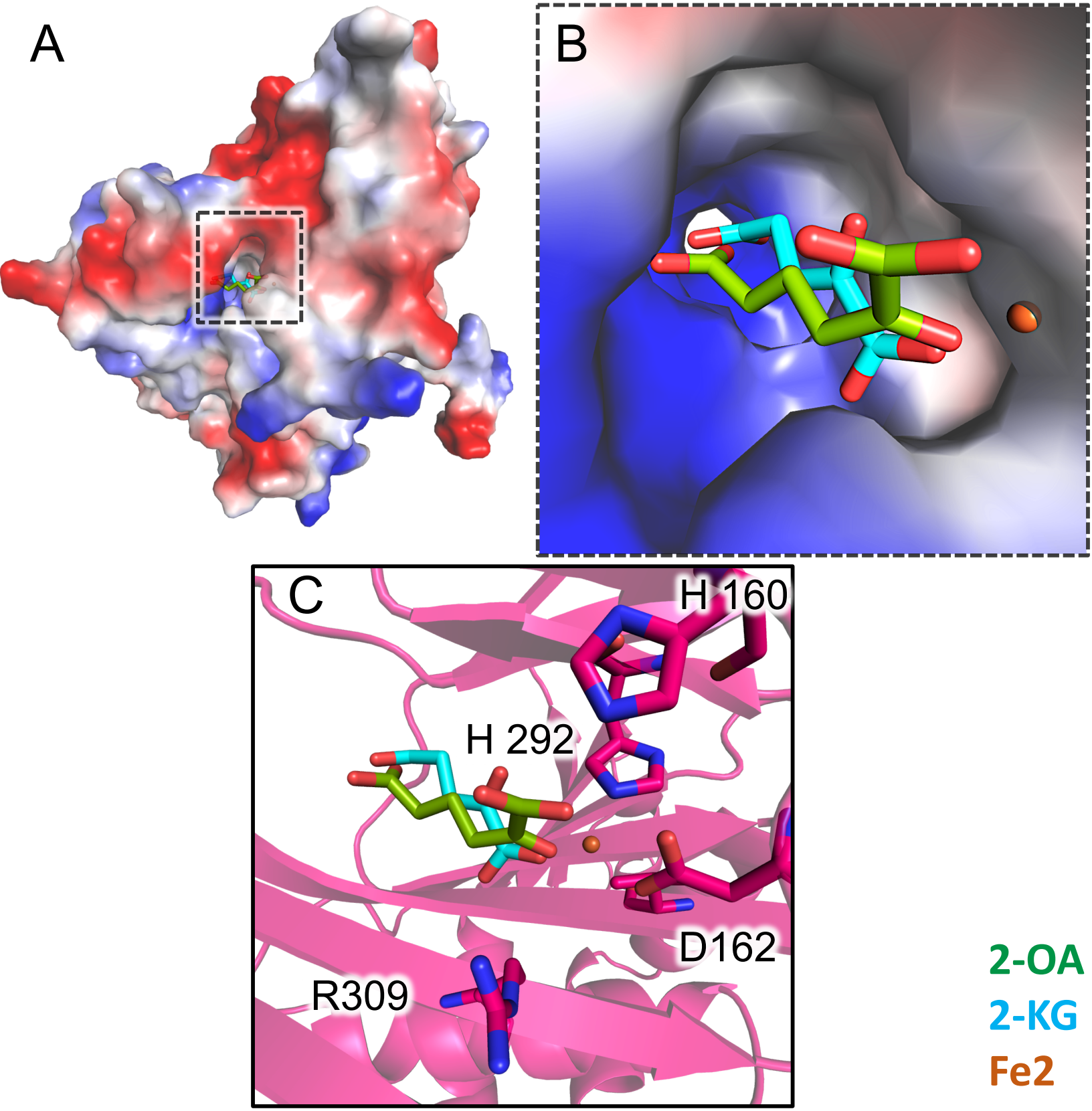


Figure S7: Docking of 2-OA and 2-KG into CsiD homology model. A) Surface electrostatic representation of CsiD docked with 2-OA (green) and 2-KG (cyan). B) A zoomed representation of the active site cavity from (A) depicting 2-OA and 2-KG coordinating with the Fe2 (brown). C) All residues 4 Å away from the Fe2 and around the docked 2-OA and 2-KG ligands.

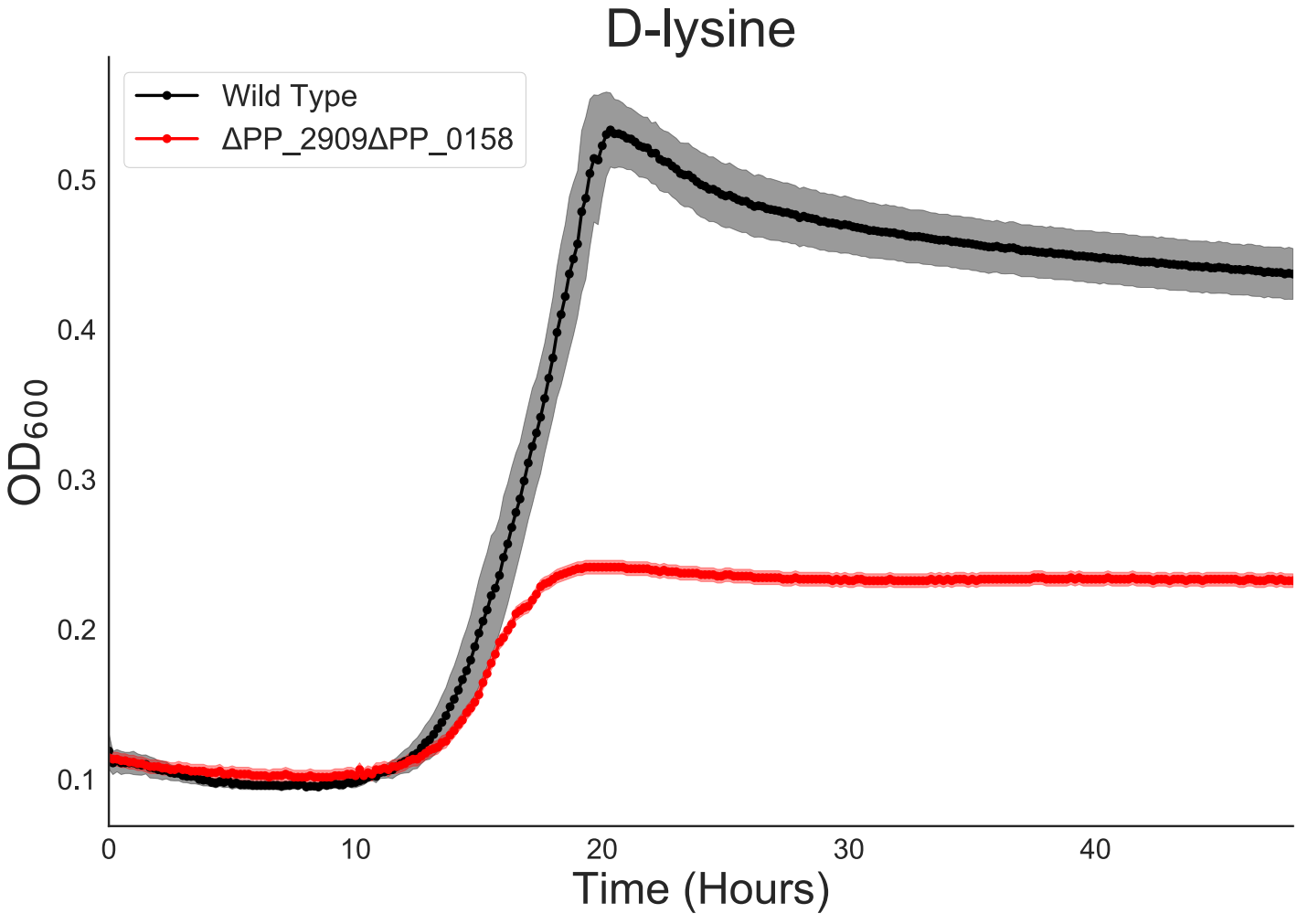


Figure S8: Growth of PP_2909/PP_0158 mutant on D-lysine. Growth of wild-type KT2440 and PP_2909/P-P_0158 mutant on D-lysine as a sole carbon source. Shaded area represents 95% CI, n=3.