1   **Rapid, parallel identification of pathways for catabolism of lignin-derived**

2   **aromatic compounds**

3

4   Jacob H. Cecil[1,2], Joshua K. Michener[1, 2, #]

5   (1) Biosciences Division and (2) BioEnergy Science Center, 1 Bethel Valley Road, Oak Ridge

6   National Laboratory, Oak Ridge, TN, 37830

7

8   (#) Correspondence should be sent to JKM (michenerjk@ornl.gov)

9

18

Introduction

Transposon mutagenesis is a powerful technique in microbial genetics for the identification of genes in uncharacterized pathways[1–3]. However, transposon library screening does not scale well, requiring a new screen for each substrate and extensive oversampling to completely identify all the genes in a given metabolic pathway. A recent technique, dubbed BarSeq[4], overcomes these limitations by introducing random 20-nucleotide barcodes into each transposon mutant, which can then be tracked using high-throughput DNA sequencing to quantitatively measure the fitness effect of each knockout. Here we show that, when applied to catabolic pathways, barcoded transposon libraries can be used to distinguish redundant pathways, decompose complex pathways into substituent modules, discriminate between enzyme homologs, and rapidly identify previously-hypothetical enzymes in an unbiased genome-scale search. We use this technique to demonstrate that two genes, which we name *desC* and *desD*, are involved in the degradation of the lignin-derived aromatic compound sinapic acid in the non-model organism *Novosphingobium aromaticivorans*. This approach will be particularly useful in the identification of complete pathways suitable for heterologous expression in metabolic engineering.

35   Microbes are capable of degrading a vast array of natural and non-natural small molecules

36   using an intricate network of metabolic pathways. The characterization of these networks is often

37   the work of many years, particularly for ancestral pathways that can be scattered throughout the

38   genome[5]. Closing a pathway by identifying the entire catalytic ensemble is particularly

39   challenging using random mutagenesis methods, since a search for the last enzyme in a pathway

40   will frequently re-isolate previously-identified enzymes. This challenge is particularly acute

41   when the pathways are intended for use in metabolic engineering, since any missing enzyme

42   renders the entire pathway ineffective[6]. Pathways for the degradation of lignin-derived aromatic

43   compounds provide a clear example of this problem, as these complex pathways are important

44   targets for heterologous expression in industrial microbes and yet rarely found in nature as

45   complete operons[5,7–9].

46   We have addressed this problem of catabolic pathway identification using randomly-barcoded

47   transposon mutagenesis (Figure 1)[4]. Using this approach, an entire library of transposon mutants

48   can be screened against a panel of substrates in a single experiment. We chose *Novosphingobium*

49   *aromaticivorans* DSM12444 (hereafter 'DSM12444'), a bacterium isolated from contaminated

50   sediment, as a model system for pathway identification[10,11]. This strain has been shown to

51   catabolize a wide range of aromatic compounds, and its catalytic repertoire is predicted to

52   include enzymes that degrade model lignin-derived biaryls such as guaiacylglycerol-β-guaiacyl

53   ether[12]. To begin identifying the genes involved in aromatic degradation in this strain, we

54   constructed a barcoded transposon library in DSM12444. A single round of transposon insertion

55   sequencing was performed to map barcodes to insertion sites, identifying a total of 43,270 unique

56   barcoded insertions in 3117 out of 3897 protein-coding genes. The pooled transposon library was

57   then grown, in triplicate, in minimal medium containing protocatechuate (PCA), 4-

58   hydroxybenzoate (4-HB), coumarate, vanillate, ferulate, syringate, or sinapate as the sole source

59   of carbon and energy. The mixed barcodes were amplified from the inoculum and the final

60   cultures and then sequenced to determine the fitness effect of knocking out each identified gene
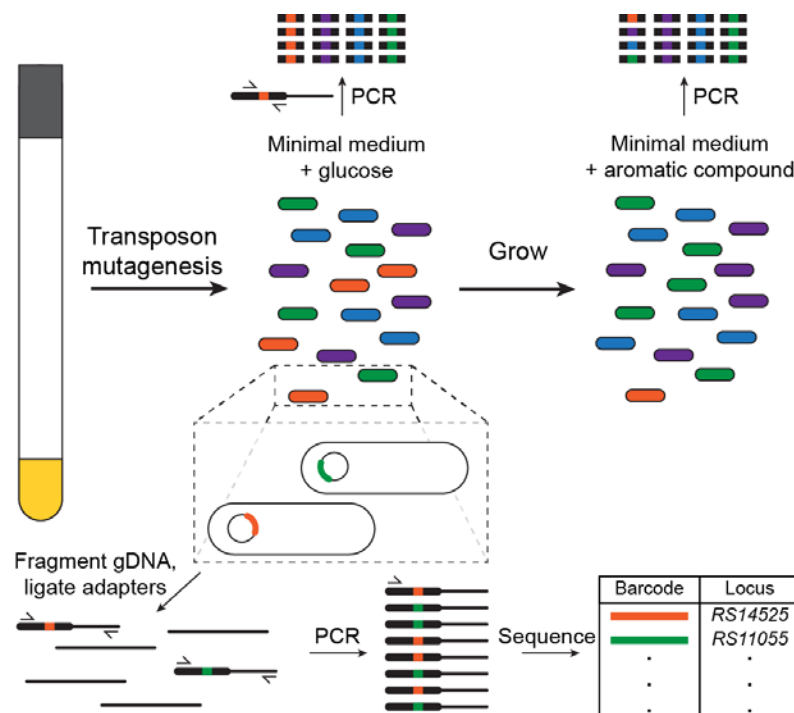
61   under the relevant conditions.

62

63   **Figure 1: Randomly barcoded transposon mutagenesis allows high-throughput**

64   **pathway identification.** A pool of barcoded transposons is introduced into the target bacterium.

65   A single round of TnSeq maps barcodes (colored inserts) to the insertion locus. The barcodes can

66   be queried through PCR and high-throughput sequencing to determine changes in the population

67   ratio during growth with a specific compound.

68

69   Sphingomonads are known to catabolize PCA through 4,5-cleavage of the aromatic ring.

70   DSM12444 contains chromosomal copies of the genes necessary for this cleavage pathway,

71   *ligABCIUJK*, as well as plasmid-encoded genes for the lower 2,3-cleavage pathway for catechol,

72 *xylEGHIJKQ*. During growth with PCA, disrupting either pathway produces a consistent

73 decrease in fitness, averaging a 76% decrease in fitness for an insertion in the *lig* pathway or a

74 39% decrease in fitness for an insertion in the *xyl* pathway (Figure 2). Replica plating can

75 obscure subtle differences in growth rate resulting from disruption of redundant pathways, but

76 these differences are readily observable in our quantitative fitness measurements. We also note

77 that disruption of the genes encoding ring-cleaving dioxygenases, *ligAB* and *xylE*, do not

78 demonstrate measurable fitness defects. DSM12444 contains at least seven ring-cleaving

79 dioxygenases[13], and we hypothesize that, when grown in the presence of high concentrations of

80 PCA, these alternate dioxygenases can substitute for the pathway-specific enzymes.
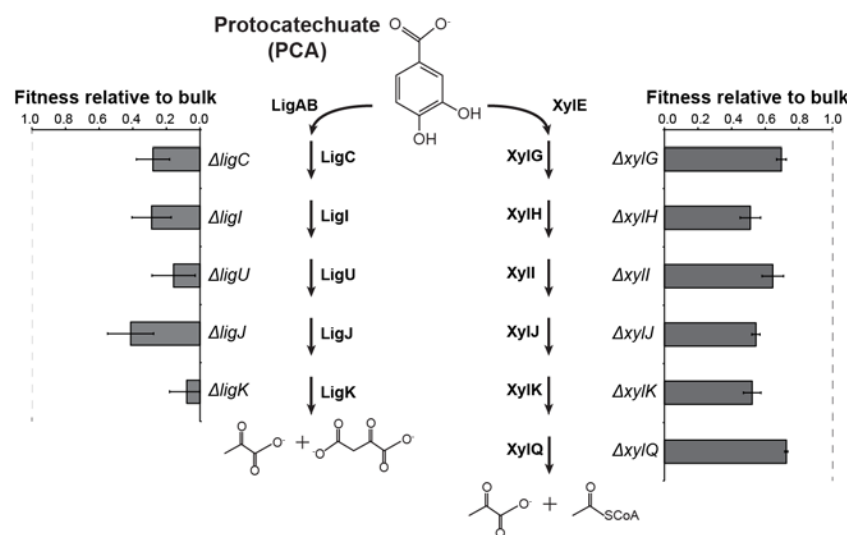

81

82 **Figure 2: Quantitative fitness measurements identify redundant catabolic pathways for**

83 **PCA.** DSM12444 has two redundant pathways for catabolism of PCA, a 4,5-cleavage pathway

84 and a 2,3-cleavage pathway. Fitness costs for each gene disruption were calculated based on the

85 change in population ratio of every mutant with a transposon insertion in the appropriate gene,

86 and are reported as the competitive fitness relative to the bulk population. Disruption of the

87 canonical 4,5-cleavage pathway produces an average 76% decrease in fitness, while disruption of

88 the 2,3-cleavage pathway decreases fitness by an average of 39%. Error bars show one standard

5

89   deviation, calculated from three biological replicates.

90

91   We next examined the genes involved in degradation of ferulate and vanillate. By measuring

92   the fitness effects during growth not just with the ultimate substrate, ferulate, but also during

93   growth with the intermediates vanillate and PCA, we can decompose the entire pathway for

94   ferulate degradation into modules for the conversion of ferulate to vanillate, vanillate to PCA,

95   and PCA to central metabolites (Figure 3). For example, it is clear from our data that *ferAB* and

96   *ligV* are needed for conversion of ferulate to vanillate, but not for further metabolism of

97   vanillate.
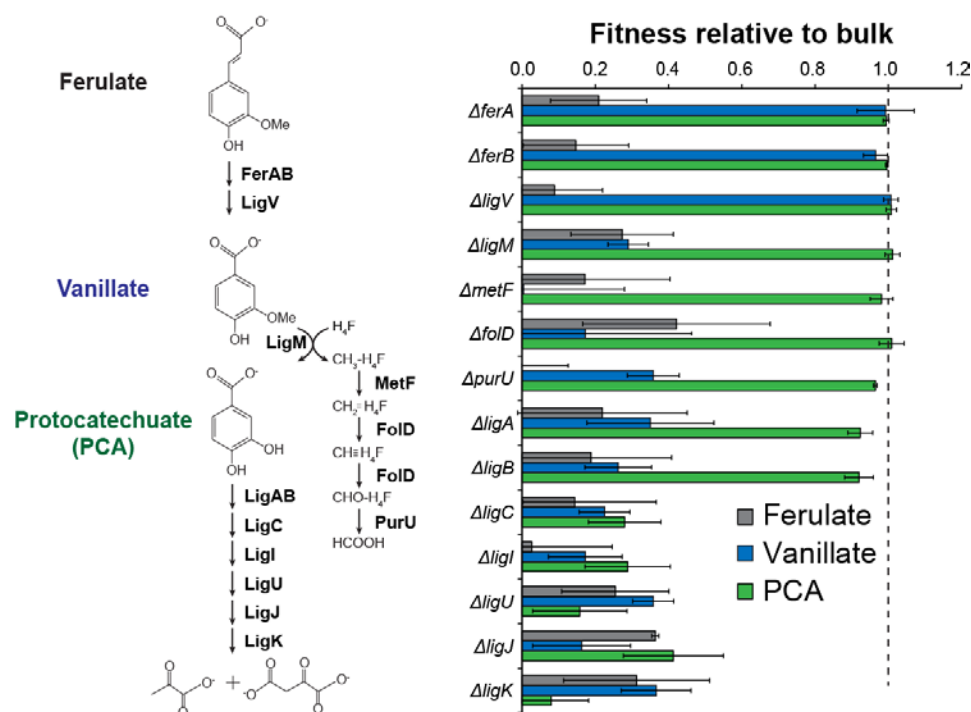

98

99   **Figure 3: BarSeq measurements enable pathway decomposition.** Fitness values were

100   measured during growth with ferulate, vanillate, and PCA, and compared to the bulk fitness.

101   Comparison of the fitness values with different substrates makes the ordering of genes in the

102   pathway readily apparent. Error bars show one standard deviation, calculated from three

103   biological replicates.

104

105   Decomposing a pathway in this manner will simplify the process of matching unknown

106   enzymes to their catalytic activities. However, and in contrast to growth with PCA, growth with

107   vanillate or ferulate requires the genes for the PCA ring-cleaving dioxygenase, *ligAB*. When

108   PCA is an intermediate rather than a substrate, its concentration is likely to be lower, which in

109   turn makes the PCA-specific ring-cleaving dioxygenase more critical. In general, we find that the

110   genetic requirements are more stringent for growth with an upper pathway substrate than an

111   intermediate, and this effect could complicate efforts at pathway elucidation. Prior knowledge of

112   the genes involved in ferulate catabolism[14], accumulated over many years' of work, certainly

113   aided in identifying relevant genes. However, this knowledge was not strictly necessary. For

114   example, eleven genes are shown in Figure 3 as being involved in vanillate degradation.

115   Disruption of these genes produced eleven of the top twelve fitness defects during growth with

116   vanillate, and the twelfth is a hypothetical gene with homology to DNA repair restriction

117   endonucleases.

118   Previous work in *Sphingobium* sp. SYK-6 has demonstrated that the vanillate demethylase

119   LigM transfers a methyl group from the aromatic substrate to tetrahydrofolate. Pathways for

120   conversion of methyltetrahydrofolate to tetrahydrofolate and formate have been characterized in

121   a variety of organisms, involving sequential oxidations by MetF, FolD, and PurU. However,

122   many organisms, including DSM12444, contain multiple homologs of these genes, and the

123   specific catabolic homologs may be necessary for proper function of a heterologously-expressed

124   pathway[15]. During growth with vanillate, specific *metF* and *folD* homologs are required for

125   growth, while the alternative homologs are entirely dispensable (Supplementary Figure 1). The

126 correct *metF* homolog is located in a gene cluster with *ligM*, consistent with its role in vanillate

127 demethylation. However, the dispensable *folD2* homolog is chromosomal, while the necessary

128 *folD* homolog is located on plasmid pNL2.

129   Two pathways are known to deacetylate cinnamic acid derivatives to the corresponding

130 benzoic acid derivatives, either through β-oxidative deacetylation to directly yield the acid or

131 through a combined hydratase/lyase that produces a free benzaldehyde, which can be further

132 oxidized to the acid. The latter pathway is present in *Sphingobium* sp. SYK-6, with FerA and

133 FerB performing the deacetylation[16], followed by oxidation by LigV or DesV[17,18]. Likewise, β-

134 oxidative routes have been described for ferulate and coumarate in *Rhodococcus jostii* RHA1[19]

135 and for ferulate, coumarate, and caffeate in *Agrobacterium fabrum*[20]. However, no β-oxidative

136 pathway has previously been demonstrated for conversion of sinapate to syringate.

137   DSM12444 contains homologs of FerA, FerB, and LigV, and disrupting any of the associated

138 genes severely decreases fitness with coumarate, ferulate, and sinapate (Supplementary Figure

139 2A). However, disruption of an additional three-gene cluster, SARO_RS14535,

140 SARO_RS14545, and SARO_RS14550, also significantly reduces growth with sinapate

141 (Supplementary Figure 2B). SARO_RS14535 and SARO_RS14545 have homology to the

142 phenylhydroxypropionyl-CoA dehydrogenase Atu1415 and the 4-hydroxyphenyl-beta-ketoacyl-

143 CoA hydrolase Atu1421 from *A. fabrum*, respectively[20]. By homology, SARO_RS14550 is

144 likely to be an aromatic aldehyde dehydrogenase, presumably with similar functionality as LigV

145 or DesV [17]. We hypothesize that these enzymes act in parallel with FerA, FerB, and LigV to

146 deacetylate cinnamic acid derivatives. SARO_RS14535 and SARO_RS14550 appear to be

147 specific to sinapate, while SARO_RS14545 contributes equally to catabolism of ferulate,

148 coumarate, and sinapate. *Sphingobium* sp. SYK-6 contains homologs to SARO_RS14535 and

149   SARO_RS14545, SLG_RS06215 and SLG_RS06210, respectively. No additional CoA-ligases,

150   aside from FerA, were identified in our analysis as being involved in activation of sinapate,

151   ferulate, or coumarate.

152   DSM12444 grows efficiently with both syringate and sinapate. In *Sphingobium* sp. SYK-6,

153   three pathways have been identified for syringate degradation. The first step in each pathway is

154   demethylation by DesA to produce 3-*O*-methylgallate (3MGA)[21]. A *desA* homolog is present in

155   DSM12444 and required for growth with sinapate, reducing fitness by 91% when disrupted

156   (Supplementary Figure 3). Conversion of 3MGA into 4-oxalomesaconate (OMA) requires two

157   steps, ring-opening and demethylation. In the first pathway, LigM demethylates 3MGA to

158   gallate, followed by oxidation by DesB or LigAB to yield OMA[22–24]. In the second pathway,

159   DesZ or LigAB oxidizes and demethylates 3MGA to 2-pyrone-4,6-dicarboxylate (PDC),

160   followed by hydrolysis by LigI[25]. Finally, the third pathway proceeds through ring opening by

161   LigAB, followed by at least one uncharacterized reaction to yield OMA[26]. No homologs of *desB*

162   or *desZ* are present in DSM12444. Growth with sinapate requires *ligABUJK* but not *ligI*,

163   suggesting that conversion through PDC is not a major route (Supplementary Figure 3).

164   Since our fitness measurements could efficiently recapitulate known pathways for aromatic

165   catabolism, we sought to use these data to identify the uncharacterized enzymes in this pathway

166   for syringate degradation. Cleavage of 3-*O*-methylgallate produces a mixture of stereoisomers of

167   4-carboxy-2-hydroxy-6-methoxy-6-oxohexa-2,4-dienoate (CHMOD)[25]. The conversion of

168   mixed isomers of CHMOD into OMA is expected to involve a methylesterase and a cis-trans

169   isomerase. Based on homology, SARO_RS14525 is predicted to encode a methylesterase and

170   SARO_RS14530 a glutathione S-transferase. Glutathione S-transferases can function as cis-trans

171   isomerases, such as in the case of maleylpyruvate isomerase[27]. Disruption of either gene is

9

172   detrimental during growth with sinapate, while being dispensable during growth with PCA, 4-

173   HB, coumarate, vanillate, and ferulate (Supplementary Figure 4). Accordingly, we propose that

174   SARO_RS14525, which we name *desC*, demethylates CHMOD, and SARO_RS14530, which

175   we name *desD*, isomerizes one of the intermediates. Notably, *Sphingobium* sp. SYK-6 contains

176   homologs of both *desC* and *desD*, and *desC* occurs in a cluster with the homologs of RS14535,

177   RS14545, *ligM*, *metF*, and *ftfL*.

178   To confirm these results, we constructed clean deletions of *ligA*, *ferA*, *ligM*, *desA*, *desC*, and

179   *desD* in DSM12444. The deletion strains were grown in minimal medium containing glucose, 4-

180   HB, ferulate, or sinapate as the sole source of carbon and energy, and their growth rates were

181   compared to the wild type (Figure 4). As expected, none of the mutations affected growth with

182   glucose, and only the *ligA* deletion decreased growth with 4-HB. Deletion of *ligA*, *ligM*, or *ferA*

183   significantly decreased growth with ferulate, while deletion of *desA*, *desC*, or *desD* had no effect.

184   In contrast, deletion of *ligM* had no impact on growth with sinapate, while deletion of *desA* and

185   *desC* severely decreased growth and deletion of *desD* had a moderate effect. Since *ligM* is

186   dispensable but *desC* is required for growth with sinapate, we conclude that DSM12444 degrades

187   sinapate solely through CHMOD. We hypothesize that, as shown in SYK-6, cleavage of the 3-

188   MGA ring by LigAB produces a mixture of stereoisomers, which are resolved by DesD. In the

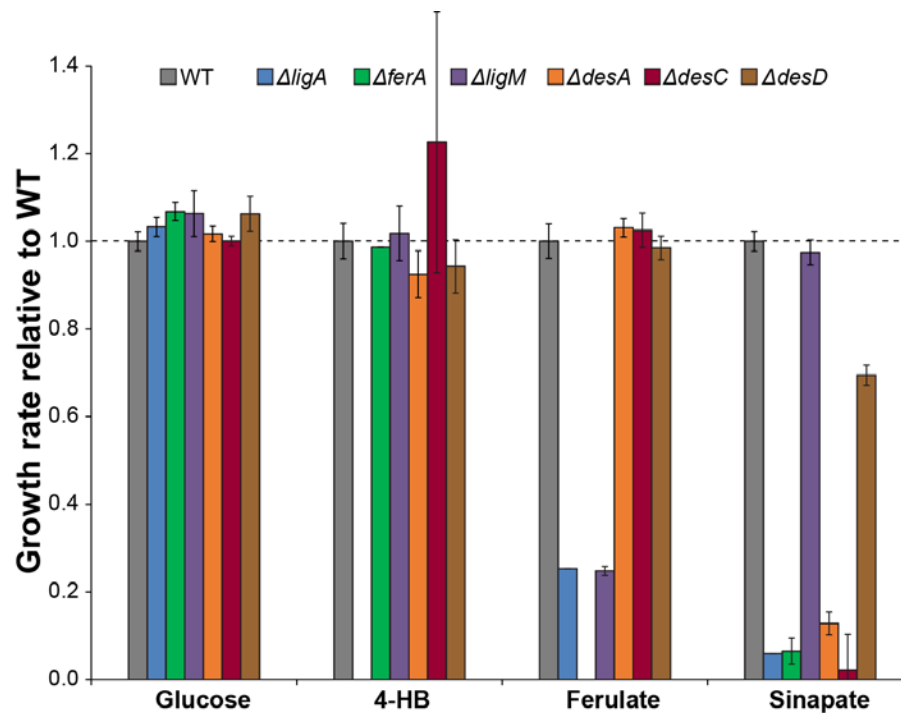189   absence of *desD*, growth is permitted, but inefficient.

Figure 4: Gene deletions confirm bulk fitness measurements and identify enzymes for sinapate degradation. Mutant strains with clean deletions of the indicated genes were grown in minimal medium containing the indicated carbon source. Disruptions of the newly-identified genes *desC* and *desD* decrease growth with sinapate, but not with any of the other substrates tested. Error bars show one standard deviation, calculated from three biological replicates.

Due to the inherent heterogeneity of lignin, its degradation produces a wide range of aromatic compounds[9]. Engineering microbes to productively convert these aromatic compounds into valuable fuels and chemicals will require the identification, characterization, and integration of multiple, complex metabolic pathways. To meet this challenge, the tools employed must scale with the scope of the networks[6]. In this work, we have shown that randomly-barcoded transposon mutagenesis provides the necessary scale for pathway discovery, by rapidly and quantitatively identifying genes that affect catabolism across panels of substrates.

11

Materials and Methods

Media and chemicals

All chemicals were purchased from Sigma-Aldrich (St. Louis, MO) or Fisher Scientific (Fairlawn, NJ) and were molecular grade. All oligonucleotides were ordered from IDT (Coralville, IA). *E. coli* were routinely cultivated at 37 ºC in LB, while *N. aromaticivorans* were grown at 30 ºC in LB or DSM Medium 457 containing the indicated carbon source. Aromatic monomers were dissolved in DMSO at 100 g/L and added at a working concentration of 1 g/L. As needed, cultures were supplemented with kanamycin at 50 mg/L, streptomycin at 100 mg/L, or diaminopimelic acid (DAP) at 60 mg/L.

Strains and plasmids

N. aromaticivorans DSM12444 was purchased from the DSMZ (Braunschweig. Germany). Strain APA766, comprising the conjugation donor WM3064 containing the barcoded Tn5 vector pKMW7, was a gift from Adam Deutschbauer. The parental deletion plasmid, pAK405, was a gift from Julia Vorholt (Addgene plasmid # 37114). Deletion plasmids pJM269-280 were synthesized *de novo* and cloned into pAK405 by Genscript (Piscataway, NJ).

Transposon library construction and indexing

DSM12444 and APA766 were grown overnight in LB, containing kanamycin and DAP as appropriate. DSM12444 was then diluted 1:100 into 50 mL of fresh LB and regrown to mid-log phase. Both cultures were pelleted by centrifugation, mixed at a ratio of 5:1 recipient:donor, and plated on sterile filter paper placed on an LB+DAP agar plate. After overnight incubation at 30 ºC, the filter paper was removed, vortexed with 5 mL of fresh LB, pelleted by centrifugation, resuspended in 1 mL of fresh LB, and plated on 20 LB+kanamycin plates. Colonies were observed after 3 days of incubation at 30 ºC. The colonies were scraped from the plates,

227   combined into a single mixed culture, and grown overnight in 50 mL of LB+kanamycin at 30 ºC.

228   Aliquots of the combined library were stored in 7% DMSO at -80 ºC for further analysis.

229   Additionally, genomic DNA was prepared using a DNeasy Blood and Tissue kit (Qiagen,

230   Valencia, CA) according to the manufacturer's directions.

231   Barcode indexing was performed based on published protocols[4,28]. Briefly, genomic DNA

232   was fragmented by sonication using a Bioruptor Plus (Diagenode, Ougrée, Belgium) and size-

233   selected to approximately 300bp using Agencourt AMPure XP beads (Beckman Coulter,

234   Indianapolis, IN). Size selection was confirmed using a Bioanalyzer DNA1000 chip (Agilent,

235   Santa Clara, CA). End repair, dA-tailing, and adaptor ligation were performed using the

236   NEBNext kit according to the manufacturer's directions (NEB, Ipswitch, MA). Custom adaptors

237   were synthesized by IDT following published sequences[4]. In lieu of a final size selection, two

238   sequential purifications using AMPure XP beads were performed instead[28]. The final transposon

239   enrichment PCR was performed using KAPA HiFi HotStart ReadyMix[28] using custom indexing

240   and transposon-specific primers[4]. The library was analyzed using a DNA1000 chip, quantified

241   using a Qubit fluorimeter (ThermoFisher, Waltham, MA), and sequenced on an Illumina MiSeq,

242   using v2 chemistry with paired-end 150bp reads (Illumina, San Diego, CA). Reads were

243   analyzed using custom scripts as described previously[4].

244   **Pooled fitness measurements**

245   An aliquot of the DSM12444 library was thawed, diluted into 50mL of LB+kanamycin, and

246   grown overnight at 30 ºC. The culture was then diluted 1:100 into three flasks, each containing

247   MM457+kanamycin+2 g/L glucose and grown to saturation. Time-zero samples were taken from

248   each flask, and each culture was then diluted to an optical density of 0.005 in 5mL of

249   MM457+kanamycin containing the appropriate carbon sources. Cultures were grown to

13

250   saturation at 30 ºC and then frozen for later analysis. Genomic DNA was isolated and barcodes

251   were amplified using custom indexing primers as described previously [4]. Barcode amplicons

252   were quantified using a Qubit fluorimeter, pooled, and sequenced at the Joint Genome Institute

253   on a HiSeq 2500 lane. Barcode frequencies and gene fitness values were calculated using custom

254   scripts as described previously[4].

Gene deletions

256   In-frame gene deletions were constructed using a kanamycin/*rpsL* suicide vector[29]. Briefly,

257   appropriate deletion cassettes were synthesized and cloned into pAK405. The resulting plasmids

258   were transformed into strain WM6026, a conjugation-proficient DAP auxotroph. Plasmids were

259   transferred to DSM12444 by conjugation following the same protocol as for library construction.

260   After selecting for single crossovers on LB+kanamycin plates, colonies were patched on

261   LB+kanamycin to confirm plasmid integration, grown in the absence of selection, and then

262   plated on LB+streptomycin plates. Colonies were patched on LB+streptomycin and replica

263   plated on LB+streptomycin+kanamycin to confirm double recombination. Deletions were

264   identified by colony PCR using locus-specific primers and confirmed by Sanger sequencing of

265   the amplicons.

Growth rate measurements

267   Strains were grown to saturation overnight in minimal medium with 2 g/L glucose. They

268   were then diluted 100x into fresh medium containing the appropriate carbon source and grown as

269   triplicate 100 µL cultures in a Bioscreen C plate reader (Oy Growth Curves Ab Ltd, Helsinki,

270   Finland). Growth was monitored using optical density at 600 nm (OD_{600}). Growth rates were

271   calculated during exponential growth using CurveFitter software[30].

272

Acknowledgements

Author Contributions

JKM conceived of the project. JC and JKM designed and performed experiments, analyzed the data, and wrote the paper.

Funding Information

References

296

297 1.   Simon, R., Priefer, U. & P?hler, A. A Broad Host Range Mobilization System for In Vivo

298   Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria.

299   *Bio/Technology* **1,** 784–791 (1983).

300 2.   Don, R. H., Weightman, A. J., Knackmuss, H. J. & Timmis, K. N. Transposon

301   mutagenesis and cloning analysis of the pathways for degradation of 2,4-

302   dichlorophenoxyacetic acid and 3-chlorobenzoate in Alcaligenes eutrophus

303   JMP134(pJP4). *J. Bacteriol.* **161,** 85–90 (1985).

304 3.   Larsen, R., Wilson, M., Guss, A. & Metcalf, W. Genetic analysis of pigment biosynthesis

305   in Xanthobacter autotrophicus Py2 using a new, highly efficient transposon mutagenesis

306   system that is functional in a wide variety of bacteria. *Arch. Microbiol.* **178,** 193–201

307   (2002).

308 4.   Wetmore, K. M. *et al.* Rapid Quantification of Mutant Fitness in Diverse Bacteria by

309   Sequencing Randomly Bar-Coded Transposons. *MBio* **6,** e00306-15 (2015).

310 5.   Masai, E., Katayama, Y. & Fukuda, M. Genetic and Biochemical Investigations on

311   Bacterial Catabolic Pathways for Lignin-Derived Aromatic Compounds. (2007).

312   doi:10.1271/bbb.60437

313 6.   Hanson, A. D., Pribat, A., Waller, J. C. & Crécy-Lagard, V. de. 'Unknown' proteins and

314   'orphan' enzymes: the missing half of the engineering parts list – and how to find it.

315   *Biochem. J.* **425,** (2010).

316 7.   Johnson, C. W. & Beckham, G. T. Aromatic catabolic pathway selection for optimal

317   production of pyruvate and lactate from lignin. *Metab. Eng.* **28,** 240–247 (2015).

318 8.   Clarkson, S. M., Kridelbaugh, D. M., Elkins, J. G., Guss, A. M. & Michener, J.

319   Construction and optimization of a heterologous pathway for protocatechuate catabolism

320   in Escherichia coli enables rapid bioconversion of model lignin monomers. *bioRxiv*

321   (2017).

322 9.   Bugg, T. D. H., Ahmad, M., Hardiman, E. M. & Rahmanpour, R. Pathways for

323   degradation of lignin in bacteria and fungi. *Nat. Prod. Rep.* **28,** 1883 (2011).

324 10.   Fredrickson, J., Balkwill, D., Romine, M. & Shi, T. Ecology, physiology, and phylogeny

325   of deep subsurface Sphingomonas sp. *J. Ind. Microbiol. Biotechnol.* **23,** 273–283 (1999).

326 11.   Fredrickson, J. K. *et al.* Aromatic-degrading Sphingomonas isolates from the deep

327   subsurface. *Appl. Environ. Microbiol.* **61,** 1917–22 (1995).

328 12.   Sato, Y. *et al.* Identification of three alcohol dehydrogenase genes involved in the

329   stereospecific catabolism of arylglycerol-beta-aryl ether by Sphingobium sp. strain SYK-

330   6. *Appl. Environ. Microbiol.* **75,** 5195–201 (2009).

331 13.   D'Argenio, V. *et al.* Complete sequencing of Novosphingobium sp. PP1Y reveals a

332   biotechnologically meaningful metabolic pattern. *BMC Genomics* **15,** 384 (2014).

333 14.   Kamimura, N. & Masai, E. in *Biodegradative Bacteria* 207–226 (Springer Japan, 2014).

334   doi:10.1007/978-4-431-54520-0_10

335 15.   Michener, J. K., Vuilleumier, S., Bringel, F. & Marx, C. J. Transfer of a Catabolic

336   Pathway for Chloromethane in Methylobacterium Strains Highlights Different Limitations

337   for Growth with Chloromethane or with Dichloromethane. *Front. Microbiol.* **7,** 1116

338   (2016).

339 16.   Masai, E. *et al.* Cloning and characterization of the ferulic acid catabolic genes of

340   Sphingomonas paucimobilis SYK-6. *Appl. Environ. Microbiol.* **68,** 4416–24 (2002).

341 17.   Kamimura, N. *et al.* A bacterial aromatic aldehyde dehydrogenase critical for the efficient

342   catabolism of syringaldehyde. *Sci. Rep.* **7,** 44422 (2017).

343   18.   Masai, E. *et al.* Characterization of ligV essential for catabolism of vanillin by

344   Sphingomonas paucimobilis SYK-6. *Biosci. Biotechnol. Biochem.* **71,** 2487–92 (2007).

345   19.   Otani, H., Lee, Y.-E., Casabon, I. & Eltis, L. D. Characterization of p-hydroxycinnamate

346   catabolism in a soil Actinobacterium. *J. Bacteriol.* **196,** 4293–303 (2014).

347   20.   Campillo, T. *et al.* Analysis of hydroxycinnamic acid degradation in Agrobacterium

348   fabrum reveals a coenzyme A-dependent, beta-oxidative deacetylation pathway. *Appl.*

349   *Environ. Microbiol.* **80,** 3341–9 (2014).

350   21.   Masai, E. *et al.* A novel tetrahydrofolate-dependent O-demethylase gene is essential for

351   growth of Sphingomonas paucimobilis SYK-6 with syringate. *J. Bacteriol.* **186,** 2757–65

352   (2004).

353   22.   Kasai, D., Masai, E., Miyauchi, K., Katayama, Y. & Fukuda, M. Characterization of the

354   gallate dioxygenase gene: three distinct ring cleavage dioxygenases are involved in

355   syringate degradation by Sphingomonas paucimobilis SYK-6. *J. Bacteriol.* **187,** 5067–74

356   (2005).

357   23.   Nogales, J. *et al.* Unravelling the gallic acid degradation pathway in bacteria: the gal

358   cluster from Pseudomonas putida. *Mol. Microbiol.* **79,** 359–374 (2011).

359   24.   Mazurkewich, S., Brott, A. S., Kimber, M. S. & Seah, S. Y. K. Structural and Kinetic

360   Characterization of the 4-Carboxy-2-hydroxymuconate Hydratase from the Gallate and

361   Protocatechuate 4,5-Cleavage Pathways of Pseudomonas putida KT2440. *J. Biol. Chem.*

362   **291,** 7669–86 (2016).

363   25.   Kasai, D., Masai, E., Miyauchi, K., Katayama, Y. & Fukuda, M. Characterization of the 3-

364   O-methylgallate dioxygenase gene and evidence of multiple 3-O-methylgallate catabolic

365   pathways in Sphingomonas paucimobilis SYK-6. *J. Bacteriol.* **186,** 4951–9 (2004).

366 26. Kasai, D., Masai, E., Katayama, Y. & Fukuda, M. Degradation of 3- *O* -methylgallate in

367   *Sphingomonas paucimobilis* SYK-6 by pathways involving protocatechuate 4,5-

368   dioxygenase. *FEMS Microbiol. Lett.* **274,** 323–328 (2007).

369 27. Vuilleumier, S. *et al.* The glutathione S-transferase OrfE3 of the dioxin-degrading

370   bacterium Sphingomonas sp. RW1 displays maleylpyruvate isomerase activity. *Chem.-*

371   *Biol. Interact* **133,** 265–267 (2001).

372 28. Barquist, L. *et al.* The TraDIS toolkit: sequencing and analysis for dense transposon

373   mutant libraries. *Bioinformatics* **32,** 1109–11 (2016).

374 29. Kaczmarczyk, A., Vorholt, J. A. & Francez-Charlot, A. Markerless Gene Deletion System

375   for Sphingomonads. *Appl. Environ. Microbiol.* **78,** 3774–3777 (2012).

376 30. Delaney, N. F. *et al.* Development of an Optimized Medium, Strain and High-Throughput

377   Culturing Methods for *Methylobacterium extorquens*. *PLoS One* **8,** e62957 (2013).

378