AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

BIODEGRADATION

Rapid, Parallel Identification of Catabolism Pathways of Lignin-Derived Aromatic Compounds in *Novosphingobium aromaticivorans*

Jacob H. Cecil,[a,b*] David C. Garcia,[a,e] Richard J. Giannone,[b,c,d] Joshua K. Michener[a,b,d]

[a]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
[b]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
[c]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
[d]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
[e]The Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, Tennessee, USA

ABSTRACT Transposon mutagenesis is a powerful technique in microbial genetics for the identification of genes in uncharacterized pathways. Recently, the throughput of transposon mutagenesis techniques has been dramatically increased through the combination of DNA barcoding and high-throughput sequencing. 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 used this technique to identify two genes, *desC* and *desD*, which are involved in the degradation of the lignin-derived aromatic compound sinapic acid in the nonmodel bacterium *Novosphingobium aromaticivorans*. We show that DesC is a methyl esterase acting on an intermediate formed during sinapic acid catabolism, providing the last enzyme in a proposed catabolic pathway. This approach will be particularly useful in the identification of complete pathways suitable for heterologous expression in metabolic engineering.

IMPORTANCE The identification of the genes involved in specific biochemical transformations is a key step in predicting microbial function from nucleic acid sequences and in engineering microbes to endow them with new functions. We have shown that new techniques for transposon mutagenesis can dramatically simplify this process and enable the rapid identification of genes in uncharacterized pathways. These techniques provide the necessary scale to fully elucidate complex biological networks such as those used to degrade mixtures of lignin-derived aromatic compounds.

KEYWORDS aerobic catabolism, genetics, metabolic engineering, transposons

Microbes are capable of degrading a vast array of natural and nonnatural small molecules by using an intricate network of metabolic pathways. Many techniques in modern microbiology rely on predicting microbial phenotypes such as catabolic potential on the basis of nucleic acid sequencing, but these predictions are only as good as our functional annotations (1). Similarly, metabolic engineers depend on accurate pathway predictions to reliably modify microbial metabolism (2). In both cases, improved methods for enzyme discovery and pathway annotation would be highly beneficial. Traditionally, the elucidation of metabolic networks can take many years, particularly for ancestral pathways scattered throughout the genome (3, 4). New

Received 21 May 2018 **Accepted** 5 September 2018

Accepted manuscript posted online 14 September 2018

Citation Cecil JH, Garcia DC, Giannone RJ, Michener JK. 2018. Rapid, parallel identification of catabolism pathways of lignin-derived aromatic compounds in *Novosphingobium aromaticivorans*. Appl Environ Microbiol 84:e01185-18. https://doi.org/10.1128/AEM.01185-18.

Editor Harold L. Drake, University of Bayreuth

Address correspondence to Joshua K. Michener, michenerjk@ornl.gov.

* Present address: Jacob H. Cecil, Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona, USA.

methods, however, offer the opportunity to speed up and simplify the process of pathway elucidation.

In the case of catabolic pathways, enzymes were historically identified by replica plating of induced or spontaneous mutants followed by the screening for colonies that had lost the ability to grow with the compound of interest (5). Random transposon mutagenesis simplified the processes of both mutant generation, since transposon insertion into a gene typically results in a loss of function, and the mapping of the disrupted gene, through amplification of the transposon insertion junction (6, 7). The barcoding of transposon mutant pools enabled the rapid identification of depleted essential genes; although initially, the mapping of the barcode to the insertion locus was challenging (8). More recently, the introduction of high-throughput sequencing enabled genome-wide measurements of transposon mutant frequencies by amplifying and sequencing the transposon insertion junction (TnSeq); yet, the difficulty of sample preparation still limited the throughput (9–12). By combining the throughput of barcode sequencing with the genome-wide scope of transposon insertion sequencing, randomly barcoded transposon insertion sequencing (RB-TnSeq) can rapidly test the effects of transposon insertions across the entire ensemble of nonessential genes under multiple conditions in a single experiment (13, 14). These properties can be exploited to simplify the process of pathway identification.

Sphingomonads are known to catabolize a broad range of xenobiotic aromatic compounds, using a core set of catabolic pathways that likely evolved to metabolize naturally occurring aromatic compounds such as those derived from lignin (15). Detailed studies, with a particular focus on *Sphingobium* sp. strain SYK-6, identified the enzymes that comprise many of these catabolic pathways (3, 16). However, less is known about the similarity of the catabolic pathways in closely related sphingomonads. We chose *Novosphingobium aromaticivorans* DSM12444, a bacterium isolated from contaminated sediment, as a model system for pathway identification (17, 18). This strain has been shown to catabolize a wide range of aromatic compounds, and its catalytic repertoire is predicted to include enzymes that degrade model lignin-derived biaryls, such as guaiacylglycerol-β-guaiacyl ether (19).

Traditional pathway identification by transposon mutagenesis has four major challenges. First, random mutagenesis is inherently a statistical process, and finding a sufficient number of random mutants to fully complete a pathway can be difficult. Second, the demonstration that certain genes are not involved in a pathway of interest requires additional targeted genetic modifications. Third, redundant pathways can limit the effects of single gene knockouts. Fourth, the identification of the order in which genes function in a pathway can be challenging, particularly for long pathways.

In this work, we show that we can overcome the first challenge as a result of the large libraries enabled by TnSeq. Using this approach, we recovered a 14-gene pathway for the catabolism of ferulic acid in a single experiment. We also demonstrated a solution to the second challenge, by differentiating between enzyme homologs that are and those that are not required for the catabolism of vanillate. For the third challenge, we quantitatively measured the effect of disrupting two redundant pathways for catabolism of protocatechuate, showing that both pathways contribute to catabolism but to different degrees. Fourth, taking advantage of the parallelism enabled by barcode sequencing, we decomposed a long pathway for ferulate catabolism into smaller metabolic modules that are more readily analyzed. Finally, we applied these techniques to an uncharacterized pathway for sinapate catabolism and identified two novel enzymes that complete this pathway. In combination, these capabilities offer the opportunity to dramatically accelerate enzyme and pathway discovery (Fig. 1).

RESULTS

Transposon library generation and sequencing. To begin identifying the genes involved in aromatic degradation in *N. aromaticivorans* DSM12444, we constructed a barcoded transposon library for this strain. A single round of transposon insertion sequencing was performed to map barcodes to insertion sites, identifying a total of

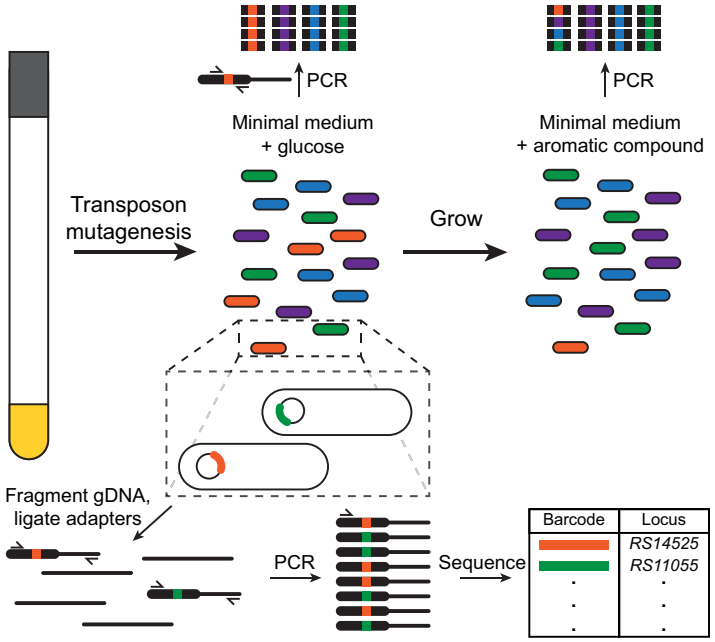

FIG 1 Randomly barcoded transposon mutagenesis enables high-throughput pathway identification. A pool of barcoded transposons is introduced into the target bacterium. A single round of TnSeq maps barcodes (colored inserts) to the insertion locus. The barcodes can be queried through PCR and high-throughput sequencing to determine changes in the population ratio during growth with a specific compound.

43,270 unique barcoded insertions in 3,117 of 3,897 protein-coding genes. The pooled transposon library was then grown, in triplicates, in minimal medium containing protocatechuate (PCA), 4-hydroxybenzoate (4-HB), coumarate, vanillate, ferulate, syringate, or sinapate as the sole source of carbon and energy. The mixed barcodes were amplified before and after growth and then sequenced to determine the fitness effect of disrupting each nonessential gene under the relevant conditions. Fitness costs for each gene disruption were calculated on the basis of the change in population ratio of every mutant with a transposon insertion in the appropriate gene and are reported as the competitive fitness relative to that of the bulk population (13). Specific gene disruptions had substantial substrate-dependent fitness costs, and the top genes for each condition are summarized in Data Set S1 in the supplemental material.

Identifying redundant pathways for PCA catabolism. Sphingomonads are known to catabolize PCA through 4,5-cleavage of the aromatic ring (20). Strain DSM12444 contains chromosomal copies of the genes necessary for this cleavage pathway, *ligABCIUJK*, as well as plasmid-borne genes for the lower 2,3-cleavage pathway for catechol, *xylEGHIJKQ* (21). During growth with PCA, a disruption of either pathway produces a consistent decrease in fitness, averaging a 76% decrease in fitness for an insertion in the pathway encoded by *lig* genes or a 39% decrease in fitness for an insertion in the pathway encoded by *xyl* (Fig. 2). We confirmed this result by constructing clean deletions of *xylG* and *ligU*. As predicted, these strains showed a growth defect during growth with PCA (see Fig. S2). By measuring the effect of transposon insertions quantitatively, rather than qualitatively, we can identify redundant pathways that have additive effects.

Pathway decomposition for analysis of ferulate catabolism. We next examined the genes involved in the degradation of ferulate and vanillate. By measuring the fitness effects during growth not just with the ultimate substrate, ferulate, but also with the intermediates vanillate and PCA, we decomposed the entire pathway for ferulate degradation into modules for the conversions of ferulate to vanillate, vanillate to PCA, and PCA to central metabolites (Fig. 3). For example, it was clear from our data that

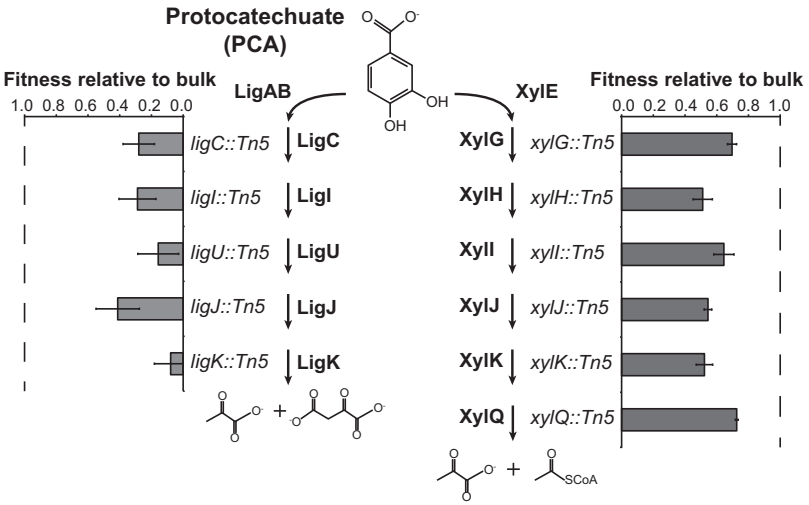

FIG 2 Quantitative fitness measurements identify redundant catabolic pathways for PCA. Strain DSM12444 has two redundant pathways for catabolism of PCA: a 4,5-cleavage pathway and a 2,3-cleavage pathway. Disruption of the canonical 4,5-cleavage pathway produces an average 76% decrease in fitness, while disruption of the 2,3-cleavage pathway decreases fitness by an average of 39%. Error bars show one standard deviation, calculated from three biological replicates.

ferAB and *ligV* are needed for the conversion of ferulate to vanillate but not for further metabolism of vanillate.

Differentiating between enzyme homologs in vanillate catabolism. Previous work in *Sphingobium* sp. strain SYK-6 demonstrated that the vanillate demethylase LigM transfers a methyl group from the aromatic substrate to tetrahydrofolate, producing PCA (22). The pathways for the conversion of methyltetrahydrofolate to tetrahydrofolate and formate, involving sequential oxidations by MetF, FolD, and PurU, have

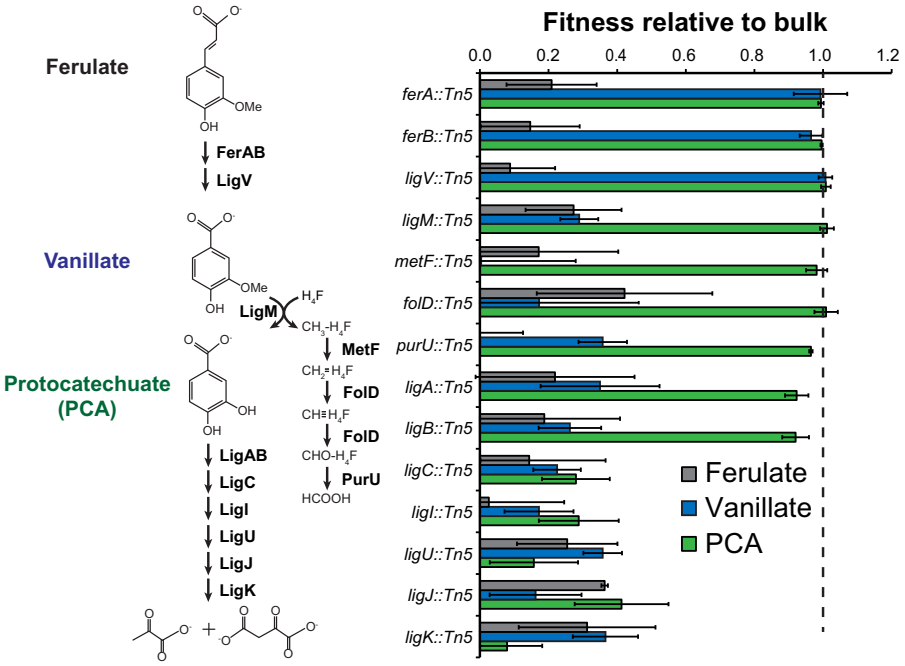

FIG 3 Barcoded transposon insertion measurements enable pathway decomposition. Fitness values were measured during growth with ferulate, vanillate, and PCA and compared to the bulk fitness. Comparison of the fitness values with different substrates makes the ordering of genes in the pathway readily apparent. Error bars show one standard deviation, calculated from three biological replicates.

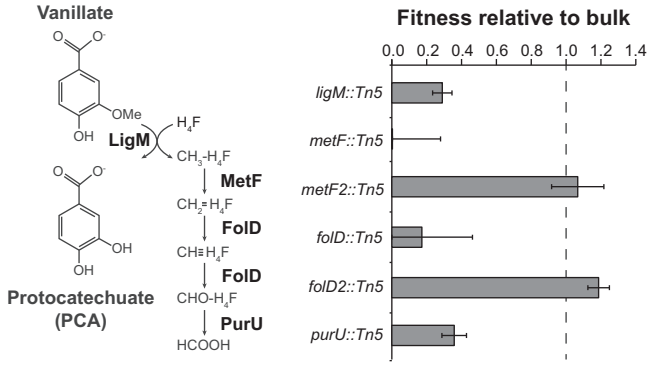

FIG 4 Vanillate demethylation requires specific accessory homologs. Vanillate degradation involves the transfer of a methyl group from vanillate to tetrahydrofolate, which is then oxidized to formate. Two homologs each of *metF* and *folD* are present in the strain DSM12444 genome. Pooled fitness measurements readily identify the specific homologs that are required for vanillate demethylation. Error bars show one standard deviation, calculated from three biological replicates.

been characterized in a variety of organisms (23, 24). However, many organisms contain multiple homologs of these genes, and the specific catabolic homologs may be necessary for the proper function of a heterologously expressed pathway (25). Strain DSM12444 contains two homologs each of *metF* and *folD*. During growth with vanillate, one *metF* and *folD* homolog is required for growth, while the alternative homolog is entirely dispensable (Fig. 4). The correct *metF* homolog is located in a gene cluster with *ligM*, consistent with its role in vanillate demethylation. However, the dispensable *folD2* homolog is chromosomal, while the necessary *folD* homolog is located on plasmid pNL2.

Identification of a novel pathway for syringate catabolism. Strain DSM12444 grows efficiently with both syringate and sinapate. The conversion of sinapate to syringate is hypothesized to follow the same route as other phenylpropanoids, by using the proteins encoded by *ferAB* and *ligV* to convert sinapate first to syringaldehyde and then to syringate. Consistent with this hypothesis, transposon insertions in any of those genes decreases the fitness during growth with sinapate but not with syringate (see Fig. S3). In *Sphingobium* sp. SYK-6, three pathways have been identified for syringate degradation. The first step in each pathway is demethylation by DesA to produce 3-*O*-methylgallate (3MGA) (26). A *desA* homolog is present in strain DSM12444 and required for growth with sinapate, reducing fitness by 91% when disrupted (Fig. 5). The conversion of 3MGA into 4-oxalomesaconate (OMA) requires two steps, namely, ring-opening and demethylation. In the first pathway, LigM demethylates 3MGA to gallate, followed by oxidation by DesB or LigAB to yield OMA (27–29). In the second pathway, DesZ or LigAB oxidizes and demethylates 3MGA to 2-pyrone-4,6-dicarboxylate (PDC), followed by ring hydrolysis using LigI (30). Finally, the third pathway proceeds through ring opening by LigAB, followed by at least one uncharacterized reaction to yield OMA (31). No homologs of *desB* or *desZ* are present in strain DSM12444. Growth with sinapate requires *ligABUJK* but not *ligI*, suggesting that the conversion through PDC is not a major route (Fig. 5).

Since our fitness measurements efficiently recapitulated known pathways for aromatic catabolism, we sought to use these data to identify the uncharacterized enzymes in this pathway for syringate degradation. The cleavage of 3MGA produces a mixture of stereoisomers of 4-carboxy-2-hydroxy-6-methoxy-6-oxohexa-2,4-dienoate (CHMOD) (30). The conversion of mixed isomers of CHMOD into OMA is expected to involve a methylesterase and a *cis-trans* isomerase. On the basis of homology, SARO_RS14525 is predicted to encode a methylesterase and SARO_RS14530 a glutathione *S*-transferase. Glutathione *S*-transferases can function as *cis-trans* isomerases, such as in the case of maleylpyruvate isomerase (32). The disruption of either gene is detrimental during growth with sinapate, but the genes are dispensable during growth with PCA, 4-HB,

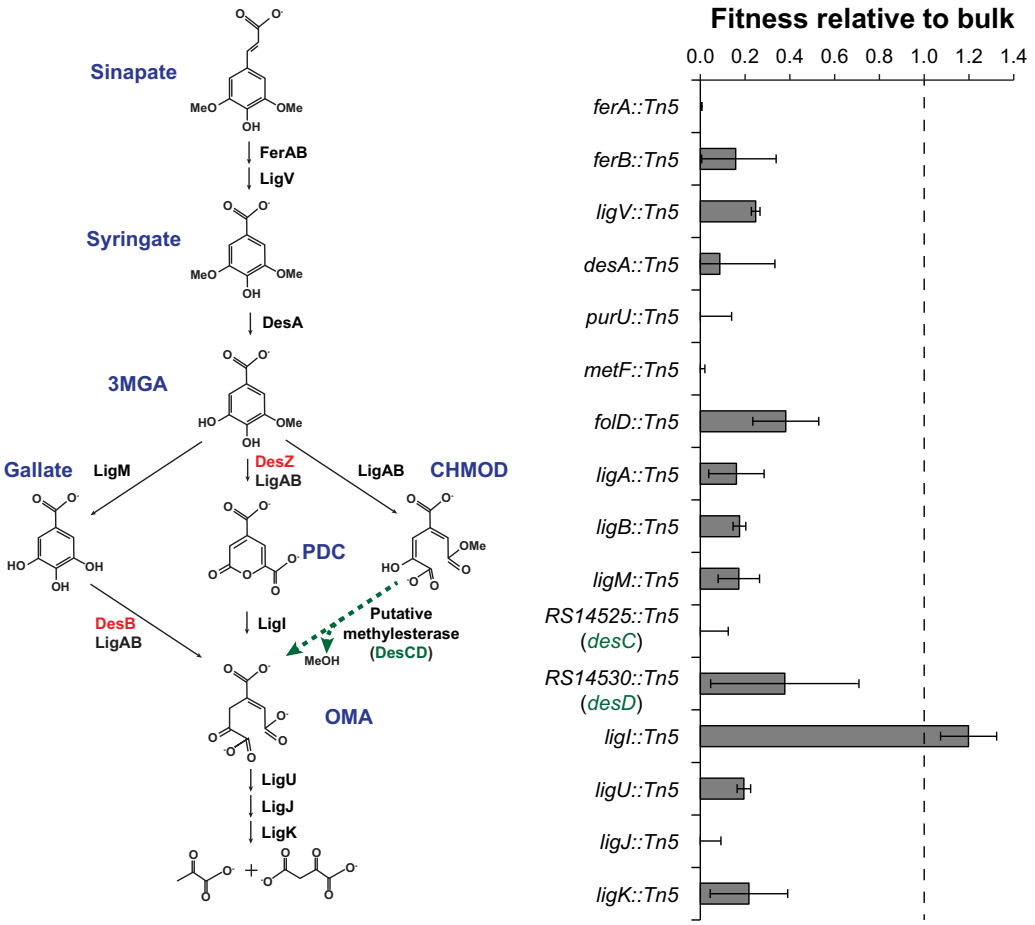

FIG 5 Identification of genes involved in sinapate degradation. Three pathways are known to be involved in the degradation of sinapate in *Sphingobium* sp. strain SYK-6. Homologs of genes in two of those pathways, *desB* and *desZ* (red), are not present in strain DSM12444. Newly identified genes, *desC* and *desD*, are predicted to perform the missing reactions (green). Error bars show one standard deviation, calculated from three biological replicates.

coumarate, vanillate, and ferulate (Fig. S3). Accordingly, we propose that SARO_RS14525, which we named *desC*, demethylates CHMOD, and SARO_RS14530, which we named *desD*, isomerizes one of the intermediates.

To confirm these results, we constructed clean deletions of *ligA*, *ferA*, *ligM*, *desA*, *desC*, and *desD* in strain DSM12444. The deletion strains were grown in minimal medium containing glucose, 4-HB, ferulate, or sinapate as the sole source of carbon and energy, and their growth rates were compared to the that of the wild type (Fig. 6A; see also Fig. S5). As expected, none of the mutations affected growth with glucose, and only the *ligA* deletion decreased growth with 4-HB. The deletion of *ligA*, *ligM*, or *ferA* significantly decreased growth with ferulate, while the deletion of *desA*, *desC*, or *desD* had no effect. In contrast, the deletion of *ligM* had no impact on growth with sinapate, while the deletion of *desA* and *desC* severely decreased growth and the deletion of *desD* had a moderate effect. Since *ligM* is dispensable but *desC* is required for growth with sinapate, we conclude that strain DSM12444 degrades sinapate solely through CHMOD. We hypothesize that, as shown in strain SYK-6, cleavage of the 3MGA ring by LigAB produces a mixture of stereoisomers, which are resolved by DesD. In the absence of *desD*, growth is permitted but is inefficient.

Biochemical characterization of DesCD. We next grew the knockout strains in minimal medium containing glucose and sinapate. Supernatant samples were collected and analyzed by liquid chromatography-mass spectrometry (LC-MS) (Fig. 6B). As expected, the deletion of *ferA* prevented any conversion of sinapate. The deletion of *desA*

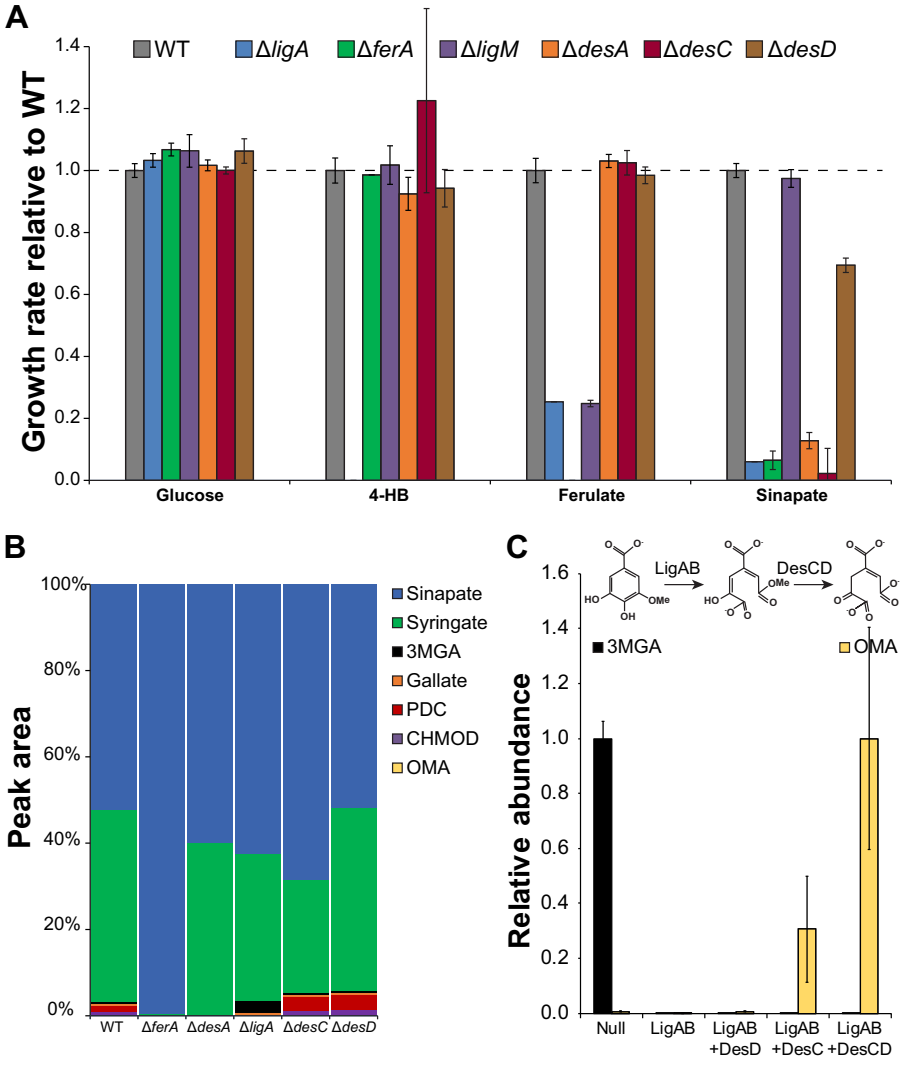

FIG 6 Gene deletions identify enzymes for sinapate degradation. (A) Mutant strains with clean deletions of the indicated genes were grown in minimal medium containing the indicated carbon source. Disruptions of the newly identified genes *desC* and *desD* decrease growth with sinapate but not with any of the other substrates tested. Error bars show one standard deviation, calculated from three biological replicates. (B) Mutant strains were grown in minimal medium containing glucose and sinapate. Accumulation of downstream metabolites was monitored by LC-MS. Results show the averages from two technical and three biological replicates. (C) Crude lysates expressing the indicated enzymes were combined and incubated with 3MGA. Conversion was monitored by LC-MS. Abundance is calculated relative to the highest measurement for a given compound. Error bars show the standard deviations from three biological replicates.

enabled production of syringate but no further metabolism. Similarly, the deletion of *ligA* led to the accumulation of 3MGA and gallate but blocked the production of the ring cleavage products. The deletion of *desC* or *desD* led to an increased accumulation of CHMOD ($P = 0.05$, two-tailed *t* test), consistent with our hypothesis that DesCD are responsible for further metabolism of CHMOD.

To confirm the role of DesCD in the demethylation of CHMOD, we heterologously expressed LigAB and DesCD in *Escherichia coli*. We incubated 3MGA with crude lysates containing LigAB alone or in combination with DesC and/or DesD and monitored the conversion of the substrate (Fig. 6C). The 3MGA was consumed in the presence of LigAB, though we were unable to uniquely detect all the intermediates, most notably, CHMOD. As predicted, the addition of DesC was required for the production of OMA, while the addition of DesD enhanced OMA production when combined with DesC.

DISCUSSION

In this work, we used barcoded transposon mutant libraries to rapidly identify the pathways for lignin-derived aromatic compounds. This approach has several advantages over traditional library screening based on replica plating. First, replica plating can identify mutations that abolish growth but is challenging to use for the identification of subtle differences in growth rate, such as those resulting from the disruption of redundant pathways. These differences are readily observable in our quantitative fitness measurements. Second, the throughput of replica plating is limited by the number of colonies and plates that can conveniently be screened, and each new substrate requires a new screen. In our approach, an entire library can be screened in replicates, in a single experiment, against an entire panel of substrates. This increased throughput enables the identification of entire pathways, such as the 14-gene pathway for ferulate catabolism, as well as the comparative analysis of multiple pathways, separating the genes required for ferulate catabolism from those required for vanillate catabolism.

Transposon mutagenesis still has several limitations, most notably, that it can only be easily applied to phenotypes that are directly connected to growth. We found that a rapid growth of the parental strain under the relevant conditions was important for identifying clear fitness effects. For example, strain DSM12444 grows much better with sinapate than with syringate, and as a result, the fitness effects measured for syringate are smaller in magnitude and more variable (see Fig. S4 and Data Set S1 in the supplemental material). Additionally, we found several false positives, where gene disruptions that appeared to have a fitness cost during the transposon screen did not result in a growth defect when clean deletions were constructed. For example, transposon insertions into *ligM* produced a severe fitness defect during growth with sinapate (Fig. 5), while the growth rate of a Δ*ligM* strain was virtually unchanged from that of the wild type (Fig. 6A). This result is likely due to polar effects, where a transposon insertion in *ligM* also disrupts the expression of the cotranscribed *metF* (Fig. S4). However, false positives are more easily tolerated than false negatives, and the moderate rate of false positives did not interfere with our pathway identification.

We also note that a disruption of the genes encoding ring-cleaving dioxygenases, *ligAB* and *xylE*, does not produce measurable fitness defects. Strain DSM12444 contains at least seven ring-cleaving dioxygenases (33). We hypothesize that, when grown in the presence of high concentrations of PCA, some of these alternate dioxygenases can substitute for the pathway-specific enzymes. Pathway decomposition, through a comparative analysis of the genes required for growth with closely related substrates, can help overcome pathway redundancy. For example, the growth with vanillate or ferulate requires the genes for the PCA ring-cleaving dioxygenase, *ligAB*, in contrast to the growth with PCA. When PCA is an intermediate rather than a substrate, its concentration is likely to be lower, which in turn, makes the PCA-specific ring-cleaving dioxygenase more critical. In general, we find that the genetic requirements are more stringent for growth with an upper pathway substrate than with an intermediate, and this effect might complicate efforts for pathway elucidation.

Prior knowledge of the genes involved in ferulate catabolism (20), accumulated over many years of work, certainly aided in identifying relevant genes. However, this knowledge was not strictly necessary. For example, 11 genes are shown in Fig. 3 as being involved in vanillate degradation. The disruption of these genes produced 11 of the top 12 fitness defects during growth with vanillate, and the twelfth is a hypothetical gene with homology to DNA repair restriction endonucleases (Data Set S1).

DesC is necessary for the production of OMA from 3MGA (Fig. 6C), whereas the function of DesD is more ambiguous. The deletion of *desD* had a moderate effect on growth (Fig. 6A), while the addition of DesD enhanced OMA production only when DesC was also present (Fig. 6C). More detailed biochemistry will be required to elucidate the precise function of DesD. In addition, the deletion of *desC* or *desD* increased the accumulation of PDC (Fig. 6B). The disruption of *ligI* had no effect during

TABLE 1 Strains used in this study

Strain	Description	Reference or source
APA766	WM3064 plus pKMW7	13
WM3064	*thrB1004 pro thi rpsL hsdS lacZ*ΔM15 *RP4-1360* Δ(*araBAD*)*567* Δ*dapA1341*::[*erm pir*(wt)]	
WM6026	*lacI*^{q}, *rrnB3*, Δ*lacZ4787*, *hsdR514*, Δ*araBAD567*, Δ*rhaBAD568*, *rph-1*, *attλ*::pAE12(Δ*oriR6K-cat*::Frt5), Δ*endA*::Frt, *uidA*(ΔMluI)::*pir*, *attHK*::pJK1006D(*oriR6K-cat*::Frt5; *trfA*::Frt) *dap*	40
Novosphingobium aromaticivorans DSM12444	Type strain	
JMN3	DSM12444 Δ*desC*	This work
JMN4	DSM12444 Δ*desD*	This work
JMN8	DSM12444 Δ*desA*	This work
JMN9	DSM12444 Δ*ferA*	This work
JMN11	DSM12444 Δ*ligA*	This work
JMN12	DSM12444 Δ*ligM*	This work
JMN48	DSM12444 Δ*ligU*	This work
JMN55	DSM12444 Δ*xylG*	This work

growth with sinapate, suggesting that the conversion of 3MGA to PDC is not a significant catabolic route *in vivo* (Fig. 5). We hypothesize that PDC is an overflow metabolite when the preferred pathway for sinapate catabolism is blocked.

Including *desC*, strain DSM12444 has 11 genes annotated as alpha/beta hydrolases, and many other enzyme classes are similarly well populated (33, 34). As a result, attempts to use homology-based methods to narrow the list of potential enzymes for a particular transformation still result in long lists of possibilities that require further validation (34). Homology-based methods also prevent the identification of novel enzymes that perform a given reaction using unexpected chemistry.

Due to the inherent heterogeneity of lignin, its degradation produces a wide range of aromatic compounds (35). The engineering of 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 (36). 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. We then used this approach to identify previously uncharacterized genes for syringate degradation in *Novosphingobium aromaticivorans*. These tools will have broad utility in the discovery of novel catabolic pathways from nonmodel bacteria.

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* was routinely cultivated at 37°C in LB, while *N. aromaticivorans* was grown at 30°C in LB or DSM medium 457 containing the indicated carbon source. Aromatic monomers were dissolved in dimethyl sulfoxide (DMSO) at 100 g/liter and added at a working concentration of 1 g/liter, corresponding to 6.5 mM PCA, 7.2 mM 4-HB, 5.9 mM vanillate, 5.0 mM ferulate, 6.1 mM coumarate, 5.0 mM syringate, or 4.5 mM sinapate. The cultures were supplemented with kanamycin at 50 mg/liter, streptomycin at 100 mg/liter, carbenicillin at 50 mg/liter, or diaminopimelic acid (DAP) at 60 mg/liter as needed.

Strains and plasmids. *N. aromaticivorans* DSM12444 was purchased from the DSMZ (Braunschweig. Germany). Strain APA766, comprising the conjugation donor WM3064 containing the barcoded Tn*5* vector pKMW7, was a gift from Adam Deutschbauer. The parental deletion plasmid, pAK405 (Addgene plasmid 37114), was a gift from Julia Vorholt. The deletion plasmid pJM269-280 was synthesized *de novo* and cloned into pAK405 by GenScript (Piscataway, NJ). Codon-optimized versions of *ligA*, *ligB*, *desC*, and *desD*, prepared using OptimumGene, were synthesized and cloned into the NdeI/SalI site of pET-22b by GenScript. These genes were transformed into *E. coli* BL21 Star (DE3) using standard techniques and grown with carbenicillin. The complete lists of strains and plasmids are provided in Tables 1 and 2.

Transposon library construction and indexing. Strains DSM12444 and APA766 were grown overnight in LB containing kanamycin and DAP as appropriate. Strain DSM12444 was then diluted 1:100 in 50 ml of fresh LB and regrown to mid-log phase. Both cultures were pelleted by centrifugation, mixed at a recipient-to-donor ratio of 5:1, and plated on sterile filter paper placed on an LB plus DAP agar plate. After an overnight incubation at 30°C, the filter paper was removed and vortexed with 5 ml of fresh LB. The solution was pelleted by centrifugation, and the pellet was resuspended in 1 ml of fresh LB and

TABLE 2 Plasmids used in this study

Plasmid	Description	Reference or source
pKMW7	Barcoded Tn*5* transposon	13
pAK405	Sphingomonad suicide vector	38
pJM269	pAK405 Δ*desC*	This work
pJM270	pAK405 Δ*desD*	This work
pJM274	pAK405 Δ*desA*	This work
pJM275	pAK405 Δ*ferA*	This work
pJM277	pAK405 Δ*ligA*	This work
pJM278	pAK405 Δ*ligM*	This work
pJM332	pAK405 Δ*ligU*	This work
pJM385	pAK405 Δ*xylG*	This work

plated on 20 LB plus kanamycin plates. Colonies were observed after 3 days of incubation at 30°C. The colonies were scraped from the plates, combined into a single mixed culture, and grown overnight in 50 ml of LB plus kanamycin at 30°C. Aliquots of the combined library were stored in 7% DMSO at −80°C for further analysis. Additionally, genomic DNA was prepared using a DNeasy blood and tissue kit (Qiagen, Valencia, CA) according to the manufacturer's directions.

Barcode indexing was performed according to published protocols (13, 37). Briefly, genomic DNA was fragmented by sonication using a Bioruptor Plus (Diagenode, Ougrée, Belgium) and size-selected to approximately 300 bp using Agencourt AMPure XP beads (Beckman Coulter, Indianapolis, IN). Size selection was confirmed using a Bioanalyzer DNA1000 chip (Agilent, Santa Clara, CA). End repair, deoxyribosyladenine (dA) tailing, and adaptor ligation were performed using the NEBNext kit according to the manufacturer's directions (NEB, Ipswich, MA). Custom adaptors containing published sequences (13) were synthesized by IDT. In lieu of a final size selection, two sequential purifications using AMPure XP beads were performed (37). The final transposon enrichment PCR was performed using KAPA HiFi HotStart ReadyMix (37) with custom indexing and transposon-specific primers (13). The library was analyzed using a DNA1000 chip, quantified using a Qubit fluorimeter (Thermo Fisher, Waltham, MA), and sequenced on an Illumina MiSeq, using v2 chemistry with paired-end 150-bp reads (Illumina, San Diego, CA). The reads were analyzed using custom scripts as described previously (13).

Pooled fitness measurements. An aliquot of the strain DSM12444 library was thawed, diluted in 50 ml of LB plus kanamycin, and grown overnight at 30°C. The culture was then diluted 1:100 into three flasks, each containing DSM medium 457 plus kanamycin plus 2 g/liter glucose, and grown to saturation. Time-zero samples were taken from each flask, and each culture was then diluted to an optical density of 0.005 in 5 ml of DSM medium 457 plus kanamycin containing the appropriate carbon sources. Cultures were grown to saturation at 30°C and then frozen for later analysis. Genomic DNA was isolated and barcodes were amplified using custom indexing primers, BarSeq_P1 and BarSeq_P2_IT001 to BarSeq_P2_IT0024, as described previously (13). Barcode amplicons were quantified using a Qubit fluorimeter, pooled, and sequenced at the Joint Genome Institute on a HiSeq 2500 lane. The barcode frequencies and gene fitness values were calculated using custom scripts as described previously, where fitness is calculated on the basis of the $\log_2$ change in barcode frequencies for all insertions in the middle 80% of the gene of interest (13).

Gene deletions. In-frame gene deletions were constructed using a kanamycin-*rpsL* suicide vector (38). Briefly, appropriate deletion cassettes were synthesized and cloned into pAK405 (see Fig. S1 in the supplemental material). The resulting plasmids were transformed into strain WM6026, a conjugation-proficient DAP auxotroph. The plasmids were transferred to strain DSM12444 by conjugation according to the protocol used for library construction. After selecting for single crossovers on LB plus kanamycin plates, the colonies were patched on LB plus kanamycin to confirm plasmid integration, grown in the absence of selection, and then plated on LB plus streptomycin plates. The colonies were patched on LB plus streptomycin and replica plated on LB plus streptomycin plus kanamycin to confirm double recombination. Deletions were identified by colony PCR using locus-specific primers and confirmed by Sanger sequencing of the amplicons.

Growth rate measurements. Strains were grown to saturation overnight in minimal medium with 2 g/liter glucose. They were then diluted 100× in fresh medium containing the appropriate carbon source and grown as triplicate 100-μl cultures in a Bioscreen C plate reader (Oy Growth Curves Ab Ltd., Helsinki, Finland). The growth was monitored by measuring the optical density at 600 nm (OD_{600}). The growth rates were calculated during exponential growth using CurveFitter software (39).

Metabolite accumulation assays. Strains were grown to saturation overnight in minimal medium with 2 g/liter glucose. They were then diluted 50× in fresh medium containing 0.5 g/liter glucose and 1 g/liter sinapate and regrown for 6 h. The supernatant was collected by centrifugation and filtered to 0.22 μm.

Filtered supernatant samples were then analyzed by high-resolution LC tandem mass spectrometry (LC-MS/MS) using a Vanquish UHPLC plumbed directly in-line with a Q Exactive Plus mass spectrometer (Thermo Scientific) outfitted with an in-house pulled nanospray ionization (NSI) emitter (75-μm inside diameter [i.d.]) packed with 30 cm of 5-μm Kinetex C_{18} RP resin (Phenomenex). Mobile phases were prepared from LC-MS-grade solvents (OmniSolv; Millipore Sigma) and include solvent A (95% H_2O, 5% isopropanol, 1 mM glacial acetic acid) and solvent B (30% H_2O, 65% acetonitrile, 5% isopropanol, 1 mM glacial acetic acid). To achieve nanoliter flow rates at the NSI emitter, the UHPLC solvent path was split

before the C_{18} column, effectively reducing the UHPLC output flow of 150 μl/min to 300 nl/min (~500×). This split-flow setup enabled nanoliter-scale sample injections that are more compatible with the i.d. of the C_{18}-packed NSI emitter. Twenty microliters of each sample was autoinjected before the split, leading to the separation and analysis of 40 nl of supernatant over a 35-min organic gradient (0% to 2% B over 5 min; 2% to 10% B over 5 min, 10% to 100% B over 5 min, hold at 100% B for 1 min, 100% to 0% B over 2 min, and hold at 0% B for 17 min to reequilibrate the column). Eluted analytes were measured directly by the Q Exactive operating in negative ion mode with a duty cycle that included a full scan (150 to 250 *m/z* range; resolution, 35,000; 3 microscan spectrum averaging) followed by targeted fragmentation (MS/MS) of pathway analytes/intermediates via parallel reaction monitoring (resolution, 17,500; higher-energy collisional dissociation [HCD] fragmentation at 35% normalized collision energy [NCE]). Each supernatant sample was run in technical duplicates and biological triplicates. Peak areas were extracted for each pathway analyte/intermediate via Xcalibur software and the areas were compared across samples.

Cell lysate preparation. Cell extracts were prepared by growing *E. coli* BL21 Star (DE3) in 2× YPTG (16 g/liter tryptone, 10 g/liter yeast extract, 5 g/liter NaCl, 7 g/liter KH_2PO_4, 3 g/liter K_2HPO_4, 18 g/liter glucose). The cells were cultured in 50-ml volumes using 250-ml baffled flasks in a 37°C incubator with vigorous shaking at 250 rpm. The cultures were induced with 0.1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) upon reaching an OD_{600} of 0.6 to 0.8. After induction, the cultures were grown for 4 h at 30°C. No antibiotics were used during growth. These cultures were harvested by centrifugation at 5,000 × *g* for 10 min in 50-ml volumes and washed twice with S30 buffer (2 g/liter magnesium acetate, 14.05 g/liter potassium glutamate, 0.154 g/liter dithiothreitol [DTT], and 1.81 g/liter Tris-acetate, pH 8.2) by resuspension and centrifugation. After the final centrifugation, the pellets were weighed, flash-frozen in liquid nitrogen, and stored at −80°C. The extracts were prepared by thawing and resuspending the pellet in 0.8 ml of S30 buffer per milligram of cell wet weight before sonicating with 530 J per ml of suspension at 50% tip amplitude with ice-water cooling. After sonication, the cell-slurry was centrifuged twice for 10 min at 21,100 × *g* at 4°C, aliquoted, flash-frozen, and stored at −80°C. Crude extract concentrations were obtained via a standard curve using the Quick Start Bradford protein assay reagent (Bio-Rad). The concentrations of the LigA, LigB, DesC, DesD, and wild-type extracts were 17.80 ± 0.77 mg/ml, 22.93 ± 1.40 mg/ml, 15.49 ± 1.01 mg/ml, 19.91 ± 0.79 mg/ml, and 17.09 ± 1.07 mg/ml, respectively.

In vitro reactions were carried out in triplicates using 1.5-ml Eppendorf tubes in 50-μl volumes at 30°C for 6 h. Each reaction consisted of 10 μl of up to four different lysates, each expressing one of the overexpressed enzymes, along with 3MGA (0.4 g/liter [2.2 mM]) and glutathione (0.61 g/liter [2.0 mM]); the final 10 μl consisted of S30 buffer. Reaction mixtures using fewer enzymes were supplemented with nonexpressing *E. coli* BL21 Star (DE3) lysate as a substitute for the missing lysate. The reaction mixtures were immediately placed on ice after incubation and passed through a 10-kDa microspin filter prior to analysis by LC-MS/MS as described above, with 5 μl/nl injected per sample.

Accession number(s). The raw reads are available from the SRA as BioProject PRJNA478570.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.01185-18.

SUPPLEMENTAL FILE 1, PDF file, 0.5 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

The strains for barcoded transposon library construction were provided by Adam Deutschbauer at the Lawrence Berkeley National Laboratory. DNA shearing was performed at the UT-Knoxville genomics core, with assistance from Joe May and Veronica Brown. Dawn Klingeman performed the index sequencing reaction. The barcode sequencing was performed by Christa Pennacchio and Matthew Blow at the Joint Genome Institute.

The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science user facility, is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the DOE under contract no. DE-AC05-00OR22725. This work was supported by the BioEnergy Science Center and Center for Bioenergy Innovation, U.S. Department of Energy Bioenergy Research Centers supported by the Office of Biological and Environmental Research in the DOE Office of Science, and by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

The funders had no involvement in the design, analysis, or interpretation of these experiments.

REFERENCES

1. Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall DH, Caporaso JG. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc Natl Acad Sci U S A 109:21390–21395. https://doi.org/10.1073/pnas.1215210110.
2. Nielsen J, Keasling JD. 2016. Engineering cellular metabolism. Cell 164:1185–1197. https://doi.org/10.1016/j.cell.2016.02.004.
3. Masai E, Katayama Y, Fukuda M. 2007. Genetic and biochemical investigations on bacterial catabolic pathways for lignin-derived aromatic compounds. Biosci Biotechnol Biochem 71:1–15. https://doi.org/10.1271/bbb.60437.
4. Arias-Barrau E, Olivera ER, Luengo JM, Fernández C, Galán B, García JL, Díaz E, Miñambres B. 2004. The homogentisate pathway: a central catabolic pathway involved in the degradation of L-phenylalanine, L-tyrosine, and 3-hydroxyphenylacetate in *Pseudomonas putida*. J Bacteriol 186:5062–5077. https://doi.org/10.1128/JB.186.15.5062-5077.2004.
5. Lederberg J, Lederberg EM. 1952. Replica plating and indirect selection of bacterial mutants. J Bacteriol 63:399–406.
6. Shuman HA, Silhavy TJ. 2003. The art and design of genetic screens: *Escherichia coli*. Nat Rev Genet 4:419–431. https://doi.org/10.1038/nrg1087.
7. Kleckner N, Chan RK, Tye B-K, Botstein D. 1975. Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. J Mol Biol 97:561–575. https://doi.org/10.1016/S0022-2836(75)80059-3.
8. Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW. 1995. Simultaneous identification of bacterial virulence genes by negative selection. Science 269:400–403. https://doi.org/10.1126/science.7618105.
9. van Opijnen T, Bodi KL, Camilli A. 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods 6:767–772. https://doi.org/10.1038/nmeth.1377.
10. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI. 2009. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe 6:279–289. https://doi.org/10.1016/j.chom.2009.08.003.
11. Langridge GC, Phan M-D, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK. 2009. Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. Genome Res 19:2308–2316. https://doi.org/10.1101/gr.097097.109.
12. Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ. 2009. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. Proc Natl Acad Sci U S A 106:16422–16427. https://doi.org/10.1073/pnas.0906627106.
13. 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. https://doi.org/10.1128/mBio.00306-15.
14. Price MN, Zane GM, Kuehl JV, Melnyk RA, Wall JD, Deutschbauer AM, Arkin AP. 2018. Filling gaps in bacterial amino acid biosynthesis pathways with high-throughput genetics. PLoS Genet 14:e1007147. https://doi.org/10.1371/journal.pgen.1007147.
15. Stolz A. 2009. Molecular characteristics of xenobiotic-degrading sphingomonads. Appl Microbiol Biotechnol 81:793–811. https://doi.org/10.1007/s00253-008-1752-3.
16. Kamimura N, Takahashi K, Mori K, Araki T, Fujita M, Higuchi Y, Masai E. 2017. Bacterial catabolism of lignin-derived aromatics: new findings in a recent decade: update on bacterial lignin catabolism. Environ Microbiol Rep 9:679–705. https://doi.org/10.1111/1758-2229.12597.
17. Fredrickson J, Balkwill D, Romine M, Shi T. 1999. Ecology, physiology, and phylogeny of deep subsurface *Sphingomonas* sp. J Ind Microbiol Biotechnol 23:273–283. https://doi.org/10.1038/sj.jim.2900741.
18. Fredrickson JK, Balkwill DL, Drake GR, Romine MF, Ringelberg DB, White DC. 1995. Aromatic-degrading *Sphingomonas* isolates from the deep subsurface. Appl Environ Microbiol 61:1917–1922.
19. Sato Y, Moriuchi H, Hishiyama S, Otsuka Y, Oshima K, Kasai D, Nakamura M, Ohara S, Katayama Y, Fukuda M, Masai E. 2009. Identification of three alcohol dehydrogenase genes involved in the stereospecific catabolism of arylglycerol-beta-aryl ether by *Sphingobium* sp. strain SYK-6. Appl Environ Microbiol 75:5195–5201. https://doi.org/10.1128/AEM.00880-09.
20. Kamimura N, Masai E. 2014. The protocatechuate 4,5-cleavage pathway: overview and new findings, p 207–226. *In* Nojiri H, Tsuda M, Fukuda M, Kamagata Y (ed), Biodegradative bacteria. Springer Japan, Tokyo, Japan.
21. Romine MF, Stillwell LC, Wong KK, Thurston SJ, Sisk EC, Sensen C, Gaasterland T, Fredrickson JK, Saffer JD. 1999. Complete sequence of a 184-kilobase catabolic plasmid from *Sphingomonas aromaticivorans* F199. J Bacteriol 181:1585–1602.
22. Abe T, Masai E, Miyauchi K, Katayama Y, Fukuda M. 2005. A tetrahydrofolate-dependent *O*-demethylase, LigM, is crucial for catabolism of vanillate and syringate in *Sphingomonas paucimobilis* SYK-6. J Bacteriol 187:2030–2037. https://doi.org/10.1128/JB.187.6.2030-2037.2005.
23. Studer A, McAnulla C, Büchele R, Leisinger T, Vuilleumier S. 2002. Chloromethane-induced genes define a third C_1 utilization pathway in *Methylobacterium chloromethanicum* CM4. J Bacteriol 184:3476–3484. https://doi.org/10.1128/JB.184.13.3476-3484.2002.
24. Vannelli T, Messmer M, Studer A, Vuilleumier S, Leisinger T. 1999. A corrinoid-dependent catabolic pathway for growth of a *Methylobacterium* strain with chloromethane. Proc Natl Acad Sci U S A 96:4615–4620.
25. Michener JK, Vuilleumier S, Bringel F, Marx CJ. 2016. Transfer of a catabolic pathway for chloromethane in *Methylobacterium* strains highlights different limitations for growth with chloromethane or with dichloromethane. Front Microbiol 7:1116. https://doi.org/10.3389/fmicb.2016.01116.
26. Masai E, Sasaki M, Minakawa Y, Abe T, Sonoki T, Miyauchi K, Katayama Y, Fukuda M. 2004. A novel tetrahydrofolate-dependent *O*-demethylase gene is essential for growth of *Sphingomonas paucimobilis* SYK-6 with syringate. J Bacteriol 186:2757–2765. https://doi.org/10.1128/JB.186.9.2757-2765.2004.
27. Kasai D, Masai E, Miyauchi K, Katayama Y, Fukuda M. 2005. Characterization of the gallate dioxygenase gene: three distinct ring cleavage dioxygenases are involved in syringate degradation by *Sphingomonas paucimobilis* SYK-6. J Bacteriol 187:5067–5074. https://doi.org/10.1128/JB.187.15.5067-5074.2005.
28. Nogales J, Canales Á, Jiménez-Barbero J, Serra B, Pingarrón JM, García JL, Díaz E. 2011. Unravelling the gallic acid degradation pathway in bacteria: the *gal* cluster from *Pseudomonas putida*. Mol Microbiol 79:359–374. https://doi.org/10.1111/j.1365-2958.2010.07448.x.
29. Mazurkewich S, Brott AS, Kimber MS, Seah SYK. 2016. Structural and kinetic characterization of the 4-carboxy-2-hydroxymuconate hydratase from the gallate and protocatechuate 4,5-cleavage pathways of *Pseudomonas putida* KT2440. J Biol Chem 291:7669–7686. https://doi.org/10.1074/jbc.M115.682054.
30. Kasai D, Masai E, Miyauchi K, Katayama Y, Fukuda M. 2004. Characterization of the 3-*O*-methylgallate dioxygenase gene and evidence of multiple 3-*O*-methylgallate catabolic pathways in *Sphingomonas paucimobilis* SYK-6. J Bacteriol 186:4951–4959. https://doi.org/10.1128/JB.186.15.4951-4959.2004.
31. Kasai D, Masai E, Katayama Y, Fukuda M. 2007. Degradation of 3-*O*-methylgallate in *Sphingomonas paucimobilis* SYK-6 by pathways involving protocatechuate 4,5-dioxygenase. FEMS Microbiol Lett 274:323–328. https://doi.org/10.1111/j.1574-6968.2007.00855.x.
32. Vuilleumier S, Ucurum Z, Oelhafen S, Leisinger T, Armengaud J, Wittich R-M, Timmis KN. 2001. The glutathione *S*-transferase OrfE3 of the dioxin-degrading bacterium *Sphingomonas* sp. RW1 displays maleylpyruvate isomerase activity. Chem Biol Interact 133:265–267.
33. D'Argenio V, Notomista E, Petrillo M, Cantiello P, Cafaro V, Izzo V, Naso B, Cozzuto L, Durante L, Troncone L, Paolella G, Salvatore F, Di Donato A. 2014. Complete sequencing of *Novosphingobium* sp. PP1Y reveals a biotechnologically meaningful metabolic pattern. BMC Genomics 15:384. https://doi.org/10.1186/1471-2164-15-38.
34. Kamimura N, Goto T, Takahashi K, Kasai D, Otsuka Y, Nakamura M, Katayama Y, Fukuda M, Masai E. 2017. A bacterial aromatic aldehyde dehydrogenase critical for the efficient catabolism of syringaldehyde. Sci Rep 7:44422. https://doi.org/10.1038/srep44422.
35. Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R. 2011. Pathways for degradation of lignin in bacteria and fungi. Nat Prod Rep 28:1883. https://doi.org/10.1039/c1np00042j.
36. Hanson AD, Pribat A, Waller JC, de Crécy-Lagard V. 2009. "Unknown" proteins and "orphan" enzymes: the missing half of the engineering parts list–and how to find it. Biochem J 425:1–11. https://doi.org/10.1042/BJ20091328.

37. Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ, Page AJ, Langridge GC, Quail MA, Keane JA, Parkhill J. 2016. The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. Bioinformatics 32:1109–1111. https://doi.org/10.1093/bioinformatics/btw022.
38. Kaczmarczyk A, Vorholt JA, Francez-Charlot A. 2012. Markerless gene deletion system for sphingomonads. Appl Environ Microbiol 78:3774–3777. https://doi.org/10.1128/AEM.07347-11.
39. Delaney NF, Kaczmarek ME, Ward LM, Swanson PK, Lee M-C, Marx CJ. 2013. Development of an optimized medium, strain and high-throughput culturing methods for *Methylobacterium extorquens*. PLoS One 8:e62957. https://doi.org/10.1371/journal.pone.0062957.
40. Blodgett JAV, Thomas PM, Li G, Velasquez JE, van der Donk WA, Kelleher NL, Metcalf WW. 2007. Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. Nat Chem Biol 3:480–485. https://doi.org/10.1038/nchembio.2007.9.