



# Vanillin Production in *Pseudomonas*: Whole-Genome Sequencing of *Pseudomonas* sp. Strain 9.1 and Reannotation of *Pseudomonas putida* CalA as a Vanillin Reductase

Javier García-Hidalgo,<sup>a</sup>  Daniel P. Brink,<sup>a</sup> Krithika Ravi,<sup>b</sup> Catherine J. Paul,<sup>a,c</sup> Gunnar Lidén,<sup>b</sup> Marie F. Gorwa-Grauslund<sup>a</sup>

<sup>a</sup>Applied Microbiology, Department of Chemistry, Lund University, Lund, Sweden

<sup>b</sup>Department of Chemical Engineering, Lund University, Lund, Sweden

<sup>c</sup>Water Resources Engineering, Department of Building and Environmental Technology, Lund University, Lund, Sweden

Javier García-Hidalgo and Daniel P. Brink contributed equally to this work. Author order was determined by order of seniority.

**ABSTRACT** Microbial degradation of lignin and its related aromatic compounds has great potential for the sustainable production of chemicals and bioremediation of contaminated soils. We previously isolated *Pseudomonas* sp. strain 9.1 from historical waste deposits (forming so-called fiber banks) released from pulp and paper mills along the Baltic Sea coast. The strain accumulated vanillyl alcohol during growth on vanillin, and while reported in other microbes, this phenotype is less common in wild-type pseudomonads. As the reduction of vanillin to vanillyl alcohol is an undesired trait in *Pseudomonas* strains engineered to accumulate vanillin, connecting the strain 9.1 phenotype with a genotype would increase the fundamental understanding and genetic engineering potential of microbial vanillin metabolism. The genome of *Pseudomonas* sp. 9.1 was sequenced and assembled. Annotation identified oxidoreductases with homology to *Saccharomyces cerevisiae* alcohol dehydrogenase ScADH6p, known to reduce vanillin to vanillyl alcohol, in both the 9.1 genome and the model strain *Pseudomonas putida* KT2440. Recombinant expression of the *Pseudomonas* sp. 9.1 FEZ21\_09870 and *P. putida* KT2440 PP\_2426 (*calA*) genes in *Escherichia coli* revealed that these open reading frames encode aldehyde reductases that convert vanillin to vanillyl alcohol, and that *P. putida* KT2440 PP\_3839 encodes a coniferyl alcohol dehydrogenase that oxidizes coniferyl alcohol to coniferyl aldehyde (i.e., the function previously assigned to *calA*). The deletion of PP\_2426 in *P. putida* GN442 engineered to accumulate vanillin resulted in a decrease in by-product (vanillyl alcohol) yield from 17% to ~1%. Based on these results, we propose the reannotation of PP\_2426 and FEZ21\_09870 as *areA* and PP\_3839 as *calA-II*.

**IMPORTANCE** Valorization of lignocellulose (nonedible plant matter) is of key interest for the sustainable production of chemicals from renewable resources. Lignin, one of the main constituents of lignocellulose, is a heterogeneous aromatic biopolymer that can be chemically depolymerized into a heterogeneous mixture of aromatic building blocks; those can be further converted by certain microbes into value-added aromatic chemicals, e.g., the flavoring agent vanillin. We previously isolated a *Pseudomonas* sp. strain with the (for the genus) unusual trait of vanillyl alcohol production during growth on vanillin. Whole-genome sequencing of the isolate led to the identification of a vanillin reductase candidate gene whose deletion in a recombinant vanillin-accumulating *P. putida* strain almost completely alleviated the undesired vanillyl alcohol by-product yield. These results represent an important step toward biotechnological production of vanillin from lignin using bacterial cell factories.

**Citation** García-Hidalgo J, Brink DP, Ravi K, Paul CJ, Lidén G, Gorwa-Grauslund MF. 2020. Vanillin production in *Pseudomonas*: whole-genome sequencing of *Pseudomonas* sp. strain 9.1 and reannotation of *Pseudomonas putida* CalA as a vanillin reductase. Appl Environ Microbiol 86:e02442-19. <https://doi.org/10.1128/AEM.02442-19>.

**Editor** Robert M. Kelly, North Carolina State University

**Copyright** © 2020 García-Hidalgo et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Marie F. Gorwa-Grauslund, [marie-francoise.gorwa@tmb.lth.se](mailto:marie-francoise.gorwa@tmb.lth.se).

**Received** 22 October 2019

**Accepted** 21 December 2019

**Accepted manuscript posted online** 10 January 2020

**Published** 2 March 2020

**KEYWORDS** NAD(P)H-dependent oxidoreductases, *Pseudomonas*, *calA*, *de novo* assembly, gene reannotation, vanillyl alcohol

Lignocellulosic biomass is a renewable feedstock of choice for the sustainable production of fine and bulk chemicals. Lignin, one of the three main components of lignocellulose together with hemicellulose and cellulose, is the most abundant aromatic biopolymer on Earth and thus a desirable feedstock for renewable aromatics (1, 2). Its high heterogeneity and recalcitrance have, however, limited its current industrial use, and large volumes of industrial lignin streams are incinerated for process energy rather than upgraded to value-added products (3, 4). From a biorefinery perspective, microbial bioconversion has emerged as a promising method for lignin valorization and gained much attention in recent years (5, 6). A common strategy is to combine the bioconversion with a pretreatment step where larger lignin polymers are broken down into smaller aromatic fragments by means of chemical depolymerization and then fed to microbes with relevant substrate specificities (7–9).

In nature, the conversion of depolymerized lignin is mainly performed by bacteria, although fungi and archaea have also been identified (10). While substrate specificities vary widely between species (and can even be strain dependent), the aromatic fragments are normally converted to monomers with a decreasing degree of substitution by means of so-called funneling pathways (11) ending with a ring fission that shunts the carbon into the central metabolism.

A number of bacteria capable of catabolism of lignin and/or lignin-derived aromatic compounds have been isolated (see, e.g., references 12–18). However, whereas characterization is a straightforward process, many isolates lack studies that elucidate the underlying molecular biology of their aromatic metabolic traits. We previously isolated *Pseudomonas* sp. strain 9.1 (DSM 105530) by culture-dependent screening of fibrous waste deposits from the pulp and paper industry collected on the northern Baltic coast in Sweden (19). The strain grew on seven out of 10 lignin model compounds tested as a sole carbon source (ferulate, *p*-coumarate, benzoate, vanillin, 4-hydroxybenzoate, vanillate, and vanillyl alcohol), whereas it could not sustain growth on syringate, guaiacol, or *cis,cis*-muconate (19). The present study continues the characterization of this isolate from a molecular point of view.

In comparison to a previous study on the *Pseudomonas* model organism *P. putida* KT2440 (20), *Pseudomonas* sp. 9.1 (here referred to as strain 9.1) displayed a distinct phenotype in that the strain excreted and reconsumed several intermediate metabolites of the lignin funneling pathways (19). The majority of the excreted substrates were metabolites produced directly downstream of the different substrates fed to the bacteria (e.g., 4-hydroxybenzoic acid from *p*-coumarate, and catechol and *cis,cis*-muconate from benzoic acid), which implies that certain reaction steps for aromatic assimilation in strain 9.1 are bottlenecks limiting the overall flux rate through the pathway. However, when grown on vanillin, the isolate excreted vanillyl alcohol, which is a less common trait for *Pseudomonas* spp. since they typically oxidize vanillin to vanillic acid as the first step for its assimilation. When grown on ferulate (a metabolite closely upstream of vanillin), no vanillyl alcohol was observed in strain 9.1 (19). The uptake rate of ferulate was very similar to that of vanillate, whereas vanillin was taken up much faster. Many microbes rapidly detoxify vanillin to its corresponding acid or alcohol to cope with its high reactivity (21). Given the above-described phenotype, it is likely that strain 9.1 relieves itself of the toxicity of vanillin by reducing it to vanillyl alcohol, which can then be later consumed once vanillin is depleted. While generally uncommon in the literature on pseudomonads, this particular phenotype has been reported in *Pseudomonas fluorescens* B56 (22) and certain fungi, e.g., *Phanerochaete chrysosporium* (23) and *Saccharomyces cerevisiae* (24).

Vanillin is one of the most utilized flavoring agents in food and fragrances and a precursor for the synthesis of polymers and fine chemicals (25). There is a high interest in replacing the chemical synthesis of vanillin from petro-based compounds (such as

**TABLE 1** Statistics of the sequencing, assembly, and annotation of *Pseudomonas* sp. 9.1 with Illumina MiSeq paired-end whole-genome sequencing<sup>a</sup>

Metric	Data
Sequencing and assembly results	
No. of paired-end reads	6,829,359
Mean read length	151 bp
Avg sequencing coverage	400×
No. of contigs	30
Total length (all contigs)	4,978,116 bp
Assembly $N_{50}$	728,368 bp
Assembly $L_{50}$	3
Assembly GC content	58.7%
Annotation results	
Total no. of genes	4,522
No. of CDSs with protein <sup>b</sup>	4,390
No. of RNA genes	75 (10 rRNAs, 61 tRNAs, 4 ncRNAs) <sup>c</sup>
No. of pseudogenes (without protein)	57

<sup>a</sup>The annotation was performed with the NCBI Prokaryotic Genome Annotation Pipeline.

<sup>b</sup>CDSs, coding sequences.

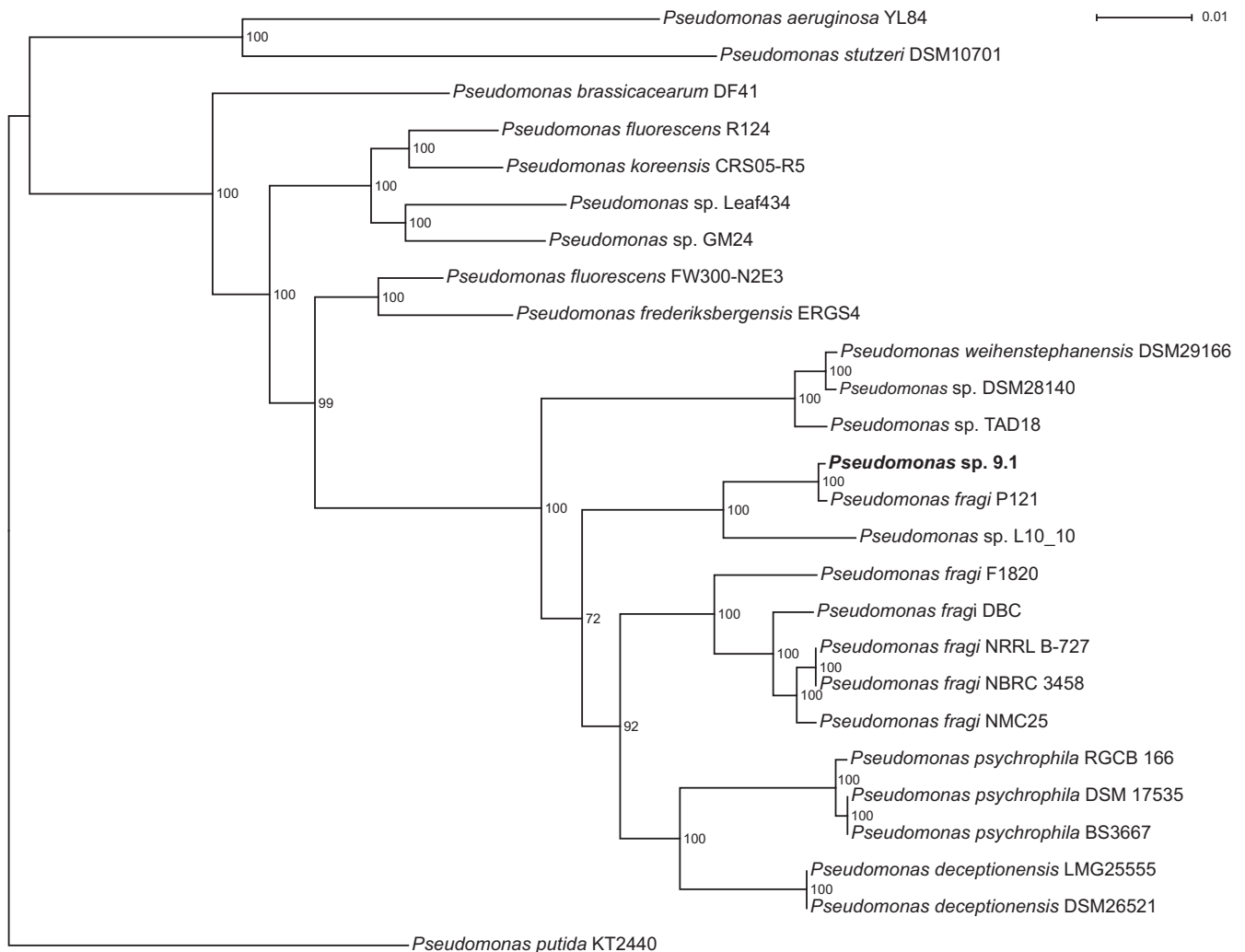
<sup>c</sup>ncRNAs, noncoding RNAs.

guaiacol) with biotechnological alternatives, such as production based on lignin (21, 25). A recent study on the engineering of *P. putida* KT2440 for bioconversion and accumulation of vanillin from ferulate reported issues with undesired vanillyl alcohol by-product formation (26). This is likely to also be the case for strain 9.1, meaning that elucidation of the enzymes responsible for the reduction of vanillin to vanillyl alcohol would not only contribute to an increased understanding of aromatic metabolism in *Pseudomonas* spp. but could also improve cell factory design for use in biorefineries (e.g., vanillin production). In the present study, the genome sequence of strain 9.1 was determined in order to relate the vanillyl alcohol-accumulating phenotype to a specific genotype. Experiments then based on this information were used to investigate the genetic basis of this phenotype, and together, these studies could be used to generate new *Pseudomonas* strains with alleviated vanillyl alcohol production.

## RESULTS

**Whole-genome sequencing and phylogeny.** The genome of *Pseudomonas* sp. 9.1 was sequenced with the aim of predicting the genetic basis of its aromatic funneling pathways and, in particular, to elucidate genetic clues regarding the oxidoreductase activities behind the vanillyl alcohol by-product formation. Whole-genome sequencing was performed with a paired-end Illumina MiSeq platform and a custom assembly pipeline designed using the SPAdes assembly algorithm (27) with iterative gap-filling steps (28–31). The final assembly consisted of 4,978,116 bp distributed over 30 contigs, with a GC content of 58.7%. This genome size was within the range for this genus, which is between 4 and 7 Mbp (Spiers et al. [32]). Statistics describing the sequencing, assembly and annotation are found in Table 1.

Previously, strain 9.1 was tentatively identified by combined 16S rRNA and *gyrB* sequencing as *Pseudomonas deceptionensis* (19), which placed it in the *P. fluorescens* lineage and *Pseudomonas fragi* subgroup (33). To further resolve the taxonomy of strain 9.1, a genome-wide phylogeny approach (34), using multiple common loci across different *Pseudomonas* genomes and the 9.1 assembly, was used to calculate species relationships. Previous phylogeny studies have suggested that branches with >70% bootstrap support correspond to a 95% probability of representing a true clade (35), and this was therefore used as a threshold in the current project for selecting the final tree. Although it shared a common relative with *P. deceptionensis*, *Pseudomonas* sp. 9.1 rather clustered within the *P. fragi* taxon (Fig. 1) when more loci than just 16S rRNA and *gyrB* genes were considered. It was decided to no longer refer to the isolate as a *P. deceptionensis* as was previously suggested (19) but rather as *Pseudomonas* sp. 9.1.



**FIG 1** Phylogram of nonclinical *Pseudomonas* genomes in relation to strain 9.1, rooted at the model strain *P. putida* KT2440. Whole-genome phylogeny was calculated with the RealPhy pipeline, and trees were generated with RAxML (a maximum likelihood method). One hundred bootstrap iterations were used to construct the final tree. Bootstrap values are given at each branch. Strain 9.1 clearly clusters with other strains in the so-called *P. fragi* lineage (33). *Pseudomonas* sp. strain Lz4W clustered with *Pseudomonas* sp. 9.1 and *P. fragi* P121 throughout the analysis, but it had to be removed from the data set due to its high similarity to these strains resulting in bootstrap values below the 70% threshold (35).

The annotated 9.1 assembly was analyzed for open reading frames (ORFs) related to known aromatic funneling pathways commonly found in *Pseudomonas* spp. Known genes from the model strain *P. putida* KT2440 and a few other annotated species from the genus were used to predict the corresponding ORFs in strain 9.1 based on homology (see Table S1 in the supplemental material for BLASTp results). For almost every reaction, a clear gene candidate in 9.1 was predicted, with the exception of the catechol 1,2-dioxygenase *catA* gene, which had three highly similar hits. A number of strains of *P. putida* (including KT2440) have been reported to have two catechol 1,2-dioxygenases with functional redundancy (CatA and CatA2) in order to better cope with excess catechol levels (36, 37); it is likely that the multiple hits represent a similar arrangement of homologs in strain 9.1.

**Identification of putative oxidoreductases with activity on vanillin.** *Saccharomyces cerevisiae* ADH6p is an NADPH-dependent alcohol dehydrogenase with broad substrate range that, among other reactions, has been described to reduce vanillin to vanillyl alcohol (24). The amino acid sequence of ScADH6p was queried against the genome assemblies of strain 9.1 and the model strain *P. putida* KT2440 using stand-alone BLASTp. The best hit for strain 9.1 and the two best hits for KT2440 were selected

**TABLE 2** Top BLASTp results in strain 9.1 and the model strain *Pseudomonas putida* KT2440 with *S. cerevisiae* ADH6p as a query protein<sup>a</sup>

Strain	Gene	Identity (%)	Coverage (%)	E value	Locus annotation
<i>Pseudomonas</i> sp. 9.1	FEZ21_09870	40.5	92.2	1.0e−61	NAD(P)-dependent alcohol dehydrogenase
	FEZ21_15000	36.3	88.4	1.0e−59	NAD(P)-dependent alcohol dehydrogenase
<i>P. putida</i> KT2440	PP_2426 ( <i>calA</i> )	39.3	87.3	2.9e−56	<i>calA</i> coniferyl alcohol dehydrogenase
	PP_3839 ( <i>adhP</i> )	25.9	86.4	3.8e−27	<i>adhP</i> alcohol dehydrogenase

<sup>a</sup>Query protein number is SGDID:S000004937. Only the top hit was selected for strain 9.1 (FEZ21\_09870).

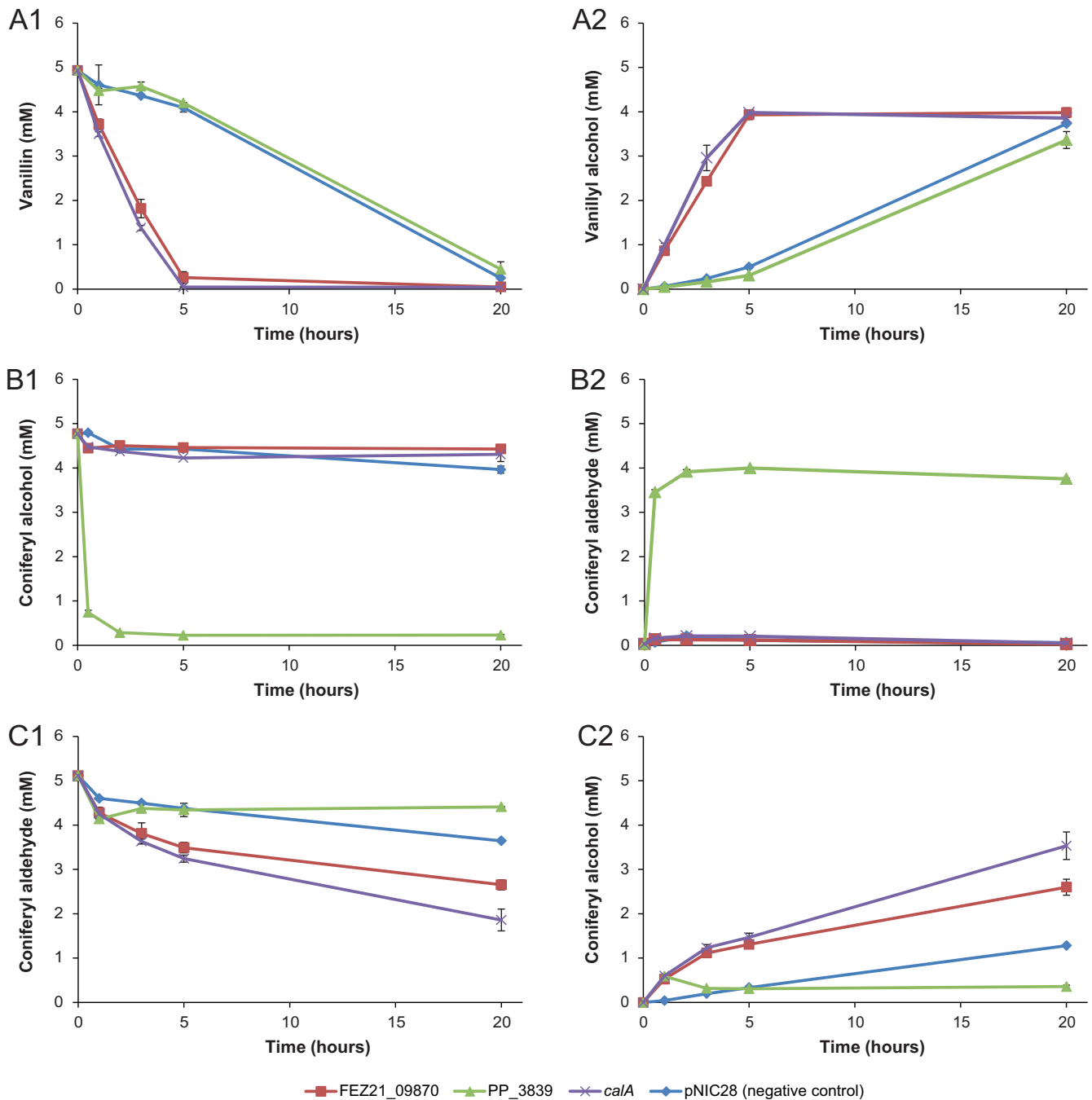
as candidate genes; these were FEZ21\_09870 in strain 9.1 and PP\_2426 (*calA*) and PP\_3839 (*adhP*) in KT2440 (Table 2). Analysis with the HMMER software (38) revealed that all three protein sequences for these genes contained the same functional domains as ScADH6 (Fig. S1), as follows: an N-terminal catalytic alcohol dehydrogenase GroES-like domain which should/could also act as the substrate-binding domain (ADH\_N; Pfam accession no. PF08240) and a C-terminal zinc-containing domain presenting a classical Rossmann fold to bind NAD(P)H (ADH\_zinc\_N; Pfam accession no. PF00107). PP\_2426 had been putatively annotated *in silico* as a gene for coniferyl dehydrogenase catalyzing the oxidation of coniferyl alcohol to coniferyl aldehyde (*calA*; the first reaction of the coniferyl branch), although this prediction was not experimentally validated (39).

**Oxidoreductase characterization in *E. coli*.** To assess the ability of the three selected candidate enzymes to reduce vanillin to vanillyl alcohol, the corresponding genes were amplified by PCR and introduced via plasmid pNIC28-Bsa4 to the expression host *E. coli* BL21(DE3). The resulting clones were grown in LB medium, and expression of the genes was induced by the addition of isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG). Overexpression of the recombinant genes as soluble proteins was confirmed by lysing cells by sonication and analyzing the resulting soluble intracellular extracts as well as the cell debris fractions by SDS-PAGE (Fig. S2).

Clones overexpressing each one of the candidate genes and the negative control [*E. coli* BL21(DE3) carrying the empty vector] were used in a whole-cell enzyme assay carried out in M9 medium supplemented with glucose to allow the regeneration of redox cofactors and 5 mM vanillin as the substrate (Fig. 2). The gene products of both *calA* and FEZ21\_09870 presented very similar and clear vanillin reductase activities, with a complete conversion of vanillin into vanillyl alcohol in the first 5 h of the assay (Fig. 2A1 and A2). The reverse conversion was also assessed in separate assays with 5 mM vanillyl alcohol to evaluate the ability of these enzymes to oxidize vanillyl alcohol to vanillin. However, this activity was not detected from any of the strains tested (Fig. S3).

Since the functional analysis of *P. putida* KT2440 *calA* suggested that it could encode an enzyme able to oxidize coniferyl alcohol to coniferyl aldehyde (39), all three candidate enzymes were also evaluated for this reaction. Surprisingly, the gene products expressed from PP\_2426 (*calA*) and FEZ21\_09870 did not exhibit any activity on coniferyl alcohol (Fig. 2B1). The gene product from PP\_3839, however, clearly displayed a very high oxidative activity on coniferyl alcohol, depleting the substrate in the first hour of the assay (Fig. 2B1 and B2). For the reverse reaction, enzymes encoded by PP\_2426 (*calA*) and FEZ21\_09870 instead exhibited some level of coniferyl aldehyde reductase activity (Fig. 2C1 and C2), i.e., the opposite reaction to the previously predicted function of PP\_2426 (39). These results suggested that the genes PP\_2426 and FEZ21\_09870 encode aromatic aldehyde reductases *in vivo*, with broad substrate specificity.

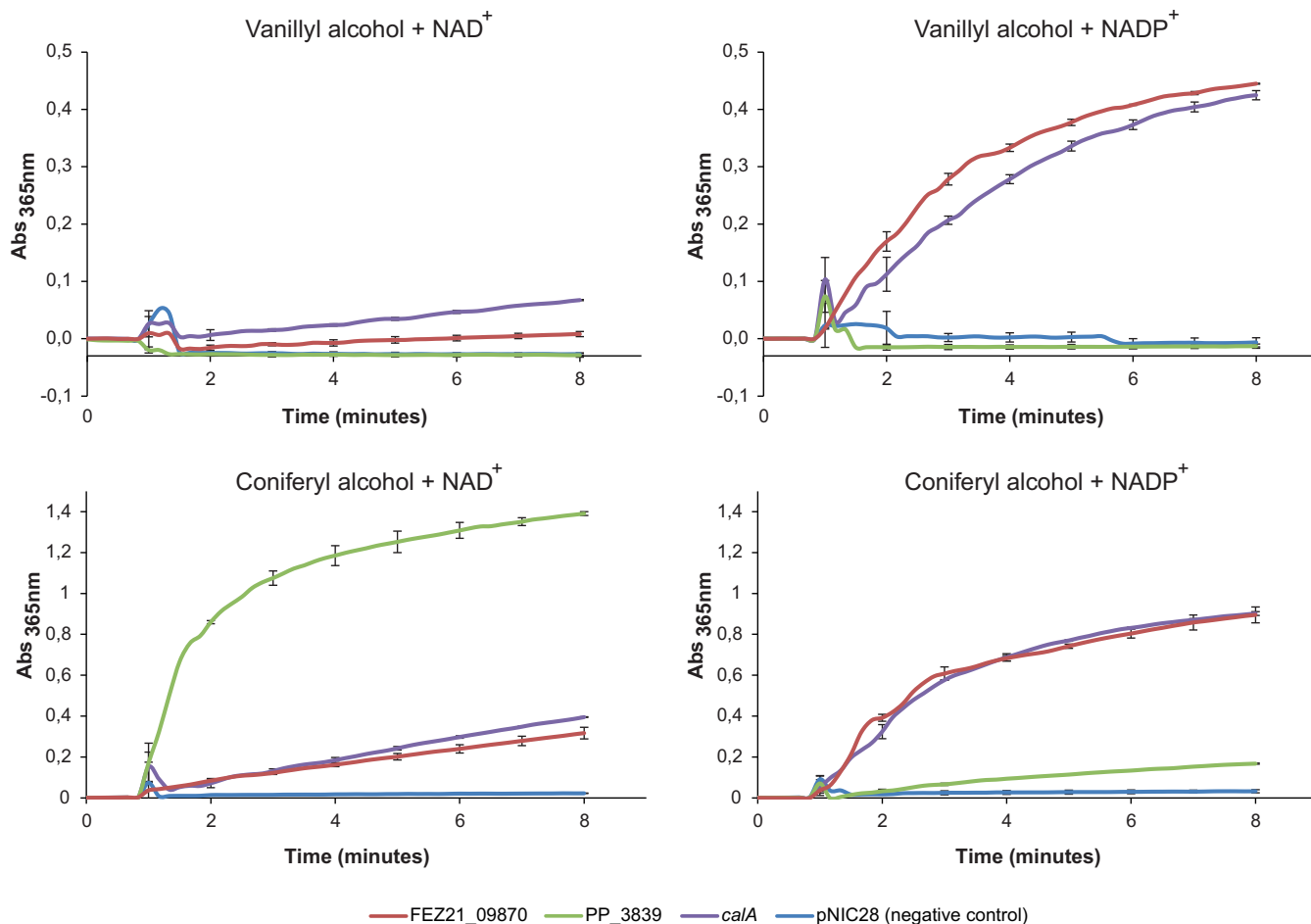
While the FEZ21\_09870 and PP\_2426 gene products clearly reduced their substrates, there was additional background reductase activity on vanillin (Fig. 2A1 and 2) and coniferyl aldehyde (Fig. 2C1 and 2) in the *E. coli* BL21(DE3) clones expressing PP\_3839 and in the negative control (empty plasmid). This is likely due to endogenous *E. coli* aldehyde reductases such as YahK, YjgB (*ahr*), and YqhD (40), which are present in the host strain BL21(DE3) genome (GenBank accession no. CP001509; Jeong et al. [41]). The



**FIG 2** Whole-cell oxidoreductase assays with *E. coli* BL21(DE3) cells expressing the three candidate genes FEZ21\_09870 (red squares), PP\_3839 (green triangles), and *calA* (purple crosses) and the negative-control pNIC28-Bsa4KpnI (blue diamonds). (A to C) Three assays with different substrates were performed and evaluated with HPLC, as follows: A1 and A2 show concentrations of vanillin (substrate) and vanillyl alcohol (product), respectively; B1 and B2 show concentrations of coniferyl alcohol (substrate) and coniferyl aldehyde (product), respectively; and C1 and C2 show concentrations of coniferyl aldehyde (substrate) and coniferyl alcohol (product), respectively. Experiments were performed in duplicates, and the standard deviations are displayed with an error bar.

lower rate of background activity compared to that observed in the clones expressing the FEZ21\_09870 and PP\_2426 gene products suggests that the enzymes responsible for the background activity are probably not specific for these substrates, and their level of expression is likely much lower than that of the overexpressed enzymes.

To determine the redox cofactor preference for each one of the candidate enzymes (encoded by FEZ21\_09870, PP\_2426 [*calA*], PP\_3839), an *in vitro* spectrophotometric

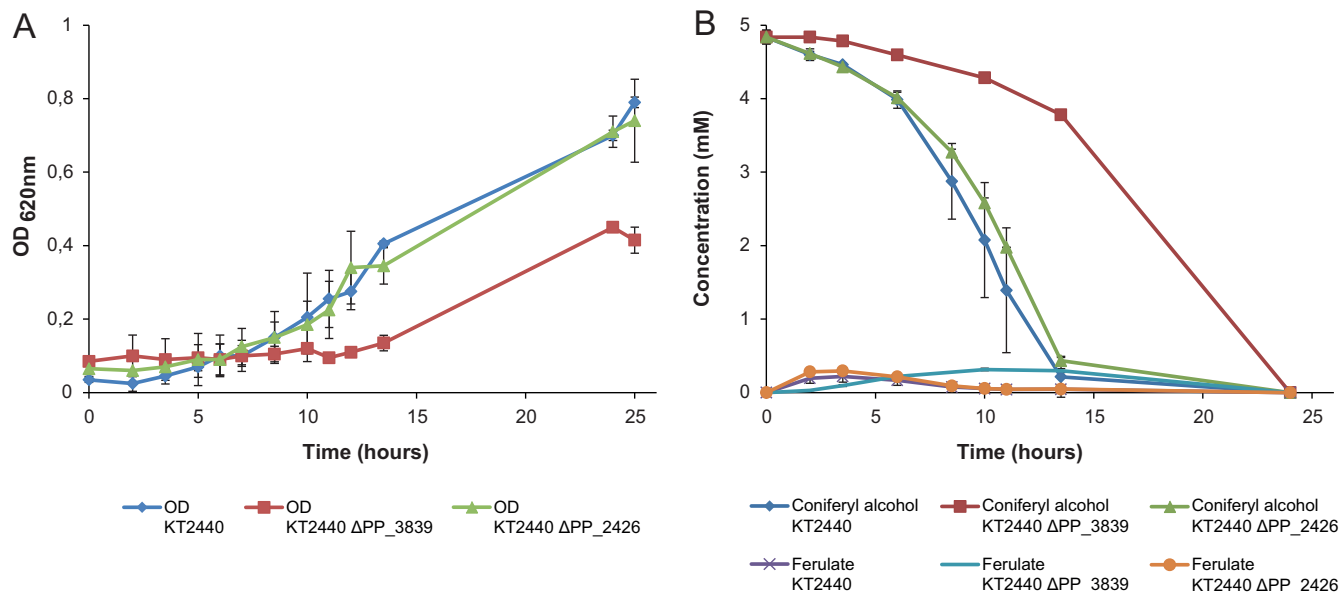


**FIG 3** *In vitro* enzyme assays with intracellular extracts of *E. coli* BL21(DE3) cells expressing the three candidate genes FEZ21\_09870 (red lines), PP\_3839 (green lines), and *calA* (PP\_2426) (purple lines) and the negative-control pNIC28-Bsa4KpnI (blue lines), with two different substrates (vanillyl alcohol and coniferyl alcohol) and two different redox cofactors (NAD<sup>+</sup> and NADP<sup>+</sup>) in order to determine the preference of each enzyme for substrates and cofactors. Due to overlap in the absorbance (Abs) of products and reduced cofactors, the monitored absorbance was shifted from 340 to 365 nm, and no activity units were calculated. Experiments were performed in duplicate, and the standard deviations are displayed with error bars.

enzyme assay was developed with two different substrates, vanillyl alcohol and coniferyl alcohol. Both assayed reactions involve the oxidation of alcohols to aldehydes, given that oxidoreductases can catalyze their reactions in both directions depending on their architecture, substrate/product concentration, and cofactor availability. Activities were evaluated in the presence of NAD<sup>+</sup> and NADP<sup>+</sup>. For these assays, soluble intracellular extracts of the *E. coli* BL21(DE3) clones overexpressing FEZ21\_09870 and PP\_2426 were employed. Both the PP\_2426 and FEZ21\_09870 gene products showed clear oxidation of vanillyl alcohol in the presence of NADP<sup>+</sup> (Fig. 3), despite being the opposite reaction as observed in the *in vivo* whole-cell assays. They also showed low activity in the presence of NAD<sup>+</sup> (Fig. 3). When assayed in the presence of coniferyl alcohol, the two enzymes confirmed their preference for NADP<sup>+</sup>, while presenting a certain degree of activity with NAD<sup>+</sup> (Fig. 3). Taking together the outcomes of the whole-cell and *in vitro* assays, these results confirmed both enzymes as NADPH-dependent reductases.

The gene product of *adhP* (PP\_3839) did not show any oxidative activity whatsoever with vanillyl alcohol as a substrate (Fig. 3), whereas coniferyl alcohol was rapidly oxidized in the presence of NAD<sup>+</sup>. A very low activity with NADP<sup>+</sup> was also measured, being barely above that detected for the negative control.

***adhP* gene (PP\_3839) encodes a coniferyl alcohol dehydrogenase.** Enzyme assays performed with the gene product of PP\_3839 (*adhP*) clearly pointed to a



**FIG 4** Fermentations of *P. putida* KT2440 and deletion strains with coniferyl alcohol 5 mM as the sole carbon source. (A) OD<sub>620</sub> measurements of *P. putida* KT2440 (blue diamonds), *P. putida* KT2440 ΔPP<sub>3839</sub> (red squares), and *P. putida* KT2440 ΔPP<sub>2426</sub> (green triangles). (B) HPLC-determined concentration of coniferyl alcohol with *P. putida* KT2440 (dark-blue diamonds), *P. putida* KT2440 ΔPP<sub>3839</sub> (red squares), and *P. putida* KT2440 ΔPP<sub>2426</sub> (green triangles) and concentration of ferulate detected with *P. putida* KT2440 (purple crosses), *P. putida* KT2440 ΔPP<sub>3839</sub> (light-blue line), and *P. putida* KT2440 ΔPP<sub>2426</sub> (orange circles). Experiments were performed in duplicate, and the standard deviations are displayed with error bars.

physiological role in coniferyl alcohol oxidation for this enzyme, which is precisely the function previously proposed for *calA* (PP<sub>2426</sub>). The latter gene product, on the other hand, seemed to have a much higher reductase activity. To clarify this discrepancy, genomic deletion of both candidate genes was performed in the parental strain KT2440, followed by the analysis of the growth profile and substrate consumption of each deletion strain in the presence of 5 mM coniferyl alcohol.

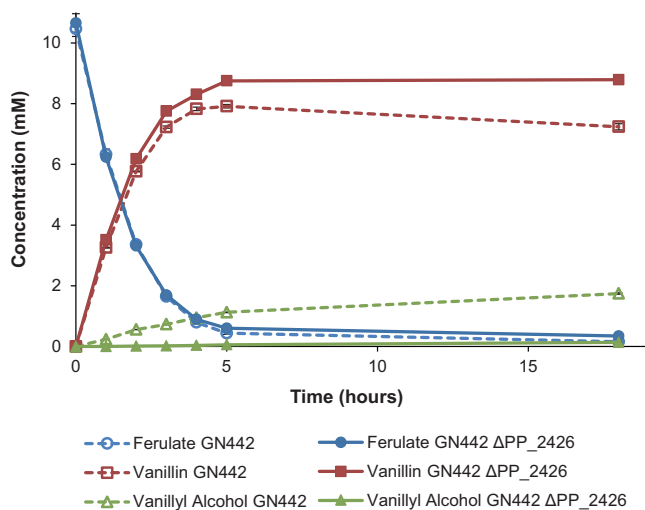
The growth behavior of the strain carrying the *calA* deletion (KT2440 ΔPP<sub>2426</sub>) under these conditions was almost identical to that of the wild-type strain (Fig. 4), consuming all of the substrate in the first 14 h, whereas the deletion strain KT2440 ΔPP<sub>3839</sub> showed a clearly different phenotype, with delayed growth and slower uptake of coniferyl alcohol. This result strongly supports the role of *adhP*, and not *calA*, as the gene encoding the main coniferyl alcohol dehydrogenase in *P. putida* KT2440, although it is clear that there are other alcohol dehydrogenases in this strain that are to oxidize this substrate with lower specificity.

#### Deletion of PP<sub>2426</sub> in *Pseudomonas* decreases vanillyl alcohol formation.

Vanillin catabolism is operating at higher rates in *P. putida* KT2440 (4.87 mmol-gram of cell dry weight [g<sub>CDW</sub>]<sup>-1</sup>·h<sup>-1</sup>; Ravi et al. [20]) than in strain 9.1 (3.19 mmol·g<sub>CDW</sub><sup>-1</sup>·h<sup>-1</sup>; Ravi et al. [19]), and vanillyl alcohol is not found as an excreted intermediate in KT2440. However, strain GN442, a KT2440-derived strain with deletions in genes directly and indirectly involved in the oxidation of vanillin, formed up to 15% (mol/mol) vanillyl alcohol during vanillin production from ferulate (26). This made GN442 an ideal background strain to evaluate the proposed vanillin reductase activity of the PP<sub>2426</sub> gene product.

The formation of the vanillyl alcohol by-product was reproduced in *P. putida* GN442 using the established bioconversion protocol (26), with a final vanillyl alcohol yield of ~17% (mol/mol) after 18 h, (dashed lines, Fig. 5). In the deletion strain (GN442 ΔPP<sub>2426</sub>), however, the by-product formation was reduced to only ~1% (mol/mol) vanillyl alcohol, detected after 18 h of bioconversion, and the vanillin yield increased from 69% in GN442 to 82% (mol/mol) in GN442 ΔPP<sub>2426</sub> (Fig. 5). The vanillin yield loss in GN442 appears to mainly result from the reduction of formed vanillin to vanillyl alcohol in the later stages of the process (Fig. 5).





**FIG 5** Bioconversion of ferulate into vanillin with the engineered *P. putida* strains GN442 (dashed lines) and GN442  $\Delta$ PP\_2426 (solid lines). Concentrations of ferulate (blue circles), vanillin (red squares), and vanillyl alcohol (green triangles) were determined by HPLC. Experiments were performed in duplicate, and the standard deviations are displayed with error bars.

## DISCUSSION

In the present study, whole-genome sequencing of *Pseudomonas* sp. 9.1, an isolate from lignin-rich waste material and that accumulates vanillyl alcohol from vanillin metabolism (19), led to the identification and reassessment of key genes participating in the metabolism of aromatic aldehydes. This knowledge was further used to obtain significant improvement in vanillin production from ferulate using genetic engineering.

The use of the fungal protein sequence from ScADH6p to query the 9.1 and KT2440 genomes for potential vanillin reductases that could explain the vanillyl alcohol formation proved to be a successful strategy. Contrary to the results from homology searches using genes/open reading frames associated with known participants in the funneling pathway (Table S1), the candidate open reading frames in the 9.1 and KT2440 genomes identified using the ScADH6p query only had sequence identities of around 25 to 40% (Table 2). It has been proposed as a rule of thumb that  $\geq 30\%$  sequence identity is a useful threshold for defining protein homology (42), although it has been debated that this cutoff might be too conservative and that E values are preferred (43). This study shows that low-identity hits should not be discarded when comparing proteins between different species, or belonging to different domains of life, as in the current case, between yeast and bacteria.

In contrast, the results from the experimental characterization of the three putative vanillin reductase genes, FEZ21\_09870, *calA* (PP\_2426), and *adhP* (PP\_3839), highlighted that *in silico* predictions should be considered putative until experimental validation has been performed because they are, to a high degree, dependent on previous annotations in databases (44). Consequently, they are not as good at predicting less well-annotated reactions, such as the aromatic aldehyde reductions in the present case.

A main finding of the present study is that the function of the PP\_2426 (*calA*) gene needs to be reannotated. One of the first descriptions of *calA* as a coniferyl alcohol dehydrogenase was from *Pseudomonas* sp. strain HR199 (45). However, the corresponding homologue in *P. putida* has been elusive; during the first genome sequencing project of KT2440, no *calA* orthologue was found (only *calB*, which catalyzes the next step in the pathway, being the conversion of coniferyl aldehyde to ferulate) (16, 46). It was only in a recent genome reassessment that a *calA* gene (PP\_2426) was proposed in KT2440 (39); however, this prediction lacked experimental verification, and, as the present study has demonstrated, PP\_2426 does not encode a coniferyl alcohol dehydrogenase.

**TABLE 3** Summary of the findings of the three candidate genes

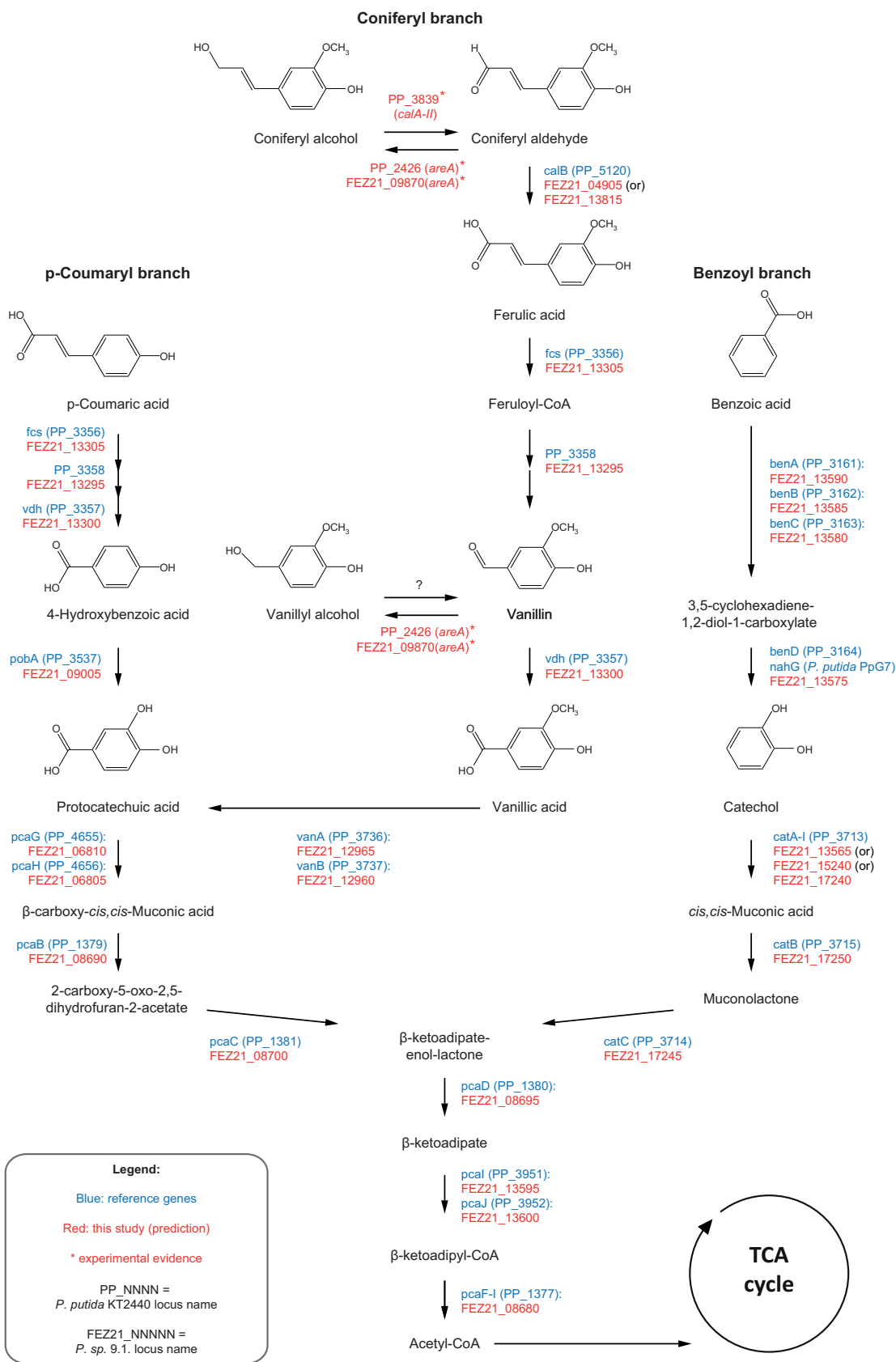
Gene	Observations	Suggested reannotation
<i>Pseudomonas</i> sp. 9.1 FEZ21_09870	Clear reductase activity with vanillin and coniferyl aldehyde; no oxidizing activity with aromatic alcohols detected; confirmation of <i>in silico</i> predicted activity	Psp91_areA (vanillin reductase/aromatic aldehyde reductase gene)
<i>P. putida</i> KT2440 PP_2426 ( <i>calA</i> )	High sequence identity to FEZ21_09870 (82.3%; BLASTp); identical activity to FEZ21_09870; unexpected reductase activity on coniferyl aldehyde, instead of the proposed oxidative activity (39); reannotation needed	areA (vanillin reductase/aromatic aldehyde reductase gene)
<i>P. putida</i> KT2440 PP_3839 ( <i>adhP</i> )	No activity with aromatic aldehydes; rapid conversion of coniferyl alcohol into coniferyl aldehyde; likely the true <i>P. putida</i> coniferyl alcohol dehydrogenase, a function previously incorrectly attributed to PP_2426	calA-II (coniferyl alcohol dehydrogenase gene)

drogenase but rather an aromatic aldehyde reductase able to reduce coniferyl aldehyde to coniferyl alcohol. Instead, it is *P. putida* KT2440 PP\_3839 that should be annotated as a *cal* gene encoding coniferyl alcohol dehydrogenase activity, and we suggest the name *calA-II* in order not to cause conflict with older literature where PP\_2426 was annotated as *calA*. We also suggest that the PP\_2426 gene should be reannotated as *areA*, since the gene product from this open reading frame catalyzed the reduction of aromatic aldehydes to alcohols. It thus follows that FEZ21\_09870 from *Pseudomonas* sp. 9.1 should also be annotated as Psp91\_areA, as it displayed the same activity as PP\_2426 (*areA*). The enzyme that allowed strain 9.1 to consume the accumulated vanillyl alcohol after vanillin was consumed (19) is yet to be identified. A summary of the findings and the proposed reannotations can be found in Table 3.

Using the new information about the functionality of the three candidate proteins, we propose an improved map of the aromatic funneling pathways in *P. putida* KT2440 (Fig. 6; gene names beginning with PP). The figure (Fig. 6) also contains the result of the analysis of the annotation of strain 9.1 (gene names beginning with FEZ21). The systematic locus names of the genes show that many of the funneling pathway genes in 9.1 are found in close proximity within the genome (just like in other *Pseudomonas* species [16]), suggesting that they belong to putative operons (Fig. S4). The annotation of FEZ21\_09870 (Psp91\_areA) as an aldehyde reductase made us wonder whether there was a PP\_3839 (*calA-II*) gene equivalent in the strain 9.1 genome. Although strain 9.1 has not been assayed for growth on coniferyl alcohol, the growth on ferulate (downstream in the funneling branch) and the putative hits for *calB* in the assembly (Fig. 6) hinted to the possibility of a coniferyl alcohol dehydrogenase phenotype. The top BLASTp hits for PP\_3839 (*calA-II*) equivalent in the 9.1 genome were FEZ21\_15000 and FEZ21\_09870 (Psp91\_areA) (Tables 2 and S1), the latter having already been demonstrated to catalyze the reverse reaction. FEZ21\_15000 was the second highest hit for ScADH6p in 9.1 (Table 2), suggesting the possibility that it encodes an oxidoreductase, although as we have demonstrated in the current study, experimental assessment is required to propose a more specific physiological role for FEZ21\_15000.

While the NAD(P)H-binding motif GxGxxG/A (47, 48) is present in the three investigated proteins and the reference protein (ScADH6), any further extrapolation on cofactor usage was not possible from sequence analysis alone because accurate distinction of NADH sites from NADPH sites from a given primary polypeptide sequence is nontrivial (49). It is widely assumed that NAD(H)-dependent proteins are generally related to oxidative metabolic processes (48, 50), and NADP(H)-dependent proteins are involved in the reduction of xenobiotic carbonyls in yeasts (51, 52) and in lignin biosynthesis in plants (53, 54). In the current study, the *calA-II* (PP\_3839) gene product involved in the assimilation of coniferyl alcohol was indeed NAD<sup>+</sup> dependent, whereas PP\_2426 (*areA*) and FEZ21\_09870 (Psp91\_areA) were mostly NADPH dependent and indeed fulfilled a detoxification role.

One of the few other studies on *Pseudomonas* spp. that mentions the same vanillyl alcohol phenotype as strain 9.1 concerns the engineering of *P. putida* KT2440 to accumulate vanillin from ferulate; the authors were unable to find a deletion that could



**FIG 6** *In silico* predictions for the aromatic funneling pathway genes in *Pseudomonas* sp. 9.1 and *P. putida* KT2440. Growth has previously been demonstrated for strain 9.1 on ferulic acid, *p*-coumaric acid, benzoic acid and vanillin (19), which suggests that known *P. putida* funneling pathways are also likely to be found in 9.1. The suggested reannotations of the three candidate genes

(Continued on next page)

alleviate vanillyl alcohol by-product formation (26). We applied our findings about PP\_2426 (*areA*) to their final strain GN442 and were able to demonstrate that the GN442  $\Delta$ PP\_2426 strain presented greatly decreased vanillyl alcohol formation compared to that of the parental strain (17% and 1%, respectively). This also shows that there is little to no background aldehyde reductase activity in KT2440 that acts on vanillin once PP\_2426 (*areA*) has been deleted, which reinforces the proposed functional reannotation of this gene, and by extension Psp91\_2426 (FEZ21\_09870), which shares the same activity.

Besides helping us determine the *in vivo* role of PP\_2426, the resulting mutant strain GN442  $\Delta$ PP\_2426 represents an interesting platform for the conversion of ferulate into vanillin with very high yield and short conversion times. This specific result is also a valuable advance for any endeavor aiming for the biological production of vanillin or related aldehydes using engineered *Pseudomonas* strains or any other microbial platform.

The present study demonstrates that the characterization of novel isolates not only provides information about the diversity of metabolic phenotypes in nature but also describes how this information can be applied to address unresolved traits of model strains. By investigating the peculiar trait of *Pseudomonas* sp. 9.1 to accumulate vanillyl alcohol during growth on vanillin (only to consume it again after vanillin is depleted), we have identified the gene encoding the aldehyde reductase in this strain (FEZ21\_09870). PP\_2426 (here proposed to be reannotated as *areA*) in *P. putida* KT2440 had the same activity on vanillin, a trait that is completely masked in wild-type KT2440 due to different tuning of activities in the coniferyl funneling pathway branch compared to strain 9.1. This finding resulted in the improved KT2440-derived cell factory (GN442  $\Delta$ PP\_2426) designed to overproduce vanillin by the deletion of PP\_2426 (*areA*). Additionally, we proved that PP\_3839, unspecifically annotated as alcohol dehydrogenase gene *adhP*, is the gene encoding the dehydrogenase responsible for the oxidation of coniferyl alcohol to coniferyl aldehyde (a function inaccurately attributed to the gene product of PP\_2426). We propose the reannotation of this gene as coniferyl alcohol dehydrogenase gene *calA-II*.

This study thus reinforces the importance of exploring novel isolates together with comparisons to the well-understood model organisms. A strain that may be undesirable as an industrial cell factory (e.g., in terms of rates and yields compared to the model strain) may reveal, through targeted genome sequence analysis, novel traits that can both contribute valuable knowledge to the understanding of the genus and lead to improved engineered strains for diverse biotechnological applications.

## MATERIALS AND METHODS

**Genome extraction and sequencing.** *Pseudomonas* sp. isolate 9.1 (DSM 105530; Ravi et al. [19]) and *P. putida* KT2440 (DSM 6125) were cultivated in liquid LB medium (Sambrook and Russell [55]; see "Culture media and conditions," below), and genomic DNA from strain 9.1 was extracted using the Invitrogen PureLink genomic DNA minikit (Thermo Scientific, Carlsbad, CA, USA) and eluted with Invitrogen UltraPure DNase/RNase-free distilled water (Thermo Fisher Scientific, Waltham, MA, USA). Genomic DNA from *P. putida* KT2440 was purified using the GeneJET genomic DNA purification kit from Thermo Fisher Scientific Baltics (Vilnius, Lithuania). Paired-end whole-genome sequencing was performed using an Illumina MiSeq (2 × 150 bp) platform at GATC Biotech AG (Cologne, Germany). The raw data from the sequencer consisted of 6,829,359 paired reads with an average read length of 151 bp and 58% GC content (Table 1). Read quality was assessed with FastQC (v0.11.5; Andrews [56]).

**De novo assembly.** The raw reads were subjected to quality trimming with sickle (v1.210; Joshi and Fass [57]), followed by an additional round of quality control with FastQC. *De novo* assembly of the trimmed reads was performed with SPAdes (v3.11.1; Bankevich et al. [27]), and assembly metrics were analyzed with QUAST (v4.5.4; Gurevich et al. [58]). To improve the assembly, a custom iterative gap-filling pipeline was designed where the assembly and trimmed reads were in turn subjected to SSPACE\_Standard (v3.0; Boetzer et al. [31]) with bwa (v0.7.17; Li and Durbin [59]), AlignGraph (28) with bowtie2 (v2.3.3.1; Langmead and Salzberg [60]) and blat (v34; Kent [61]), and GapFiller (v1.10; Boetzer and Pirovano [29]) and ABySS-sealer (v2.0.2; Paulino et al. [30]). QUAST was used to measure assembly improvements after each iteration of the pipeline; after three iterations of gap filling, the improvement

### FIG 6 Legend (Continued)

evaluated in this study (FEZ21\_09870, PP\_2426, and PP\_3839) are referred to by the suggested reannotations (cf. the Discussion). The reaction in strain 9.1 to consume the accumulated vanillyl alcohol is unknown, and thus, the proposed reaction to oxidize it back to vanillin is denoted by a question mark. CoA, coenzyme A; TCA, tricarboxylic acid.

in assembly metrics reached stagnation, and thus, it was decided to proceed with that version of the assembly. Contigs were sorted using Mauve (v2015-02-13; Darling et al. [62]) in a biologically relevant order using the closely related *Pseudomonas deceptionensis* LMG 25555 assembly (GenBank accession number GCA\_900106095.1) as a reference genome. The final assembly was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (63).

**Phylogeny.** Genome-wide phylogeny was performed with a set of genomes from 108 nonclinical isolates in the *Pseudomonas* genome database (Winsor et al. [64]). The number of genomes was, however, reduced in an iterative process to find trees with good statistical basis (i.e., removing genomes that were too similar to their closest relatives). The phylogeny was calculated using RealPhy (v1.12; Bertels et al. [34]), which was run with bowtie2 (v2.3.3.1; Langmead and Salzberg [60]) and SAMtools (v1.6; Li et al. [65]) as a read mapper. Strain 9.1 was compared to the other genomes using fastq paired-end reads after trimming with sickle (see "De novo assembly," above). The set of genomes that was used in the final version of the RealPhy analysis used 26 genomes (including strain 9.1) and inferred phylogeny from 4,223 patterns out of a total of 53,115 polymorphic sites in the set. All trees were built with the maximum likelihood method using RAxML (v8.2.10; Stamatakis [66]) with the GTRGAMMA model and 100 bootstrap iterations. Dendroscope (v3.5.9; Huson and Scornavacca [67]) was used to visualize the phylogenetic trees.

**Identification of funneling pathways and putative oxidoreductases.** Stand-alone BLASTp using the NCBI Genome Workbench (v2.12.10) was used to compare known funneling pathway proteins from *Pseudomonas* species to the genome of strain 9.1. The *Pseudomonas* sp. 9.1 assembly was converted into a stand-alone BLAST database with makeblastdb (BLAST v2.7.1). Query protein amino acid sequences (Table S1) were downloaded from NCBI GenBank (68). The amino acid sequence of *S. cerevisiae* S288c ADH6p was taken from The *Saccharomyces* Genome Database (<https://www.yeastgenome.org/>) (69). Putative vanillin reductases with homology to ScADH6p were identified with the same stand-alone BLASTp procedure using the 9.1 assembly as well as the latest version of the *P. putida* KT2440 genome (GenBank accession no. AE015451.2) (39). The HMMER Web server (<https://www.ebi.ac.uk/Tools/hmmer/>) was used to analyze the candidate proteins for structural features (38), and protein domain data were taken from the Pfam database (<http://pfam.xfam.org/>) (El-Gebali et al. [70]).

**Culture media and conditions.** *E. coli* BL21(DE3) (71) cells were routinely cultured in lysogeny broth (LB) medium (10 g/liter tryptone, 5 g/liter yeast extract, 10 g/liter NaCl) at 37°C. *Pseudomonas* strains were cultured in LB medium or in mineral M9 medium (55) supplemented with trace element solution (72) and appropriate carbon sources at 30°C. Antibiotics were added to the media when necessary at the following concentrations: 150 µg/ml ampicillin, 50 µg/ml kanamycin, and 30 µg/ml chloramphenicol; the concentration of ampicillin was increased to 500 µg/ml for *Pseudomonas* strains. IPTG was added at 1 mM to induce the expression of genes under the T7 promoter (pNIC28 plasmid) or 5 mM to induce the expression of the *ech-fcs* operon controlled by the *tac* promoter in *P. putida* GN442 and GN442 ΔPP\_2426 strains before the bioconversion assays. All the strains and plasmids used in the study are listed in Table 4.

**Shake flask fermentations.** Seed cultures of *P. putida* KT2440 (DSM 6125), *P. putida* KT2440 ΔPP\_3839, and *P. putida* KT2440 ΔPP\_2426 were grown overnight in 10 ml M9 medium with 10 g/liter glucose in 50-ml conical tubes. Cells from these seed cultures were harvested, washed with 0.9% sterile saline solution, and used to inoculate 250-ml shake flasks with 20 ml M9 medium with 5 mM coniferyl alcohol at an optical density at 620 nm (OD<sub>620</sub>) of around 0.1. Flasks were further incubated at 30°C with 200 rpm orbital shaking. Samples were taken at different time points for OD<sub>620</sub> measurement and high-performance liquid chromatography (HPLC) analysis. All fermentations were carried out in duplicate.

**Ferulate bioconversion assays.** Bioconversion assays of ferulate to vanillin were performed with strains *P. putida* GN442 (26) and *P. putida* GN442 ΔPP\_2426, as described by Graf and Altenbuchner (26), with slight modifications. Briefly, overnight LB cultures of both strains were diluted 50 times and further cultured in LB shake flasks to generate biomass, with overexpression of the enzymes responsible for the conversion of ferulate (Fcs and Ech) induced by the addition of IPTG. After this phase, the OD<sub>620</sub> was measured, and cells were harvested, aiming for a final OD<sub>620</sub> of 7.2 (approximately 25 × 10<sup>9</sup> cells). These cells were gently washed with 50 mM sterile sodium phosphate buffer (pH 7.2) and resuspended in 5 ml of the same buffer containing 10 mM ferulate in 50-ml conical tubes. Tubes were shaken at 200 rpm and 30°C for 18 h; samples were taken at several time points for HPLC analysis.

**Cloning and expression of recombinant oxidoreductases.** Candidate genes *calA* (PP\_2426) and PP\_3839 from *P. putida* KT2440 and FEZ21\_09870 from *Pseudomonas* sp. 9.1 were amplified by PCR with Phusion high-fidelity DNA polymerase (Thermo Fisher Scientific), according to the instructions of the manufacturer, using the primers *calA* LIC Fw and *calA* LIC Rv, PP\_3839 LIC Fw and PP\_3839 LIC Rv, and *peg1955* LIC Fw and *peg1955* LIC Rv, respectively (see full primer list and sequences in Table S2). The resulting DNA fragments were purified using the GeneJET PCR purification kit (Thermo Fisher Scientific) and sent for sequencing with the same primers to Eurofins Genomics (Ebersberg, Germany). Verified fragments were inserted into BsaI-linearized pNIC28-Bsa4 plasmid by ligation-independent cloning with the In-Fusion HD cloning kit (Clontech, Mountain View, CA, USA) and introduced into chemically competent *E. coli* BL21(DE3) cells. In order to generate a negative-control plasmid not producing levansucrase *SacB*, pNIC28-Bsa4 was digested with KpnI and then religated to lose the negative selection marker *sacB*, obtaining the plasmid pNIC28-Bsa4KpnI, which was also introduced into *E. coli* BL21(DE3).

For the overexpression of the cloned genes, overnight LB cultures of the corresponding *E. coli* BL21(DE3) clones and negative-control strain *E. coli* BL21(DE3)/pNIC28-Bsa4KpnI were diluted to an OD<sub>620</sub> of 0.1 in 30 ml fresh LB medium with kanamycin and further incubated at 37°C until reaching OD<sub>620</sub> of

**TABLE 4** List of plasmids and strains used in the current study

Strain or plasmid	Features or genotype <sup>a</sup>	Reference
<b>Bacterial strains</b>		
<i>E. coli</i> DH5 $\alpha$ $\lambda$ pir	Strain used for preparation and conjugative delivery of pSEVA212S-derived plasmids; F <sup>-</sup> $\phi$ dlacZ $\Delta$ M15 $\Delta$ lacZYA-argF U169 deoR supE44 hsdR17 recA1 endA1 gyrA96 thi-1 relA1	Dunn et al. (78)
<i>E. coli</i> HB101	Helper strain carrying pRK600 plasmid, used in triparental conjugation; F <sup>-</sup> thi-1 hsdS20 (r <sub>B</sub> <sup>-</sup> , m <sub>B</sub> <sup>-</sup> ) supE44 recA13 ara-14 leuB6 proA2 lacY1 galK2 rpsL20 (Str <sup>r</sup> ) xyl-5 mtl-1	Boyer and Roulland-Dussoix (79)
<i>E. coli</i> BL21(DE3)	Strain used for high-level T7 expression of cloned oxidoreductases; fhuA2 [lon] ompT gal ( $\lambda$ DE3) [dcm] $\Delta$ hsdS $\lambda$ sBamHlo $\Delta$ EcoRI-B int:: (lac::PlacUV5::T7 gene1) i21 $\Delta$ lin5	Studier and Moffatt (71)
<i>Pseudomonas putida</i> KT2440	Model organism	DSM 6125
<i>P. putida</i> KT2440 $\Delta$ PP_2426	Deletion mutant without <i>calA</i> (PP_2426) ORF	This study
<i>P. putida</i> KT2440 $\Delta$ PP_3839	Deletion mutant without PP_3839 ORF	This study
<i>P. putida</i> GN442	Engineered strain for production of vanillin from ferulate; $\Delta$ upp $\Delta$ PP_0166-0168 $\Delta$ vdh $\Delta$ PP_3827-3832 $\Delta$ PP_2680 $\Delta$ PP_0545 $\Delta$ PP_1948 <i>lacI<sup>q</sup>-P<sub>tac</sub>-ech-fcs</i>	Graf and Altenbuchner (26)
<i>P. putida</i> GN442 $\Delta$ PP_2426	Optimized strain for high-yield production of vanillin from ferulate; $\Delta$ upp $\Delta$ PP_0166-0168 $\Delta$ vdh $\Delta$ PP_3827-3832 $\Delta$ PP_2680 $\Delta$ PP_0545 $\Delta$ PP_1948 <i>lacI<sup>q</sup>-P<sub>tac</sub>-ech-fcs</i> $\Delta$ PP_2426	This study
<i>Pseudomonas</i> sp. 9.1	Environmental isolate	Ravi et al. (19), DSM 105530
<b>Plasmids</b>		
pNIC28-Bsa4	Plasmid for ligation-independent cloning and T7 expression of cloned genes, Km <sup>r</sup>	Savitsky et al. (80)
pNIC28-Bsa4KpnI	KpnI-digested and religated version of pNIC28-Bsa4 used as negative-control plasmid, Km <sup>r</sup>	This study
pNIC28-Bsa4- <i>calA</i>	Clone of PP_2426 in pNIC28-Bsa4, Km <sup>r</sup>	This study
pNIC28-Bsa4-PP_3839	Clone of PP_3839 in pNIC28-Bsa4, Km <sup>r</sup>	This study
pNIC28-Bsa4-FEZ21_09870	Clone of FEZ21_09870 in pNIC28-Bsa4, Km <sup>r</sup>	This study
pSEVA212S	Suicide plasmid for integration in <i>Pseudomonas</i> genomes, Km <sup>r</sup>	Martínez-García and de Lorenzo (75)
pSEVA212S- $\Delta$ <i>calA</i>	pSEVA212S containing the upstream and downstream regions of <i>calA</i> assembled together, Km <sup>r</sup>	This study
pSEVA212S- $\Delta$ PP_3839	pSEVA212S containing the upstream and downstream regions of PP_3839 assembled together, Km <sup>r</sup>	This study
pRK600	Helper plasmid used in triparental mating between <i>E. coli</i> and <i>Pseudomonas</i> strains, Cm <sup>r</sup>	Finan et al. (81)
pSEVA128S	Plasmid carrying the gene encoding I-SceI endonuclease, Ap <sup>r</sup>	Martínez-García and de Lorenzo (75)

<sup>a</sup>Str<sup>r</sup>, streptomycin resistance; Km<sup>r</sup>, kanamycin resistance; Cm<sup>r</sup>, chloramphenicol resistance; Ap<sup>r</sup>, apramycin resistance.

around 0.6, and then 1 mM IPTG was added and incubation was extended for 5 additional hours. Cells were subsequently harvested and washed with sterile 0.9% NaCl and resuspended in 1 ml lysis buffer (25 mM Tris-HCl [pH 8] with NaCl 100 mM) prior to their lysis by sonication with a Branson Sonifier 150 equipped with a microprobe. Finally, lysed cells were centrifuged at 4°C and 21,130  $\times$  *g* for 20 min, and supernatants were collected as soluble intracellular extracts for SDS-PAGE analysis and *in vitro* enzyme assays. SDS-PAGE analysis of cell extracts was performed in a 4 to 20% Mini-Protean TGX precast gel (Bio-Rad, Hercules, CA, USA) using a PageRuler prestained 10- to 180-kDa protein ladder (Thermo Fisher Scientific Baltics, Vilnius, Lithuania) as a molecular mass marker.

**Whole-cell oxidoreductase assays.** *E. coli* BL21(DE3) cells overexpressing the cloned genes from *Pseudomonas* spp. as well as the negative-control strain *E. coli* BL21(DE3)/pNIC28-Bsa4KpnI were prepared from 5 ml of LB culture as described in "Culture media and conditions" above. Following the wash with saline solution, cells were resuspended in 5 ml of M9 medium with 50  $\mu$ g/ml kanamycin, 1 mM IPTG, 10 g/liter D-glucose, and 5 mM aromatic substrates (vanillin, vanillyl alcohol, coniferyl aldehyde, or coniferyl alcohol). These cells were incubated at 30°C and 250 rpm orbital shaking for 20 h; samples were taken at several time points for HPLC analysis. All assays were done in duplicate.

**In vitro enzyme assays.** Relative alcohol dehydrogenase activity of the different overexpressed enzymes in *E. coli* BL21(DE3) soluble intracellular extracts was assessed in the presence of two different substrates (vanillyl alcohol and coniferyl alcohol) with two different redox cofactors (NAD<sup>+</sup> and NADP<sup>+</sup>) by a spectrophotometric assay based on that described by Larroy et al. (73). Assays were carried out in duplicate in a total volume of 600  $\mu$ l at 30°C in an Ultrospec 2100 pro UV-Vis spectrophotometer (GE Healthcare, Little Chalfont, UK) equipped with a Peltier cell holder. The reaction mixture contained 33 mM sodium phosphate buffer (pH 7) with 1 mM substrate and 0.5 mM cofactor. Reactions were started by adding the corresponding cell extract (40  $\mu$ g total protein) after 1 min of baseline reading. Progression of the reactions was followed by monitoring the absorbance at 365 nm every 10 s for 8 min.

**Genomic deletions in *Pseudomonas* strains.** Genomic deletions of individual genes in *P. putida* KT2440 and *P. putida* GN442 were carried out following previous protocols (74–76). Briefly, the adjacent regions of the target genes were amplified by PCR, assembled by overlap-extension PCR (see list of primers used in Table S2), and cloned into the suicide plasmid pSEVA212S carrying a kanamycin resistance gene and two sites for the homing endonuclease I-SceI. The resulting plasmids were introduced and maintained in *E. coli* DH5 $\alpha$  *lambda*pir and subsequently transferred to a recipient *Pseudomonas* strain by triparental mating using the aforementioned *E. coli* strain together with the helper strain *E. coli* HB101/pRK600. Since this plasmid cannot replicate in *Pseudomonas* spp., the kanamycin resistance can only be acquired by chromosomal integration of the whole plasmid. After this event, the plasmid pSEVA128S containing ampicillin resistance and the gene encoding the endonuclease I-SceI were delivered by electroporation to a kanamycin-resistant *Pseudomonas* culture, leading to the generation of two lethal double-strand breaks which must be repaired by RecA-dependent homologous recombination with the adjacent regions inserted in the pSEVA212S plasmid. Electroporation of *P. putida* electrocompetent cells was done according to the protocols of Martínez-García and de Lorenzo (74), using 0.2-cm-wide gap cuvettes with a Gene Pulser apparatus equipped with a pulse controller (Bio-Rad, Hercules, CA, USA). Deletion of the genes had to be confirmed by PCR amplification and sequencing of the surrounding region with the primers calA Fw, calA Rv, PP\_3839 Fw, and PP\_3839 Rv (Table S2). Finally, the plasmid pSEVA128S was cured by several consecutive passes on LB without antibiotic until the ampicillin resistance was lost. Verified mutant strains were stored at  $-80^{\circ}\text{C}$  until further use.

**Chromatographic analysis.** The culture aliquots were centrifuged ( $20,000 \times g$ ) for 3 min, and the supernatants were filtered before analysis. A Waters Acquity HPLC system coupled with a UV detector (Milford, MA, USA) was used for the analysis of aromatic compounds. An Agilent InfinityLab Poroshell 120 EC-C<sub>18</sub> column with an internal diameter of 4.6 mm, length of 100 mm, and 4- $\mu\text{m}$  particle size was used for separation. The column heater was set to  $50^{\circ}\text{C}$ . The mobile phases consisted of water and acetonitrile, with acetic acid as an additive. A gradient elution method previously described (77) was used for analysis. The sample injection volume was 5  $\mu\text{l}$ . Peaks were quantified using the area under the curve against their authentic calibration standards.

**Data availability.** The sequence reads from this article have been deposited at the NCBI Sequence Read Archive under the accession no. PRJNA543218. The assembly data set supporting the results of this article has been deposited at DDBJ/EMBL/GenBank under the accession no. VBWF00000000. The version described in this paper is VBWF00000000.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 0.8 MB.

## ACKNOWLEDGMENTS

The computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project snic2017-7-381. Plasmid pNIC28-Bsa4 and the *E. coli* BL21(DE3) strain were donated by Claes Von Wachenfeldt from the department of Biology of Lund University. Plasmids pSEVA212S, pSEVA128S, and pRK600 and the *E. coli* DH5 $\alpha$  *lambda*pir strain were a gift from Esteban Martínez-García and Tomás Aparicio from the National Centre for Biotechnology (CNB, Madrid, Spain). The *P. putida* GN442 strain was kindly provided by Josef Altenbuchner (Institute of Industrial Genetics, University of Stuttgart).

This work was financed by the Swedish Foundation for Strategic Research through grant contract RBP14-0052.

J.G.-H. performed the genomic DNA extraction, data analysis, genetic engineering, fermentations, and enzyme assays and collaborated in the writing of the manuscript. D.P.B. designed and performed the bioinformatics setup (assembly and phylogeny pipeline and subsequent analyses) and collaborated in the writing of the manuscript. K.R. performed the strain characterization and HPLC analysis. C.J.P. contributed to the original isolation of strain 9.1. G.L. and M.F.G.-G. conceived the study. C.J.P., G.L., and M.F.G.-G. revised the manuscript. All authors read and approved the final manuscript.

We declare no competing interests.

## REFERENCES

- Holladay JE, White JF, Bozell JJ, Johnson D. 2007. Top value added chemicals from biomass. Volume II: results of screening for potential candidates from biorefinery lignin. Pacific Northwest National Lab (PNNL), Richland, WA. [https://www.pnnl.gov/main/publications/external/technical\\_reports/PNNL-16983.pdf](https://www.pnnl.gov/main/publications/external/technical_reports/PNNL-16983.pdf).
- Ayyachamy M, Cliffe FE, Coyne JM, Collier J, Tuohy MG. 2013. Lignin: untapped biopolymers in biomass conversion technologies. *Biomass Conv Bioref* 3:255–269. <https://doi.org/10.1007/s13399-013-0084-4>.
- Li T, Takkellapati S. 2018. The current and emerging sources of technical

- lignins and their applications. *Biofuel Bioprod Biorefin* 12:756–787. <https://doi.org/10.1002/bbb.1913>.
4. Stewart D. 2008. Lignin as a base material for materials applications: chemistry, application and economics. *Ind Crops Prod* 27:202–207. <https://doi.org/10.1016/j.indcrop.2007.07.008>.
  5. Beckham GT, Johnson CW, Karp EM, Salvachúa D, Vardon DR. 2016. Opportunities and challenges in biological lignin valorization. *Curr Opin Biotechnol* 42:40–53. <https://doi.org/10.1016/j.copbio.2016.02.030>.
  6. Abejón R, Pérez-Acebo H, Clavijo L. 2018. Alternatives for chemical and biochemical lignin valorization: hot topics from a bibliometric analysis of the research published during the 2000–2016 period. *Processes* 6:98. <https://doi.org/10.3390/pr6080098>.
  7. Abdelaziz OY, Brink DP, Prothmann J, Ravi K, Sun M, García-Hidalgo J, Sandahl M, Hultberg CP, Turner C, Lidén G, Gorwa-Grauslund MF. 2016. Biological valorization of low molecular weight lignin. *Biotechnol Adv* 34:1318–1346. <https://doi.org/10.1016/j.biotechadv.2016.10.001>.
  8. Linger JG, Vardon DR, Guarnieri MT, Karp EM, Hunsinger GB, Franden MA, Johnson CW, Chupka G, Strathmann TJ, Pienkos PT, Beckham GT. 2014. Lignin valorization through integrated biological funneling and chemical catalysis. *Proc Natl Acad Sci U S A* 111:12013–12018. <https://doi.org/10.1073/pnas.1410657111>.
  9. Rodriguez A, Salvachúa D, Katahira R, Black BA, Cleveland NS, Reed M, Smith H, Baidoo EE, Keasling JD, Simmons BA. 2017. Base-catalyzed depolymerization of solid lignin-rich streams enables microbial conversion. *ACS Sustain Chem Eng* 5:8171–8180. <https://doi.org/10.1021/acssuschemeng.7b01818>.
  10. Brink DP, Ravi K, Lidén G, Gorwa-Grauslund MF. 2019. Mapping the diversity of microbial lignin catabolism: experiences from the eLignin database. *Appl Microbiol Biotechnol* 103:3979–4002. <https://doi.org/10.1007/s00253-019-09692-4>.
  11. Harwood CS, Parales RE. 1996. The beta-ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol* 50:553–590. <https://doi.org/10.1146/annurev.micro.50.1.553>.
  12. Tian JH, Pourcher AM, Bouchez T, Gelhaye E, Peu P. 2014. Occurrence of lignin degradation genotypes and phenotypes among prokaryotes. *Appl Microbiol Biotechnol* 98:9527–9544. <https://doi.org/10.1007/s00253-014-6142-4>.
  13. Shinoda E, Takahashi K, Abe N, Kamimura N, Sonoki T, Masai E. 2019. Isolation of a novel platform bacterium for lignin valorization and its application in glucose-free *cis,cis*-muconate production. *J Ind Microbiol Biotechnol* 46:1071–1080. <https://doi.org/10.1007/s10295-019-02190-6>.
  14. 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>.
  15. Pérez-Pantoja D, De la Iglesia R, Pieper DH, González B. 2008. Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium *Cupriavidus necator* JMP134. *FEMS Microbiol Rev* 32:736–794. <https://doi.org/10.1111/j.1574-6976.2008.00122.x>.
  16. Jimenez JI, Minambres B, García JL, Diaz E. 2002. Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ Microbiol* 4:824–841. <https://doi.org/10.1046/j.1462-2920.2002.00370.x>.
  17. Ravi K, García-Hidalgo J, Brink DP, Skyvell M, Gorwa-Grauslund MF, Lidén G. 2019. Physiological characterization and sequence analysis of a syringate-consuming Actinobacterium. *Bioresour Technol* 285:121327. <https://doi.org/10.1016/j.biortech.2019.121327>.
  18. McLeod MP, Warren RL, Hsiao WWL, Araki N, Myhre M, Fernandes C, Miyazawa D, Wong W, Lillquist AL, Wang D, Dosanjh M, Hara H, Petrescu A, Morin RD, Yang G, Stott JM, Schein JE, Shin H, Smailus D, Siddiqui AS, Marra MA, Jones SJM, Holt R, Brinkman FSL, Miyauchi K, Fukuda M, Davies JE, Mohn WW, Eltis LD. 2006. The complete genome of *Rhodococcus* sp RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci U S A* 103:15582–15587. <https://doi.org/10.1073/pnas.0607048103>.
  19. Ravi K, García-Hidalgo J, Nöbel M, Gorwa-Grauslund MF, Lidén G. 2018. Biological conversion of aromatic monolignol compounds by a *Pseudomonas* isolate from sediments of the Baltic Sea. *AMB Express* 8:32. <https://doi.org/10.1186/s13568-018-0563-x>.
  20. Ravi K, García-Hidalgo J, Gorwa-Grauslund MF, Lidén G. 2017. Conversion of lignin model compounds by *Pseudomonas putida* KT2440 and isolates from compost. *Appl Microbiol Biotechnol* 101:5059–5070. <https://doi.org/10.1007/s00253-017-8211-y>.
  21. Priefert H, Rabenhorst J, Steinbüchel A. 2001. Biotechnological production of vanillin. *Appl Microbiol Biotechnol* 56:296–314. <https://doi.org/10.1007/s002530100687>.
  22. Asm H, Onozaki H, Imaseki H. 1988. Vanillylamine metabolism in *Pseudomonas fluorescens*. *Agric Biol Chem* 52:2741–2746. <https://doi.org/10.1080/00021369.1988.10869155>.
  23. Stentelaire C, Lesage-Meesen L, Delattre M, Haon M, Sigoillot J, Ceccaldi BC, Asther M. 1997. By-passing of unwanted vanillyl alcohol formation using selective adsorbents to improve vanillin production with *Phanerochaete chrysosporium*. *World J Microbiol Biotechnol* 14:285–287. <https://doi.org/10.1023/A:1008811019693>.
  24. Hansen EH, Møller BL, Kock GR, Büchner CM, Kristensen C, Jensen OR, Okkels FT, Olsen CE, Motawia MS, Hansen J. 2009. *De novo* biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). *Appl Environ Microbiol* 75:2765–2774. <https://doi.org/10.1128/AEM.02681-08>.
  25. Fache N, Boutevin B, Caillol S. 2015. Vanillin production from lignin and its use as a renewable chemical. *ACS Sustain Chem Eng* 4:35–46. <https://doi.org/10.1021/acssuschemeng.5b01344>.
  26. Graf N, Altenbuchner J. 2014. Genetic engineering of *Pseudomonas putida* KT2440 for rapid and high-yield production of vanillin from ferulic acid. *Appl Microbiol Biotechnol* 98:137–149. <https://doi.org/10.1007/s00253-013-5303-1>.
  27. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
  28. Bao E, Jiang T, Girke T. 2014. AlignGraph: algorithm for secondary *de novo* genome assembly guided by closely related references. *Bioinformatics* 30:i319–i328. <https://doi.org/10.1093/bioinformatics/btu291>.
  29. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with Gap-Filler. *Genome Biol* 13:R56. <https://doi.org/10.1186/gb-2012-13-6-r56>.
  30. Paulino D, Warren RL, Vandervalk BP, Raymond A, Jackman SD, Biro I. 2015. Sealer: a scalable gap-closing application for finishing draft genomes. *BMC Bioinformatics* 16:230. <https://doi.org/10.1186/s12859-015-0663-4>.
  31. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <https://doi.org/10.1093/bioinformatics/btq683>.
  32. Spiers AJ, Buckling A, Rainey PB. 2000. The causes of *Pseudomonas* diversity. *Microbiology* 146:2345–2350. <https://doi.org/10.1099/00221287-146-10-2345>.
  33. Mulet M, Lalucat J, García-Valdés E. 2010. DNA sequence-based analysis of the *Pseudomonas* species. *Environ Microbiol* 12:1513–1530. <https://doi.org/10.1111/j.1462-2920.2010.02181.x>.
  34. Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E. 2014. Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Mol Biol Evol* 31:1077–1088. <https://doi.org/10.1093/molbev/msu088>.
  35. Hillis DM, Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* 42:182–192. <https://doi.org/10.2307/2992540>.
  36. Jiménez JI, Pérez-Pantoja D, Chavarría M, Diaz E, de Lorenzo V. 2014. A second chromosomal copy of the *catA* gene endows *Pseudomonas putida* mt-2 with an enzymatic safety valve for excess of catechol. *Environ Microbiol* 16:1767–1778. <https://doi.org/10.1111/1462-2920.12361>.
  37. Van Duuren J, Wijte D, LePrince A, Karge B, Puchałka J, Wery J, Dos Santos VM, Eggink G, Mars A. 2011. Generation of a *catR* deficient mutant of *P. putida* KT2440 that produces *cis, cis*-muconate from benzoate at high rate and yield. *J Biotechnol* 156:163–172. <https://doi.org/10.1016/j.jbiotec.2011.08.030>.
  38. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER Web server: 2018 update. *Nucleic Acids Res* 46:W200–W204. <https://doi.org/10.1093/nar/gky448>.
  39. Belda E, van Heck RGA, José Lopez-Sanchez M, Cruveiller S, Barbe V, Fraser C, Klenk H-P, Petersen J, Morgat A, Nikel PI, Vallenet D, Rouy Z, Sekowska A, Martins Dos Santos VAP, de Lorenzo V, Danchin A, Médigue C. 2016. The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic chassis. *Environ Microbiol* 18:3403–3424. <https://doi.org/10.1111/1462-2920.13230>.
  40. Pick A, Rühmann B, Schmid J, Sieber V. 2013. Novel CAD-like enzymes from *Escherichia coli* K-12 as additional tools in chemical production. *Appl Microbiol Biotechnol* 97:5815–5824. <https://doi.org/10.1007/s00253-012-4474-5>.
  41. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi S-H, Couloux A, Lee



- S-W, Yoon SH, Cattolico L, Hur C-G, Park H-S, Ségurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P, Kim JF. 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J Mol Biol* 394:644–652. <https://doi.org/10.1016/j.jmb.2009.09.052>.
42. Rost B. 1999. Twilight zone of protein sequence alignments. *Protein Eng* 12:85–94. <https://doi.org/10.1093/protein/12.2.85>.
  43. Pearson WR. 2013. An introduction to sequence similarity (“homology”) searching. *Curr Protoc Bioinformatics* 42:3.1.1–3.1.8. <https://doi.org/10.1002/0471250953.bi0301s42>.
  44. Yandell M, Ence D. 2012. A beginner’s guide to eukaryotic genome annotation. *Nat Rev Genet* 13:329–342. <https://doi.org/10.1038/nrg3174>.
  45. Overhage J, Priefert H, Steinbüchel A. 1999. Biochemical and genetic analyses of ferulic acid catabolism in *Pseudomonas* sp. strain HR199. *Appl Environ Microbiol* 65:4837–4847. <https://doi.org/10.1128/AEM.65.11.4837-4847.1999>.
  46. 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, Hance I, Chris Lee P, Holtzapple E, Scanlan D, Tran K, Moazzez A, Utterback T, Rizzo M, Lee K, Kosack D, Moestl D, Wedler H, Lauber J, Stjepandic D, Hoheisel J, Straetz M, Heim S, Kiewitz C, Eisen JA, Timmis KN, Dusterhöft A, Tümmler B, Fraser CM. 2002. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* 4:799–808. <https://doi.org/10.1046/j.1462-2920.2002.00366.x>.
  47. Wierenga RK, De Maeyer MC, Hol WG. 1985. Interaction of pyrophosphate moieties with alpha-helices in dinucleotide-binding proteins. *Biochemistry* 24:1346–1357. <https://doi.org/10.1021/bi00327a012>.
  48. Xu Y, Bhargava G, Wu H, Loeber G, Tong L. 1999. Crystal structure of human mitochondrial NAD(P)<sup>+</sup>-dependent malic enzyme: a new class of oxidative decarboxylases. *Structure* 7:877–889. [https://doi.org/10.1016/S0969-2126\(99\)80115-4](https://doi.org/10.1016/S0969-2126(99)80115-4).
  49. Hua YH, Wu CY, Sargsyan K, Lim C. 2014. Sequence-motif detection of NAD(P)-binding proteins: discovery of a unique antibacterial drug target. *Sci Rep* 4:6471. <https://doi.org/10.1038/srep06471>.
  50. Sellés Vidal L, Kelly CL, Mordaka PM, Heap JT. 2018. Review of NAD(P)H-dependent oxidoreductases: properties, engineering and application. *Biochim Biophys Acta Proteins Proteom* 1866:327–347. <https://doi.org/10.1016/j.bbapap.2017.11.005>.
  51. Sybesma W, Straathof A, Jongejan J, Pronk J, Heijnen J. 1998. Reductions of 3-oxo esters by baker’s yeast: current status. *Biocatal Biotransformation* 16:95–134. <https://doi.org/10.3109/10242429809003616>.
  52. Katz M, Hahn-Hägerdal B, Gorwa-Grauslund MF. 2003. Screening of two complementary collections of *Saccharomyces cerevisiae* to identify enzymes involved in stereo-selective reductions of specific carbonyl compounds: an alternative to protein purification. *Enzyme Microb Technol* 33:163–172. [https://doi.org/10.1016/S0141-0229\(03\)00086-3](https://doi.org/10.1016/S0141-0229(03)00086-3).
  53. Park H, Kim T, Bhoo S, Lee T, Lee S-W, Cho M-H. 2018. Biochemical characterization of the rice cinnamyl alcohol dehydrogenase gene family. *Molecules* 23:2659. <https://doi.org/10.3390/molecules23102659>.
  54. Ma Q-H. 2010. Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. *J Exp Bot* 61:2735–2744. <https://doi.org/10.1093/jxb/erq107>.
  55. Sambrook J, Russell DW. 2001. *Molecular cloning: a laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
  56. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
  57. Joshi N, Fass J. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33). <https://github.com/najoshi/sickle>.
  58. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
  59. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
  60. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
  61. Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res* 12:656–664. <https://doi.org/10.1101/gr.229202>.
  62. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>.
  63. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
  64. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS. 2016. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res* 44:D646–D653. <https://doi.org/10.1093/nar/gkv1227>.
  65. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
  66. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>.
  67. Huson DH, Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol* 61:1061–1067. <https://doi.org/10.1093/sysbio/sys062>.
  68. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2013. GenBank. *Nucleic Acids Res* 41:D36–D42. <https://doi.org/10.1093/nar/gks1195>.
  69. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED. 2012. *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* 40:D700–D705. <https://doi.org/10.1093/nar/gkr1029>.
  70. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427–D432. <https://doi.org/10.1093/nar/gky995>.
  71. Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189:113–130. [https://doi.org/10.1016/0022-2836\(86\)90385-2](https://doi.org/10.1016/0022-2836(86)90385-2).
  72. Pfennig N, Lippert KD. 1966. Über das Vitamin B<sub>12</sub>-Bedürfnis Phototropher Schwefelbakterien. *Archiv Mikrobiol* 55:245–256. <https://doi.org/10.1007/BF00410246>.
  73. Larroy C, Fernández MR, González E, Parés X, Biosca JA. 2002. Characterization of the *Saccharomyces cerevisiae* YMR318C (ADH6) gene product as a broad specificity NADPH-dependent alcohol dehydrogenase: relevance in aldehyde reduction. *Biochem J* 361:163–172. <https://doi.org/10.1042/bj3610163>.
  74. Martínez-García E, de Lorenzo V. 2012. Transposon-based and plasmid-based genetic tools for editing genomes of Gram-negative bacteria, p 267–283. In Weber W, Fussenegger M (ed), *Synthetic gene networks: methods and protocols*. Springer, New York, NY.
  75. Martínez-García E, de Lorenzo V. 2011. Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440. *Environ Microbiol* 13:2702–2716. <https://doi.org/10.1111/j.1462-2920.2011.02538.x>.
  76. Aparicio T, de Lorenzo V, Martínez-García E. 2015. Broadening the SEVA plasmid repertoire to facilitate genomic editing of Gram-negative bacteria, p 9–27. In McGenity TJ, Timmis KN, Nogales B (ed), *Hydrocarbon and lipid microbiology protocols: genetic, genomic and system analyses of pure cultures*. Springer Berlin Heidelberg, Heidelberg, Germany.
  77. García-Hidalgo J, Ravi K, Kuré L-L, Lidén G, Gorwa-Grauslund M. 2019. Identification of the two-component guaiaacol demethylase system from *Rhodococcus rhodochrous* and expression in *Pseudomonas putida* EM42 for guaiaacol assimilation. *AMB Express* 9:34. <https://doi.org/10.1186/s13568-019-0759-8>.
  78. Dunn AK, Martin MO, Stabb EV. 2005. Characterization of pES213, a small mobilizable plasmid from *Vibrio fischeri*. *Plasmid* 54:114–134. <https://doi.org/10.1016/j.plasmid.2005.01.003>.
  79. Boyer HW, Roulland-Dussoix D. 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J Mol Biol* 41:459–472. [https://doi.org/10.1016/0022-2836\(69\)90288-5](https://doi.org/10.1016/0022-2836(69)90288-5).
  80. Savitsky P, Bray J, Cooper CD, Marsden BD, Mahajan P, Burgess-Brown NA, Gileadi O. 2010. High-throughput production of human proteins for crystallization: the SGC experience. *J Struct Biol* 172:3–13. <https://doi.org/10.1016/j.jsb.2010.06.008>.
  81. Finan TM, Kunkel B, De Vos GF, Signer ER. 1986. Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. *J Bacteriol* 167:66–72. <https://doi.org/10.1128/jb.167.1.66-72.1986>.