PNAS
www.pnas.org

Main Manuscript for

Characterization of alkylguaiacol-degrading cytochromes P450 for the biocatalytic valorization of lignin

Morgan M. Fetherolf[a], David J. Levy-Booth[a], Jie Liu[a], Andrew Wilson[a], Rui Katahira[b], Gregg T. Beckham[b], William W. Mohn[a], and Lindsay D. Eltis[a,*]

[a]Department of Microbiology & Immunology, Life Sciences Institute, The University of British Columbia, Vancouver, British Columbia, Canada

[b]National Bioenergy Center, National Renewable Energy Laboratory, Golden CO, United States

*Corresponding author: Lindsay D. Eltis

E-mail: leltis@mail.ubc.ca

ORCIDs: M.M.F., 0000-0003-0263-7315; D.J.L.-B. 0000-0002-2382-709X; G.T.B., 0000-0002-3480-212X: W.W.M., 0000-0002-6509-746X; L.D.E., 0000-0002-6774-8158

Classification: Biological sciences (major) and Microbiology (minor)

Keywords: cytochrome P450, lignin valorization, *O*-demethylase, guaiacol, biocatalysis

Author Contributions: M.M.F. performed growth and enzymological experiments M.M.F. wrote the manuscript. D.J.L.-B. performed transcriptomic analysis and edited the manuscript. J.L. and A.W. produced the RHA1 mutants. R.K. synthesized compounds. G.T.B. edited the manuscript. W.W.M. designed the study and edited the manuscript. L.D.E. designed the study and wrote the manuscript. M.M.F. and D.J.L.-B. contributed equally to this work.

The Main Manuscript includes:

Main Text
Table 1
Figures 1 to 5

ABSTRACT

Cytochrome P450 enzymes have tremendous potential as industrial biocatalysts, including in biological lignin valorization. Here we report two P450s that catalyze the *O*-demethylation of lignin-derived guaiacols with different ring substitution patterns. Bacterial strains *Rhodococcus rhodochrous* EP4 and *Rhodococcus jostii* RHA1 both utilized alkylguaiacols as sole growth substrates. Transcriptomics of EP4 grown on 4-propylguaiacol (4PG) revealed the up-regulation of *agcA*, encoding a CYP255A1 family P450, and the *aph* genes, previously shown to encode a *meta*-cleavage pathway responsible for 4-alkylphenol catabolism. The function of the homologous pathway in RHA1 was confirmed: deletion mutants of *agcA* and *aphC*, encoding the *meta*-cleavage alkylcatechol dioxygenase, grew on guaiacol but not 4PG. By contrast, deletion mutants of *gcoA* and *pcaL*, encoding a CYP255A2 family P450 and an *ortho*-cleavage enzyme, respectively, grew on 4-propylguaiacol but not guaiacol. CYP255A1 catalyzed the *O*-demethylation of 4-alkylguaiacols to 4-alkylcatechols with the following apparent specificities (k_{cat}/K_M): ethyl > propyl >> methyl. AgcA bound guaiacol with relatively high affinity (K_d = 22 ± 1 μM) but did not detectably transform it. The biocatalytic potential of AgcA was demonstrated by the ability of EP4 to grow on lignin-derived products obtained from the reductive catalytic fractionation of corn stover, depleting alkylguaiacols and alkylphenols. By identifying related P450s with complementary specificities for lignin-relevant guaiacols, this study facilitates the design of these enzymes for biocatalytic applications. We further demonstrated that the metabolic fate of the guaiacol depends on its substitution pattern, a finding that has significant implications for engineering biocatalysts to valorize lignin.

SIGNIFICANCE STATEMENT

Upgrading lignin, an underutilized component of biomass, is essential for the sustainability of biorefineries. Biocatalysis has considerable potential for upgrading lignin, but our lack of knowledge of relevant enzymes and pathways has limited its application. Herein, we describe a microbial pathway responsible for catabolizing alkylguaiacols, a major component of several industrial lignin streams. Catabolism is initiated by a cytochrome P450, with related P450s catalyzing the *O*-demethylation of different lignin-derived guaiacols and subsequent catabolism depending on the substitution pattern of the guaiacol. Importantly, the alkylguaiacol catabolic pathway enables bacterial growth on corn stover lignin produced by reductive catalytic fractionation. Overall, these insights greatly facilitate the engineering of P450s and bacteria to biocatalytically upgrade lignin.

INTRODUCTION

Lignin is a complex aromatic polymer found in plant cell walls (1). Although lignin is the most abundant aromatic polymer on earth, comprising up to 40% of lignocellulosic biomass, it is underutilized in biorefineries, typically being slated for combustion to power the extraction of carbohydrates (2-5). Technoeconomic analysis and life cycle assessment have highlighted the importance of lignin valorization to the economic viability and sustainability of biorefineries (6, 7). To this end, considerable effort has been invested in developing lignin depolymerization technologies (2-5). Such technologies produce a mixture of aromatic compounds, depending on the chemistry and on the proportions of *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) subunits that comprise the lignin (2). A major challenge of lignin valorization is converting these complex mixtures of compounds to single chemical species in high atom yield.

Bacteria offer an attractive approach for converting lignin depolymerization products to value-added bioproducts via biological funneling (8-11). This process exploits the capacity of some bacteria to catabolize a variety of aromatic compounds as well as the convergent nature of the pathways responsible for this catabolism. A large number of "upper pathways" transform a variety of aromatic compounds to a few key intermediates, which are further transformed to central metabolism by a limited number of "lower" pathways (8, 12). In lignin-relevant catabolism, the key intermediates are typically catechol, protocatechuate and gallate, all of which bear hydroxyl groups on adjacent ring carbon atoms (8, 9). The aromatic ring is then cleaved either *ortho* (between) or *meta* (adjacent) to those hydroxyl substituents (13). The development of bacteria as lignin-valorizing biocatalysts requires the elucidation and optimization of pathways able to catabolize the major components of industrial lignin streams.

An essential reaction in the catabolism of aromatic compounds derived from G and S-type lignin is the *O*-demethylation of the aromatic methoxy groups to produce a catecholic intermediate. In the cases where bacterial strains have been engineered and optimized to produce bioproducts from lignin-derived compounds from G and S-type lignin, aromatic *O*-demethylation is a rate-limiting step (14-16). Accordingly, there has been significant interest in mechanistic understanding of the three enzyme types known to catalyze aromatic *O*-demethylation: tetrahydrofolate-dependent *O*-demethylases (17, 18); Rieske-type non-heme iron oxygenases (19-21), and cytochromes P450 (22-24). A critical unknown for all three enzyme paradigms is their structure-activity relationship on the various G and S-type lignin-derived compounds, which exhibit substantially different ring substitution patterns depending on the upstream chemistry.

One important group of lignin depolymerization products whose catabolism has not been elucidated are the 4-alkylguaiacols. Alkylguaiacols are produced in the reductive catalytic fractionation (RCF) of all lignins as well as in various depolymerization processes utilizing Kraft, organosolv, or black liquor as the feedstock (25-31). RCF of corn stover yields up to 38% monoaromatic compounds, with the predominant ones being 4-ethylphenol, 4-propylguaiacol (4PG), 4-propanolguaiacol, 4-propylsyringol, methyl ferulate and methyl *p*-coumarate (27). This mixture lends itself to bacterial biological funneling as rhodococci naturally catabolize several of these compounds. For example, *Rhodococcus jostii* RHA1 (RHA1 hereafter) catabolizes ferulate

and *p*-coumarate via the Cou and β-ketoadipate pathways (32). Both *Rhodococcus rhodochrous* EP4 (EP4 hereafter) and RHA1 catabolize 4-ethylphenol via the Aph pathway, in which alkylphenols are hydroxylated by AphAB to form alkylcatechols (33). The latter are subject to *meta*-cleavage by AphC, an alkylcatechol 2,3-dioxygenase, and eventually funneled into central metabolism via pyruvate and either acetyl- or propionyl-CoA. Elucidation of alkylguaiacol catabolism would afford a more complete catabolism of lignin depolymerization products.

A number of rhodococci and related strains grow efficiently on guaiacol (34, 35). In these strains, catabolism is initiated by a cytochrome P450 that belongs to the Cyp255A2 family which catalyzes the *O*-demethylation of guaiacol to catechol (34). The best characterized of these P450s is $GcoA_{Amyc}$ from *Amycolatopsis* sp. ATCC 39116 which, together with its cognate reductase GcoB, transforms a range of aromatic ethers including guaiacol and anisole (22). In *R. rhodochrous* 116, the resulting catechol is subsequently catabolized via *ortho*-cleavage and presumably the β-ketoadipate pathway (34, 36). The potential of these P450s for valorizing lignin is highlighted by the heterologous expression of *gcoAB* from *R. rhodochrous* J3 in *Pseudomonas putida* EM42 to enable growth on guaiacol and the engineering of $GcoA_{Amyc}$ to enable the *O*-demethylation of syringol (23, 24).

Herein, we describe the catabolism of 4-alkylguaiacols by two previously characterized strains of *Rhodococcus*: EP4 (33) and RHA1 (37). A combination of transcriptomics, targeted gene deletion and enzymology was used to identify the catabolic pathway. Our studies established that this catabolism is initiated by AgcAB, a CYP255A1 family P450 that catalyzes the *O*-demethylation of 4-alkylguaiacols. This distinguishes AgcA from GcoA, which both strains also contain. The ability of AgcA to bind and transform a variety of guaiacols was characterized. Finally, we investigated the ability of the rhodococcal strains to grow on corn stover RCF oil, which contains a mixture of compounds, including 4PG. The results are discussed in terms of related catabolic enzymes and pathways, as well as the design and engineering of biocatalysts to valorize lignin.

RESULTS

Transcriptomic analysis revealed alkylguaiacol catabolic genes in EP4. We found that EP4, previously characterized for its ability to catabolize alkylphenols (33), could also grow on a variety of alkylguaiacols. When grown on 1 mM substrate, the growth yield reflected the length of the alkyl side chain, with yields on 4PG > 4-ethylguaiacol (4EG) > 4-methylguaiacol (4MG) > guaiacol (Fig. 1A). However, the lag phase of the culture was shortest on guaiacol and longest on 4MG. Growth on 4PG was also confirmed using CFU (Fig. S1A).

To identify genes potentially involved in the catabolism of 4-alkylguaiacols, the transcriptomes of cells growing exponentially on 4PG relative to a succinate control were compared using RNAseq. The most highly upregulated genes in 4PG-grown cells included the recently identified *aph* genes, encoding a *meta*-cleavage pathway for alkylphenols, together with three additional genes in a putative operon located immediately downstream of *aphC* (Fig. 1B). These three genes were predicted to encode a P450, its cognate reductase, and a small protein of unknown function. Based on their involvement in the *O*-demethylation of 4-alkylguaiacols, established below, these genes were annotated as *agcA* and *agcB* (Table S1). None of the *cat* and *pca* genes,

encoding the *ortho*-cleavage of catechol and the β-ketoadipate pathway, respectively, were up-regulated during growth on 4PG.

AgcA belongs to the CYP255A1 family. Bioinformatic analysis revealed that AgcA belongs to the CYP255A1 family, and shares 65% amino acid sequence identity with $GcoA_{Amyc}$, a CYP255A2 that catalyzes the *O*-demethylation of guaiacol (Fig. 2A (22)). However, the reciprocal best hit of $GcoA_{Amyc}$ in EP4 is encoded by C6369_RS07265, with whose gene product it shares 76% sequence identity. C6369_RS07265 and the adjacent C6369_RS07270, predicted to encode a P450 reductase, lie in close proximity to *catA*, encoding the catechol 1,2-dioxygenase associated with the β-ketoadipate pathway. These genes, annotated as *gcoAB*, were not upregulated in EP4 during growth on 4PG.

Our analysis further revealed that RHA1 also harbors both *agcABC* and *gcoAB*. As in EP4, these genes lie in close proximity to the *aph* and *cat* genes, respectively. In a structure-based phylogenetic analysis, $AgcA_{RHA1}$ and $AgcA_{EP4}$ belong to the same clade (Fig. 2B) despite sharing only 58% amino acid sequence identity. The above gene organization and the phylogeny of AgcA and GcoA suggest that these enzymes have different substrate specificities.

Genetic validation of the AgcA and GcoA in RHA1. We used RHA1 for molecular genetic analysis of alkylguaiacol catabolism. This is because RHA1 is much more genetically tractable than EP4(33), and the relevant genes (*agcABC, aph, gcoAB, cat* and *pca*) have high sequence identity and synteny between the two strains. Consistent with the bioinformatic analyses, RHA1 grew on guaiacol, 4MG, and 4PG as sole growth substrates (Fig. 3A, Fig. S2A, Fig. S1B). As observed in EP4, the growth yield on the different guaiacols reflected the length of the side chain. However, unlike EP4, the longest lag phase occurred during growth on 4PG: RHA1 only grew on 1 mM 4PG after a lag phase of 20 hours during which the CFUs decreased by ~50%.

To further evaluate the respective catabolic roles of the P450s, the expression of select genes was assessed in RHA1 growing exponentially on 4PG and guaiacol. RT-qPCR analyses revealed that *agcA* and *aphC* were up-regulated during growth on 4PG while neither *gcoA* nor *catA* were up-regulated under these conditions (Fig. S1D). By contrast, *gcoA* and *catA* were upregulated during growth on guaiacol while *agcA*, *agcB,* and *aphC* were not.

To validate the two guaiacol catabolic pathways of RHA1, we constructed a Δ*agcA* strain and tested the ability of this mutant and others in our collection to grow on different guaiacols. Neither the Δ*agcA* mutant nor our previously constructed Δ*aphC* mutant grew on 4PG (Fig. 3B,C). Both mutants grew on guaiacol with a similar kinetics and yields as wild-type RHA1 (Fig. 3). This is consistent with the bioinformatic predictions that 4PG but not guaiacol is catabolized by AgcAB and the Aph pathway. Importantly, the Δ*agcA* mutant but not Δ*aphC* was able to utilize 4-ethylphenol, establishing that the Aph pathway is intact in the Δ*agcA* mutant (Fig. S2). In addition, the Δ*gcoA* and Δ*pcaL* mutants grew normally on 4PG but neither grew on guaiacol (Fig. 3D, E). These data are consistent with the prediction that guaiacol is catabolized by GcoA, *ortho*-cleavage and the β-ketoadipate pathway. This conclusion was substantiated by RT-qPCR data, which demonstrated that *agcA* and *aphC* transcripts were much more abundant than *gcoA* and *catA*

transcripts in RHA1 growing on 4PG (Fig. S1D). Similarly, *gcoA* and *catA* transcripts were much more abundant than *agcA* and *aphC* transcripts in RHA1 growing on guaiacol (Fig. 3F).

The data for 4MG were more complex. Thus, of the tested mutants, only Δ*aphC* was unable to grow on this compound (Fig. 3). This suggests that regardless of how 4MG is *O*-demethylated, the resulting 4-methylcatechol is only degraded via the Aph *meta*-cleavage pathway. Further, the growth of the Δ*gcoA* and Δ*pcaL* mutants on 4MG was similar to that of wild-type RHA1, suggesting that AgcAB is the primary 4MG *O*-demethylase. This was substantiated by RT-qPCR analyses, which revealed that *agcA* transcripts were over an order of magnitude more abundant than *gcoA* transcripts in cells growing on 4MG (Fig. 3F). Somewhat unexpectedly, the Δ*agcA* mutant was able to grow on 4MG, albeit with a lag phase of ~20 hours (Fig. 3B). RT-qPCR analyses revealed that *gcoA* trancripts were ~13-fold more abundant in Δ*agcA* cells than in wild-type cells growing on 4MG (Fig. 3F). These results suggest that in the mutant, GcoA *O*-demethylates 4MG and partially compensates for the loss of AgcA.

AgcA binds a variety of aromatic ethers with high affinity. To characterize AgcA's substrate specificity and ligand-binding ability, both $AgcA_{EP4}$ and $AgcB_{EP4}$ were heterologously produced in *E. coli* and purified to apparent homogeneity. As isolated, AgcA did not contain a full complement of heme (Rz ~0.52, where Rz = A_{417}/A_{280}). Reconstitution with hemin yielded preparations with Rz values of ~ 1.05 (Supplemental Fig. 4A). AgcA had a Soret peak at 417 nm, as well as α and β bands at 567 and 536 nm, respectively, typical of the low-spin state of the ferric ion in ligand-free P450s (38). Reduction of the reconstituted AgcA with dithionite and addition of CO, caused the Soret peak to shift to 450 nm, indicating that the reconstituted heme was in the proper coordination environment with an axial cysteine ligand (Supplementary Fig. 4B). Preparations of AgcB absorbed maximally at 454 nm, which corresponds to a FAD cofactor. An additional peak at 425 nm together with shoulders around 475 nm and 550 are diagnostic of the predicted 2Fe-2S cluster (Fig. S5) (39). Preparations of AgcB reduced cytochrome *c* with a specific activity of 22 ± 4 mmol cytochrome *c* min^{-1} mg^{-1}.

The affinity of AgcA for a variety of aromatic ligands was examined by monitoring their abilities to induce a Type I absorption spectrum (40). In this assay, ligand-binding induces a shift in the Soret peak as the ferric ion transitions from a low spin to a high spin state (Fig. 4A). Among the tested compounds, AgcA bound 4EG with the highest affinity (K_d ~0.3 μM; Table 1). Increasing or decreasing the length of the alkyl group by a single carbon decreased the affinity by ~4-fold, and removing the alkyl group altogether (guaiacol) further lowered the affinity by an order of magnitude. AgcA also bound anisole and 4-methylanisole with similar affinity as guaiacol, but bound 3-methylanisole with ~3-fold higher affinity. Finally, AgcA bound syringol and 3-hydroxy 4-methoxybenzyl alcohol with an order of magnitude lower affinity than guaiacol. These results indicate that the hydroxyl, methoxy, and alkyl substituents of the 4-alkylguaiacol all contribute to the recognition of the ligand by AgcA.

AgcAB catalyzes the *O*-demethylation of 4-alkylguaiacols. To establish that the P450 catalyzes the *O*-demethylation of 4PG, AgcA and AgcB were incubated overnight with 100 μM 4PG and NADH at 30 °C (20 mM MOPS, 90 mM NaCl, pH 7.2). HPLC analyses of the reaction mixture

revealed that 4PG was converted to a compound that eluted with a retention time corresponding to that of an authentic standard of 4-propylcatechol (Fig. 4B). In the absence of AgcA, 4PG was not transformed. This demonstrates that AgcA catalyzes the *O*-demethylation of 4PG.

To establish the apparent specificity (k_{cat}/K_M) of AgcAB for various substrates, we established a coupled assay using $AphC_{RHA1}$, the alkylcatechol 2,3-dioxygenase from the Aph pathway of RHA1. This enzyme cleaves the alkylcatechol resulting from the AgcAB reaction to a yellow-colored MCP whose formation was continuously monitored at 400 nm. Using this assay, AgcAB had highest apparent specificity for 4EG, followed by the 4-propyl and 4-methyl derivatives (Table 1). AgcAB did not detectably transform guaiacol, a result confirmed using the HPLC-based assay, or any of the anisoles, despite AgcAB binding those compounds with relatively high affinity.

EP4 grows on corn stover RCF oil. To investigate the potential of EP4 and RHA1 to transform an industrially relevant lignin stream, these strains were tested for growth on corn stover RCF oil. EP4 grew on media containing 250 ppm RCF oil, as measured by CFU (Fig. 5A). Analysis of the culture supernatant by GC-MS revealed that of the six major components of the RCF oil, 4-ethylphenol, 4PG and 4-propanolguaiacol were efficiently removed (Fig. 5B). These compounds appear to be catabolized via AgcAB and the Aph pathway, as *agcA*, *aphA* and *aphC* were up-regulated during growth on corn stover RCF oil relative to a succinate control (Fig. 5C). The other three RCF components, 4-propanolsyringol, methyl ferulate and methyl coumarate, were transformed to compounds with retention times of 13.06, 14.22 and 12.93, respectively (Fig. S7). Indeed, previous efforts to grow EP4 on the latter three RCF components were unsuccessful, consistent with the strain's lack of the *cou* genes, responsible for *p*-coumarate and ferulate catabolism. Interestingly, *gcoA* and *catA* were significantly down-regulated during growth on RCF oil relative to the succinate control (Fig. 5C). Finally, RHA1 did not grow well on corn stover RCF, even though the strain carries the *agc*, *aph* and *cou* genes (Fig. 5A).

DISCUSSION

This study describes AgcA, a P450 that catalyzes the *O*-demethylation of 4-alkylguaiacols to initiate their catabolism in Actinobacteria via the Aph *meta*-cleavage pathway. AgcA belongs to the CYP255A1 family and was unable to transform guaiacol. By contrast, GcoA belongs to the CYP255A2 family and has a high specificity (k_{cat}/K_M) for guaiacol (22). The genetic analysis of *agcA* and *gcoA* in RHA1 is consistent with the different specificities of these enzymes and establishes the different physiological roles of these two P450s. Further, the genetic association and co-regulation of the *agcAB* and *gcoAB* genes with *meta*- and *ortho*-cleavage pathways, respectively, is consistent with the tendency of alkylated aromatic compounds to be degraded by *meta*-cleavage pathways (33).

Recent interest in GcoA stems from its ability to transform lignin depolymerization products such as guaiacol and syringol (22, 23). The identification of an enzyme able to catalyze the *O*-demethylation of alkylguaiacols significantly expands these efforts, given the prevalence of such alkylated compounds in a variety of industrially-relevant mixtures of lignin depolymerization products. More generally, the different substrate specificities of AgcA and GcoA, despite their high amino acid sequence identity, facilitates the engineering of these enzymes for lignin

transformation. Inspection of the $GcoA_{Amyc}$:guaiacol binary complex reveals that C-4 of the guaiacol is within 4 Å of Ile81 and Thr296, and within 5 Å of Val241, Ile292, Ala295, and Phe395 (22). These residues are conserved in $AgcA_{EP4}$ with the exception of Ile81 and Thr296, which are leucine and alanine, respectively. Thr296 is particularly interesting as its side chain is orientated towards the C-4 carbon of guaiacol in the $GcoA_{Amyc}$:guaiacol complex. The smaller side chain of alanine at this position is consistent with AgcA's higher affinity for 4EG relative to guaiacol (Table 1). Currently, we are investigating the structure of $AgcA_{EP4}$:alkylguaiacol complexes to elucidate the basis of the enzyme's specificity. Such studies could further inform the engineering of these enzymes to catalyze the *O*-demethylation a variety of other compounds, such as 4-propylsyringol, a major component of corn stover and hardwood RCF oils (25, 27).

The Aph pathway represents another case of convergent catabolism, as both 4-alkylguaiacols and 4-alkylphenols are funneled into this pathway (33). This is highlighted by the growth of EP4 on corn stover RCF oil with concomitant removal of 4-ethylphenol and 4PG. More specifically, the biochemical and molecular genetic data indicate that the catabolism converges at 4-alkylcatechol. The ability of the Aph pathway to catabolize a variety of 4-alkylcatechols implies that AphC and the subsequent enzymes have a relatively relaxed substrate specificity. This is in contrast to the β-ketoadipate pathway where compounds such as *p*-coumarate, ferulate, vanillin, and 4-hydroxybenzoate are converted to protocatechuate prior to ring opening (32, 41, 42).

The growth rates of EP4 on the various alkylguaiacols correlate with the substrate specificity of AgcA. Thus, EP4 grew fastest on 4EG (Fig. 1A), the best substrate of AgcA, and slowest on 4MG, the worst substrate of AgcA. Nevertheless, it is possible that this difference reflects other factors, such as the induction of the catabolic genes. In this respect, the RHA1 Δ*agcA* mutant was unable to grow on 4PG, but grew on 4MG more slowly than the wild-type (Fig. 3A,B), suggesting the ability of GcoA to transform 4MG. This is corroborated by the RT-qPCR data, which demonstrate that *gcoA* was up-regulated in the Δ*agcA* mutant growing on 4MG with respect to WT RHA1 (Fig. 3F). Nevertheless, the strong growth of the Δ*gcoA* strain on 4MG and RT-qPCR data further indicate that 4MG is exclusively catabolized by AgcA in RHA1. Irrespective of which P450 catalyzes the *O*-demethylation of 4MG, the inability of the Δ*aphC* strain to grow on 4MG indicates that 4-methylcatechol is exclusively catabolized by the Aph *meta*-cleavage pathway.

The mechanism by which 4-methylcatechol is routed to *meta*-cleavage in RHA1 is unclear. Generally, alkylaromatic compounds are catabolized by *meta*-cleavage pathways, with the *ortho*-cleavage of methylcatechols leading to dead-end products. For example, the *ortho*-cleavage of 4-methylcatechol by CatA in *Bacillus pumilus* and *Pseudomonas desmolyticum* yields 3-methyl-muconate with subsequent accumulation of 4-methylmuconolactone as a dead-end product (43, 44). However, in RHA1 cells growing on 4MG, *catA* transcripts were present at ~1% of the level as *aphC* transcripts (Fig. 3F). Thus, 4MG does not appear to repress the *cat* genes the way other Aph pathway substrates apparently do (Fig. 5C). The specificities of CatA and AphC could help determine the fate of 4-methylcatechol in RHA1, but these k_{cat}/K_M values are not known. Interestingly, *R. opacus* 1CP contains several CatAs, all of which cleave 4-methylcatechol relatively efficiently (45). Indeed, this strain harbors an *ortho*-cleavage pathway that is adapted to *p*-cresol. This pathway is characterized by a CatA that has highest specificity for 4-

methylcatechol and a cycloisomerase that transforms the resulting 3-methylmuconate to prevent the accumulation of 4-methylmuconolactone. Further studies, including metabolomics, are required to determine the true extent of *ortho*-cleavage of 4-methylcatechol in RHA1 as well as the mechanism underlying its routing to *meta*-cleavage.

The characterization of an alkylguaiacol catabolic pathway greatly facilitates engineering bacterial strains to transform a broader range of lignin depolymerization products. Alkylguaiacols represent a significant fraction of the monoaromatic compounds present in RCF oils and other streams of depolymerized lignin (27, 29-31). Going forward, exploitation of the *agc* and *aph* genes for biocatalyst development will require improved understanding of the regulation of these genes and the physiology of the strains containing them. For example, it is unclear why EP4 is better able than RHA1 to grow on corn stover RCF oil. Indeed, this result was somewhat unexpected given that RHA1 can catabolize *p*-coumarate and ferulate but EP4 cannot (33). This may be related to the poorer growth of RHA1 relative to EP4 on 4PG. Of note, 4PG had a significant bactericidal effect on a fresh inoculum of RHA1. Moreover, there may be additional toxic compounds in RCF oil to which RHA1 is more sensitive than EP4. Finally, the *cou* genes, which encode *p*-coumarate and ferulate metabolism through *ortho*-cleavage (32), may be subject to catabolite repression, as discussed above. Clearly, understanding the basis of RHA1's relatively poor growth on RCF oil is important to engineering strains to degrade mixtures of lignin depolymerization products.

MATERIALS AND METHODS

Chemicals. Chemicals and reagents were of analytical grade and used without further purification unless otherwise noted. 4-Propanol guaiacol and 4-propylcatechol were synthesized as described in SI Appendix. Corn stover RCF oil was produced using established methods (27). Water for buffers was purified to a resistance of at least 18 megaohms.

Bacterial strains and growth conditions. Strains used in this study are listed in SI Appendix. Strains were grown on LB agar, LB broth or M9 minimal media supplemented with the appropriate growth substrate as described in SI Appendix. For growth on corn stover RCF oil (27), a stock of solution of 50,000 ppm oil was made in DMSO and added to M9-Goodies (250 ppm final). The medium was incubated at 30 °C for 30 minutes to maximize solubilization of the oil, and inoculated with cells to a final OD_{600} of ~0.05.

Growth substrate depletion. Culture supernatants were analyzed using a gas chromatograph coupled to a mass-selective detector (GC-MS) as described in SI Appendix.

Transcriptomic and RT-qPCR analyses. Cells were grown and RNA was extracted as previously described (33). RNA-seq and RT-qPCR were performed as described in SI Appendix.

Gene cloning and deletion. DNA was purified, manipulated, and propagated using standard procedures (46). Specific genes were cloned and deleted as described in SI Appendix.

Protein production and purification. AgcA, AgcB and AphC were produced heterologously as polyHis-tagged (Ht-) proteins in *E. coli* and purified using chromatography as described in SI Appendix. Upon purification, proteins were flash-frozen as beads in liquid N_2 and stored at -80 °C until use.

Characterization of P450 ligand binding. CO binding and K_d values for various aromatic ligands were evaluated spectrophotometrically as described in SI Appendix.

Characterization of P450 reaction and kinetics. The products of the AgcAB-catalyzed reaction were evaluated using high-pressure liquid chromatography (HPLC) as described in SI Appendix. Steady-state kinetic parameters for AgcAB were evaluated spectrophotometrically by coupling the *O*-demethylation of the alkyguaiacol to the *meta*-cleavage of the resulting alkylcatechol using AphC as described in SI Appendix.

Conflict of Interest. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

FUNDING. This study was supported by a research contract from Genome BC (SIP004) and a grant from the Natural Sciences and Engineering Research Council of Canada (STPGP 506595-17). LDE is the recipient of a Canada Research Chair. This work was authored in part by Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. RK and GTB thank the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies Office for funding. GTB also acknowledges support from The Center for Bioenergy Innovation, a U.S. Department of Energy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science

ACKNOWLEDGEMENTS. We thank Nicholas Thornburg and Camille Amador for generating the corn stover RCF oil and Pavneet Kalsi for her contribution to lab work.

TABLES AND FIGURES

Table 1. Affinity and apparent specificity of $AgcA_{EP4}$ for aromatic compounds[a]

Substrate	K_d (μM)	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (mM^{-1} s^{-1})
Guaiacol	22 (1)	N.D.[b]	N.D.[b]	N.D.
4-Methylguaiacol	1.2 (0.1)	0.011 (0.0004)	44 (5)	0.25 (0.03)
4-Ethylguaiacol	0.28 (0.02)	0.017 (0.001)	3.9 (0.5)	4.5 (0.6)
4-Propylguaiacol	1.2 (0.1)	0.01 (0.0005)	4.4 (0.8)	2.2 (0.4)
4-Propenylguaiacol	3.0 (0.3)	N.D.	N.D.	N.D.
Anisole	23 (3)	N.D.	N.D.	N.D.
4-Methylanisole	29 (0.9)	N.D.	N.D.	N.D.
3-Methylanisole	7.2 (0.5)	N.D.	N.D.	N.D.
3-Hydroxy 4-methoxybenzyl alcohol	350 (36)	N.D.	N.D.	N.D.
4-Ethylphenol	10.0 (0.4)	N.D.	N.D.	N.D.
Syringol	200 (20)	N.D.	N.D.	N.D.

[a]Affinities and apparent steady-state kinetic parameters were measured at 25 °C in 20 mM MOPS, 90 mM NaCl pH 7.2.

[b]Not detected

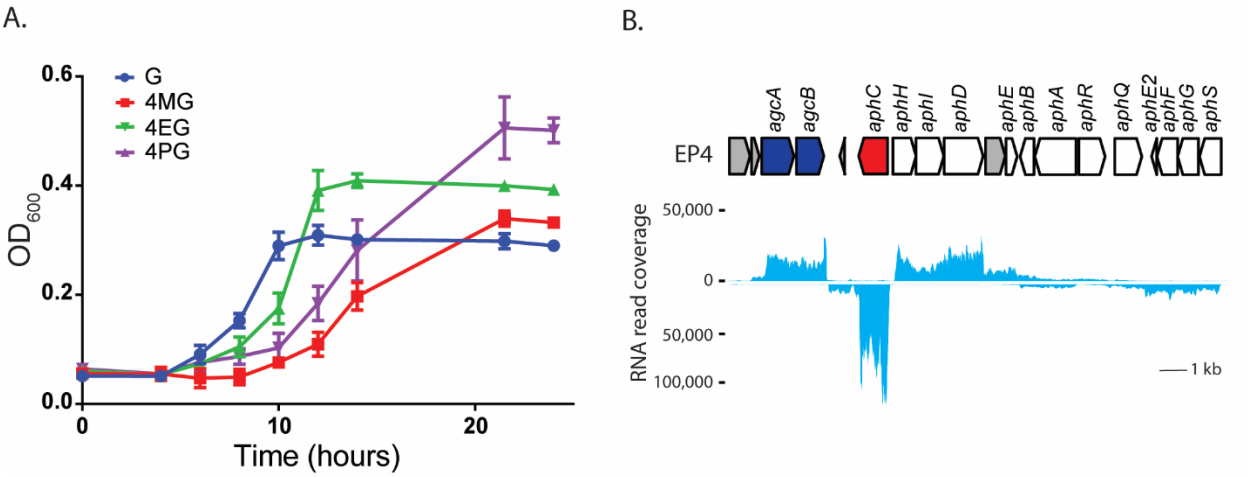


Figure 1. Growth of *R. rhodochrous* EP4 on alkylguaiacols and guaiacol (A). The gene cluster encoding the Aph *meta*-cleavage pathway of EP4 with transcript reads from cells growing on 4PG are plotted below the cluster (B).

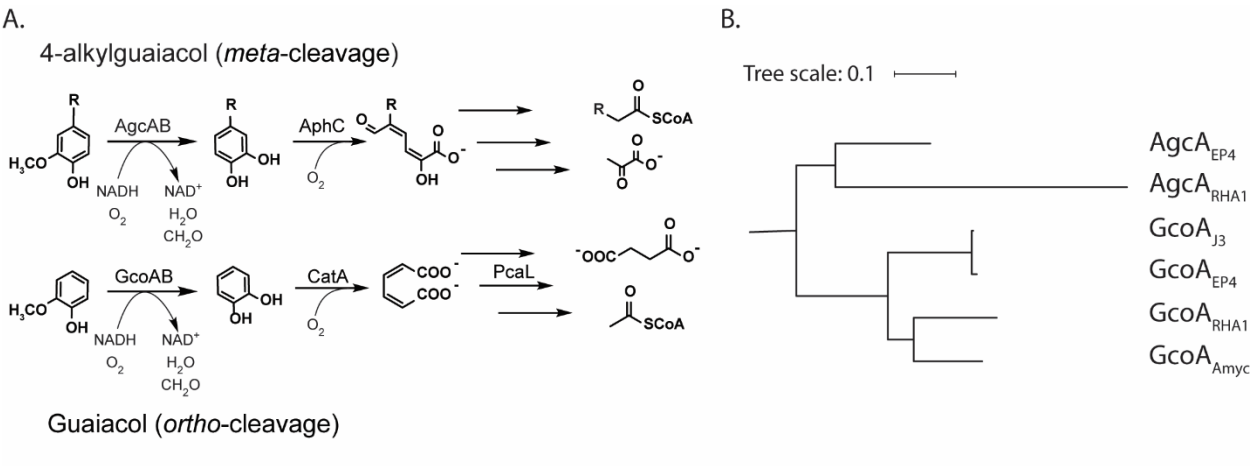


Figure 2. Relationship between the rhodococcal CYP255As. (A) Reactions catalyzed by AgcAB and GcoAB, respectively, and the subsequent catabolism of their respective reaction products via *meta* and *ortho*-cleavage, respectively. R = CH_3, CH_2CH_3, $CH_2CH_2CH_3$. (B) Phylogenetic tree of CYP255As. The tree includes characterized enzymes $GcoA_{Amyc}$ from *Amycolatopsis* ATCC 39116 and $GcoA_{J3}$ from *R*. *rhodochrous* J3 as well as the homologs present in EP4 and RHA1.

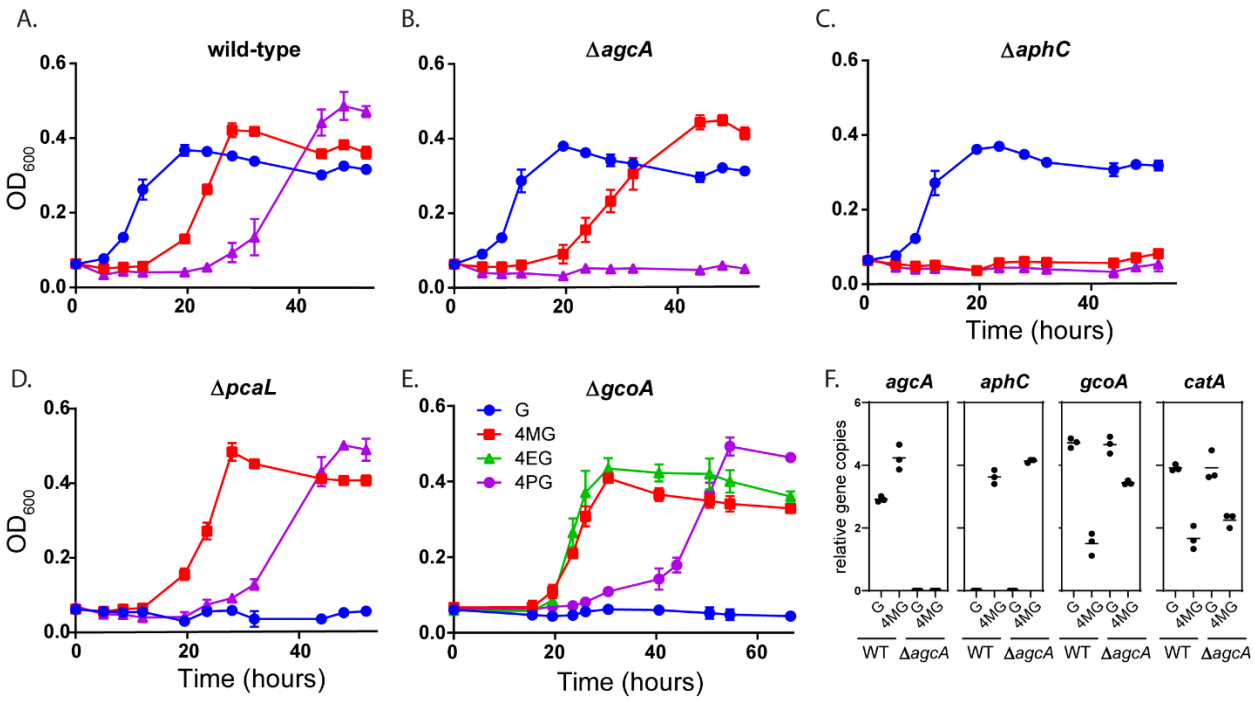


Figure 3. Growth of wild-type and mutant *R. jostii* RHA1 strains on various guaiacols and expression of selected genes. Growth of wild type (A), Δ*agcA* (B), Δ*aphC* (C), Δ*pcaL* (D), and Δ*gcoA* (E) mutants on guaiacol (G), 4-methylguaiacol (4MG), 4-propylguaiacol (4PG) and, for Δ*gcoA* , 4-ethylguaiacol (4EG). Growth curves color-coded according to growth substrate as indicated in panel E. (F) RT-qPCR of *agcA*, *aphC*, *gcoA* and *catA* in wild-type and Δ*agcA* cells grown on G or 4MG. Lines indicate the means of triplicates. Genes copies were normalized to *sigA*.

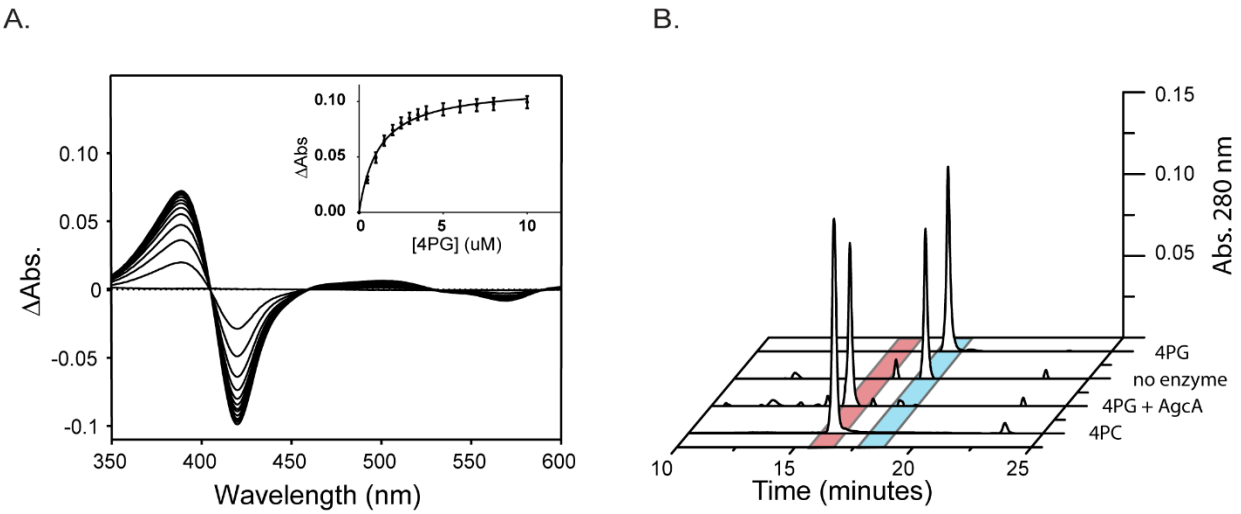


Figure 4. The binding and transformation of 4PG by AgcA. (A) Difference spectra obtained upon titration of 0.2 μM AgcA with 4PG (0.5 to 10 μM). The resulting binding curve is inset with error bars representing the standard error calculated from triplicate data; the solid line represents the best fit of a quadratic binding equation to the data, with fitted parameters: K_D = 1.2 ± 0.1 μM, and ΔA_{421Max} = 0.11 ± 0.002(20 mM MOPS, 90 mM NaCl, pH 7.2 at 25.0 °C). (B) Turnover of 4PG by AgcA. Reactions were performed in the same buffer as A at 30 °C containing 100 μM 4PG, 1 mM NADH, 2 μM AgcA and 10 μM AgcB. HPLC traces (from back to front) are of 4PG standard, reaction containing 4PG without AgcA, a reaction containing AgcA plus 4PG, and 4-propylcatechol (4PC) standard. Additions of 4PG and 4-propylcatecol were 100 μM.

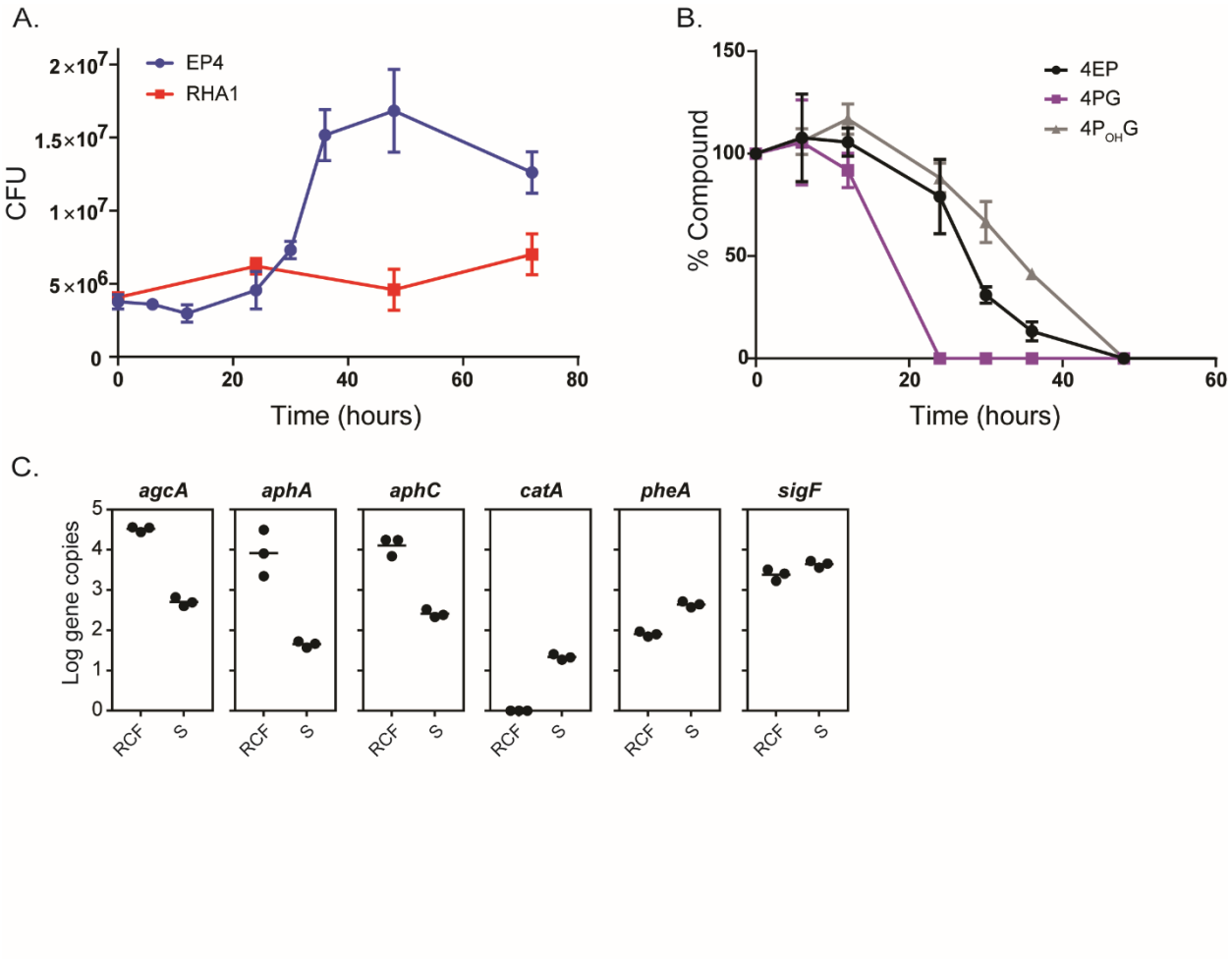


Figure 5. Growth of *R. rhodochrous* EP4 and *R. jostii* RHA1 on corn stover RCF oil. (A) Growth of strains in M9 media containing 250 ppm RCF oil (n = 3 for EP3, n = 2 for RHA1). (B) Depletion of 4EP, 4PG, and 4-propanolguaiacol from cultures of *R. rhodochrous* EP4. Plotted values are the percentage of peak area at t = 0 divided by the peak area of the internal standard (3-chlorobenzoate). (C) Transcript levels of key genes in *R. rhodochrous* EP4 cells harvested at 32 hours of growth on RCF oil (RCF) versus on succinate (S). Growth of *R. rhodochrous* EP4 on succinate shown in Figure S1A.

REFERENCES

1. Ragauskas AJ*, et al.* (2014) Lignin valorization: improving lignin processing in the biorefinery. *Science* 344(6185):1246843.
2. Zakzeski J, Bruijnincx PC, Jongerius AL, & Weckhuysen BM (2010) The catalytic valorization of lignin for the production of renewable chemicals. *Chemical Reviews* 110(6):3552-3599.
3. Rinaldi R*, et al.* (2016) Paving the Way for Lignin Valorisation: Recent Advances in Bioengineering, Biorefining and Catalysis. *Angew. Chem. Int. Ed. Engl.* 55(29):8164-8215.
4. Schutyser W*, et al.* (2018) Chemicals from lignin: an interplay of lignocellulose fractionation, depolymerisation, and upgrading. *Chem. Soc. Rev.* 47:852-908.
5. Sun Z, Fridrich B, de Santi A, Elangovan S, & Barta K (2018) Bright Side of Lignin Depolymerization: Toward New Platform Chemicals. *Chem. Rev.* 118(2):614-678.
6. Davis R*, et al.* (2013) Process Design and Economics for the Conversion of Lignocellulosic Biomass to Hydrocarbons: Dilute-Acid and Enzymatic Deconstruction of Biomass to Sugars and Biological Conversion of Sugars to Hydrocarbons. (NREL).
7. Corona A*, et al.* (2018) Life cycle assessment of adipic acid production from lignin. *Green Chem.* 20(16):3857-3866.
8. Eltis LD & Singh R (2018) Biological Funneling as a Means of Transforming Lignin-derived Aromatic Compounds into Value-added Chemicals. *Lignin Valorization: Emerging Approaches*, (The Royal Society of Chemistry), pp 290-313.
9. 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):1-15.
10. Beckham GT, Johnson CW, Karp EM, Salvachua D, & Vardon DR (2016) Opportunities and challenges in biological lignin valorization. *Curr. Opin. Biotechnol.* 42:40-53.
11. Becker J & Wittmann C (2019) A field of dreams: Lignin valorization into chemicals, materials, fuels, and health-care products. *Biotechnol. Adv.*:Ahead of Print.
12. Linger JG*, et al.* (2014) Lignin valorization through integrated biological funneling and chemical catalysis. *Proc. Natl. Acad. Sci. U.S.A.* 111(33):12013-12018.
13. Vaillancourt FH, Bolin JT, & Eltis LD (2006) The ins and outs of ring-cleaving dioxygenases. *Crit. Rev. Biochem.Mol. Biol.* 41(4):241-267.
14. Vardon DR*, et al.* (2015) Adipic acid production from lignin. *Energy Environ. Sci.* 8(2):617-628.
15. Johnson CW*, et al.* (2016) Enhancing muconic acid production from glucose and lignin-derived aromatic compounds via increased protocatechuate decarboxylase activity. *Metab. Eng. Commun.* 3:111-119.
16. Salvachua D*, et al.* (2018) Bioprocess development for muconic acid production from aromatic compounds and lignin. *Green Chem.* 20(21):5007-5019.
17. Harada A*, et al.* (2017) The crystal structure of a new *O*-demethylase from *Sphingobium* sp. strain SYK-6. *FEBS J.* 284(12):1855-1867.
18. Kohler AC, Mills MJL, Adams PD, Simmons BA, & Sale KL (2017) Structure of aryl *O*-demethylase offers molecular insight into a catalytic tyrosine-dependent mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 114(16):E3205-E3214.

19. Brunel F & Davison J (1988) Cloning and sequencing of *Pseudomonas* genes encoding vanillate demethylase. *J. Bacteriol.* 170(10):4924-4930.
20. Buswell JA & Ribbons DW (1988) Vanillate *O*-demethylase from *Pseudomonas* species. *Methods Enzymol.* 161(Biomass, Pt. B):294-301.
21. Meyer F, Pupkes H, & Steinbuchel A (2017) Development of an Improved System for the Generation of Knockout Mutants of *Amycolatopsis* sp. Strain ATCC 39116. *Appl. Environ. Microbiol.* 83(3):02660-02616.
22. Mallinson SJB*, et al.* (2018) A promiscuous cytochrome P450 aromatic *O*-demethylase for lignin bioconversion. *Nat. Commun.* 9(1):2487.
23. Machovina MM*, et al.* (2019) Enabling microbial syringol conversion through structure-guided protein engineering. *Proc. Natl. Acad. Sci. U.S.A.*
24. García-Hidalgo J, Ravi K, Kuré L-L, Lidén G, & Gorwa-Grauslund M (2019) Identification of the two-component guaiacol demethylase system from *Rhodococcus rhodochrous* and expression in *Pseudomonas putida* EM42 for guaiacol assimilation. *AMB Express* 9(1):34.
25. Van den Bosch S*, et al.* (2015) Reductive lignocellulose fractionation into soluble lignin-derived phenolic monomers and dimers and processable carbohydrate pulps. *Energy Environ. Sci.* 8(6):1748-1763.
26. Schutyser W (2015) Influence of bio-based solvents on the catalytic reductive fractionation of birch wood. *Green Chem.* 17(11):5035-5045.
27. Anderson EM*, et al.* (2016) Reductive Catalytic Fractionation of Corn Stover Lignin. *ACS Sustain. Chem. & Eng.* 4(12):6940-6950.
28. Parsell TH*, et al.* (2013) Cleavage and hydrodeoxygenation (HDO) of C–O bonds relevant to lignin conversion using Pd/Zn synergistic catalysis. *Chem. Sci.* 4(2):806-813.
29. Torr KM, van de Pas DJ, Cazeils E, & Suckling ID (2011) Mild hydrogenolysis of in-situ and isolated *Pinus radiata* lignins. *Bioresour. Technol.* 102(16):7608-7611.
30. Lazaridis PA, Fotopoulos AP, Karakoulia SA, & Triantafyllidis KS (2018) Catalytic Fast Pyrolysis of Kraft Lignin With Conventional, Mesoporous and Nanosized ZSM-5 Zeolite for the Production of Alkyl-Phenols and Aromatics. *Front. Chem.* 6(295).
31. Guo D (2017) Effect of molecular weight on the pyrolysis characteristics of alkali lignin. *Fuel (Guildford)* 193:45-53.
32. Otani H, Lee YE, Casabon I, & Eltis LD (2014) Characterization of *p*-hydroxycinnamate catabolism in a soil Actinobacterium. *J. Bacteriol.* 196(24):4293-4303.
33. Levy-Booth DJ*, et al.* (2019) Catabolism of Alkylphenols in *Rhodococcus* via a *meta*-Cleavage Pathway Associated with Genomic Islands. *Front. Microbiol.* 10:1862.
34. Karlson U*, et al.* (1993) Two independently regulated cytochromes P-450 in a *Rhodococcus rhodochrous* strain that degrades 2-ethoxyphenol and 4-methoxybenzoate. *J. Bacteriol.* 175(5):1467-1474.
35. Henson WR*, et al.* (2018) Multi-omic elucidation of aromatic catabolism in adaptively evolved *Rhodococcus opacus*. *Metab. Eng.* 49:69-83.
36. Eltis LD, Karlson U, & Timmis KN (1993) Purification and characterization of cytochrome P450RR1 from *Rhodococcus rhodochrous*. *Eur. J. Biochem.* 213(1):211-216.
37. McLeod MP*, et al.* (2006) The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc. Natl. Acad. Sci. U.S.A.* 103(42):15582-15587.

38. Luthra A, Denisov IG, & Sligar SG (2011) Spectroscopic features of cytochrome P450 reaction intermediates. *Arch. Biochem. Biophys.* 507(1):26-35.
39. Eby DM, Beharry ZM, Coulter ED, Kurtz DM, Jr., & Neidle EL (2001) Characterization and evolution of anthranilate 1,2-dioxygenase from *Acinetobacter* sp. strain ADP1. *J. Bacteriol.* 183(1):109-118.
40. Lewis DF (1996) *Cytochromes P450: structure, function and mechanism* (Taylor & Francis Ltd).
41. Chen HP*, et al.* (2012) Vanillin catabolism in *Rhodococcus jostii* RHA1. *Appl. Environ. Microbiol.* 78(2):586-588.
42. Patrauchan MA*, et al.* (2005) Catabolism of benzoate and phthalate in *Rhodococcus* sp. strain RHA1: redundancies and convergence. *J. Bacteriol.* 187(12):4050-4063.
43. Guenther K, Schlosser D, & Fritsche W (1995) Phenol and cresol metabolism in *Bacillus pumilus* isolated from contaminated groundwater. *J. Basic Microbiol.* 35(2):83-92.
44. Catelani D, Fiecchi A, & Galli E (1971) (+)-γ-Carboxymethyl-γ-methyl-Δα-butenolide. 1,2-Ring-fission product of 4-methylcatechol by *Pseudomonas desmolyticum*. *Biochem. J.* 121(1):89-92.
45. Kolomytseva MP, Baskunov BP, & Golovleva LA (2007) Intradiol pathway of *para*-cresol conversion by *Rhodococcus opacus* 1CP. *Biotechnol. J.* 2(7):886-893.
46. Ausubel FM (2002) *Short protocols in molecular biology : a compendium of methods from Current protocols in molecular biology* (Wiley, New York) 5th Ed.

Supplementary Information for

Characterization of alkylguaiacol-degrading cytochromes P450 for the biocatalytic valorization of lignin

Morgan M. Fetherolf[a,1], David J. Levy-Booth[a,1], Jie Liu[a], Andrew Wilson[a], Rui Katahira[b], Gregg T. Beckham[b], William W. Mohn[a], and Lindsay D. Eltis[a,*]

[a]Department of Microbiology & Immunology, Life Sciences Institute, The University of British Columbia, Vancouver, British Columbia, Canada

[b]National Bioenergy Center, National Renewable Energy Laboratory, Golden CO, United States

*Email: leltis@mail.ubc.ca

[1]These authors contributed equally to this work

This Supplementary Materials file includes:

Supplementary Materials and Methods

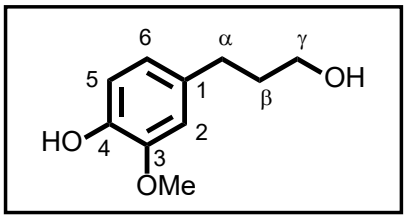
Tables S1 to S3

Figs. S1 to S8

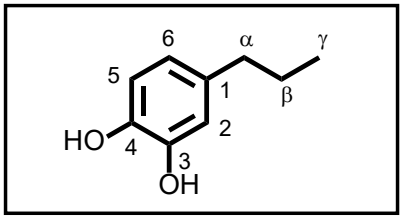
References for SI citations

Supplementary Methods

Synthesis of 4-propylcatechol and 4-propylsyringol. Eugenol, 4-allyl-2,6-dimethoxyphenol, borane-THF complex solution (1 M), iodocyclohexane, and 30% H_2O_2 were purchased from Sigma-Aldrich Co (St. Louis, USA). All chemicals were used without additional purification. Flash chromatography was performed using RediSep Rf silica cartridges on a CombiFlash (Teledyne, Lincoln, Nebraska, USA). NMR spectra were acquired at 25°C on a Bruker AVANCE III 400 MHz nuclear magnetic resonance (NMR) spectrometer (9.39 T) with a 5 mm BBO probe with a Z gradient. The standard Bruker implementations of 1D (^{1}H and ^{13}C) and 2D [^{1}H-^{1}H correlation spectroscopy (COSY), heteronuclear single-quantum coherence (HSQC)] Chemical shifts were referenced to tetramethylsilane (TMS; 0.0 ppm).



4-Propanol guaiacol was synthesized according to the method from Pepper and co-workers (1). To the solution of eugenol (3.58 g, 21.8 mM) in THF (50 mL), borane-THF complex solution (1 M) (21.8 mL, 21.8 mM) was added dropwise at 0°C. After stirring for 1.5 hours, deionized H_2O (7.5 mL) was added slowly to obtain a transparent solution. This was followed by addition of 3 M NaOH (9 mL) slowly and the solution was allowed to warm to ambient temperature. Subsequently, 30% H_2O_2 aq (5.2 mL) was added dropwise at room temperature. After stirring for 2 hours, THF in the reaction mixture was removed under reduced pressure until almost dryness, diluted in deionized H_2O, and then extracted with petroleum ether (2 × 100 mL). The aqueous layer was acidified with 2 N HCl (10 mL), and then extracted with petroleum ether (2 × 100 mL). The combined organic layer was washed with brine and then dried over Na_2SO_4 to obtain a crude solid. The crude residue was purified by silica gel chromatography with ethyl acetate/hexane (1:1, v/v) to yield a white powder (1.042 g, 26.2 mol%). ^{1}H NMR (400 MHz, acetone-d_6): δ = 6.68 (d, J = 2.0 Hz, 1H, H2), 6.59 (d, J = 8.0 Hz, 1H, H5), 6.51 (dd, J = 8.0, 2.0 Hz, 1H, H6), 3.68 (s, 3H, OCH_3), 3.42 (m, 2H, Hγ), 2.46 (t, J = 7.56 Hz, 2H, Hα), 1.64 (m, 2H, Hβ) ppm; ^{13}C NMR (75 MHz, Acetone-d_6): δ = 146.7 (C3), 144.0 (C4), 133.1 (C1), 120.1 (C6), 114.1 (C5), 111.4 (C2), 60.4 (Cγ), 54.7 (OCH_3), 34.4 (Cβ), 31.0 (Cα) ppm.



To synthesize 4-propylcatechol, iodocyclohexane (12.5 mL, 94.9 mM) was added at room temperature to a solution of 4-propylguaiacol (3.16 g, 19.0 mM) in DMF (5 mL). The reaction mixture was refluxed for 8 hours. After cooling down to room temperature, the reagent (iodocyclohexane) was removed under reduced pressure. The reaction mixture was diluted with deionized water (30 mL) and then extracted with ethyl acetate (3 × 30mL). The combined organic layer was washed with brine and dried over Na_2SO_4 to obtain a crude oil. The crude residue was purified by flash chromatography with MeOH/dichloromethane to yield a white powder (1.15 g, 39.8 mol%). ^{1}H NMR (400 MHz, $CDCl_3$) of acetylate: δ = 7.08 (d, J = 8.24 Hz, 1H, H5), 7.04 (dd, J = 8.28, 1.6 Hz, 1H, H6), 6.99 (d, J = 1.6 Hz, 1H, H2), 2.57 (t, J = 7.52 Hz, 2H, Hα), 2.28 (s, 3H, -$OCOCH_3$), 2.27 (s, 3H, -$OCOCH_3$), 1.63 (m, 2H, Hβ), 0.94 (t, J = 7.32 Hz, 3H, Hγ) ppm; ^{13}C NMR (75 MHz, $CDCl_3$): δ = 168.7 (-O$\underline{C}$OCH_3), 168.6 (-O$\underline{C}$OCH_3), 141.9 (C3), 141.8 (C4), 140.1 (C1), 126.7 (C6), 123.4 (C2), 123.1 (C5), 37.6 (Cα), 24.4 (Cβ), 20.8 (-OCO$\underline{C}H_3$), 14.0 (Cγ) ppm.

Bacterial strains and growth conditions. *E. coli* strains were grown at 37 °C on LB agar or LB broth supplemented with the appropriate antibiotic. Rhodococcal strains were routinely grown on LB agar

or LB broth at 30 °C. For growth on specific substrates and transcriptomic analysis, single colonies of EP4 or RHA1 were inoculated in 5 mL LB broth and grown overnight at 30 °C (200 rpm). Cells were pelleted at 4000 × g, washed twice with M9 media, then used to inoculate M9-Goodies containing 1.0 mM growth substrate at an optical density at 600 nm (OD_{600}) ~0.05 unless otherwise stated.

Transcriptomic analyses. RNA reads were subject to quality control, filtered and trimmed using Trimmomatic 0.3.6 defaults (2). Transcripts were quantified using Salmon 0.8.1 (3). Differential expression was analyzed using *DeSeq2* 1.18.1 (4) in R 3.4.4 (5).

RT-qPCR analyses. RNA was reverse transcribed to cDNA using SuperScript VILO (Life Technologies) and Turbo DNase (Thermo Fisher Scientific). Amplification reactions were performed using aStepOnePlus Real-Time PCR System (Thermo Fisher Scientific) and the oligonucleotides shown in Table S1. The cycling conditions were: 95 °C for 5 min followed by 40 cycles of: 95 °C for 30s, 52 °C for 20s, 72 °C for 30s. Standard curves for all genes were made using 10× serial dilutions of pCR2.1-TOPO TA plasmids containing the target amplicon from 10^9 to 10^2. RT-qPCR transcript abundances were compared statistically using two-tailed Student's t-tests with Bonferroni correction.

Phylogenetic analysis. Structure-based sequence alignments were performed using the Expresso version of Tcoffee (6). Gaps and poorly conserved structural elements were removed using Jalview version 2.1 (7). A maximum likelihood tree was built using the PhyML algorithm (8, 9). Trees were visualized using iTOL version 4 (10).

Gene cloning and deletion. Restriction enzymes and the Phusion PCR system were from New England Biolabs. The *agcA* and *agcB* genes were amplified from DNA isolated from EP4 using primers listed in Table S1. Similarly, *aphC* was amplified from RHA1 DNA. The amplicons were cloned into pET15b using the *Nde*I and *Bam*HI restriction sites to yield pETAgcA, pETAgcB and pETAphC, which encode proteins with N-terminal polyhistidine tags that can be proteolytically removed using TEV^{pro}. The nucleotide sequences of all constructs were verified. The *agcA* and *gcoA* genes were deleted in RHA1 utilizing *sacB* counterselection as previously described (11). Briefly, flanking regions of the target genes were amplified using the oligonucleotides indicated in Table S1 to generate pK18ΔagcA and pK18Δ*gcoA* which were then conjugated into RHA1. Deletion mutants were verified by PCR (Fig. S9).

Protein production and purification. AgcA, AgcB and AphC were produced heterologously as polyHis-tagged (Ht-) proteins using *E. coli* BL-21 λ (DE3) containing pETAgcA, pETAgcB and pETAphC, respectively. Freshly transformed cells were grown at 37 °C in LB broth containing 100 mg/L ampicillin to OD_{600} ~0.5. Heterologous gene expression was induced by adding isopropyl β-D-thiogalactopyranoside to 0.5 mM and the cells were incubated at 25 °C for an additional 16 h. Cells were harvested by centrifugation. Cells collected from 2 L of culture were suspended in 20 mL buffer (20 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 8) and lysed at 4 °C using an EmulsiFlex-C5 homogenizer (Avestin, Ottawa, ON, Canada). Cellular debris was removed by centrifugation. Ht-proteins were purified from the cell extract using immobilized metal affinity chromatography (Ni sepharose 6 fast flow resin, Qiagen) according to the manufacturer's protocol.

Ht-AgcA eluted from the Ni resin was reconstituted by dropwise addition of 50 mM hemin in 0.1 N NaOH while stirring to a final 2:1 molar concentration of hemin:AgcA. The polyHis tag was removed incubating Ht-AgcA with TEV protease (50:1 molar ratio) and dialyzing the mixture overnight against 20 mM Tris, pH 8.0 at 4 °C. The protease and uncleaved Ht-AgcAwere removed by passing the

dialyzed mixture over the Ni resin. AgcA was further purified using a MonoQ 10/100 GL column and an ÄKTA Purifier (GE Healthcare). The protein was eluted with a linear gradient from 0 to 1 M NaCl in 120 mL of 20 mM Tris, pH 8.0. Fractions containing AgcA were pooled, dialyzed into 20 mM Tris, pH 8.0 concentrated to ~15 mg/ml, flash-frozen as beads in liquid N_2, and stored at -80 °C until further use. Ht-AgcB eluted from the Ni resin was exchanged into 20 mM sodium phosphate, 150 mM NaCl, pH 7.2 and flash-frozen as beads in liquid N_2. Ht-AphCwas purified under the same chromatography conditions as Ht-AgcA.

CO binding. CO binding of the reconstituted AgcA was assessed as described previously (12). Briefly, 0.2 µM AgcA in 1 mL of 20 mM MOPS, 90 mM NaCl, pH 7.2 (I = 0.1 M) was reduced with a few grains of sodium dithionite. A spectrum of the reduced enzyme was recorded between 300-500 nm. CO was bubbled into the reduced enzyme mixture for 1 minute and a spectrum of the CO bound species was immediately recorded.

Determination of dissociation constants. K_d values were evaluated by titrating AgcA with ligands of interest and monitoring the resulting Type I difference spectrum using a Cary 60 UV-vis spectrophotometer equipped with a thermostatted cuvette holder. Solutions contained 0.3 µM AgcA in 20 mM MOPS, 90 mM NaCl, pH 7.2 (I = 0.1 mM) at 25 °C. Ligands in the same buffer were added in aliquots at concentrations up to 1.5 mM, and spectra (350 to 450 nm) were recorded after the addition of each aliquot. K_d values were determined by fitting a single site binding model to the data using Origin 8.1 software (Northampton, MA).

HPLC analysis of reaction products. Reaction products were evaluated using a Waters 2695 Separation high-pressure liquid chromatography (HPLC) module (Milford, MA) equipped with a Luna 5 µm C18(2) column 250 × 4.6 mm (Phenomenex, Torrance, CA, USA) operated at 0.7 mL/min and a Waters 2996 photodiode array detector. Reactions containing 2 µM AgcA, 10 µM AgcB, 100 µM substrate, 1000 U/mL of catalase, and 350 µM NADH, in 20 mM MOPS, 90 mM NaCl, pH 7.2 (I = 0.1 mM) were incubated overnight at 30 °C. Reactions were quenched by adding acetic acid to 1%, centrifuged (16,000 × g for 5 min) and passed through a 0.2 µm filter. Substrates and products were eluted using a 16.8 mL linear gradient of 1% formic acid in H_2O to 100% methanol. The eluate was monitored at 280 nm.

Steady-state Kinetics. Steady-state kinetic parameters for AgcAB were evaluated by coupling the *O*-demethylation of the alkyguaiacol to the *meta*-cleavage of the resulting alkylcatechol using AphC. Assays were performed using a Cary 5000 spectrophotometer equipped with a thermostatted cuvette holder. The standard assay was performed in 0.2 mL of air-saturated 20 mM MOPS, 90 mM NaCl, pH 7.2 (I = 100 mM) at 25 °C containing 2 µM AgcA, 6 µM AgcB, 0.5 µM AphC, 1000 U/mL of catalase, 350 µM NADH, and 200 µM 4-propylguaiacol. Assays were initiated by adding the guaiacol, and the reaction progress was monitored at 400 nm. For parameter evaluation, alkylguaiacol concentrations were varied between 1 and 300 µM. Extinction coefficients of the methyl-, ethyl-, and propylcatechol MCPs at 400 nm were 18,600, 15,100, and 19,400 M^{-1} cm^{-1}, respectively. Steady-state kinetic parameters were evaluating by fitting the Michaelis-Menten equation to the data using Origin 8.1 software (Northampton, MA).

The reductase activity of AgcB was measured spectrophotometrically by monitoring the reduction of cytochrome *c* as previously described (12). The standard assay contained 25 nM AgcB and 40 µM bovine cytochrome *c* in 300 mM phosphate, pH 7.7 at 25 °C. The reaction was initiated by adding

NADH to 100 µM and the increase in absorbance at 550 nm was measured. Specific activity was calculated using $\Delta\varepsilon_{550} = 0.021\ mM^{-1}\ cm^{-1}$.

Determination of extinction coefficients. Extinction coefficients of *meta*-cleavage products (MCPs) were evaluated by determining the amount of oxygen consumed upon cleavage of the alkylcatechol by AphC using a Clark-type polarographic O_2 electrode OXYG1 (Hansatech, Pentney, UK) connected to a circulating water bath. The absorbance of the MCP was measured at 400 nm using a Cary5000 spectrophotometer. Reactions were performed in 1.0 mL of air-saturated 20 mM MOPS, 90 mM NaCl, pH 7.2 (I = 0.1 mM) at 25 °C. Measurements were initiated with the addition of ~0.2 nmol of AphC and the reactions were allowed to proceed to completion. The amount of O_2 consumed was taken to be the amount of MCP generated, and was used to calculate the extinction coefficient from the absorbance and the Beer-Lambert equation.

GC-MS analysis of culture supernatants. Culture supernatants were analyzed using an Agilent Technologies (Santa Clara, U.S.A.) 6890N gas chromatograph equipped with a 30-m Agilent 190915-433 capillary column and an Agilent 5973 mass-selective detector. Samples of 300 µL were spiked with 5.0 nmol of 3-chlorobenzoic acid as an internal standard, acidified with a final concentration of 1% acetic acid, and extracted with an equal volume of ethyl acetate. The extracted compounds were dried and derivatized using a 50/50 mixture of BSTFA + TMCS and pyridine. The run was held at 90 °C for 3 min, then ramped to 290 °C at 12 °C min^{-1} with a 10 min final hold. Peak areas were exported and normalized to the internal standard to measure compound depletion.

Supplementary Tables

Table S1. CYP255A systems in *R. rhodochrous* EP4 and *R. jostii* RHA1.

Gene	EP4[a]	RHA1[b]	Product	Best Hit[c]	% ID[d]
agcA	RS01540	RS18805	alkylguaiacol *O*-demethylase, oxygenase	$GcoA_{Amyco}$, P0DPQ7	65
agcB	RS01545	RS18800	alkylguaiacol *O*-demethylase, reductase	$GcoB_{Amyco}$, P0DPQ8	49
gcoA	RS07265	RS11640	guaiacol *O*-demethylase, oxygenase	$GcoA_{Amyco}$, P0DPQ7	76
gcoB	RS07270	RS11645	guaiacol *O*-demethylase, reductase	$GcoB_{Amyco}$, P0DPQ8	58

[a]Locus in EP4 (C6369_RSXXXXX).
[b]Locus in RHA1 (RHA1_RSXXXXX).
[c]Gene name and Uniprot identifier of closest characterized homolog.
[d]Percent amino acid sequence identity of best hit and EP4 homolog.

Table S2. Oligonucleotides used in this study

Primer name	Sequence (5′ to 3′)	Purpose
agcA_F	GTGTGTGTCATATGATGACTTCCACCCACAGC	clone *agcA*$_{EP4}$ into pET15b
agcA_R	ACACACACGGATCCTTACACTTCCCATTTGACAAACAG	clone *agcB*$_{EP4}$ into pET15b
agcB_F	GTGTGTGTCATATGATGTCTGATAAATATTTGGTCCGTATAG	clone *agcB*$_{EP4}$ into pET15b
agcB_R	ACACACACGGATCCTCAGAGCACTGCCGC	clone *agcB*$_{EP4}$ into pET15b
aphC_F	ATATATCATATGATGAGTGACGCTCGTTTCGACATC	clone *agcA*$_{RHA1}$ into pET15b
aphC_R	TATATAGGATCCTCAGAGCGACGC	clone *aphC*$_{RHA1}$ into pET15b
agcA_up_F	CTAGTCTAGAGCTGCACCGTCCGGAG	deletion of *agcA* $_{RHA1}$
agcA_up_R	CGGGGTACCCCTCGCGCCGGAGC	deletion of *agcA* $_{RHA1}$
agcA_down_F	CGGGGTACCATCCCGAACATCGAGCCC	deletion of *agcA* $_{RHA1}$
agcA_down_R	CCCAAGCTTCTCACTGAGGACCGGATGATAGT	deletion of *agcA* $_{RHA1}$
gcoA_up_F	CTAGTCTAGATCGAGCACTCGTTTGACGAT	deletion of *gcoA* $_{RHA1}$
gcoA_up_R	CGGGGTACCCTCGTCGATCCAAGACAGGG	deletion of *gcoA* $_{RHA1}$
gcoA_down_F	CGGGGTACCGAACCTCGAACGTGACTCCG	deletion of *gcoA* $_{RHA1}$
gcoA_down_R	CCCAAGCTTCGACAGGTGTCGAAGTCCTC	deletion of *gcoA* $_{RHA1}$
agcA_E_F5	GTCTCCGTTGGTATTCAGCA	RT-qPCR of *agcA*$_{EP4}$
agcA_E_R5	AAATTGCGCGAGTGGTTTC	RT-qPCR of *agcA*$_{EP4}$
aphA_E_F4	GATTGAGATAGAGCACCCGC	RT-qPCR of *aphA*$_{EP4}$
aphA_E_R4	CTTCACTGTTCCGATGGACTC	RT-qPCR of *aphA*$_{EP4}$
pheA1_E_F2	TTGTAGTTGGTGATCGCTGAG	RT-qPCR of *pheA1*$_{EP4}$
pheA1_E_R2	CTTCATCTTCGACAAGGTCCTC	RT-qPCR of *pheA1*$_{EP4}$
aphC_E_F4	ATACAGGCAACTTCGTGACC	RT-qPCR of *aphC*$_{EP4}$
aphC_E_R4	CTATCTGCATCTCGATCCTGAC	RT-qPCR of *aphC*$_{EP4}$
catA_E_F3	TGGATGATGTCGGCGAAAG	RT-qPCR of *catA*$_{EP4}$
catA_E_R3	GAGAAGGACAAGGTTGCTCC	RT-qPCR of *catA*$_{EP4}$
sigF_E_250_F	AAGGAGACGAAATCCGAACC	RT-qPCR of *sigF*$_{EP4}$
sigF_E_366_R	ATGAAGGAACTGCACCTGTC	RT-qPCR of *sigF*$_{EP4}$
agcA_R_F2	TTTCTCGTGCTGTTCACCG	RT-qPCR of *agcA*$_{RHA1}$
agcA_R_R2	CTGATGGGTCTCACCGAAATC	RT-qPCR of *agcA*$_{RHA1}$
aphC_R_F4	TGACTGGTACTCGAAGGTCT	RT-qPCR of *aphC*$_{RHA1}$
aphC_R_R4	GAGATGTTCCGCGACTACG	RT-qPCR of *aphC*$_{RHA1}$
gcoA_F	TGAACAGCAGGATGGAGTTG	RT-qPCR of *gcoA*$_{RHA1}$
gcoA_R	ATCGTGGACATCGAAATCGG	RT-qPCR of *gcoA*$_{RHA1}$
catA_R_F5	GTAGTAGCCGTCGTTGTCG	RT-qPCR of *catA*$_{RHA1}$
catA_R_R5	TGATCCTCGACCCGAAGA	RT-qPCR of *catA*$_{RHA1}$
sigF_R_F1	AAATACCGGCGAACCTCG	RT-qPCR of *sigF*$_{RHA1}$
sigF_R_R1	GAAGAACGCCTACCAAACCT	RT-qPCR of *sigF*$_{RHA1}$

sigA_F	CTGATCCAGGAAGGCAACC	RT-qPCR of *sigA*$_{RHA1}$
sigA_R	AGAACTTGTAGCCCTTGGTG	RT-qPCR of *sigA*$_{RHA1}$
sigA_Probe	CGAACTTCTCGACGGCACGGAT	RT-qPCR of *sigA*$_{RHA1}$
gcoA_F	AGGCATTCCGACGGAC	*gcoA* $_{RHA1}$ knockout confirmation
gcoA_R	TGAACTTCTCCCGGAAGATG	*gcoA* $_{RHA1}$ knockout confirmation

Table S3. Strains and plasmids used in this study

	Description	Source
***E. coli* strains**		
DH5α	DNA propagation	Invitrogen
BL21 (DE3)	Protein production	Invitrogen
***Rhodococcus* strains**		
R. jostii RHA1	Wild-type	(13)
Δ*agcA*	RHA1 deletion mutant of *agcA*	This study
Δ*gcoA*	RHA1 deletion mutant of *gcoA*	This study
Δ*aphC*	RHA1 deletion mutant of *aphC*	(14)
Δ*pcaL*	RHA1 deletion mutant of *pcaL*	(15)
R. rhodochrous EP4		(14)
Plasmids		
pK18*mobsacB*	*sacB, oriT, aphII*	(16)
pK18Δ*agcA*	pK18*mobsacB*-derived construct to delete *agcA*	This study
pK18Δ*gcoA*	pK18*mobsacB*-derived construct to delete *gcoA*	This study
pETAgcA	pET15b harboring *agcA* of EP4	This study
pETAgcB	pET15b harboring *agcB* of EP4	This study
pETAphC	pET15b harboring *aphC* of RHA1	This study

Supplementary Figures

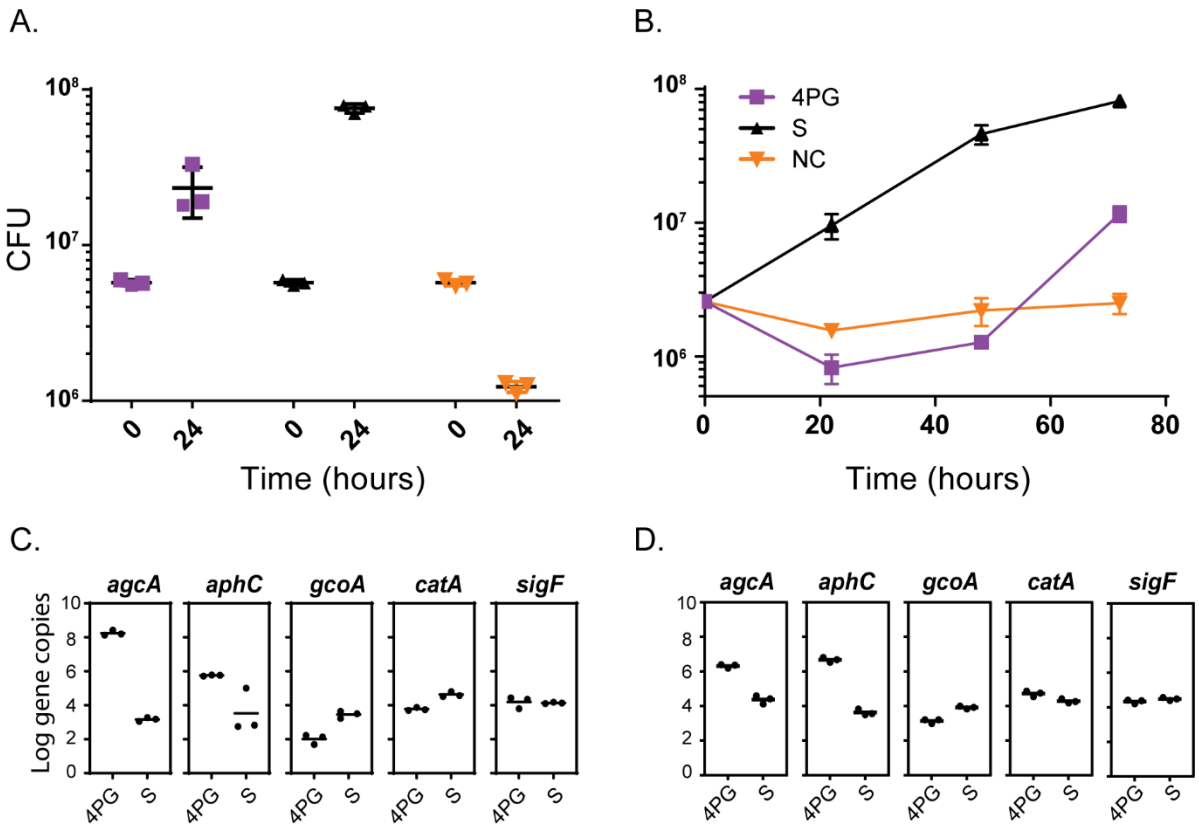


Figure S1. Growth of *R. rhodochrous* EP4 (A) and *R. jostii* RHA1 (B) on M9 minimal medium containing 1 mM 4PG. Transcript copies from key genes in EP4 (C) and wild-type RHA1 (D) growing on 4PG.

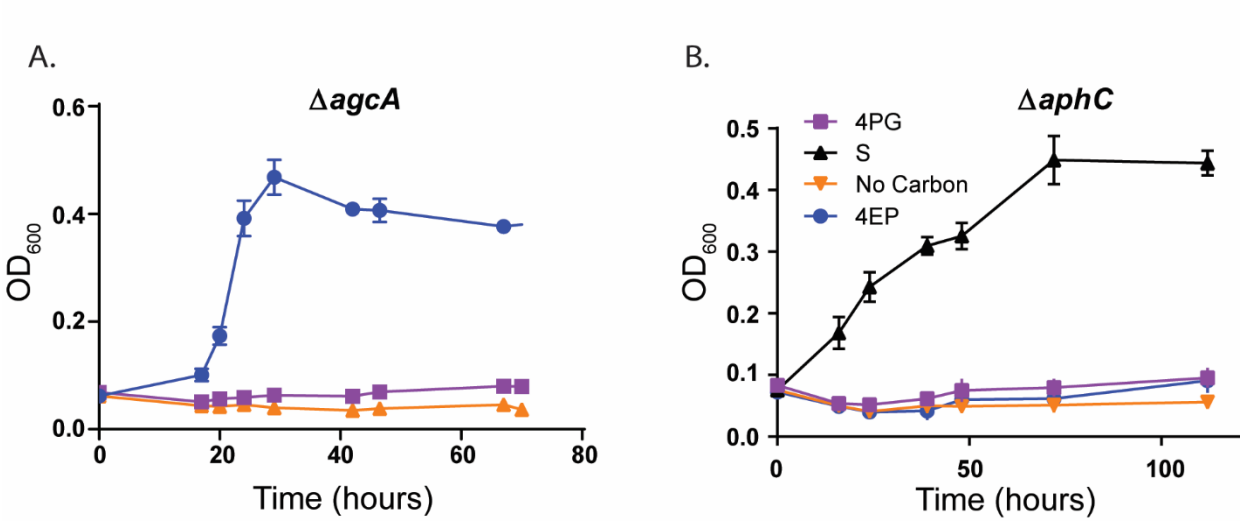


Figure S2. Growth of *R. jostii* RHA1 mutant Δ*agcA* (A) and Δ*aphC* (B)strains on various alkylaromatics.

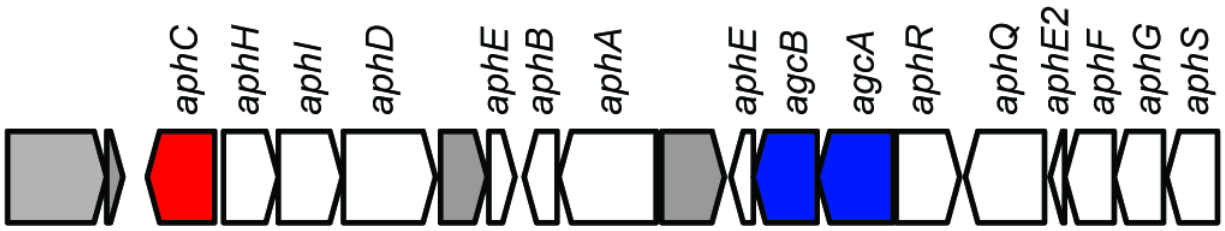


Figure S3. Cluster of genes encoding Aph *meta* cleavage pathway and CYP255A1 in *R. jostii* RHA1. Genes colored in grey have unknown functions.

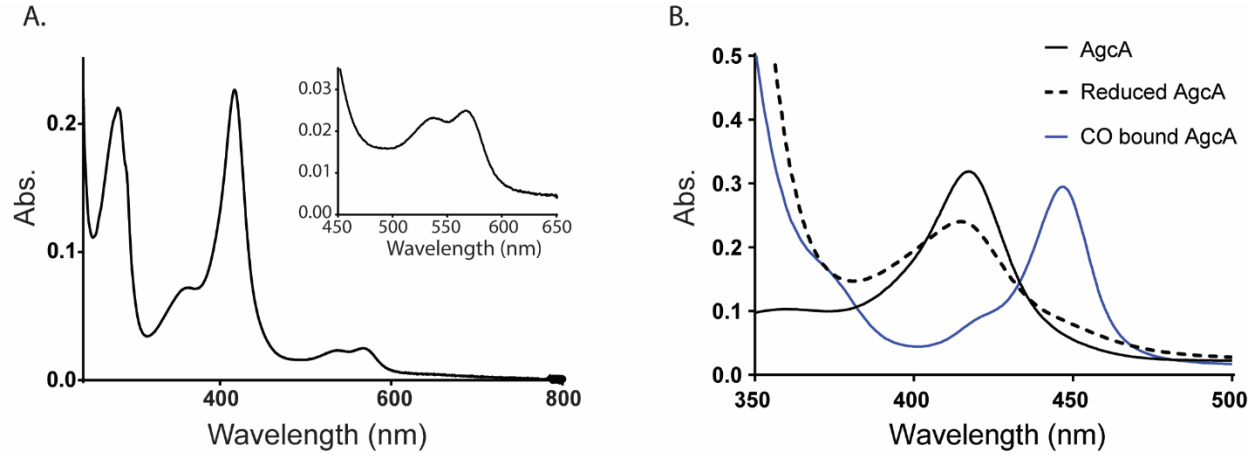


Figure S4. UV-vis spectra of AgcA from *R. rhodochrous* EP4. (A) Spectrum of AgcA purified from *E. coli* and reconstituted with hemin (50 mM phosphate, 300 mM NaCl, pH 7.4 at 25 °C). The inset is of the visible region. (B) Spectra of Soret region of AgcA (20 mM Tris, pH 8.0 at 25 °C): resting state (black), after reduction with sodium dithionite (dashed), and after gentle bubbling with CO (blue).

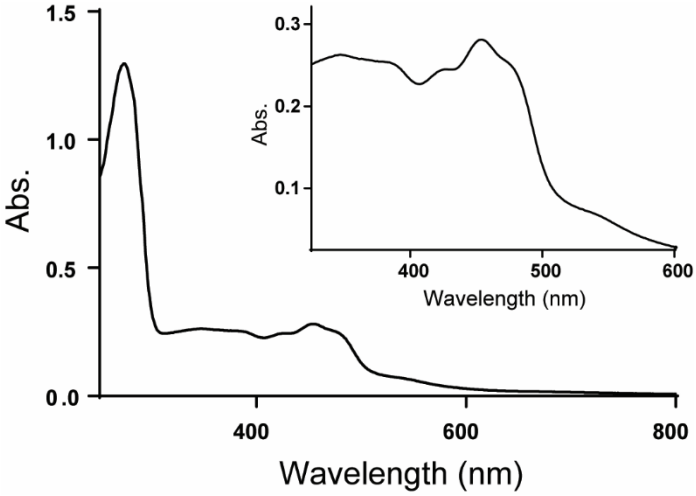


Figure S5. UV-vis spectrum of AgcB from *R. rhodochrous* EP4. Spectrum of AgcB purified from *E. coli* (50 mM phosphate, 300 mM NaCl, pH 7.4). The inset is of the visible region.

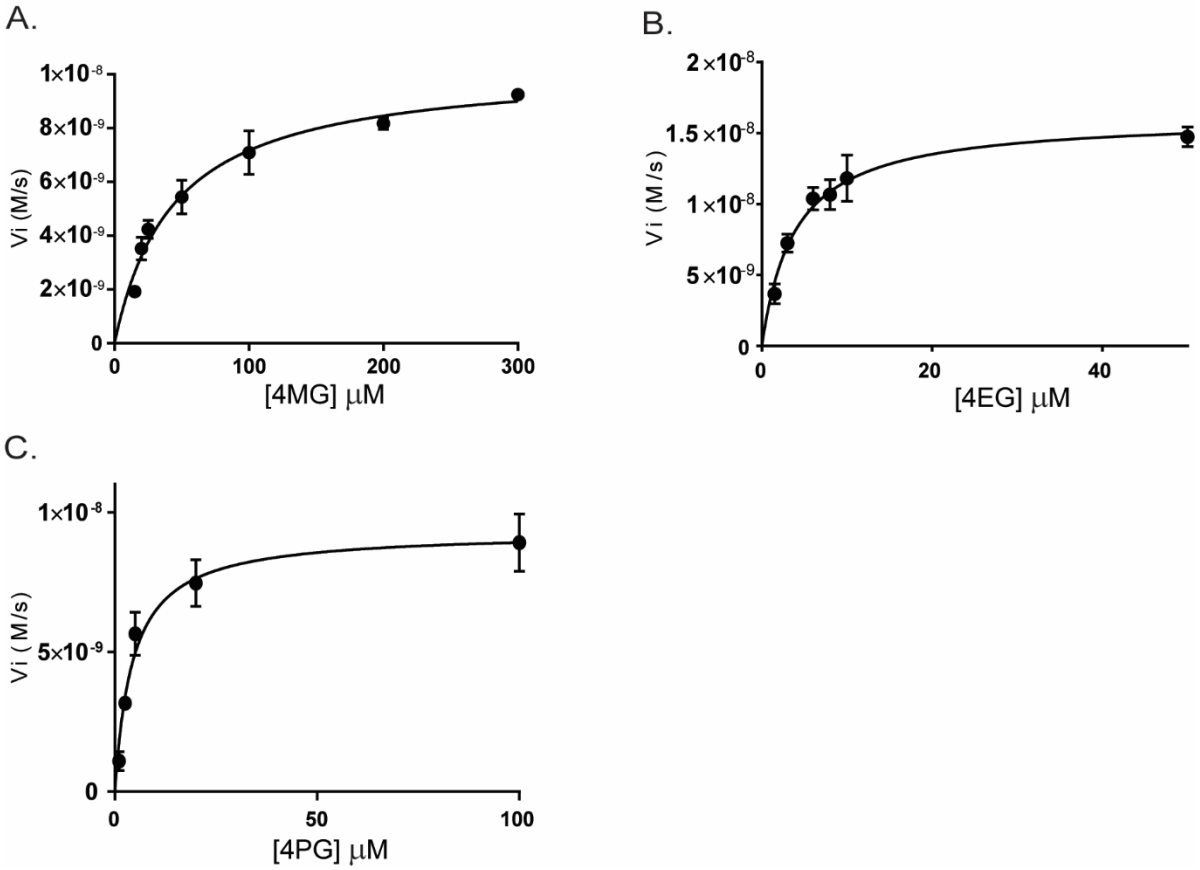


Figure S6. Steady-state kinetic analysis of $AgcA_{EP4}$. Substrates were 4MG (A), 4EG (B), and 4PG (C). The assay was performed in 0.2 ml of air-saturated 20 mM MOPS, 90 mM NaCl, pH 7.2 (*I* = 100 mM) at 25 °C containing

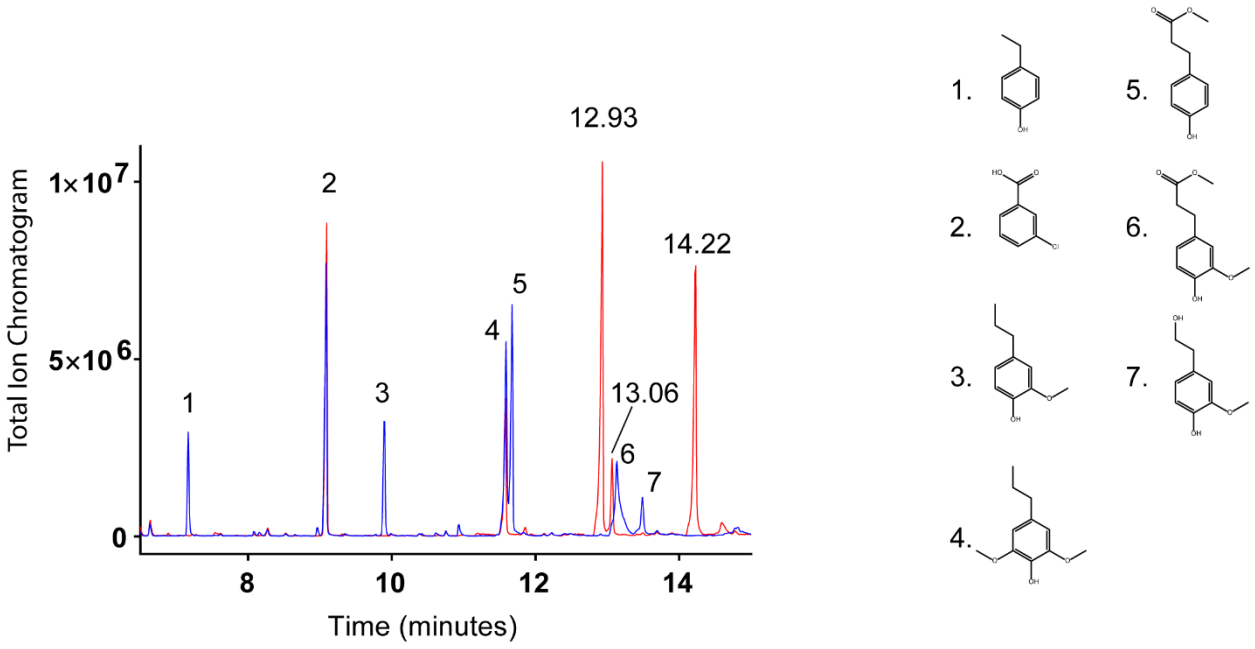


Figure S7. Aromatic compounds in culture of *R. rhodochrous* EP4 growing on corn stover RCF oil. GC trace of culture supernatant sampled at 0 (blue) and 72 (red) hours. Peaks correspond to the compounds on the right: (1) 4-ethylphenol, (2) 3-chlorobenzoate, (3) 4-propylguaiacol, (4) 4-propylsyringol, (5) methyl coumarate, (6) methyl ferulate, and (7) 4-propanolguaiacol. Peaks only present at 72 hours (t_R = 12.93, 13.06, and 14.22 min) were not identified.

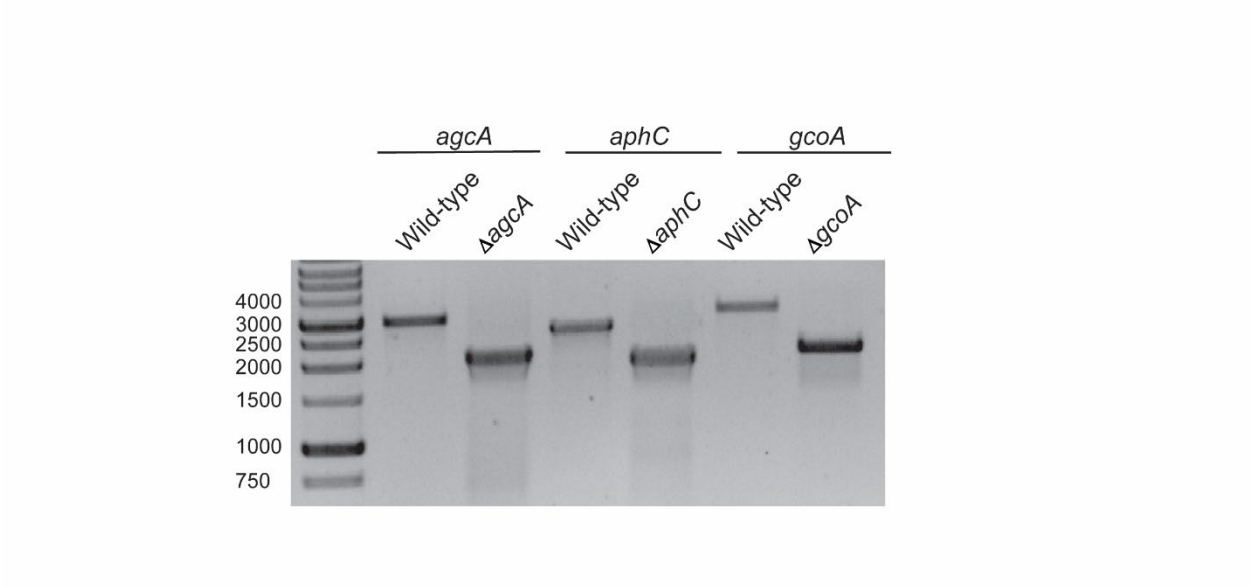


Figure S8. PCR analysis of gene deletion strains of *R. jostii* RHA1. Genes analyzed are indicated above the line, and the strain used for the PCR are located directly above the gel. Expected sizes of the amplicons in the wild-type and mutant strains, respectively are: 3.2 and 2.2 kb for Δ*agcA*; 2.9 and 1.9 kb for Δ*aphC*; and 3.2 and 2.2 kb for Δ*gcoA*.

REFERENCES

1. Pepper JM, Sundaram GS, & Dyson G (1971) Lignin and Related Compounds. III. An Improved Synthesis of 3-(4-Hydroxy-3-methoxyphenyl)-1-propanol and 3-(4-Hydroxy-3,5-dimethoxyphenyl)-1-propanol. *Canadian Journal of Chemistry* 49(20):3394-3395.
2. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.
3. Patro R, Duggal G, Love MI, Irizarry RA, & Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14(4):417-419.
4. Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15(12):550.
5. Team RC (2016) A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
6. Armougom F*, et al.* (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* 34(Web Server issue):W604-608.
7. Waterhouse AM, Martin DMA, Barton GJ, Procter JB, & Clamp M (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189-1191.
8. Dereeper A*, et al.* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36(Web Server issue):W465-469.
9. Guindon S*, et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59(3):307-321.
10. Letunic I & Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments.
11. van der Geize R, Hessels GI, van Gerwen R, van der Meijden P, & Dijkhuizen L (2001) Unmarked gene deletion mutagenesis of kstD, encoding 3-ketosteroid Δ1-dehydrogenase, in *Rhodococcus erythropolis* SQ1 using sacB as counter-selectable marker. *FEMS Microbiol. Lett.* 205(2):197-202.
12. Guengerich FP, Martin MV, Sohl CD, & Cheng Q (2009) Measurement of cytochrome P450 and NADPH–cytochrome P450 reductase. *Nat. Protoc.* 4:1245.
13. Seto M*, et al.* (1995) A Novel Transformation of Polychlorinated Biphenyls by *Rhodococcus* sp. Strain RHA1. *Appl. Environ. Microbiol.* 61(9):3353-3358.
14. Levy-Booth DJ*, et al.* (2019) Catabolism of Alkylphenols in *Rhodococcus* via a *meta*-Cleavage Pathway Associated with Genomic Islands. *Front. Microbiol.*
15. Patrauchan MA*, et al.* (2005) Catabolism of benzoate and phthalate in *Rhodococcus* sp. strain RHA1: redundancies and convergence. *J. Bacteriol.* 187(12):4050-4063.
16. Schafer A*, et al.* (1994) Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* 145(1):69-73.