Catabolism of alkylphenols in *Rhodococcus* via a *meta*-cleavage pathway associated with genomic islands

Running title: Alkylphenol catabolism in *Rhodococcus*

Authors: David J. Levy-Booth[a], Morgan M. Fetherolf[a], Gordon Stewart[a], Jie Liu[a], Lindsay D. Eltis[a] and William W. Mohn[a#]

Affiliations:

1. Department of Microbiology & Immunology, Life Sciences Institute, The University of British Columbia, Vancouver, British Columbia, Canada

D.L.B and M.M.F. contributed equally to this work.

[#]Correspondence:

WW Mohn, Department of Microbiology & Immunology, Life Sciences Institute, The University of British Columbia, 2350 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada. E-mail: wmohn@mail.ubc.ca

Abstract

The bacterial catabolism of aromatic compounds has considerable promise to convert lignin depolymerization products to commercial chemicals. Alkylphenols are a key class of depolymerization products whose catabolism is not well elucidated. We isolated *Rhodococcus rhodochrous* EP4 on 4-ethylphenol and applied genomic and transcriptomic approaches to elucidate alkylphenol catabolism in EP4 and *Rhodococcus jostii* RHA1. RNA-Seq and RT-qPCR revealed a pathway encoded by the *aphABCDEFGHIQRS* genes that degrades 4-ethylphenol via the *meta*-cleavage of 4-ethylcatechol. This process was initiated by a two-component alkylphenol hydroxylase, encoded by the *aphAB* genes, which were up-regulated ~3,000-fold. Purified AphAB from EP4 had highest specific activity for 4-ethylphenol and 4-propylphenol (~2000 U/mg) but did not detectably transform phenol. Nevertheless, a Δ*aphA* mutant in RHA1 grew on 4-ethylphenol by compensatory up-regulation of phenol hydroxylase genes (*pheA1-3*). Deletion of *aphC*, encoding an extradiol dioxygenase, prevented growth on 4-alkylphenols but not phenol. Disruption of *pcaL* in the β-ketoadipate pathway prevented growth on phenol but not 4-alkylphenols. Thus, 4-ethylphenol and 4-propylphenol are catabolized exclusively via *meta*-cleavage in rhodococci while phenol is subject to *ortho*-cleavage. Putative genomic islands encoding *aph* gene*s* were identified in EP4 and several other rhodococci. Overall, this study identifies a 4-alkylphenol pathway in rhodococci, demonstrates key enzymes involved, and presents evidence that the pathway is encoded in a genomic island. These advances are of particular importance for wide-ranging industrial applications of rhodococci, including upgrading of lignocellulose biomass.

Importance

Elucidation of bacterial alkylphenol catabolism is important for the development of biotechnologies to upgrade the lignin component of plant biomass. We isolated a new strain, *Rhodococcus rhodochrous* EP4, on 4-ethylphenol, an alkylphenol that occurs in lignin-derived streams, including reductive catalytic fractionation products of corn stover. We further demonstrated its degradation via a *meta*-cleavage pathway (Aph) with transcriptomics. A new class of Actinobacterial hydroxylase, AphAB, acts specifically on alkylphenols. Phylogenomic analysis indicated that the *aph* genes occur on putative genomic islands in several rhodococcal strains. These genes were identified in the genetically-tractable strain *Rhodococcus jostii* RHA1. Strains missing this element cannot metabolise 4-ethylphenol and 4-propylphenol. Overall, we advanced the understanding of how aromatic compounds are degraded by environmental bacteria and identified enzymes that can be employed in the transition away from petro-chemicals towards renewable alternatives.

Introduction

Lignin, a heterogeneous aromatic polymer, accounts for up to 40% dry weight of terrestrial plant biomass (Ragauskas et al., 2014). It is primarily composed of *p*-hydroxyphenyl (H), guaiacyl (G), and sinapyl (S) subunits, polymerized by ether and C-C bonds (Boerjan et al., 2003). Lignin's heterogeneity and recalcitrant bonds create substantial barriers to its efficient microbial and chemical degradation. Industrial lignin depolymerization is gaining traction as a means to produce fuels and chemicals historically derived from petroleum (Beckham et al., 2016; Ragauskas et al., 2014). Yet heterogeneous depolymerization products can require extensive separation and purification (Linger et al., 2014; Ragauskas et al., 2014). Bacterial biocatalysts provide a means of transforming mixtures of aromatic compounds to single compounds due to the convergent nature of their catabolic pathways, whereby substrates are transformed to central metabolites via shared intermediates, such as catechols (Eltis and Singh, 2018; Linger et al., 2014). Harnessing this biological funneling to refine lignin to high-value chemicals (Beckham et al., 2016; Eltis and Singh, 2018; Linger et al., 2014) is limited in part by a lack of knowledge of the catabolism of lignin-derived monomers.

Alkylphenols are a major class of aromatic compounds generated by a variety of lignin depolymerization technologies. For example, solvolysis of corn lignin produced 24 wt % alkylated monolignins, 46% of which was 4-ethylphenol derived from H-subunits (Jiang et al., 2014). Alkylphenols were also major pyrolysis products of wheat straw black liquor lignin fractions (Guo et al., 2017). Existing depolymerization strategies can require multiple stages of preprocessing and depolymerization, high heat or corrosive chemicals, and can produce dozens of alkylphenol and aromatic products (Asawaworarit et al., 2019; Kim et al., 2015; Ye et al.,

2012). One promising depolymerization strategy that produces a narrow stream of alkylphenols is reductive catalytic reduction (RCF) (Pepper and Lee, 1969). 4-Ethylphenol was a major RCF product of corn stover, comprising up to 16.4% of the resulting aromatic monomers (Anderson et al., 2016).

Two bacterial pathways for the aerobic catabolism of 4-ethylphenol have been reported, initially involving either oxidation of the alkyl-side chain or hydroxylation of the aromatic ring. In *Pseudomonas putida* JD1, the alkyl-side chain is oxidized by 4-ethylphenol methylhydroxylase to eventually yield hydroquinone (Darby et al., 1987; Hopper and Cottrell, 2003). In contrast, *Pseudomonas* sp. KL28 hydroxylates 4-ethylphenol to 4-ethylcatechol (Jeong et al., 2003). In these pathways, the hydroquinone and 4-ethylcatechol undergo *meta*-cleavage (Darby et al., 1987; Jeong et al., 2003). In KL28, the alkylphenol hydroxylase is a six-component enzyme encoded by genes organized in a single co-linear transcriptional unit within a *meta*-cleavage pathway gene cluster (Jeong et al., 2003). A homologous pathway degrades phenol in *Comamonas testosteroni* TA441 (Arai et al., 2000).

Rhodococcus is a genus of mycolic acid-producing Actinobacteria that catabolize a wide variety of aromatic compounds (Yam et al., 2010), including phenols (Gröning et al., 2014; Kolomytseva et al., 2007). These bacteria also have considerable potential as biocatalysts for the industrial production of compounds ranging from nitriles, to steroids and high-value lipids (Alvarez et al., 1996; Round et al., 2017; Sengupta et al., 2019; Shields-Menard et al., 2017). In *Rhodococcus*, phenol catabolism is initiated by a two-component flavin-dependent monooxygenase (PheA1A2) (Saa et al. 2010) to generate a catechol. PheA1A2 homologs in *Rhodococcus opacus* 1CP can also hydroxylate chlorophenols and 4-methylphenol (Gröning et al., 2014) to produce the

97 corresponding catechols, which undergo *ortho*-cleavage (Kolomytseva et al., 2007; Maltseva et
98 al., 1994) and subsequent transformation to central metabolites via the β-ketoadipate pathway. In
99 *rhodococci*, the β-ketoadipate pathway is confluent, with branches responsible for
100 protocatechuate and catechol catabolism converging at PcaL, a β-ketoadipate enol-lactonase
101 (Patrauchan et al., 2005; Yam et al., 2010). However, some *rhodococci* appear to have pathways
102 responsible for the catabolism of alkylated aromatic compounds via *meta*-cleavage (Jang et al.,
103 2005). Elucidating 4-alkylphenol metabolism in *rhodococci* will improve our understanding of
104 Actinobacterial aromatic degradation and support the development of *Rhodococcus* strains as
105 platforms for industrial lignin upgrading.

106 Genomic islands (GIs) are DNA segments likely to have been acquired by horizontal gene
107 transfer. They are characterized by altered nucleotide characteristics (e.g., GC content), syntenic
108 conservation, and frequent presence of mobility genes (transposases, insertion sequences (IS),
109 and integrases) (Hacker and Kaper, 2000; Juhas et al., 2009). They can be further identified by
110 the absence of genomic regions in closely-related strains (Hacker et al., 1990). GIs can confer
111 resistance, virulence, symbiosis, and catabolic pathways (Dobrindt et al., 2004; Juhas et al.,
112 2009). For example, the self-transferable *clc* element enabling 3- and 4-chlorocatechol and 2-
113 aminophenol catabolism was identified as a GI in several *Gamma*- and *Betaproteobacteria*
114 strains (Gaillard et al., 2006). Recent horizontal gene transfer may have played less of a role in
115 shaping the *Rhodococcus jostii* RHA1 genome than in other bacteria such as *Burkholderia*
116 *xenovorans* LB400, which has a similarly-sized genome (McLeod et al., 2006). Further, although
117 RHA1 contains a high number of aromatic pathways, genes encoding these pathways are slightly
118 underrepresented in the identified genomic islands. GIs can ameliorate in host genomes through

119 nucleotide optimization or loss of mobility elements (Juhas et al., 2009; Lawrence and Ochman,
120 1997), reducing our effectiveness at predicting ancestral genomic additions. However,
121 examination of GIs in multiple related genomes with an ensemble of predictive software can
122 improve our understanding of the role of GIs in the evolution of bacterial catabolic pathways.

123 This study sought to identify catabolic pathways required for 4-alkylphenol catabolism. We
124 report on the genomic, transcriptomic and enzymatic characterization of 4-alkylphenol
125 catabolism in a newly-isolated 4-ethylphenol-degrading bacterium, *Rhodococcus rhodochrous*
126 EP4 (Figure 1A), as well as RHA1. The activity of a novel two-component alkylphenol
127 monooxygenase (AphAB) was characterized. Gene deletion analysis was employed to identify
128 the subsequent route of catechol catabolism. Genomic analysis identified a putative *aph* GI,
129 providing new evolutionary insight to the *aph meta*-cleavage pathway in *Rhodococcus*.
130 Knowledge gained by this study will facilitate the efficient valorization of lignin following its
131 depolymerization.

132 Materials and Methods

133 Bacterial strains and growth conditions

134 Liquid enrichment cultures were inoculated with either ~4 month aged agricultural compost (~25
135 cm depth) from The University of British Columbia farm (49° 14' 57.8904" N, 123° 14' 0.0492"
136 W) or forest soil from Pacific Spirit Park in Vancouver, Canada. The cultures contained 1.0 mM
137 4-ethylphenol (≥97.0% Sigma-Aldrich, St. Louis, U.S.A.) as sole organic substrate in M9-
138 Goodies (Bauchop and Elsden, 1960; Elder, 1983). The cultures were incubated at 30ºC with
139 shaking at 200 rpm for 2 weeks. Removal of 4-ethylphenol was monitored by GC-MS, after
140 which cultures were transferred to fresh medium, 0.5% inocula. After 3 serial transfers, isolates

were obtained by plating on homologous medium solidified with 1.5% purified agar. Colonies appeared in 10 days. Individual colonies were transferred to liquid media and replated on solid media for colony isolation. *Rhodococcus rhodochrous* strain DSM43241 was purchased from DSMZ (Braunschweig, Germany).

EP4 and RHA1 cultures for RNA extraction were grown overnight at 30°C on LB broth (200 rpm), diluted to 0.05 OD_{600} and washed thrice in M9 with no supplement, and grown to mid-log phase on 50 ml M9-Goodies with 1.0 mM 4-ethylphenol or succinate. EP4 was additionally grown on 1.0 mM benzoate (99%, Sigma-Aldrich). Additional screening of EP4, *R. rhodochrous* DSM43241, RHA1, and mutant strains used 5 ml M9-Goodies with 1.0 mM phenol (≥99%, VWR International, Ltd., Mississauga, Canada), 3-methylphenol (*m*-cresol; 99% Sigma-Aldrich), 4-methylphenol (99% Sigma-Aldrich), 4-propylphenol (>99%, TCI), 3,4-dimethylphenol (DMP) (99% Sigma-Aldrich), 2,4-DMP (98% Sigma-Aldrich), 4-hydroxyphenylacetate (HPA) (98% Sigma-Aldrich), 4-hydroxybenzioic acid (HBA) (99% Sigma-Aldrich), or 0.5 mM 4-nitrophenol (NP) (≥99%, Sigma-Aldrich), incubated for 24 h. Cells were lysed by boiling in 1M NaOH and protein quantified using the Micro BCA™ Protein Assay (Thermo-Fisher Scientific Inc., Waltham, U.S.A.) and a VersaMax™ microplate reader (Molecular Devices LLC, San Jose, U.S.A.).

DNA manipulation, plasmid construction and gene deletion

DNA was isolated, manipulated, and analyzed using standard protocols (Sambrock and W. Russel, 2001). *E. coli* and RHA1 were transformed with DNA by electroporation using a MicroPulser with GenePulser cuvettes (Bio-Rad). To produce N-terminal His_6-tagged $AphA_{EP4}$, $AphB_{EP4}$ and $AphC_{RHA1}$ (See Table 1) C6369_RS01585, C6369_RS01550 and RHA1_RS18750 were amplified from genomic DNA using Phusion Polymerase™ with the oligonucleotides listed

164 in Supplementary Table 1. The nucleotide sequence of the cloned genes was verified. The *ΔaphA*
165 and *ΔaphC* mutants were constructed using a *sacB* counter selection system (van der Geize et al.,
166 2007). Five hundred bp flanking regions of RHA1_RS18785 and RHA1_RS18750 were
167 amplified from RHA1 genomic DNA using the primers listed in Supplementary Table 1. The
168 resulting amplicons were inserted into pK18mobsacB linearized with EcoR1 using Gibson
169 Assembly. The nucleotide sequences of the resulting constructs were verified. Kanamycin-
170 sensitive/sucrose-resistant colonies were screened using PCR and the gene deletion was
171 confirmed by sequencing.

172 ## Enzyme production and purification

173 The production and purification of $AphA_{EP4}$, $AphB_{EP4}$ and $AphC_{RHA1}$ are described in
174 Supplementary Methods.

175 ## GC/MS analysis

176 Growth substrate depletion was analyzed in culture supernatants using an Agilent Technologies
177 (Santa Clara, U.S.A.) 6890N gas chromatograph equipped with a 30-m Agilent 190915-433
178 capillary column and a 5973 mass-selective detector (GC/MS). Briefly, 400-µl samples of
179 culture supernatant were amended with 3-chlorobenzoic acid (as internal standard), extracted
180 with ethyl acetate (1:1 v/v) and dried under a nitrogen stream. The extract was suspended in
181 pyridine and derivatized with trimethylsilyl for 1 h at 60°C. One-µl injections were analyzed
182 using the following parameters: transfer line temperature of 325°C, run temperature of 90°C for
183 3 min, then ramped to 290°C at 12°C min^{-1} with a 10 min final hold. Peaks from raw trace files
184 were aligned and integrated using *xcms* in R 3.4.4 (R Core Team, 2016) against 4-ethylphenol

and succinate standards. Values were normalized to the area of the internal standard and expressed as a percent of maximum peak area.

Nucleic acid extraction and sequencing

RNA was extracted from cellular pellets from 15 ml of EP4 and RHA1 cultures using TRIzol™ (Thermo-Fisher) and Turbo™ DNase (Thermo-Fisher). Quality and quantity of nucleic acids were assessed using 1% [w/v] agarose gel electrophoresis and Qubit fluorometric quantitation (Thermo-Fisher), prior to storage at -80ºC. Approximately 1 µg RNA underwent Ribo-Zero rRNA removal (Illumina, San Diego, U.S.A.), TruSeq LT (Illumina) library preparation, and sequencing using HiSeq4000 2x100bp. Genomic DNA was extracted using CTAB. Fifteen µg was pulse-field electrophoresis size-selected and sequenced with one Pacific Biosciences (PacBio) RS II SMRT cell.

Bioinformatics

A *de novo* draft genome was assembled with HGAP in the SMRT Analysis v2.3 pipeline (Chin et al., 2013) and MeDuSa 1.6 scaffolding (Bosi et al., 2015). Circularizing the genome sequence was attempted using Circlator 1.5.5 (Hunt et al., 2015), and plasmid detection was attempted using PlasmidFinder 2.0.1 (Carattoli et al., 2014). Annotation used the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) 4.4 and BLASTp against the Protein Data Bank (e-value 10^{-3}). Quality control, filtering and trimming of RNA reads used Trimmomatic 0.3.6 defaults (Bolger et al., 2014). Assembly used Trinity 2.4.0 (Grabherr et al., 2011). Transcript quantification used HTSeq 0.9.1 (Anders et al., 2015), FeatureCounts 1.5.0-p3 (Liao et al., 2014) and Salmon 0.8.1 (Patro et al., 2017). Differentially-expression was analyzed using *DeSeq2* 1.18.1 (Love et al., 2014) with false discovery rate (*fdr*) correction. RT-qPCR conditions are in in

Supplementary Methods. All data visualization used *ggplot2* 3.1.0 unless otherwise noted. To download all reference data and re-create the transcriptomic analysis herein, refer to: https://github.com/levybooth/Rhodococcus_Transcriptomics.

Phylogenomic characterization

Protein sequences were structurally aligned with T-Coffee 11.00 Expresso (Armougom et al., 2006); maximum-likelihood trees were generated using the best of 100 RAxML 8.0.0 iterations using the PROTGAMMALG model (Stamatakis, 2014) and visualized with iTOL v4 (Letunic and Bork, 2016). Genomic island regions were predicted with IslandViewer 4 (Bertelli et al., 2017; Hsiao et al., 2003). All available *Rhodococcus* genomes (n = 325) were downloaded from NCBI RefSeq. Genomic alignment with a subset of genomes (Supplementary Table 2) used *nucmer* 3.1 (Marçais et al., 2018) and visualized with *circlize* 0.4.5 (Gu et al., 2014). Ribosomal protein trees were constructed as in Hug et al. (2016).

Enzyme assays

$AphAB_{EP4}$ activity was measured spectrophotometrically by following the *meta*-cleavage of the produced 4-ethylcatechol in a coupled assay with $AphC_{RHA1}$ at 25 ± 0.5 °C. The standard assay was performed in 200µl air- saturated 20 mM MOPS, 90 mM NaCl (I = 0.1 mM, pH 7.2) containing 0.5 mM 4-ethylphenol, 5 µM $AphA_{EP4}$, 1 µM $AphB_{EP4}$, 0.2 µM$AphC_{RHA1}$, 1000 U mL^{-1} of catalase, 1 mM NADH, and 2.5 µM FAD. Components were incubated for 30 s, then the reaction was initiated by adding NADH and was monitored at 400 nm. Absorbance was monitored using a Varian Cary 5000 spectrophotometer controlled by WinUV software. One unit of activity, U, was defined as the amount of enzyme required to hydroxylate of 1 nmol substrate per minute. Extinction coefficients for methyl-, ethyl-, and propylcatechol cleavage at 400 nm

were 18,600, 15,100, 19,400 M^{-1} cm^{-1}, respectively, calculated by differences in liberation of O_2 from alkylcatechol cleavage by 0.2 nmol $AphC_{RHA1}$ monitored using a Clark- type polarographic O_2 electrode OXYG1 (Hansatech, Pentney, UK) connected to a circulating water bath. Details of additional enzyme end point assays are provided in the Supplementary Methods.

Results

Isolation and genomic characterization of a 4-ethylphenol-degrading *Rhodococcus* strain

Enrichment cultures with 4-ethylphenol as a sole organic growth substrate were inoculated with either forest soil or compost and incubated at 30°C. Those cultures inoculated with compost demonstrated superior potential for 4-ethylphenol degradation and were used for subsequent isolation of strain EP4. The 16S rRNA gene (27F-1492R; Lane, 1991) of EP4 shared 100% sequence identity with that of *R. rhodochrous* NBRC 16069. *De novo* assembly produced a 5.72-Mb, high-quality, single-scaffold EP4 genome sequence (Figure 1A) containing 5,198 predicted genes: 4,942 protein coding sequences, 12 rRNAs, 54 tRNAs, three other RNAs, and 187 pseudogenes. Only a single origin of replication (*oriC*) was found, at 1,863,144 bp, and no plasmids were detected (Supplementary Table 2). The lack of PacBio long reads overlapping the 5' and 3' genome regions indicated that the EP4 genome is linear.

EP4 grew on 1.0 mM 4-ethylphenol to stationary phase within 14 hours in shake flasks (Figure 1B). Growth on 4-ethylphenol was verified by plating CFUs (Figure 1C). GC-MS analysis indicated that 4-ethylphenol was completely removed from the medium during growth (Figure 1D), with no metabolites detected.

Quasi-mapping-based quantification of prokaryotic gene expression

We used transcriptomics to identify 4-ethylphenol catabolic genes in EP4 without *a priori* bias. Transcriptome reads were aligned strand-wise to predict transcriptional start sites in the EP4 genome (Figure 2A). Read alignment is a common but time-consuming step in prokaryotic RNA-Seq pipelines (Supplementary Figure 1A). We therefore compared quasi-mapping to genomic coding regions using Salmon (Patro et al., 2017) with alignment-based read counting software. Salmon results were numerically (Supplementary Figure 1B) and statistically equivalent ($p_{adj} = 0.21$) (Supplementary Figure 1C) to FeatureCounts, with strong correlation to RT-qPCR expression ($R^2_{adj} = 0.91$, $p < 0.001$) (Supplementary Figure 1D). Salmon was about eight times faster than FeatureCounts, and superior to htseq in terms of total counts and accuracy, and was therefore used for gene quantification prior to differential expression analysis using *DESeq2*.

Transcriptomic analysis of 4-ethylphenol metabolism via *meta*-cleavage

Growth on 4-ethylphenol versus succinate significantly modulated expression of 559 genes with $p_{fdr} < 0.001$. Nine of the 16 most upregulated genes occurred in a cluster encoding a proposed alkylphenol catabolic pathway, *aphABCDEFGHIQRS* (Table 1). This cluster includes *aphAB*, encoding a two-component alkylphenol hydroxylase discussed below, and *aphC* (Figure 2B), encoding an extradiol dioxygenase that we subsequently identified as alkylcatechol 2,3-dioxygenase. The gene cluster is organized as four putative operons based on transcriptomic data and operon prediction with BPROM: *aphAB*, *aphHIDE*, *aphE2FGS*, *aphG2H2I2* (Figure 2A). The deduced Aph pathway catabolizes 4-ethylphenol to pyruvate and butyryl-CoA (Figure 2C), similar to the Dmp pathway of *Pseudomonas* sp. strain CF600 that catabolizes dimethylphenols (Shingler et al., 1992) and phenol (Powlowski and Shingler, 1994). Based on the transcriptomic

data, the resultant butyryl-CoA is degraded to central metabolites by an aerobic fatty acid degradation pathway (Jimenez-Diaz et al., 2017) encoded by butyryl-CoA dehydrogenase genes (locus tags: C6369_RS06395, C6369_RS20140, C6369_RS07820, C6369_RS05465), enoyl-CoA hydratase (C6369_RS03325, C6369_RS19860), 3−hydroxybutyryl−CoA dehydrogenase (C6369_RS03325, C6369_RS06400) and acetyl-CoA acyltransferase (C6369_RS17095, C6369_RS15900, C6369_RS19850) (Supplementary Figure 2).

The *catABC* cluster encoding catechol 1,2-dioxygenase and other enzymes feeding into the β-ketoadipate pathway was also significantly upregulated during growth on 4-ethylphenol), although much less highly than the *aph* genes. No *ortho*-cleavage metabolites were detected in the culture supernatants (Figures 1D), and the genes encoding the downstream β-ketoadipate pathway, *pcaBLIJ*, were not up-regulated (Figure 2B). Overall, the data suggest that 4-ethyphenol is catabolized via *meta*-cleavage.

Characterization of a two-component akylphenol hydroxylase, AphAB

We hypothesized that the highly upregulated *aphA* gene ($L_2FC = 11.5$) encodes the oxygenase component of a novel alkylphenol monooxygenase, based on its location within the *aph* cluster as well as the phylogenetic and functional data presented below. The *aphB* gene, encoding a flavin reductase was co-transcribed with *aphA* (Figure 2B). The upregulation of *aphA* and *aphC* genes in EP4 during growth on 4-ethylphenol was confirmed using RT-qPCR (Supplementary Figure 3).

To establish the physiological role of AphAB from EP4, the oxygenase and reductase components were each overproduced in *E. coli* and purified to apparent homogeneity. The reconstituted $AphAB_{EP4}$ hydroxylated 4-ethylphenol to 4-ethylcatechol (Figure 3A). The enzyme

294 also catalyzed the hydroxylation of 4-methylphenol, 4-propylphenol (Figure 3B) and, to a much
295 lesser extent, 4-NP (Figure 3C). However, AphAB $_{EP4}$ did not detectably transform phenol
296 (Figure 3B) or 4-HPA (Figure 3C). In an assay measuring cytochrome *c* reduction, AphB $_{EP4}$
297 preferentially utilized NADH and flavin adenine dinucleotide (FAD) (Figure 3D), as reported for
298 PheA2 (Gröning et al., 2014; Saa et al., 2010; Straube, 1987).

299 ## Annotation of additional genes

300 During growth of EP4 on 4-ethylphenol, phenol hydroxylase genes, *pheA* and *pheB,* adjacent to
301 the *cat* gene cluster, were additionally upregulated (Figure 2B, Supplementary Figure 3). In
302 structure-based alignments, PheA$_{EP4}$ and AphA$_{EP4}$ clustered with separate 4-NP hydroxylases,
303 rather than the clade of characterized phenol hydroxylases (Figure 3E). More specifically,
304 PheA$_{EP4}$ and AphA$_{EP4}$ clustered most closely, respectively, with NphA1 from *Rhodococcus* sp.
305 BUBNP1 (WP_059382681.1) (Sengupta et al., 2019) and NphA1 from *Rhodococcus* sp. PN1
306 (Q8RQQ0) (Takeo et al., 2003). Despite 100% sequence identity with NphA1$_{BUBNP1}$, PheA $_{EP4}$
307 was annotated based on sequence similarity to known phenol hydroxylases (Figure 3E). In
308 support of this annotation, EP4 lacks a 4-NP catabolism gene cluster and was unable to grow on
309 4-NP, while it did grow on phenol (discussed below). PheA $_{EP4}$ shares 82% identity with
310 PheA1(1) (ABS30825.1) in *Rhodococcus erythropolis* UPV-1 (Saa et al., 2010) and 65% identity
311 with a chlorophenol 4-monooxygenase (Q8GMG6) from *Streptomyces globisporus* (Liu et al.,
312 2002) (Supplementary Table 4; Supplementary Figure 4). These similarities suggest that PheA$_{EP4}$
313 may have broad substrate specificity.

314 In EP4, genes encoding AraC-family transcriptional regulators (TR) were found directly adjacent
315 to and in the opposite orientation as *aphAB* and *pheAB* (Figure 2A) (Supplementary Figure 5).
316 These AraC-family TRs were annotated as AphR and PheR, respectively. Another AraC-family

TR is encoded by a gene immediately downstream of *aphR*, which has a distinct phylogeny from AphR (Supplementary Figure 5) and was annotated as AphQ. Finally, an IclR-family TR is encoded by the last gene of the *aphE2FGS* operon.

Syntenic conservation of EP4 *aph* gene cluster in rhodococci

The above comparative analyses of hydroxylase proteins revealed homologs of the EP4 *aphA* gene in several other rhodococci. In RHA1, a putative *aphA* gene (Table 1) was previously annotated as an aromatic ring hydroxylase possibly involved in steroid degradation (McLeod et al., 2006) and there were three previously-identified *pheA* homologs (Supplementary Table 4) (Gröning et al., 2014). Local alignment of the 13 Aph proteins against proteins predicted from all 325 *Rhodococcus* genomes identified 75 strains with full or partial (≥7 genes) putative Aph pathways, including RHA1 (Supplementary Figure 5A). Related pathways were also found in other Actinobacteria, but this study focused on the rhodococcal pathway. RHA1 *aph* genes displayed syntenic conservation with the EP4 *aph* cluster (Figure 4), except for an additional butyryl-CoA dehydrogenase gene (RHA1_RS18815). The *aphCHIDE* region was conserved in all *aph*-containing genomes based on *nucmer* alignment (Figure 4, Supplementary Figure 5B).

The EP4 and RHA1 Aph proteins shared 66.7% to 90.4% identity (median = 85.7%). Consistent with the occurrence of the putative Aph pathway in RHA1, this strain grew on 4-ethylphenol (Figure 5A), with concomitant upregulation of the *aph* genes (Figure 5B). Specifically, *aphA* and *aphC*) were highly upregulated (L_2FC, 14.3 and 10.0, respectively). In contrast to EP4, when RHA1 grew on 4-ethylphenol, it did not upregulate any of its three *pheA* genes or any of the ring-cleavage dioxygenase genes associated with the *pheA* genes, including *catA* (RHA1_RS11595), *catA2* (RHA1_RS35920) and a plasmid-borne catechol 2,3-dioxygenase gene (RHA1_RS35970) (McLeod et al., 2006) (Figure 5B).

Gene deletion analysis of 4-alkylphenol ring cleavage

Because RHA1 is genetically-tractable, we constructed Δ*aphA* and Δ*aphC* deletion mutants and used these together with an existing Δ*pcaL* mutant to further investigate 4-ethylphenol catabolism. The Δ*aphC* mutant did not grow on either 1.0 mM 4-ethylphenol or 4-propylphenol (Figure 5AB) demonstrating that both compounds are exclusively metabolised by *meta*-cleavage. However, Δ*aphA* did grow on 4-ethylphenol (Figure 5A). It appears that one or more of three PheA homologs from RHA1 may catalyze 4-ethylphenol hydroxylation and compensate for the deletion of *aphA*. While the corresponding *pheA* genes were not upregulated in wild-type RHA1 growing on 4-ethylphenol versus succinate, they were upregulated 7.6 to 8.8 L_2FC in the Δ*aphA* mutant (Figure 5B), while the plasmid-borne *C23D* gene was not upregulated. Finally, the Δ*pcaL* mutant grew on alkylphenols but did not grow on either phenol or 4-HBA, indicating that the latter two substrates are catabolized solely via *ortho*-cleavage pathways (Figure 5C).

Identification of a putative *aph* genomic island

The *aph* gene cluster (approximately 17 kb) occurs within 117 kb and 4.2 kb regions predicted to be two of 61 GIs (or 38 non-overlapping GI regions) identified in EP4 using IslandViewer4 (Figure 4; Supplementary Table 2). These GI elements do not include *aphAB* or *aphE* in EP4, but do in three other *Rhodococcus* strains. These putative GIs are located near the 3' end of the EP4 chromosome assembly in a 600-kb region of apparent genomic instability, as it contains high insertion sequence (IS) density (Figure 1A) and 36 predicted GIs (Figure 4). The GI-like characteristics of these elements containing the *aph* cluster include a -5.8% deviation from the mean GC content, presence of mobility genes (integrases, transposases, insertion sequences) and absence of the region in closely-related genomes following alignment (Langille et al., 2008) (Figure 4; Supplementary Figure 6). They are not located near a tRNA sequence, indicating that

it is not likely a mobile integrative and conjugative element (ICE) (Burrus and Waldor, 2004). More generally, 7.4% of the EP4 genome and 9.2% of the RHA1 genome were predicted to occur on GIs (Supplementary Table 2).

Analysis of 37 complete, full-length *Rhodococcus* genomes found that 16 carried genes encoding a complete Aph pathway. The *aph* genes are predicted to be fully or partially contained in a GI in six of these strains and to occur immediately downstream of a GI in a seventh, RHA1. This genomic region was conserved in three *Rhodococcus* clades: one containing EP4 and *R. pyridinivorans* strains, one containing *R. jostii*, *R. opacus* and *Rhodococcus wratislaviensis* strains and one containing *Rhodococcus* sp. Eu32 (Supplementary Figure 6). With the exception of a partial *aph* cluster in *R. rhodochrous* ATCC 21198, the *aph* genes were not found in any of the 13 other *R. rhodochrous* genomes including strain DSM43241. Accordingly, DSM43241 could not grow on 4-ethylphenol and 4-propylphenol, but grew on phenol, 4-HBA, 3-methylphenol and 4-methylphenol (Figure 5C).

Discussion

In this study, we used a variety of approaches to identify an Aph pathway responsible for the catabolism of alkylphenols via *meta*-cleavage in *Rhodococcus*. Catabolism is initiated by AphAB, a two-component hydroxylase that transforms the alkylphenol to the corresponding alkylcatechol (Figures 2,3). To date, only six-component proteobacterial alkylphenol hydroxylases have been reported (Arai et al., 2000; Jeong et al., 2003). The ensuing Aph pathway generates pyruvate and an acyl-CoA following *meta*-cleavage of 4-alkylcatechol. The length of the acyl-CoA produced depends on the alkyl side chain of the growth substrate. This is in

384 contrast to the Phe and Nph pathways which catabolize phenol and 4-NP, respectively, via *ortho*-
385 cleavage (Sengupta et al., 2019; Szőköl et al., 2014; Takeo et al., 2008) .

386 The activity of $AphAB_{EP4}$ is consistent with its phylogenetic relationship with two-component
387 phenolic hydroxylases. Thus, the clade containing $AphA_{EP4}$ and $AphA_{RHA1}$ includes an NphA1
388 but no characterized PheA or HpaB (Figure 3). In keeping with this classification, $AphAB_{EP4}$ had
389 weak activity with 4-NP (Figure 3C) but did not detectably transform phenol or 4-HPA.
390 However, the determinants of substrate specificity of these enzymes are not clear. The catalytic
391 residues of these hydroxylases (Chang et al., 2016; Kim et al., 2007) are conserved in AphA:
392 Arg119, Tyr123 and His161 ($AphA_{EP4}$ numbering; Supplementary Figure 4). In a structurally
393 characterized HpaB:4-HPA binary complex, the substrate's carboxylate is coordinated by Ser197
394 and Thr198 (Kim et al., 2007). In PheA, NphA and AphA, these residues are His214 and Tyr215,
395 suggesting that they do not contribute to the enzyme's substrate specificity despite their
396 predicted interaction with the *para*-substituent of the substrate.

397 The Aph pathway is similar to the Dmp pathway described in *Pseudomonas* sp. strain CF600
398 (Shingler et al., 1992). However, it is clear that the Aph pathway has a distinct substrate
399 specificity because neither EP4 nor RHA1 grew on 2,4- or 3,4-DMP (Figure 4E) and *aph*
400 pathway mutants grew on phenol. We had previously suggested that some of the *aph* genes could
401 be involved in steroid degradation (McLeod et al., 2006) due to their similarity to known steroid
402 catabolic genes (Van der Geize et al., 2007). Further, in a recently published structure of
403 $AphC_{RHA1}$, the enzyme was identified as 2,3-dihydroxybiphenyl dioxygenase (Table 1).
404 However, AphC is encoded in a gene cluster upregulated on 4-alkylphenols and is essential for
405 growth of RHA1 on those compounds, supporting annotation of this rhodococcal Aph *meta*-
406 cleavage pathway, with AphC as an alkylcatechol 2,3-dioxygenase.

4-Ethylphenol strongly induced *aphAB* expression. This is likely due to positive induction of the AphR TR, just as phenol activates *pheA2A1* expression by PheR in RHA1 (Szőköl et al., 2014), and 4-NP activates *npaA2A1* expression by NphR in *Rhodococcus* sp. PN1 (Takeo et al., 2008). AphR, PheR and NphR are all AraC-family TRs. AphR and PheR may play a role in the unexpected ability of the RHA1 Δ*aphA* mutant to grow on 4-ethylphenol. The lack of *pheA1-3* expression in wild-type RHA1 (Figure 5B) strikingly contrasts with the upregulation of these genes in the Δ*aphA* mutant (Figure 5C). Binding regions for $PheR_{RHA1}$ and predicted $AphR_{RHA1}$ binding sites overlap by 10/18 nt (Supplementary Figure 5B), suggesting the potential for non-target regulation, as Szőköl et al. (2014) argue for *R. erythropolis* CSM2595 PheR and XylS. This overlap is much less in predicted $PheR_{EP4}$ and $AphR_{EP4}$ binding regions (3/18 nt). Indeed, the *aphA* promotor (-10: CAGGAG; -35: CCGTCT) (Supplementary Figure 5C) bears more similarity to the T80 promotor of *Mycobacterium tuberculosis* (Bashyam et al., 1996) than with the rhodococcal *pheA* promotors. It is also noted that two of the *pheA* genes in RHA1 are subject to catabolite repression (Szőköl et al., 2014). It is possible that this repression is relieved in the Δ*aphA* mutant that can't metabolise 4-ethylphenol. Related to this, homologs of PheAB in *R. opacus* 1CP hydroxylated 4-methylphenol with about twice the specific activity as with phenol (Gröning et al., 2014), further suggesting that the RHA1 PheABs may hydroxylate 4-ethylphenol.

In addition to 4-ethylphenol, alkylguaiacols and alkylsyringols commonly occur in lignin depolymerization streams (Anderson et al., 2016; Asawaworarit et al., 2019; Guo et al., 2017; Jiang et al., 2014; Kim et al., 2015; Ye et al., 2012). Interestingly, genes encoding a cytochrome P450 and reductase are linked to the *aph* clusters in some rhodococci (Figures 2,4). Further, these were the second and third most highly upregulated genes in EP4 during growth on 4-

ethylphenol versus succinate (both $L_2FC = 11.1$) (Figure 2B). The P450 shares 65% sequence identity with a guaiacol *O*-demethylase (Mallinson et al., 2018), suggesting that the rhodococcal enzyme has a similar role, and that these strains may also funnel methoxylated compounds into the Aph pathway.

The ability of RHA1 and EP4 to catabolize 4-ethylphenol and other alkylphenols is of potential use in upgrading lignin streams generated by RCF and other depolymerization strategies. The Aph *meta*-cleavage pathway harboured by these strains contrasts with the *ortho*-cleavage pathways targeted to date in the design of biocatalysts for lignin valorization (Abdelaziz et al., 2016; Barton et al., 2018; Beckham et al., 2016). This is largely due to the identified economic potential of some of the *ortho*-cleavage metabolites. For example, *cis*,*cis*-muconate resulting from *ortho*-cleavage can be used to make adipic acid and terephthalic acid (Barton et al., 2018; Beckham et al., 2016; Xie et al., 2014). However, alkylphenols may be funneled through *ortho*-cleavage by oxidizing the *para*-side chain to 4-hydroxyacetophenone or hydroquinone (Darby et al., 1987). Alternatively, oleaginous *Rhodococcus* strains such as RHA1 may be modified to use the Aph pathway to produce lipid-based commodity chemicals (e.g., Round et al., 2017) offering a method for valorization of alkylphenols via fatty-acid synthesis in this genus.

We found genes encoding the Aph pathway in several *Rhodococcus* strains, including oleaginous strains such as RHA1 and *R. opacus* B4 (Figure 3A, Figure 4). However, the absence of an *aph* cluster in most *R. rhodochrous* strains (e.g., DSM43241) demonstrates that phylogenetically-related strains can have important metabolic differences. Previous studies suggested that recent horizontal gene transfer did not play a large role in generating RHA1's considerable catabolic capabilities (McLeod et al., 2006). In several rhodococci, putative GIs did not contain all of the

aph genes. This could represent the imprecision of the prediction tools, incomplete amelioration of the element, or reflect that these GIs arose from separate insertion events. Patchwork *aph* GIs are consistent with the theorized modular origins of GIs (Juhas et al., 2009). Fermentation of plant-derived aromatic compounds, including cinnamic acids, by yeasts and lactic acid bacteria can naturally produce alkylphenols (Caboni et al., 2007; Kridelbaugh et al., 2010). The apparent complete loss of the *aph* genes in other *R. rhodochrous* strains may result without selective alkylphenol exposure if it otherwise has a deleterious effect on overall fitness. Testing these regions for their excision capacity was beyond the scope of this work, but remains an intriguing prospect. We posit that the presence of aromatic compounds in compost selected for microorganisms capable of 4-ethylphenol catabolism, including EP4.

In this study we described a newly-isolated, 4-ethylphenol-catabolizing strain, EP4, a novel alkylphenol hydroxlase, AphAB, and its role in funneling alkylphenols into the Aph *meta*-cleavage pathway in some *Rhodococcus* strains. We showed that this pathway is associated with putative GIs, primarily found in strains from contaminated soil environments. Characterizing 4-ethylphenol metabolism in EP4 and RHA1 advances our capacity for bio-refinement of reductively depolymerized lignin subunits from sustainable chemical feedstocks.

Acknowledgements

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). All sequencing was performed at the McGill University and Génome Québec Innovation Centre (Montreal, CAN). We thank Andrew Wilson and Alexandra Booth for their assistance with laboratory experiments.

References

Abdelaziz, O.Y., Brink, D.P., Prothmann, J., Ravi, K., Sun, M., García-Hidalgo, J., Sandahl, M., Hulteberg, C.P., Turner, C., Lidén, G., Gorwa-Grauslund, M.F., 2016. Biological valorization of low molecular weight lignin. Biotechnol. Adv. 34, 1318–1346. https://doi.org/10.1016/j.biotechadv.2016.10.001

Alvarez, H.M., Mayer, F., Fabritius, D., Steinbüchel, A., 1996. Formation of intracytoplasmic lipid inclusions by *Rhodococcus opacus* strain PD630. Arch. Microbiol. 165, 377–386.

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Anderson, E.M., Katahira, R., Reed, M., Resch, M.G., Karp, E.M., Beckham, G.T., Román-Leshkov, Y., 2016. Reductive catalytic fractionation of corn stover lignin. ACS Sustainable Chem. Eng. 4, 6940–6950. https://doi.org/10.1021/acssuschemeng.6b01858

Arai, H., Ohishi, T., Chang, M.Y., Kudo, T., 2000. Arrangement and regulation of the genes for *meta*-pathway enzymes required for degradation of phenol in *Comamonas testosteroni* TA441. Microbiology (Reading, Engl.) 146 (Pt 7), 1707–1715. https://doi.org/10.1099/00221287-146-7-1707

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., Notredame, C., 2006. Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucleic Acids Res 34, W604–W608. https://doi.org/10.1093/nar/gkl092

496 Asawaworarit, P., Daorattanachai, P., Laosiripojana, W., Sakdaronnarong, C., Shotipruk, A.,
497 Laosiripojana, N., 2019. Catalytic depolymerization of organosolv lignin from bagasse by
498 carbonaceous solid acids derived from hydrothermal of lignocellulosic compounds.
499 Chemical Engineering Journal 356, 461–471. https://doi.org/10.1016/j.cej.2018.09.048
500 Barton, N., Horbal, L., Starck, S., Kohlstedt, M., Luzhetskyy, A., Wittmann, C., 2018. Enabling
501 the valorization of guaiacol-based lignin: Integrated chemical and biochemical
502 production of *cis,cis*-muconic acid using metabolically engineered *Amycolatopsis* sp.
503 ATCC 39116. Metabolic Engineering 45, 200–210.
504 https://doi.org/10.1016/j.ymben.2017.12.001
505 Bashyam, M.D., Kaushal, D., Dasgupta, S.K., Tyagi, A.K., 1996. A study of mycobacterial
506 transcriptional apparatus: identification of novel features in promoter elements. Journal of
507 Bacteriology 178, 4847–4853. https://doi.org/10.1128/jb.178.16.4847-4853.1996
508 Bauchop, T., Elsden, S.R., 1960. The growth of micro-organisms in relation to their energy
509 supply. Microbiology 23, 457–469. https://doi.org/10.1099/00221287-23-3-457
510 Beckham, G.T., Johnson, C.W., Karp, E.M., Salvachúa, D., Vardon, D.R., 2016. Opportunities
511 and challenges in biological lignin valorization. Current Opinion in Biotechnology,
512 Chemical biotechnology • Pharmaceutical biotechnology 42, 40–53.
513 https://doi.org/10.1016/j.copbio.2016.02.030
514 Bertelli, C., Laird, M.R., Williams, K.P., Lau, B.Y., Hoad, G., Winsor, G.L., Brinkman, F.S.,
515 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets.
516 Nucleic Acids Res 45, W30–W35. https://doi.org/10.1093/nar/gkx343
517 Boerjan, W., Ralph, J., Baucher, M., 2003. Lignin biosynthesis. Annu Rev Plant Biol 54, 519–
518 546. https://doi.org/10.1146/annurev.arplant.54.031902.134938

519 Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina
520 sequence data. Bioinformatics 30, 2114–2120.
521 https://doi.org/10.1093/bioinformatics/btu170
522 Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.-F., Lió, P., Crescenzi, P., Fani, R.,
523 Fondi, M., 2015. MeDuSa: a multi-draft based scaffolder. Bioinformatics 31, 2443–2451.
524 https://doi.org/10.1093/bioinformatics/btv171
525 Burrus, V., Waldor, M.K., 2004. Shaping bacterial genomes with integrative and conjugative
526 elements. Res. Microbiol. 155, 376–386. https://doi.org/10.1016/j.resmic.2004.01.012
527 Caboni, P., Sarais, G., Cabras, M., Angioni, A., 2007. Determination of 4-ethylphenol and 4-
528 ethylguaiacol in wines by LC-MS-MS and HPLC-DAD-fluorescence. J. Agric. Food
529 Chem. 55, 7288–7293. https://doi.org/10.1021/jf071156m
530 Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., Møller
531 Aarestrup, F., Hasman, H., 2014. *In silico* detection and typing of plasmids using
532 PlasmidFinder and plasmid multilocus sequence typing. Antimicrob. Agents Chemother.
533 58, 3895–3903. https://doi.org/10.1128/AAC.02412-14
534 Carere, J., McKenna, S.E., Kimber, M.S., Seah, S.Y.K., 2013. Characterization of an aldolase-
535 dehydrogenase complex from the cholesterol degradation pathway of *Mycobacterium*
536 *tuberculosis*. Biochemistry 52, 3502–3511. https://doi.org/10.1021/bi400351h
537 Chang, C.-Y., Lohman, J.R., Cao, H., Tan, K., Rudolf, J.D., Ma, M., Xu, W., Bingman, C.A.,
538 Yennamalli, R.M., Bigelow, L., Babnigg, G., Yan, X., Joachimiak, A., Phillips, G.N.,
539 Shen, B., 2016. Crystal structures of SgcE6 and SgcC, the two-component
540 monooxygenase that catalyzes hydroxylation of a carrier protein-tethered substrate during

the biosynthesis of the enediyne antitumor antibiotic C-1027 in *Streptomyces globisporus*. Biochemistry 55, 5142–5154. https://doi.org/10.1021/acs.biochem.6b00713

Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W., Korlach, J., 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods 10, 563–569. https://doi.org/10.1038/nmeth.2474

DARBY, J.M., TAYLOR, D.G., HOPPER, D.J., 1987. hydroquinone as the ring-fission substrate in the catabolism of 4-ethylphenol and 4-hydroxyacetophenone by *Pseudomonas putida* JD1. Microbiology 133, 2137–2146. https://doi.org/10.1099/00221287-133-8-2137

Dobrindt, U., Hochhut, B., Hentschel, U., Hacker, J., 2004. Genomic islands in pathogenic and environmental microorganisms. Nature Reviews Microbiology 2, 414–424. https://doi.org/10.1038/nrmicro884

Elder, R.T., 1983. Cloning Techniques. BioScience 33, 721–722. https://doi.org/10.2307/1309366

Eltis, L.D., Singh, R., 2018. Chapter 11:Biological funneling as a means of transforming lignin-derived aromatic compounds into value-added chemicals, in: Lignin Valorization. pp. 290–313. https://doi.org/10.1039/9781788010351-00290

Gaillard, M., Vallaeys, T., Vorhölter, F.J., Minoia, M., Werlen, C., Sentchilo, V., Pühler, A., Meer, J.R. van der, 2006. The *clc* element of *Pseudomonas* sp. Strain B13, a genomic island with various catabolic properties. Journal of Bacteriology 188, 1999–2013. https://doi.org/10.1128/JB.188.5.1999-2013.2006

Geize, R.V. der, Yam, K., Heuser, T., Wilbrink, M.H., Hara, H., Anderton, M.C., Sim, E., Dijkhuizen, L., Davies, J.E., Mohn, W.W., Eltis, L.D., 2007. A gene cluster encoding

564 cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium*
565 *tuberculosis* survival in macrophages. PNAS 104, 1947–1952.
566 https://doi.org/10.1073/pnas.0605728104
567 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X.,
568 Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A.,
569 Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N.,
570 Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a
571 reference genome. Nat Biotech 29, 644–652. https://doi.org/10.1038/nbt.1883
572 Gröning, J.A.D., Eulberg, D., Tischler, D., Kaschabek, S.R., Schlömann, M., 2014. Gene
573 redundancy of two-component (chloro)phenol hydroxylases in *Rhodococcus opacus* 1CP.
574 FEMS Microbiol Lett 361, 68–75. https://doi.org/10.1111/1574-6968.12616
575 Gu, Z., Gu, L., Eils, R., Schlesner, M., Brors, B., 2014. *circlize* Implements and enhances
576 circular visualization in R. Bioinformatics 30, 2811–2812.
577 https://doi.org/10.1093/bioinformatics/btu393
578 Guo, D., Wu, S., Lyu, G., Guo, H., 2017. Effect of molecular weight on the pyrolysis
579 characteristics of alkali lignin. Fuel 193, 45–53.
580 https://doi.org/10.1016/j.fuel.2016.12.042
581 Hacker, J., Bender, L., Ott, M., Wingender, J., Lund, B., Marre, R., Goebel, W., 1990. Deletions
582 of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in
583 various extra intestinal *Escherichia coli* isolates. Microbial Pathogenesis 8, 213–225.
584 https://doi.org/10.1016/0882-4010(90)90048-U
585 Hacker, J., Kaper, J.B., 2000. Pathogenicity islands and the evolution of microbes. Annual
586 Review of Microbiology 54, 641–679. https://doi.org/10.1146/annurev.micro.54.1.641

587 Hopper, D.J., Cottrell, L., 2003. Alkylphenol biotransformations catalyzed by 4-ethylphenol
588 methylenehydroxylase. Appl. Environ. Microbiol. 69, 3650–3652.
589 https://doi.org/10.1128/AEM.69.6.3650-3652.2003
590 Hsiao, W., Wan, I., Jones, S.J., Brinkman, F.S.L., 2003. IslandPath: aiding detection of genomic
591 islands in prokaryotes. Bioinformatics 19, 418–420.
592 https://doi.org/10.1093/bioinformatics/btg004
593 Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield,
594 C.N., Hernsdorf, A.W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D.A.,
595 Finstad, K.M., Amundson, R., Thomas, B.C., Banfield, J.F., 2016. A new view of the tree
596 of life. Nature Microbiology 1, nmicrobiol201648.
597 https://doi.org/10.1038/nmicrobiol.2016.48
598 Hunt, M., Silva, N.D., Otto, T.D., Parkhill, J., Keane, J.A., Harris, S.R., 2015. Circlator:
599 automated circularization of genome assemblies using long sequencing reads. Genome
600 Biology 16, 294. https://doi.org/10.1186/s13059-015-0849-0
601 Jang, J.Y., Kim, D., Bae, H.W., Choi, K.Y., Chae, J.-C., Zylstra, G.J., Kim, Y.M., Kim, E., 2005.
602 Isolation and characterization of a rhodococcus species strain able to grow on *ortho*- and
603 *para*-xylene. J. Microbiol. 43, 325–330.
604 Jeong, J.J., Kim, J.H., Kim, C.-K., Hwang, I., Lee, K., 2003. 3- and 4-alkylphenol degradation
605 pathway in *Pseudomonas* sp. strain KL28: genetic organization of the lap gene cluster
606 and substrate specificities of phenol hydroxylase and catechol 2,3-dioxygenase.
607 Microbiology 149, 3265–3277. https://doi.org/10.1099/mic.0.26628-0

608 Jiang, Z., He, T., Li, J., Hu, C., 2014. Selective conversion of lignin in corncob residue to
609 monophenols with high yield and selectivity. Green Chemistry 16, 4257–4265.
610 https://doi.org/10.1039/C4GC00620H
611 Jimenez-Diaz, L., Caballero, A., Segura, A., 2017. Pathways for the degradation of fatty acids in
612 bacteria, in: Rojo, F. (Ed.), Aerobic Utilization of Hydrocarbons, Oils and Lipids,
613 Handbook of Hydrocarbon and Lipid Microbiology. Springer International Publishing,
614 Cham, pp. 1–23. https://doi.org/10.1007/978-3-319-39782-5_42-1
615 Juhas, M., van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W., Crook, D.W., 2009.
616 Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS
617 Microbiol Rev 33, 376–393. https://doi.org/10.1111/j.1574-6976.2008.00136.x
618 Kim, J.-Y., Park, J., Kim, U.-J., Choi, J.W., 2015. Conversion of lignin to phenol-rich oil
619 fraction under supercritical alcohols in the presence of metal catalysts. Energy Fuels 29,
620 5154–5163. https://doi.org/10.1021/acs.energyfuels.5b01055
621 Kim, S.-H., Hisano, T., Takeda, K., Iwasaki, W., Ebihara, A., Miki, K., 2007. Crystal structure
622 of the oxygenase component (HpaB) of the 4-hydroxyphenylacetate 3-monooxygenase
623 from *Thermus thermophilus* HB8. J. Biol. Chem. 282, 33107–33117.
624 https://doi.org/10.1074/jbc.M703440200
625 Kolomytseva, M.P., Baskunov, B.P., Golovleva, L.A., 2007. Intradiol pathway of *para*-cresol
626 conversion by *Rhodococcus opacus* 1CP. Biotechnology Journal 2, 886–893.
627 https://doi.org/10.1002/biot.200700013
628 Kridelbaugh, D., Hughes, S., Allen, T., Doerner, K.C., 2010. Production of 4-ethylphenol from
629 4-hydroxycinnamic acid by *Lactobacillus* sp. isolated from a swine waste lagoon. Journal

630 of Applied Microbiology 109, 190–198. https://doi.org/10.1111/j.1365-
631 2672.2009.04642.x
632 Langille, M.G., Hsiao, W.W., Brinkman, F.S., 2008. Evaluation of genomic island predictors
633 using a comparative genomics approach. BMC Bioinformatics 9, 329.
634 https://doi.org/10.1186/1471-2105-9-329
635 Lawrence, J.G., Ochman, H., 1997. Amelioration of bacterial genomes: rates of change and
636 exchange. J. Mol. Evol. 44, 383–397.
637 Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and
638 annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242-245.
639 https://doi.org/10.1093/nar/gkw290
640 Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for
641 assigning sequence reads to genomic features. Bioinformatics 30, 923–930.
642 https://doi.org/10.1093/bioinformatics/btt656
643 Linger, J.G., Vardon, D.R., Guarnieri, M.T., Karp, E.M., Hunsinger, G.B., Franden, M.A.,
644 Johnson, C.W., Chupka, G., Strathmann, T.J., Pienkos, P.T., Beckham, G.T., 2014.
645 Lignin valorization through integrated biological funneling and chemical catalysis. PNAS
646 111, 12013–12018. https://doi.org/10.1073/pnas.1410657111
647 Liu, W., Christenson, S.D., Standage, S., Shen, B., 2002. Biosynthesis of the enediyne antitumor
648 antibiotic C-1027. Science 297, 1170–1173. https://doi.org/10.1126/science.1072110
649 Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for
650 RNA-seq data with DESeq2. Genome Biology 15, 550. https://doi.org/10.1186/s13059-
651 014-0550-8

652 Mallinson, S.J.B., Machovina, M.M., Silveira, R.L., Garcia-Borràs, M., Gallup, N., Johnson,
653 C.W., Allen, M.D., Skaf, M.S., Crowley, M.F., Neidle, E.L., Houk, K.N., Beckham,
654 G.T., DuBois, J.L., McGeehan, J.E., 2018. A promiscuous cytochrome P450 aromatic *O*-
655 demethylase for lignin bioconversion. Nature Communications 9, 2487.
656 https://doi.org/10.1038/s41467-018-04878-2
657 Maltseva, O.V., Solyanikova, I.P., Golovleva, L.A., 1994. Chlorocatechol 1,2-dioxygenase from
658 *Rhodococcus erythropolis* 1CP. Kinetic and immunochemical comparison with
659 analogous enzymes from gram-negative strains. Eur. J. Biochem. 226, 1053–1061.
660 Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., Zimin, A., 2018.
661 MUMmer4: A fast and versatile genome alignment system. PLOS Computational
662 Biology 14, e1005944. https://doi.org/10.1371/journal.pcbi.1005944
663 McLeod, M.P., Warren, R.L., Hsiao, W.W.L., Araki, N., Myhre, M., Fernandes, C., Miyazawa,
664 D., Wong, W., Lillquist, A.L., Wang, D., Dosanjh, M., Hara, H., Petrescu, A., Morin,
665 R.D., Yang, G., Stott, J.M., Schein, J.E., Shin, H., Smailus, D., Siddiqui, A.S., Marra,
666 M.A., Jones, S.J.M., Holt, R., Brinkman, F.S.L., Miyauchi, K., Fukuda, M., Davies, J.E.,
667 Mohn, W.W., Eltis, L.D., 2006. The complete genome of *Rhodococcus* sp. RHA1
668 provides insights into a catabolic powerhouse. Proc. Natl. Acad. Sci. U.S.A. 103, 15582–
669 15587. https://doi.org/10.1073/pnas.0607048103
670 Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V. a. P.,
671 Fouts, D.E., Gill, S.R., Pop, M., Holmes, M., Brinkac, L., Beanan, M., DeBoy, R.T.,
672 Daugherty, S., Kolonay, J., Madupu, R., Nelson, W., White, O., Peterson, J., Khouri, H.,
673 Hance, I., Chris Lee, P., Holtzapple, E., Scanlan, D., Tran, K., Moazzez, A., Utterback,
674 T., Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J., Stjepandic, D.,

675 Hoheisel, J., Straetz, M., Heim, S., Kiewitz, C., Eisen, J.A., Timmis, K.N., Düsterhöft,
676 A., Tümmler, B., Fraser, C.M., 2002. Complete genome sequence and comparative
677 analysis of the metabolically versatile *Pseudomonas putida* KT2440. Environ. Microbiol.
678 4, 799–808.
679 Nordlund, I., Shingler, V., 1990. Nucleotide sequences of the meta-cleavage pathway enzymes
680 2-hydroxymuconic semialdehyde dehydrogenase and 2-hydroxymuconic semialdehyde
681 hydrolase from *Pseudomonas* CF600. Biochim. Biophys. Acta 1049, 227–230.
682 Patrauchan, M.A., Florizone, C., Dosanjh, M., Mohn, W.W., Davies, J., Eltis, L.D., 2005.
683 Catabolism of benzoate and phthalate in *Rhodococcus* sp. strain RHA1: redundancies and
684 convergence. Journal of Bacteriology 187, 4050–4063.
685 https://doi.org/10.1128/JB.187.12.4050-4063.2005
686 Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and
687 bias-aware quantification of transcript expression. Nature Methods 14, 417–419.
688 https://doi.org/10.1038/nmeth.4197
689 Pepper, J.M., Lee, Y.W., 1969. Lignin and related compounds. I. A comparative study of
690 catalysts for lignin hydrogenolysis. Can. J. Chem. 47, 723–727.
691 https://doi.org/10.1139/v69-118
692 Pollard, J.R., Bugg, T.D., 1998. Purification, characterisation and reaction mechanism of
693 monofunctional 2-hydroxypentadienoic acid hydratase from *Escherichia coli*. Eur. J.
694 Biochem. 251, 98–106.
695 Pouyssegur, J., Stoeber, F., 1974. Genetic control of the 2-keto-3-deoxy-d-gluconate metabolism
696 in *Escherichia coli* K-12: *kdg* regulon. J. Bacteriol. 117, 641–651.

697 Powlowski, J., Shingler, V., 1994. Genetics and biochemistry of phenol degradation by
698 Pseudomonas sp. CF600. Biodegradation 5, 219–236.
699 Ragauskas, A.J., Beckham, G.T., Biddy, M.J., Chandra, R., Chen, F., Davis, M.F., Davison,
700 B.H., Dixon, R.A., Gilna, P., Keller, M., Langan, P., Naskar, A.K., Saddler, J.N.,
701 Tschaplinski, T.J., Tuskan, G.A., Wyman, C.E., 2014. Lignin valorization: improving
702 lignin processing in the biorefinery. Science 344, 1246843.
703 https://doi.org/10.1126/science.1246843
704 Round, J., Roccor, R., Li, S.-N., Eltis, L.D., 2017. A fatty acyl coenzyme a reductase promotes
705 wax ester accumulation in *Rhodococcus jostii* RHA1. Appl. Environ. Microbiol. 83,
706 e00902-17. https://doi.org/10.1128/AEM.00902-17
707 Saa, L., Jaureguibeitia, A., Largo, E., Llama, M.J., Serra, J.L., 2010. Cloning, purification and
708 characterization of two components of phenol hydroxylase from *Rhodococcus*
709 *erythropolis* UPV-1. Appl Microbiol Biotechnol 86, 201–211.
710 https://doi.org/10.1007/s00253-009-2251-x
711 Sambrock, J., W. Russel, D., 2001. Molecular Cloning: A Laboratory Manual(3rd edition).
712 Sengupta, K., Swain, M.T., Livingstone, P.G., Whitworth, D.E., Saha, P., 2019. Genome
713 sequencing and comparative transcriptomics provide a holistic view of 4-nitrophenol
714 degradation and concurrent fatty acid catabolism by *Rhodococcus* sp. Strain BUPNP1.
715 Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.03209
716 Shields-Menard, S.A., AmirSadeghi, M., Green, M., Womack, E., Sparks, D.L., Blake, J.,
717 Edelmann, M., Ding, X., Sukhbaatar, B., Hernandez, R., Donaldson, J.R., French, T.,
718 2017. The effects of model aromatic lignin compounds on growth and lipid accumulation

719 of *Rhodococcus rhodochrous*. International Biodeterioration & Biodegradation 121, 79–
720 90. https://doi.org/10.1016/j.ibiod.2017.03.023
721 Shingler, V., Powlowski, J., Marklund, U., 1992. Nucleotide sequence and functional analysis of
722 the complete phenol/3,4-dimethylphenol catabolic pathway of *Pseudomonas* sp. strain
723 CF600. J. Bacteriol. 174, 711–724.
724 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
725 large phylogenies. Bioinformatics 30, 1312–1313.
726 https://doi.org/10.1093/bioinformatics/btu033
727 Straube, G., 1987. Phenol hydroxylase from *Rhodococcus* sp. P 1. Journal of Basic Microbiology
728 27, 229–232. https://doi.org/10.1002/jobm.3620270415
729 Szőköl, J., Rucká, L., Šimčíková, M., Halada, P., Nešvera, J., Pátek, M., 2014. Induction and
730 carbon catabolite repression of phenol degradation genes in *Rhodococcus erythropolis*
731 and *Rhodococcus jostii*. Appl Microbiol Biotechnol 98, 8267–8279.
732 https://doi.org/10.1007/s00253-014-5881-6
733 Takeo, M., Murakami, M., Niihara, S., Yamamoto, K., Nishimura, M., Kato, D., Negoro, S.,
734 2008. Mechanism of 4-nitrophenol oxidation in *Rhodococcus* sp. strain PN1:
735 characterization of the two-component 4-nitrophenol hydroxylase and regulation of its
736 expression. J. Bacteriol. 190, 7367–7374. https://doi.org/10.1128/JB.00742-08
737 Takeo, M., Yasukawa, T., Abe, Y., Niihara, S., Maeda, Y., Negoro, S., 2003. Cloning and
738 characterization of a 4-nitrophenol hydroxylase gene cluster from *Rhodococcus* sp. PN1.
739 J. Biosci. Bioeng. 95, 139–145.
740 Tsuda, M., Iino, T., 1990. Naphthalene degrading genes on plasmid NAH7 are on a defective
741 transposon. Mol. Gen. Genet. 223, 33–39.

742 Xie, N.-Z., Liang, H., Huang, R.-B., Xu, P., 2014. Biotechnological production of muconic acid:
743 current status and future prospects. Biotechnology Advances 32, 615–622.
744 https://doi.org/10.1016/j.biotechadv.2014.04.001
745 Yam, K.C., van der Geize, R., Eltis, L.D., 2010. Catabolism of aromatic compounds and steroids
746 by Rhodococcus, in: Alvarez, H.M. (Ed.), Biology of Rhodococcus, Microbiology
747 Monographs. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 133–169.
748 https://doi.org/10.1007/978-3-642-12937-7_6
749 Ye, Y., Zhang, Y., Fan, J., Chang, J., 2012. Novel method for production of phenolics by
750 combining lignin extraction with lignin depolymerization in aqueous ethanol. Ind. Eng.
751 Chem. Res. 51, 103–110. https://doi.org/10.1021/ie202118d
752

Tables

Table 1. Genes in the alkylphenol *meta*-cleavage pathway

Gene	EP4[a]	RHA1[b]	Product[c]	Best Hit[d]	% ID[e]	Ref.
aphA	RS01585	RS18785	Alkylphenol hydroxylase, oxygenase	4-Nitrophenol 2-monooxygenase, oxygenase (NphA1) Q8RQQ0	87	Takeo et al. (2003)
aphB	RS01580	RS18780	Alkylphenol hydroxylase, reductase	NADH-dependent flavin reductase (NphB1) Q8RQP9	81	Takeo et al. (2003)
aphC	RS01550	RS18750	Alkylcatechol 2,3-dioxygenase	Biphenyl-2,3-diol 1,2-dioxygenase (BphC)[f] Q0S9X1	87	PDB entry, unpublished
aphD	RS01565	RS18765	5-Alkyl-2-hydroxy-muconate-6-semialdehyde dehydrogenase	4-hydroxymuconic-semialdehyde dehydrogenase (DmpC) P19059	45	Nordlund and Shingler (1990)
aphE	RS01575, RS01600	RS18775, RS18820	5-Alkyl-2-hydroxymuconate tautomerase	2-hydroxymuconate tautomerase (DmpI) P49172	38	Shingler et al. (1992)
aphF	RS01605	RS18825	Enol 5-alkyl-2-oxalocrotonate decarboxylase	4-oxalocrotonate decarboxylase (NahK, DmpH) Q1XGK3	85	Tsuda and Iino (1990)
aphG	RS01610	RS18830	2-Keto-4-alkylpentenoate hydratase	2-keto-4-pentenoate hydratase (MhpD) P77608	42	Pollard and Bugg (1998)
aphH	RS01560	RS18760	4-Hydroxy-2-alkylketopentenoate aldolase	4-hydroxy-2-oxovalerate aldolase (DmpG, MhpE) P51016	48	Shingler et al. (1992)
aphI	RS01555	RS18755	Alkylacetaldehyde dehydrogenase	Acetaldehyde dehydrogenase (HsaG, MphF) P9WQH3	57	Carere et al. (2013)
aphR	RS01590	RS18790	Aph transcriptional regulator	AraC family Transcriptional regulator Q88H39	39	Nelson et al. (2002)
aphQ	RS01595	RS18810	Aph transcriptional regulator	AraC family Transcriptional regulator Q88H39	36	Nelson et al. (2002)
aphS	RS01615	RS18835	Aph transcriptional regulator	IclR family transcriptional regulator Q0SH23	31	Pouyssegur and Stoeber, (1974)

[a]Locus in EP4 (C6369_RSXXXXX).
[b]Locus in RHA1 (RHA1_RSXXXXX).
[c]Where "alkyl" represents the variable-length 4-alkyl side chain.
[d]Gene name and Uniprot identifier of closest characterized homolog.
[e]Percent identity of best hit and EP4 homolog determined by Clustal Omega alignment.

760 [f]Also annotated as catechol 2,3-dioxygenase (DmpB)

761 ## Figure Legends

762 Figure 1. Isolation and growth of 4-ethylphenol-catabolizing strain *Rhodococcus rhodochrous*
763 EP4. A) Schematic of EP4 isolation from compost on 4-ethylphenol , which is produced from
764 reductive lignin depolymerization. Outer genomic track: coding sequences by strand; Inner track:
765 insertion sequences; lines: GC content (%). MC, *meta*-cleavage pathway gene cluster; OC,
766 *ortho*-cleavage pathway gene cluster. B) Growth of EP4 on 1 mM 4-ethylphenol or 2 mM
767 succinate controls as optical density at 600 nm (OD_{600}). Points and error bars reflect mean and
768 standard error (n = 3). C) Colony forming units (CFU) during growth in Panel B. Points and
769 horizontal bar indicate individual measurements and mean. D) Removal of 4-ethylphenol and
770 succinate in EP4 cultures by GC/MS.

771

772 Figure 2. Transcriptomic and genomic identification of the 4-ethylphenol catabolic pathway
773 genes in *Rhodococcus rhodochrous* EP4. A) RNA reads from cells grown on 4-ethylphenol or
774 succinate mapped to the EP4 gene clusters encoding catechol meta-cleavage and ortho-cleavage.
775 B) *Deseq2* differential-expression analysis showing log2 fold-change (FC) on 4-ethylphenol
776 versus succinate (FDR-corrected p-values: * $p_{fdr} < 0.001$). Points show values for n = 3
777 replicates; horizontal bar indicates mean. P450, cytochrome P450 gene; Red., P450 reductase. C)
778 Proposed funneling of 4-ethylphenol into the alkylcatechol *meta*-cleavage pathway (upper) and
779 not the catechol *ortho*-cleavage pathway (lower). TCA, tricarboxylic acid cycle.

780

781 Figure 3. Characterization of $AphAB_{EP4}$. A) Phylogenetic tree constructed using structural-based
782 alignment and RAxML. B) Hydroxylation of 4-ethylphenol to 4-ethlycatechol by purified

783 $AphAB_{EP4}$. Reaction mixtures contained 20 μM of each enzyme component, 100 μM substrate,
784 and were incubated overnight. C) Specific activity of $AphAB_{EP4}$ for select phenols. Activity was
785 measured using a coupled, spectrophotometric assay. D) Transformation of 4-HPA and 4-NP by
786 $AphAB_{EP4}$. Conditions as in C. E) Cofactor and substrate preference of $AphB_{EP4}$. Reductase
787 activity was measured using cytochrome *c*. HPA; hydroxyphenylacetate; NP, nitrophenol; FAD,
788 flavin adenine dinucleotide; FMN, flavin mononucleotide; RF, riboflavin.
789
790 Figure 4. Identification of a putative *aph* genomic island in rhodococci. A) Alignment of
791 *Rhodococcus* genomes to EP4 reference with *nucmer* ordered by RAxML tree calculated from
792 concatenated alignment of 16 ribosomal protein sequences; predicted RHA1 genomic islands;
793 GC content (%); and syntentic organization of *aph* gene clusters showing genomic islands
794 predicted using IslandViewer4. *Nucmer* alignment regions shown with dashed line. *aphA*, 4-
795 alkylphenol 3-monooxygenase, oxygenase gene; *aphB*, 4-alkylphenol 3-monooxygenase,
796 reductase gene; *aphC*, alkylcatechol 2,3-dioxygenase gene; *bph*, biphenyl catabolism gene
797 cluster; P450, cytochrome P450 gene; Red., P450 reductase. The single origin of replication
798 (*oriC*) shown with orange diamond. Detail of the *aph* region alignments in Supplementary Figure
799 6B.
800
801 Figure 5. Molecular genetic analysis of 4-ethylphenol catabolism in RHA1. A) Growth of WT,
802 Δ*aphA* and Δ*aphC1* RHA1 strains on 1 mM 4-ethylphenol or 2 mM succinate. B) Protein yield
803 of WT, Δ*aphC,* and Δ*pcaL* RHA1 strains as well as EP4 and *R. rhodochrous* DSM43241 on
804 phenolic substrates. Protein measured after 24 hours incubation. C) Expression of select genes in
805 WT and Δ*aphA* RHA1 strains during growth on 1 mM 4-ethylphenol or 2 mM succinate using

806 RT-qPCR. Colors indicate cleavage pathway. Points and horizontal bars show individual
807 measurements (n = 3) and mean. Significance levels following Bonferroni-corrected two-tailed
808 Student's t-tests (*, p_{bon}<0.05; **, p_{bon}<0.01; ***, p_{bon}<0.001). DMP, dimethylphenol; EP,
809 ethylphenol; HBA, hydroxybenzoic acid; HPA; hydroxyphenylacetate; MP, methylphenol; NP,
810 nitrophenol; PP, propylphenol.

811

812 ## Supplementary Figure Captions

813 Supplementary Figure 1. Quantitative mapping of prokaryotic RNASeq reads. A) Compute time
814 in seconds for read mapping steps for three methods of feature quantification: FeatureCounts
815 (FC), HTSeq and Salmon. Timing benchmarks calculated on a 256 Gb, 32-core Ubuntu 14.04.5
816 server, running eight cores per analysis. B) Total counts produced by each method. C) PCoA
817 ordination of read counts of all genomic features with results of pairwise PERMANOVA. D)
818 Linear regression of RNASeq method against RT-qPCR transcript abundance showing R2adj.
819 All *p*-values < 0.0001). 4-EP, 4-ethylphenol; 4-PG, 4-propylguaiacol; S, succinate. Full 4-PG
820 results can be found in a separate manuscript (Fetherolf et al., unpublished).

821

822 Supplementary Figure 2. Transcriptomic analysis of the metabolism of butyryl-CoA produced
823 from 4-ethylphenol catabolism by the Aph pathway in EP4. The top panel shows the predicted
824 pathway of metabolism showing the production of the acyl-CoA during the final step of the Aph
825 pathway. The bottom panels show plots of -$\log_{10}$ transformed *p*-values plotted against $\log_2$ fold
826 change for all putative homologs of the genes encoding each step of acyl-CoA metabolism
827 during growth on 1.0 mM 4-ethylphenol versus succinate calculated using *DeSeq2*.

828

Supplementary Figure 3. RT-qPCR expression of *aphA*, *aphC*, *pheA*, *catA* and *sigF* (control) genes during growth of EP4 on 1 mM 4-ethylphenol (EP), 2 mM succinate (S) and 1 mM benzoate. N = 3, log10 gene copies ng^{-1} total RNA). Table shows log_2 fold change and *p*-value following Bonferroni-corrected two-tailed Student's t-tests: *, $p_{bon} < 0.05$, **, $p_{bon} < 0.01$.

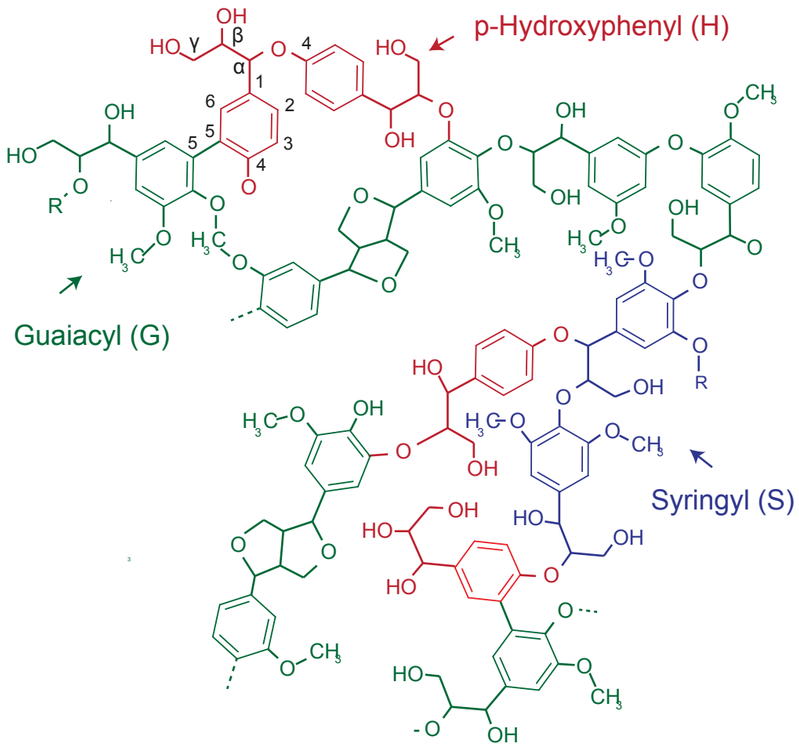
Supplementary Figure 4. TCoffee-Expresso-based structure alignment of EP4 AphA and select homolog amino acid sequences. See Figure 3 for sequence identification.

Supplementary Figure 5. Investigating the role of AphR in *aphAB* expression. A) RAxML-tree from TCoffee-aligned AraC-family transcriptional regulatory proteins. Red indicates previously identified proteins. B) Alignment of EP4 and RHA *aphA*-AphR and *pheA*-PheR binding regions; inset: organization of EP4 and RHA1 *aphBAR* and *pheRBA* genes. D) Map of the *aphR-aphA* intergenic region showing the AphR binding site predicted from alignment with CCM2595 PheR binding regions (yellow, dotted underline), *aphA* promotors (solid underline), *aphA* transcriptional start site (TSS) (purple) and predicted coding start (CS) (pink). Strains: ATCC39116, *Amycolatopsis* sp. ATCC39116; BUPNP1, *Rhodococcus* sp. BUPNP1; CCM2595, Rhodococcus *erythropolis* CCM2595; KL28, *Pseudomonas* sp. KL28.

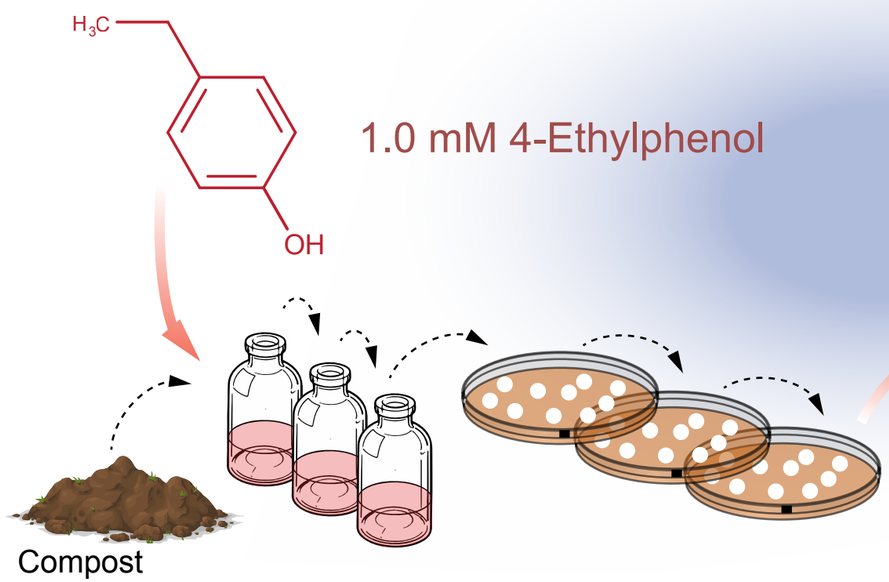
Supplementary Figure 6. The *aph* cluster in *Rhodococcus* genomes. A) RAxML phylogenetic tree calculated from MUSCLE alignment of concatenated sequences for ribosomal proteins L2, L3, L4, L5, L6, L14, L16, L18, L22, L24, S3, S8, S10, S17 and S19 from 325 *Rhodococcus* ssp. genomes showing genomes for which BLASTp evalues were < 10^{-100} for at least 7 of the 13 EP4

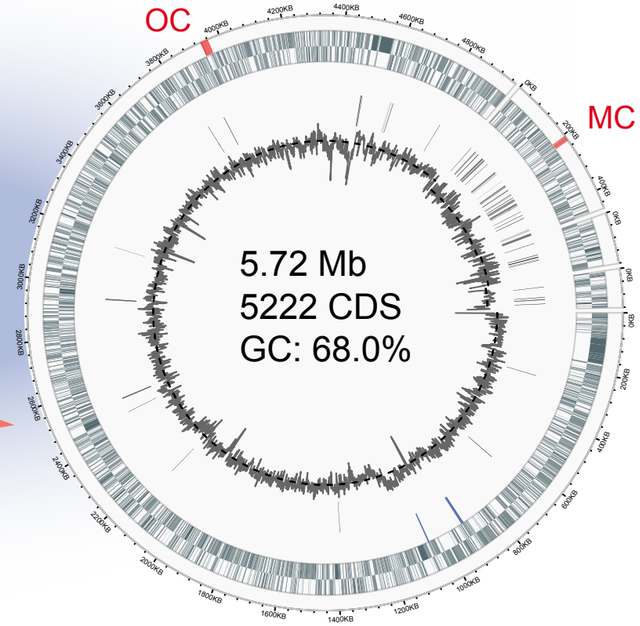
851 Aph amino acid sequences (gold boxes). Also showing the *Rhodococcus*
852 *opacus*/*jostii*/*wratislaviensis* clade (green), *Rhodococcus rhodochrous*/*pyradinivorans* clade
853 (gold), and *Rhodococcus* sp. Eu-32 (peach). B) Detail from Figure 4 showing *nucmer* alignment
854 of select *Rhodococcus* genomes to the EP4 *aph* gene cluster including a portion of the 117-kb
855 *aph* gene-containing genomic island predicted using IslandViewer4 (grey), showing *nucmer*
856 aligned regions (black dashed lines), median GC content (red dashed line) and GC content in 200
857 bp segments (solid grey line).
858
859

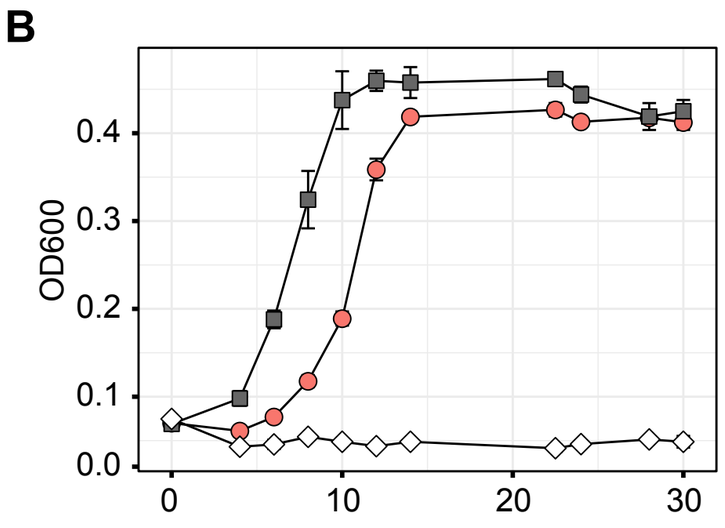
A Lignin

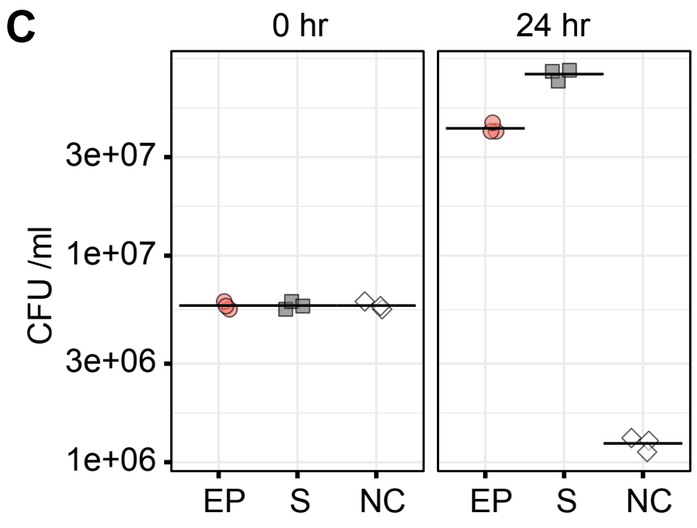


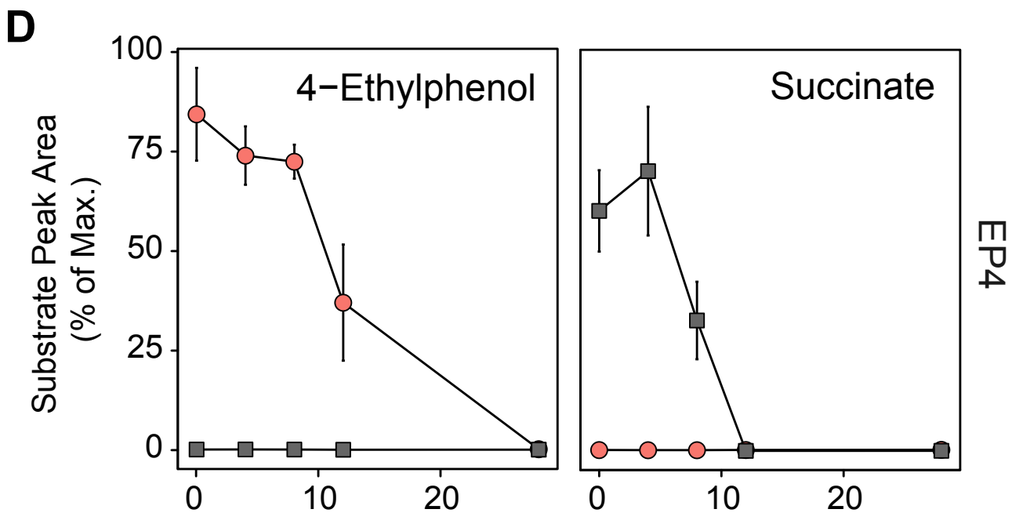
Rhodococcus rhodochrous EP4 isolation and genome sequencing



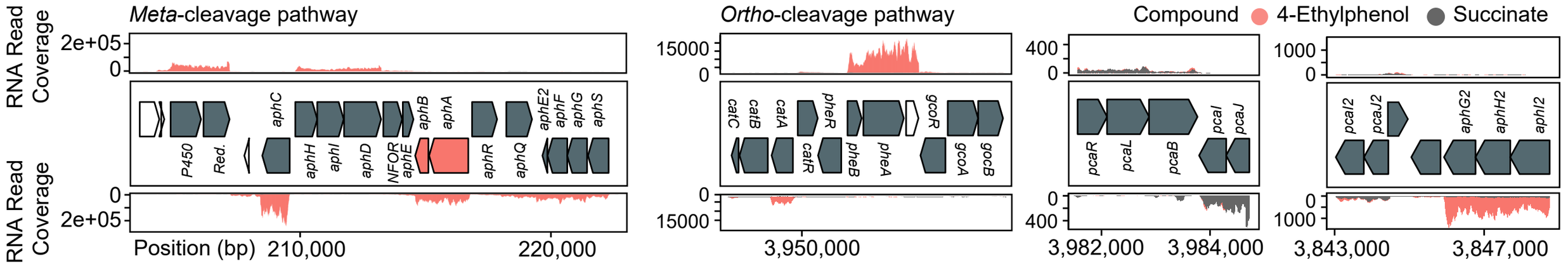




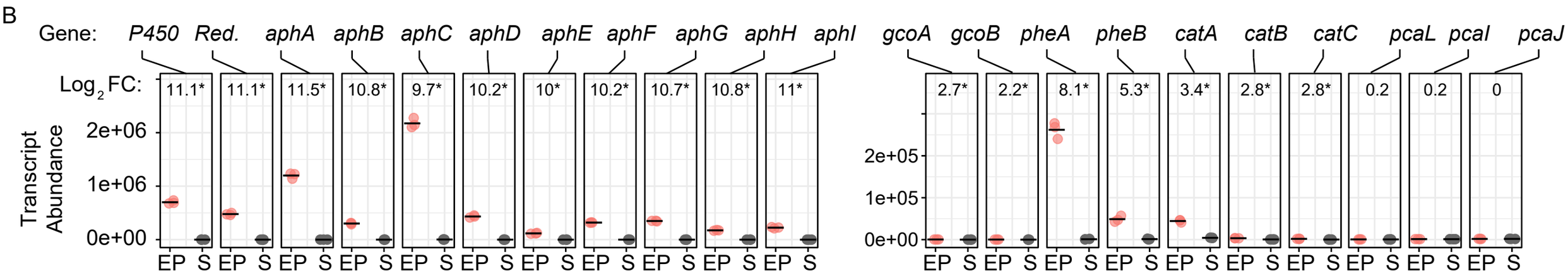




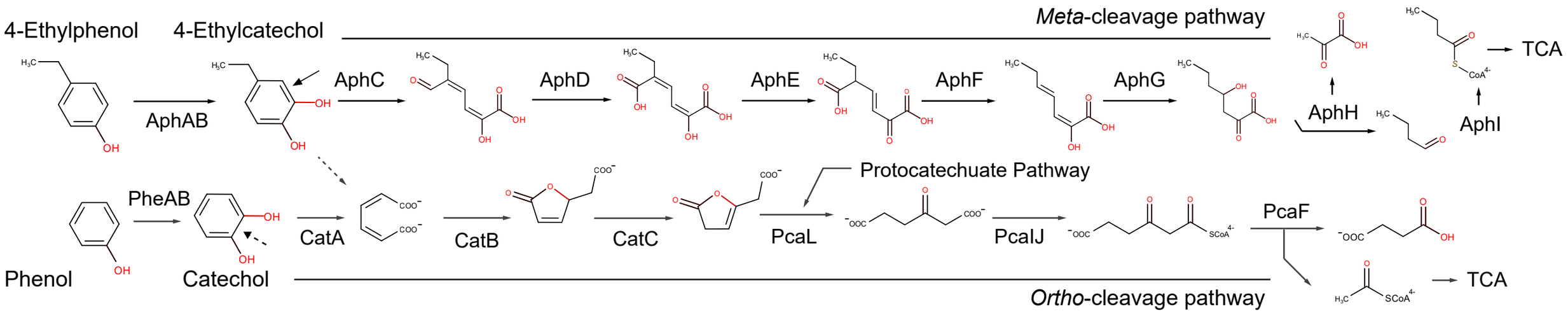
A

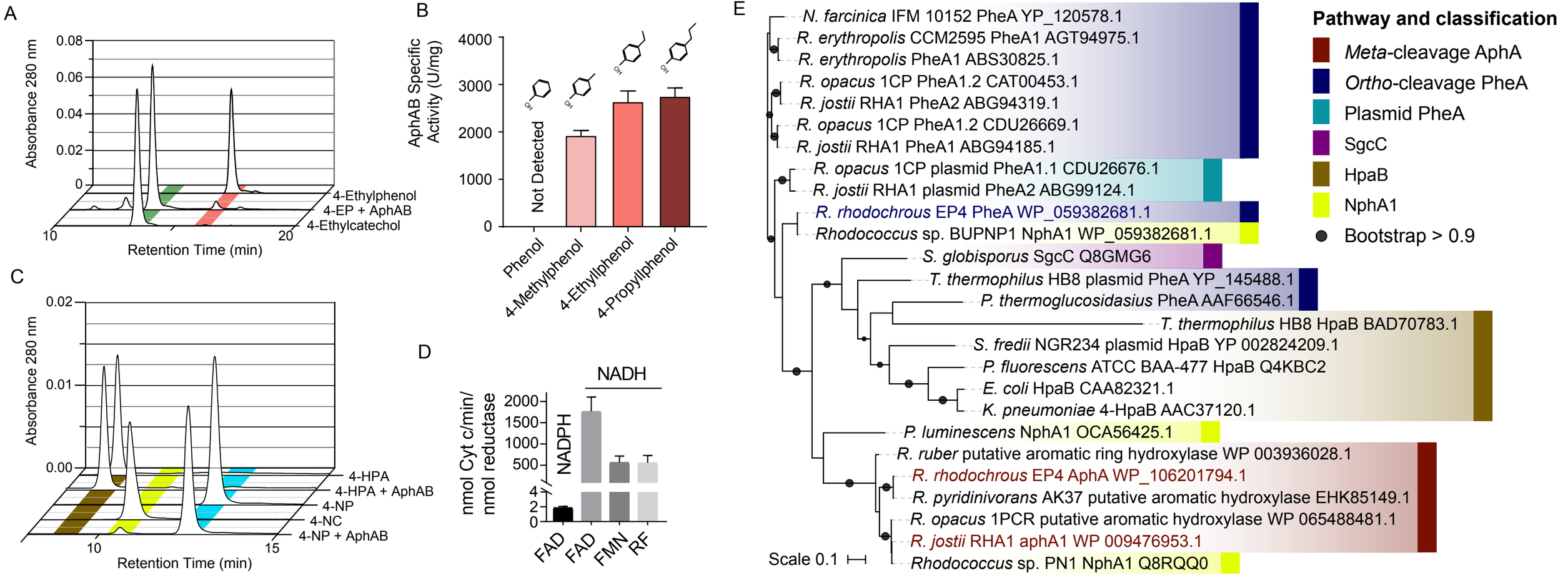

Meta-cleavage pathway
Ortho-cleavage pathway
Compound
4-Ethylphenol
Succinate
RNA Read Coverage
2e+05
0
15000
400
1000
P450
Red.
aphC
aphH
aphI
aphD
NFOR
aphE
aphB
aphA
aphR
aphQ
aphE2
aphF
aphG
aphS
catC
catB
catA
catR
pheR
pheB
pheA
gcoR
gcoA
gocB
pcaR
pcaL
pcaB
pcaI
pcaJ
pcaI2
pcaJ2
aphG2
aphH2
aphI2
Position (bp)
210,000
220,000
3,950,000
3,982,000
3,984,000
3,843,000
3,847,000

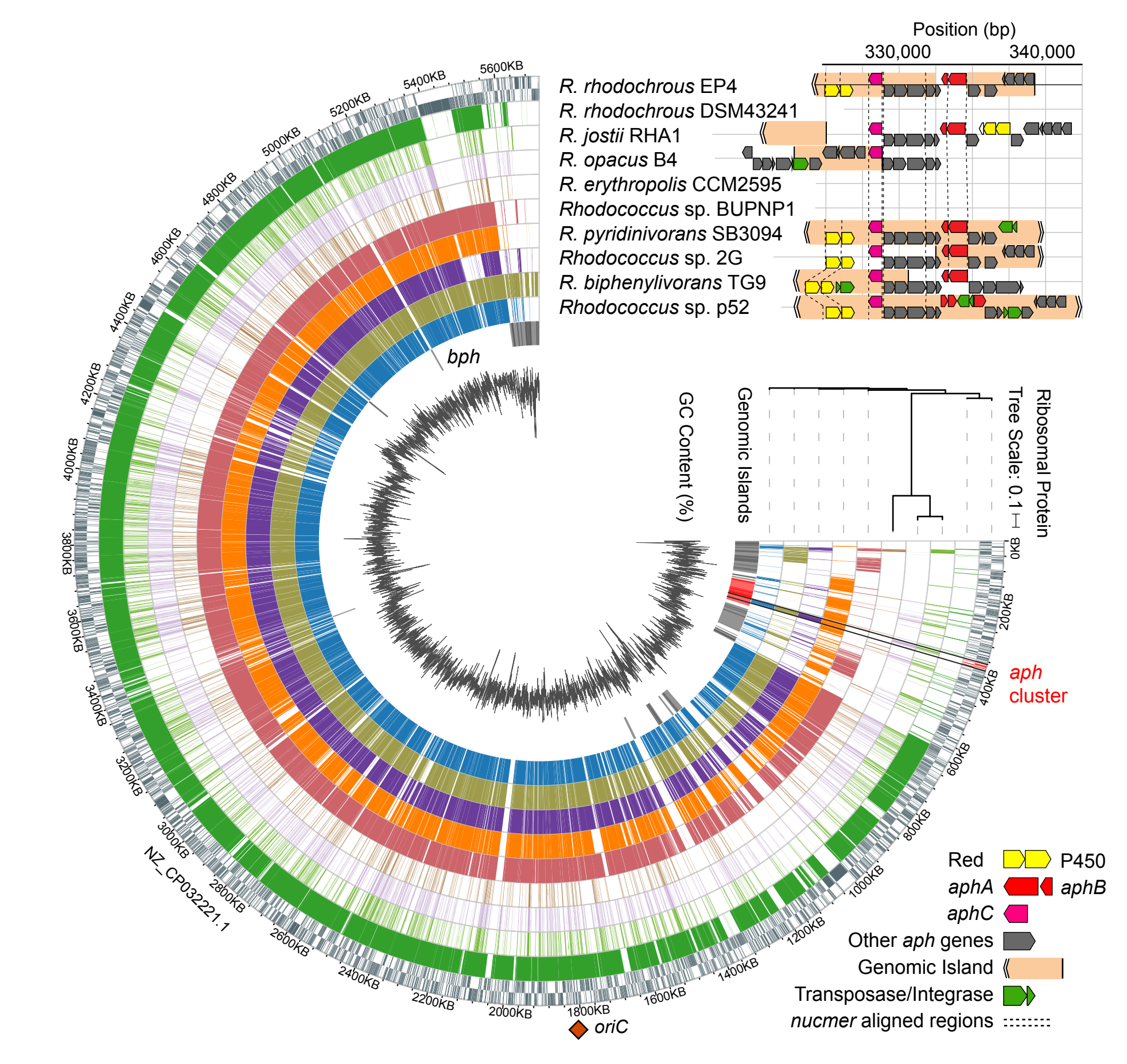
B

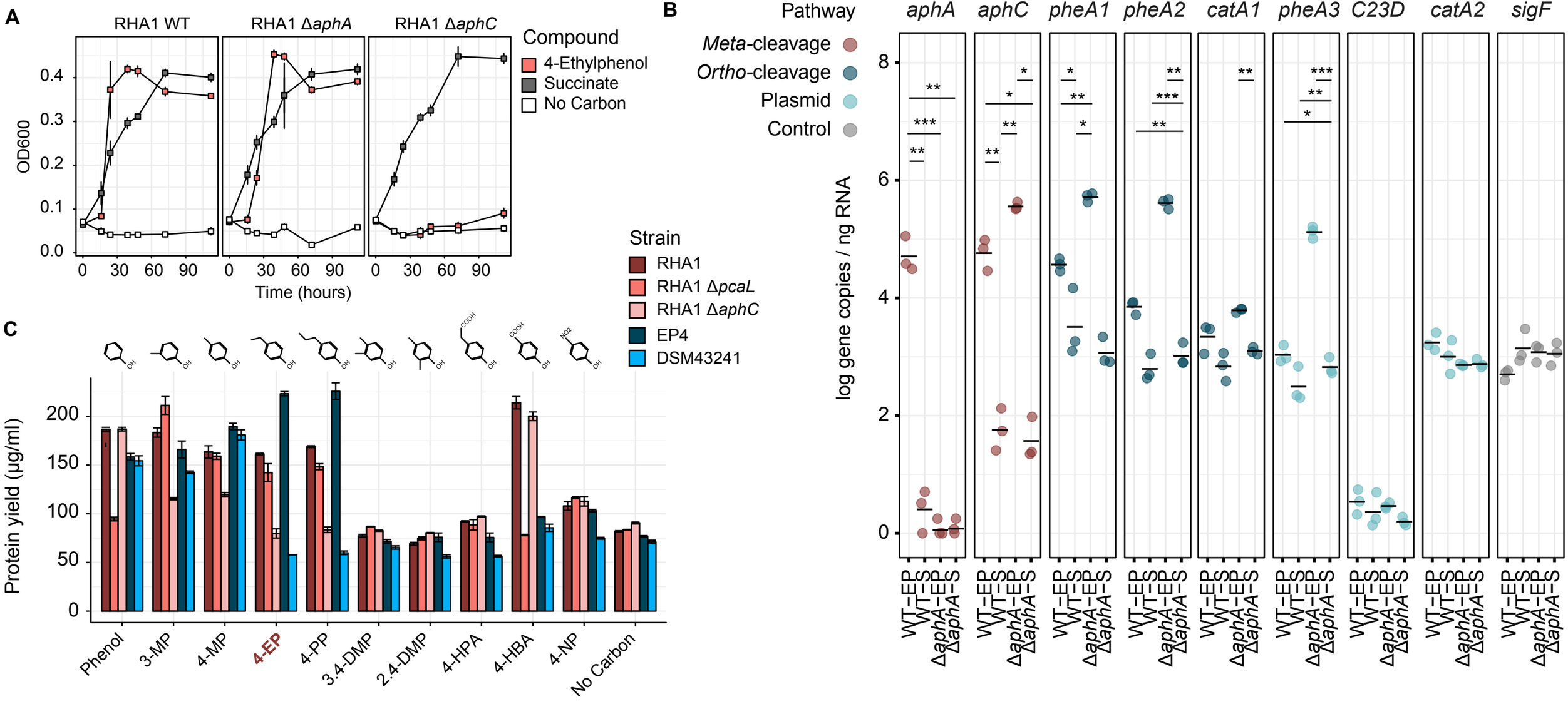

Gene:
P450
Red.
aphA
aphB
aphC
aphD
aphE
aphF
aphG
aphH
aphI
gcoA
gcoB
pheA
pheB
catA
catB
catC
pcaL
pcaI
pcaJ
Log2FC:
11.1*
11.1*
11.5*
10.8*
9.7*
10.2*
10*
10.2*
10.7*
10.8*
11*
2.7*
2.2*
8.1*
5.3*
3.4*
2.8*
2.8*
0.2
0.2
0
Transcript Abundance
2e+06
1e+06
0e+00
2e+05
1e+05
EP
S

C


4-Ethylphenol
4-Ethylcatechol
Meta-cleavage pathway
AphAB
AphC
AphD
AphE
AphF
AphG
AphH
AphI
TCA
Protocatechuate Pathway
PheAB
CatA
CatB
CatC
PcaL
PcaIJ
PcaF
Phenol
Catechol
Ortho-cleavage pathway
TCA


A
Absorbance 280 nm
0.08
0.06
0.04
0.02
0
10
20
Retention Time (min)
4-Ethylphenol
4-EP + AphAB
4-Ethylcatechol
B
AphAB Specific Activity (U/mg)
4000
3000
2000
1000
0
Not Detected
Phenol
4-Methylphenol
4-Ethylphenol
4-Propylphenol
C
Absorbance 280 nm
0.02
0.01
0.00
10
15
Retention Time (min)
4-HPA
4-HPA + AphAB
4-NP
4-NC
4-NP + AphAB
D
NADH
NADPH
nmol Cyt c/min/ nmol reductase
2000
1500
1000
500
4
2
0
FAD
FAD
FMN
RF
E
N. farcinica IFM 10152 PheA YP_120578.1
R. erythropolis CCM2595 PheA1 AGT94975.1
R. erythropolis PheA1 ABS30825.1
R. opacus 1CP PheA1.2 CAT00453.1
R. jostii RHA1 PheA2 ABG94319.1
R. opacus 1CP PheA1.2 CDU26669.1
R. jostii RHA1 PheA1 ABG94185.1
R. opacus 1CP plasmid PheA1.1 CDU26676.1
R. jostii RHA1 plasmid PheA2 ABG99124.1
R. rhodochrous EP4 PheA WP_059382681.1
Rhodococcus sp. BUPNP1 NphA1 WP_059382681.1
S. globisporus SgcC Q8GMG6
T. thermophilus HB8 plasmid PheA YP_145488.1
P. thermoglucosidasius PheA AAF66546.1
T. thermophilus HB8 HpaB BAD70783.1
S. fredii NGR234 plasmid HpaB YP 002824209.1
P. fluorescens ATCC BAA-477 HpaB Q4KBC2
E. coli HpaB CAA82321.1
K. pneumoniae 4-HpaB AAC37120.1
P. luminescens NphA1 OCA56425.1
R. ruber putative aromatic ring hydroxylase WP 003936028.1
R. rhodochrous EP4 AphA WP_106201794.1
R. pyridinivorans AK37 putative aromatic hydroxylase EHK85149.1
R. opacus 1PCR putative aromatic hydroxylase WP 065488481.1
R. jostii RHA1 aphA1 WP 009476953.1
Rhodococcus sp. PN1 NphA1 Q8RQQ0
Scale 0.1
Pathway and classification
Meta-cleavage AphA
Ortho-cleavage PheA
Plasmid PheA
SgcC
HpaB
NphA1
Bootstrap > 0.9


Position (bp)
330,000
340,000
R. rhodochrous EP4
R. rhodochrous DSM43241
R. jostii RHA1
R. opacus B4
R. erythropolis CCM2595
Rhodococcus sp. BUPNP1
R. pyridinivorans SB3094
Rhodococcus sp. 2G
R. biphenylivorans TG9
Rhodococcus sp. p52
bph
GC Content (%)
Genomic Islands
Tree Scale: 0.1
Ribosomal Protein
aph
cluster
NZ_CP032221.1
oriC
Red
P450
aphA
aphB
aphC
Other aph genes
Genomic Island
Transposase/Integrase
nucmer aligned regions


A
RHA1 WT
RHA1 ΔaphA
RHA1 ΔaphC
OD600
Time (hours)
Compound
4-Ethylphenol
Succinate
No Carbon
Strain
RHA1
RHA1 ΔpcaL
RHA1 ΔaphC
EP4
DSM43241
B
Pathway
Meta-cleavage
Ortho-cleavage
Plasmid
Control
aphA
aphC
pheA1
pheA2
catA1
pheA3
C23D
catA2
sigF
log gene copies / ng RNA
WT−EP
WT−S
ΔaphA−EP
ΔaphA−S
C
Protein yield (μg/ml)
Phenol
3-MP
4-MP
4-EP
4-PP
3.4-DMP
2.4-DMP
4-HPA
4-HBA
4-NP
No Carbon