

1Enzyme discovery for toluene synthesis in anoxic microbial communities

2

3H. R. Beller†\*<sup>1,2</sup>, A. V. Rodrigues†<sup>1</sup>, K. Zargar†<sup>1</sup>, Y.-W. Wu<sup>3</sup>, A. K. Saini<sup>1</sup>, R. M. Saville<sup>1</sup>, [J. H. Pereira<sup>1</sup>](#), [P. D. Adams<sup>1</sup>](#), S. G. Tringe<sup>4</sup>, C. J. Petzold<sup>1,5</sup>, J. D. Keasling<sup>1,5,6,7</sup>

5

6

7Joint BioEnergy Institute, 5885 Hollis Avenue, Emeryville, CA, USA<sup>1</sup>; Earth and Environmental  
8Sciences, LBNL, Berkeley, CA, USA<sup>2</sup>; Graduate Institute of Biomedical Informatics, College of  
9Medical Science and Technology, Taipei Medical University, Taipei, Taiwan<sup>3</sup>; Joint Genome  
10Institute, 2800 Mitchell Drive, Walnut Creek, CA, USA<sup>4</sup>; Biological Systems and Engineering,  
11Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA, USA<sup>5</sup>; Departments of  
12Chemical & Biomolecular Engineering and of Bioengineering, University of California,  
13Berkeley, CA, USA<sup>6</sup>; Novo Nordisk Foundation Center for Biosustainability, Technical  
14University of Denmark, Kogle Allé, DK2970-Hørsholm, Denmark<sup>7</sup>

15

16

17

18

19

20

21†These authors contributed equally to this work.

22\* Corresponding author: HRBeller@lbl.gov

23

24

25 **Microbial toluene biosynthesis was reported in anoxic lake sediments more than 3 decades**  
26 **ago, however the enzyme(s) catalyzing this biochemically challenging reaction have never**  
27 **been elucidated. Here we report the first toluene synthase, a glycy radical enzyme of**  
28 **bacterial origin that catalyzes phenylacetic acid decarboxylation (PhdB), and its cognate**  
29 **activating enzyme (PhdA, a radical S-adenosylmethionine enzyme), discovered in two**  
30 **distinct anoxic microbial communities that produced toluene. The unconventional process**  
31 **of enzyme discovery from a complex microbial community (>300,000 genes) rather than**  
32 **from a microbial isolate, involved metagenomics- and metaproteomics-enabled**  
33 **biochemistry, as well as *in-vitro* confirmation of activity with recombinant enzymes. This**  
34 **work expands the known catalytic range of glycy radical enzymes (only seven reaction**  
35 **types had been characterized previously) and aromatic hydrocarbon-producing enzymes**  
36 **(only one reaction type characterized previously), and will enable first-time biochemical**  
37 **synthesis of an aromatic fuel hydrocarbon from renewable resources, such as lignocellulosic**  
38 **biomass, rather than petroleum.**

39

40 The extraordinary metabolic diversity of microorganisms in combination with ready  
41 access to increasingly rapid and less expensive DNA sequencing technologies has revealed a  
42 well-recognized challenge in modern biology: the dearth of experimental evidence to support  
43 functional annotation of a large fraction of genes/proteins in public data repositories<sup>1-3</sup>. A related  
44 challenge, termed “orphan enzymes”<sup>4</sup>, is the abundance of unambiguously defined enzymatic  
45 activities that are not linked with specific amino acid sequences; in 2014, 22% of defined EC  
46 (Enzyme Commission) numbers were orphan enzymes<sup>5</sup>. To the extent that specific enzymes can  
47 be better linked to a broad range of chemically diverse reactions, the scope and versatility of

48biochemical transformations harnessed for biotechnological applications will be enhanced. One  
49area in which knowledge of enzymes is very limited is biosynthesis of aromatic hydrocarbons,  
50which could be useful as renewable fuels or chemicals made from non-petroleum feedstocks. To  
51our knowledge, the only aromatic hydrocarbon that can currently be synthesized wholly from  
52known enzymes is styrene, which can be produced from phenylalanine-derived *trans*-cinnamic  
53acid by enzymes displaying phenylacrylate decarboxylase activity, such as ~~PAL2~~ from  
54~~*Arabidopsis thaliana*~~ or FDC1 from *Saccharomyces cerevisiae*<sup>6</sup>.

55 We targeted the aromatic hydrocarbon toluene for enzyme discovery, as it is an important  
56petrochemical with a global market of 29 million tons per year whose uses include synthesis of  
57other aromatic feedstocks and serving as an effective octane booster in gasoline (octane number,  
58114). Microbial sources of biogenic toluene were reported more than three decades ago,  
59however, the underlying biochemistry and specific enzymes catalyzing toluene biosynthesis have  
60never been elucidated. Biogenic toluene was observed in anoxic lake sediments / hypolimnion<sup>7</sup>,  
61in anoxic enrichment cultures derived from municipal sewage sludge<sup>8</sup>, and in two bacterial  
62isolates, *Tolomonas auensis*<sup>9</sup> and *Clostridium aerofetidum*<sup>10</sup>, which were reported to synthesize  
63toluene from phenylacetate and L-phenylalanine (however, recent attempts to reproduce toluene  
64biosynthesis by these two isolates were unsuccessful<sup>8</sup>). Although a toluene synthase has not been  
65specifically identified, *in vitro* studies with cell-free extracts from a toluene-producing culture  
66suggest catalysis by a glyceryl radical enzyme (GRE)<sup>8</sup>. Evidence supporting the hypothesized role  
67of a GRE in toluene biosynthesis included (a) irreversible inactivation by O<sub>2</sub> (a characteristic of  
68GREs), (b) the ruling out of a mechanism involving successive reduction (phenylacetate to  
69phenylacetaldehyde) and decarbonylation/~~deformylation~~ (phenylacetaldehyde to toluene), which  
70would not be expected to be catalyzed by GREs<sup>11,12</sup>, and (c) the observation that the known

71enzyme with the greatest functional similarity to phenylacetate decarboxylase, namely *p*-  
72hydroxyphenylacetate decarboxylase (HpdBC or CsdBC), is a GRE<sup>13,14</sup>. Although a GRE has  
73been implicated in toluene biosynthesis, even the most detailed *in vitro* studies conducted to date  
74have not identified any specific gene candidates<sup>8</sup>.

## 75**Identification of toluene synthase candidates**

76 Studies to identify a toluene synthase (phenylacetate decarboxylase) were conducted with  
77anaerobic, toluene-producing microbial cultures that derived from two different inocula:  
78municipal sewage sludge<sup>8</sup> and lake sediments from Berkeley, CA (Extended Data Fig. 1). The  
79sewage culture, which was more amenable to cultivation and *in vitro* studies, served as the basis  
80for most of the experimental discovery studies, whereas the lake sediment culture was used  
81primarily for metagenome sequencing. We employed a metagenomics- and metaproteomics-  
82enabled protein purification approach for enzyme discovery from these microbial communities.  
83Toluene synthase activity was monitored in chromatographically separated fractions of cell-free  
84extracts from the sewage culture using *in vitro* assays that measured phenylacetic acid-2-<sup>13</sup>C  
85conversion to [*methyl*-<sup>13</sup>C]toluene. All experimental procedures, including cultivation, cell lysis,  
86protein purification by FPLC (fast protein liquid chromatography), and *in vitro* assays, were  
87performed under strictly anaerobic conditions to protect the organisms and enzymes from  
88molecular oxygen. Proteomic profiles of active FPLC fractions were compared to those of  
89adjacent inactive (or much less active) fractions to identify toluene synthase candidates (i.e.,  
90those proteins enriched in, and ideally unique to, active fractions). An unknown GRE (hereafter  
91referred to as PhdB) co-eluted with the maximal toluene synthase activity (Extended Data Fig.  
922). Although more than 650 proteins co-eluted with PhdB in these fractions (Supplementary  
93Data File 1), this protein was initially of interest because the toluene synthase in this sewage-

94derived culture had been postulated to be a GRE based upon *in vitro* studies with cell-free  
95extracts<sup>8</sup>. Notably, PhdB was one of the few glycy radical enzymes detected in active fractions  
96among the many glycy radical enzymes encoded in the sewage community metagenome (Fig.  
971). As shown in Fig. 1, only three glycy radical enzymes were detected in the active FPLC  
98fractions: (1) PhdB, (2) pyruvate formate-lyase (PflB; JGI2065J20421\_100036324; IMG Taxon  
99ID 3300001865), which had 99% sequence identity to known *Enterobacter* PflB copies], and (3)  
100an unknown glycy radical enzyme (JGI2065J20421\_10067673; IMG Taxon ID 3300001865) –  
101this protein shares ca. 47% sequence identity and key conserved residues with a known glycerol  
102dehydratase (PDB 1R8W). Of these three proteins, only PhdB and the PflB had greater  
103abundance in active than in flanking inactive fractions (Fig. 1), and PflB was among the most  
104abundant proteins in both active *and* inactive fractions (Supplementary Data File 1), which,  
105along with its well-characterized function, reduced its plausibility as a toluene synthase  
106candidate.

107       The strength of *phdB* as a candidate toluene synthase gene was enhanced by its  
108identification in metagenomes of both the anoxic, toluene-producing sewage and lake sediment  
109cultures, despite the fact that these cultures had disparate inocula and phylogenetic compositions  
110(a comparison of dominant taxa in these two cultures is shown in Extended Data Fig. 3 and  
111Supplementary Data Files 2 and 3). In sewage culture metagenomes, *phdB* occurred in a three-  
112gene cluster consisting of a putative transcription factor (Sequence 11, Supplementary Data File  
1134), *phdB* (Sequence 6, Supplementary Data File 4), and a glycy radical activating enzyme  
114(hereafter referred to as *phdA*; Sequence 1, Supplementary Data File 4) (Fig. 2). Such adjacent  
115positioning in genomes of genes encoding glycy radical enzymes and their cognate activating  
116enzymes is very common<sup>15</sup>, [as indicated in Fig. 2](#). Although assembled contigs from the lake

117sediment metagenomes (e.g., IMG Taxon ID 2100351000) were not observed to harbor the  
118complete three-gene cluster detected in the sewage metagenome, the quality of these assemblies  
119was suboptimal as a result of older sequencing methods used. Indeed, PCR amplification and  
120Sanger sequencing of this cluster from genomic DNA of the lake culture revealed an intact three-  
121gene cluster (Sequence 13-Sequence 9-Sequence 4; Supplementary Data File 4) with identical  
122length (6065 bp) and strikingly similar coding and intergenic sequences compared to the sewage  
123culture (Fig. 2). As shown in Fig. 2, the three genes share from ca. 87 to 96% sequence identity  
124(and 86 to 97% translated sequence identity) in the sewage and lake cultures and the intergenic  
125regions are ca. 82-85% identical (Sequences 15 and 16; Supplementary Data File 4).

#### 126***In vitro* confirmation of PhdB and PhdA activity**

127        Recombinant versions of PhdA and PhdB were assayed for *in vitro* activity to confirm  
128their role in catalyzing toluene biosynthesis from phenylacetate. The expected activity for PhdA  
129was based on characterization of other glycy radical activating enzymes<sup>16</sup>. In glycy radical  
130systems, the reduced [4Fe-4S]<sup>+1</sup> cluster of the activase, a radical S-adenosylmethionine (SAM)  
131enzyme, transfers an electron to SAM, resulting in homolytic cleavage of SAM to form  
132methionine and a 5'-deoxyadenosyl radical (Fig. 3a). The 5'-deoxyadenosyl radical activates the  
133GRE by stereospecific abstraction of a C-2 *pro-S* H atom from a highly conserved glycine  
134residue, which in turn abstracts an H atom from a conserved cysteine residue in the GRE to form  
135a thiyl radical. A substrate radical is formed when the thiyl radical abstracts an H atom from the  
136substrate (phenylacetic acid, in the case of PhdB; Fig. 3b).

137        *In vitro* reconstitution of the [4Fe-4S] cluster of PhdA was performed before final  
138purification (all under strictly anaerobic conditions), and the [4Fe-4S] cluster was reduced with  
139dithionite in an **anaerobic-anoxic** assay measuring methionine production from SAM using liquid

140 chromatography-mass spectrometry (LC/MS). Observed methionine production in the presence  
141 of PhdA, but not in its absence (Fig. 3a), demonstrated the expected activity of a glycy radical  
142 activating enzyme.

143 The ability of activated (enzyme-radical) PhdB to catalyze decarboxylation of  
144 phenylacetic acid-2-<sup>13</sup>C to [*methyl*-<sup>13</sup>C]toluene was tested in an anaerobic, *in vitro* assay in  
145 the presence of dithionite-reduced PhdA and SAM (Fig. 3b). Labeled toluene was detected by  
146 gas chromatography-mass spectrometry (GC/MS) in the presence of SAM but not in its absence,  
147 confirming the role of PhdB in catalyzing toluene biosynthesis *via* a radical mechanism. A series  
148 of other negative control assays also displayed negligible activity, including the following: (1)  
149 assays lacking PhdB but containing dithionite-reduced PhdA and SAM, (2) assays conducted  
150 with a mutant version of PhdB (G815A) in which the putative site of the glycy radical was  
151 modified to alanine, and (3) assays in which the assay mixture was briefly exposed to air before  
152 the substrate was added, demonstrating O<sub>2</sub> sensitivity that is characteristic of GREs (Extended  
153 Figure 4). Specific activities observed in SAM-containing assays represented in Figure 3b were  
154 relatively low (in the pmol · min<sup>-1</sup> · mg protein<sup>-1</sup> range) compared to reported values for most other  
155 GREs, which range broadly from pmol · min<sup>-1</sup> · mg protein<sup>-1</sup> (benzylsuccinate synthase<sup>17</sup>) to mmol  
156 min<sup>-1</sup> · mg protein<sup>-1</sup> (glycerol dehydratase<sup>18</sup>). In part, low PhdB activity may reflect the generally  
157 sensitive nature of GREs when purified and manipulated *in vitro*. For example, even for a given  
158 enzyme, reported specific activities have differed by orders of magnitude in various studies [e.g.,  
159 for benzylsuccinate synthase, from 0.02<sup>17</sup> to 72 nmol · min<sup>-1</sup> · mg protein<sup>-1</sup><sup>19</sup>; for *p*-  
160 hydroxyphenylacetate decarboxylase, from 0.034<sup>13</sup> to 18.45 μmol · min<sup>-1</sup> · mg protein<sup>-1</sup><sup>14</sup>]. In the  
161 present study, a likely factor affecting PhdB activity was the poor solubility of the recombinant  
162 protein when expressed in *E. coli* (Extended Figure 5); a maltose-binding protein (MBP) tag was

163used to enhance solubility but may not have fully ameliorated suboptimal folding. For  
164biotechnological application of PhdB, enhanced solubility (e.g., through protein engineering)  
165will be required.

166 While PhdB displays phenylacetate decarboxylase activity, it does not display  
167comparable *p*-hydroxyphenylacetate decarboxylase activity (characteristic of the GRE  
168HpdBC/CsdBC). During assays in which equimolar amounts of phenylacetate and *p*-  
169hydroxyphenylacetate were amended to a mixture containing PhdA, PhdB, and SAM, labeled  
170toluene production was readily observed, however, *p*-cresol (the product of *p*-  
171hydroxyphenylacetate decarboxylation) was detected at levels approximately 100-fold lower  
172than those expected if PhdB activity were comparable for phenylacetate and *p*-  
173hydroxyphenylacetate (Extended Data Fig. 46). Analogous assays with *o*- and *m*-  
174hydroxyphenylacetate similarly indicated very low (in this case, undetectable) PhdB activity for  
175these hydroxyphenylacetate isomers, whereas labeled toluene was easily detected.

#### 176**Comparison of PhdB-PhdA to other glycol radical systems**

177 The demonstration of PhdB as a phenylacetate decarboxylase adds it to the group of  
178seven characterized GREs (Fig. 4), which includes pyruvate formate-lyase (EC 2.3.1.54<sup>20</sup>),  
179anaerobic ribonucleotide reductase (EC 1.17.4.1<sup>21</sup>), benzylsuccinate synthase (EC  
1804.1.99.11<sup>17,19,22</sup>), *p*-hydroxyphenylacetate decarboxylase (EC 4.1.1.82<sup>13,14,23</sup>), B<sub>12</sub>-independent  
181glycerol (and 1,2-propanediol) dehydratase (EC 4.2.1.30<sup>18</sup>), choline trimethylamine-lyase (EC  
1824.3.99.4<sup>24,25</sup>), and the very recently discovered *trans*-4-hydroxy-L-proline dehydratase<sup>26</sup>. Note  
183that benzylsuccinate synthase, which catalyzes the first step of anaerobic toluene degradation, is  
184the best characterized representative of a larger group of aromatic- and alkylsuccinate synthase



185enzymes that activate substrates including 2-methylnaphthalene, *p*-cresol, and *n*-hexane by  
186fumarate addition and have been collectively termed “X-succinate synthases”<sup>27</sup>.

187 PhdB shares important features characteristic of all known GREs, including the  
188following: (1) a conserved glycy radical motif (RVx**G**[FWY]<sub>x6-8</sub>[IL]<sub>x4</sub>Qx<sub>2</sub>[IV]<sub>x2</sub>R —  
189modification from Selmer et al.<sup>15</sup> indicated in italics) near the C-terminus of the protein (Fig. 5a),  
190(2) a conserved cysteine residue near the middle of the protein sequence (the site of the thiyl  
191radical in the active site that initiates H atom abstraction from the substrate) (Fig. 5b), and (3) a  
192cognate activating enzyme that belongs to the radical SAM superfamily<sup>15</sup>. However, PhdB is  
193clearly distinct from the other known glycy radical enzymes in a number of ways. For example,  
194the sequence identity of PhdB (from the sewage and lake cultures) to other GREs is relatively  
195low, ranging from ca. 14 to 31% (Extended Data Fig. [75](#)). Further, PhdB does not share all of the  
196conserved residues that have been assigned for other GREs. To illustrate, in the region near the  
197conserved active-site C residue (Fig. 5b), some conserved residues not shared by PhdB include  
198an additional C adjacent to the strictly conserved active-site C (PflB<sup>20</sup>), an E located two residues  
199downstream of the active-site C (CsdB<sup>23</sup>, Gdh<sup>18</sup>, CutC<sup>24</sup>, HypD<sup>26</sup>), and M-S-P residues  
200immediately downstream of the active-site C (BssA<sup>27</sup>).

201 With respect to *p*-hydroxyphenylacetate decarboxylase in particular, differences from  
202PhdB are noteworthy, since these proteins might be expected to be very similar based on the  
203seemingly analogous reactions that they catalyze (Fig. 4). Phenylacetate decarboxylase (PhdB)  
204has only one subunit type, in contrast to *p*-hydroxyphenylacetate decarboxylase (CsdBC or  
205HpdBC), which has two ([Fig. 2](#)), and does not share conserved CsdB residues postulated to  
206interact with the *para*-hydroxy group (e.g., active-site residue E637 of CsdB<sup>23</sup>). Furthermore, *p*-  
207hydroxyphenylacetate decarboxylase (CsdBC) does not act on phenylacetate<sup>8</sup>, and conversely,

208PhdB has far lower activity on *p*-hydroxyphenylacetate than on phenylacetate (Extended Data  
209Fig. 64). Based upon the sole structural feature that differentiates the substrates of PhdB and *p*-  
210hydroxyphenylacetate decarboxylase (CsdBC/HpdBC), namely a *para*-hydroxy group, and its  
211essential role in the proposed mechanism of the latter enzyme, it is likely that PhdB and  
212CsdBC/HpdBC differ mechanistically. The Kolbe-type decarboxylation proposed for CsdBC<sup>23,28</sup>  
213involves an unprecedented mechanism for *p*-hydroxyphenylacetate activation: a concerted  
214abstraction of a proton from the *para*-hydroxy group by E637 and abstraction of an electron from  
215the carboxyl group by C503<sup>23</sup>. Together, the proton and electron abstraction constitute a *de facto*  
216H-atom abstraction, although the abstraction occurs in two distinct locations on the substrate  
217molecule. Molecular modeling of the substrate-bound active sites of PhdB (based on homology  
218modeling) and CsdBC (based on crystallographic data) indicates important conserved residues,  
219such as the sites of the thiyl radical (C482 in PhdB and C503 in CsdB) and glycy radical (G815  
220in PhdB and G873 in CsdB), but also important differences, such as a hydrophobic pocket in  
221PhdB (including W495, Y691, and V693) accommodating the unsubstituted ring of  
222phenylacetate and lacking the H536 and E637 residues in CsdB that are proposed to interact with  
223the *para*-hydroxy group of *p*-hydroxyphenylacetate (Extended Figure 8).

224       Just as PhdB represents a novel glycy radical enzyme, PhdA represents a new glycy  
225radical activating enzyme. Whereas PhdA shares some characteristics of the cognate activating  
226enzymes for the seven GREs described above, such as a conserved CxxxCxxC [4Fe-4S]-binding  
227motif near the N-terminus of the protein (Fig. 5c), its sequence identity to these activating  
228enzymes is relatively low (from ca. 23 to 42% for both the sewage and lake culture versions of  
229PhdA; Extended Data Fig. 75). To date, studies have indicated that glycy radical activating  
230enzymes are not interchangeable but rather are specific to their cognate glycy radical enzymes<sup>16</sup>.

### 231 Identity of toluene-producing bacterium

232 As toluene synthase discovery was conducted with the proteome of a complex microbial  
233 community rather than that of a microbial isolate, the task of identifying the microbe whose  
234 genome encodes *phdA* and *phdB* was challenging. Nonetheless, we were able to recover the  
235 draft genome of the bacterium in the sewage community that putatively expressed *phdA* and  
236 *phdB* (Fig. 6a). This 3.61-Mbp genome (Fig. 6a, Supplementary Data File 5), which resulted  
237 from co-assembly of Illumina reads from multiple metagenome sequences produced from the  
238 sewage culture, is estimated to be 96.35% complete and contains a 51.8-kb contig including the  
239 three-gene *phd* cluster (Fig. 2) relevant to toluene biosynthesis. In addition to *phdA* and *phdB*,  
240 the genome encodes other putative radical-related enzymes (Fig. 6a), including a GRE of  
241 unknown function (TOLSYN\_01027) and seven putative radical SAM enzymes that contain the  
242 CxxxCxxC motif near the N terminus (TOLSYN\_00781, TOLSYN\_01308, TOLSYN\_01024,  
243 TOLSYN\_00072, TOLSYN\_00941, TOLSYN\_02430, and TOLSYN\_01488).

244 The recovered genome contained a partial 16S rRNA gene indicating that the toluene-  
245 producing bacterium (hereafter referred to as *Acidobacteria* strain Tolsyn) belongs to the  
246 *Acidobacteria* phylum (Extended Data Fig. 96). The closest match among bacterial isolates is to  
247 *Candidatus* Koribacter versatilis (95% identity), which is classified in Subdivision 1 of the  
248 *Acidobacteria* but is not well characterized with respect to its physiology and metabolism<sup>29</sup>.  
249 Evaluation of the recovered genome against the available *Acidobacteria* isolate genomes using  
250 129 concatenated proteins (including 33 ribosomal proteins) indicated, as did the 16S rRNA  
251 analysis, that the closest isolated relative is *Ca.* Koribacter versatilis (Fig. 6b). However, the  
252 genomes of *Acidobacteria* strain Tolsyn and *Ca.* Koribacter versatilis are much less similar than  
253 the 16S rRNA comparison would suggest: average sequence identity for the proteins in these two

254genomes was only ca. 56%. Admittedly, there are few *Acidobacteria* isolates for comparison to  
255strain Tolsyn, as *Acidobacteria* are notoriously difficult to isolate<sup>29,30</sup>. Notably, BLASTP<sup>31</sup>  
256searches of the *Ca. Koribacter versatilis* genome did not yield any hits to PhdA or PhdB.

257 From an ecological perspective, the selective advantage conferred by toluene production  
258in strain Tolsyn is currently unknown. The metabolic advantages rendered by phenylacetate  
259conversion to toluene are not obvious, as the reaction yields only CO<sub>2</sub>, which is unlikely to be  
260limiting in environments like anoxic lake sediments or sewage sludge, and toluene, which is  
261likely lost from the cell by diffusion and not further metabolized [e.g., benzylsuccinate synthase<sup>22</sup>  
262was not found in the genome nor, indeed, in the entire sewage metagenome (IMG Taxon ID  
2633300001865)]. ~~Further, the PhdB reaction will not provide reducing equivalents to the host  
264because it is not an oxidation-reduction reaction. Here, we present two possible explanations for  
265the selective advantage offered by toluene biosynthesis. By First, by analogy to *p*-  
266hydroxyphenylacetate decarboxylation to *p*-cresol, as catalyzed by the nosocomial pathogen  
267*Peptoclostridium difficile* (formerly *Clostridium difficile*), it is possible that toluene production  
268represents a form of negative allelopathy. In *P. difficile*, production of the bacteriostatic agent *p*-  
269cresol is thought to provide a competitive advantage to the producing strain and has been  
270proposed as a virulence factor<sup>32</sup>. Just as the ultimate source of *p*-hydroxyphenylacetate to *P*.  
271*difficile* is tyrosine metabolism, the source of phenylacetate to strain Tolsyn is likely  
272phenylalanine metabolism<sup>8</sup>, potentially involving transamination of phenylalanine to  
273phenylpyruvate (e.g., via phenylalanine transaminase; EC 2.6.1.57), decarboxylation to  
274phenylacetaldehyde (e.g., via phenylpyruvate decarboxylase; EC 4.1.1.43), and oxidation to  
275phenylacetate (e.g., via phenylacetaldehyde dehydrogenase; EC 1.2.1.39)<sup>33</sup>, although other  
276pathways are possible<sup>34</sup>. Notably, BLASTP searches of the *Acidobacteria* strain Tolsyn genome~~

277did not reveal definitive copies of genes encoding any of these enzymes, suggesting that the  
278conversion of phenylalanine to phenylacetate may not occur within strain Tolsyn, but rather that  
279phenylacetate may be imported from its environment. Regardless of which microorganisms are  
280converting phenylalanine to phenylacetate, previous studies have documented that the  
281conversion of labeled phenylalanine (L-phenylalanine- $\beta$ - $^{13}\text{C}$ ) to labeled toluene ([methyl-  
282 $^{13}\text{C}$ ]toluene) definitively occurs in this sewage culture<sup>8</sup>.

283       The prospect of phenylacetate import into *Acidobacteria* strain Tolsyn introduces a  
284second possible explanation for the selective advantage offered by toluene biosynthesis:  
285intracellular pH homeostasis and/or development of a proton motive force (pmf). If the anion  
286phenylacetate were imported into the cell, the PhdB-catalyzed decarboxylation to toluene  
287consumed a proton from the cytoplasm (consistent with the balanced reaction of  $\text{C}_8\text{H}_7\text{O}_2^- + \text{H}^+ \rightarrow$   
288 $\text{C}_7\text{H}_8 + \text{CO}_2$ ), and the neutral reaction products toluene and  $\text{CO}_2$  (or  $\text{H}_2\text{CO}_3$ ) exited the cell (e.g.,  
289by diffusion), the result would be alkalinization of the cytoplasm and indirect development of a  
290pmf (by depletion of protons from the cytoplasm rather than the canonical pumping of protons  
291across the cytoplasmic membrane). Studies of tyrosine and histidine decarboxylation in  
292*Enterococcus* and *Lactobacillus* spp. have experimentally supported analogous mechanisms for  
293pmf development and intracellular pH regulation<sup>35,36</sup>. Thus, alkalinization of the cytoplasm *via*  
294phenylacetate decarboxylation could promote tolerance to the moderately acidic conditions  
295characteristic of some fermentative environments (such as those used to cultivate the sewage and  
296lake sediment cultures and likely representative of their native habitats) and could also provide a  
297source of energy to the bacterium (as pmf), even though the PhdB reaction would not provide  
298reducing equivalents to the host because it is not an oxidation-reduction reaction.

299

300

### 301**Conclusion**

302 We have discovered a GRE that catalyzes an activity heretofore unavailable to  
303biotechnology, enabling biochemical synthesis of toluene (and potentially other products of  
304aromatic acid decarboxylation) from renewable feedstocks. Furthermore, this study, like the  
305recent discovery of another GRE (*trans*-4-hydroxy-L-proline dehydratase<sup>26</sup>), provides a glimpse  
306into the untapped catalytic potential of GREs. It is likely that the catalytic diversity of GREs has  
307been widely underestimated because automated annotation pipelines routinely misidentify  
308diverse GREs as pyruvate formate-lyase (as was the case for PhdB), and there is a dearth of  
309experimental data to correct such misannotation. To illustrate the unexplored diversity of GREs,  
310consider the sewage-derived microbial community investigated in this study. In addition to  
311PhdB, we conservatively estimate that there are at least four other novel GREs represented in the  
312sewage culture metagenome (Fig. 1), as detailed in Extended Data Fig. [107](#). These GREs deviate  
313from known GREs with respect to at least one conserved residue, and share only ca. 16 to 38%  
314protein sequence identity with known GREs and each other. All four of these putatively novel  
315GREs were misannotated as pyruvate formate-lyase by an automated pipeline. Further  
316experimental characterization of the catalytic range of GREs promises to expand our  
317understanding of the metabolic diversity of anaerobic bacteria and the reach of biotechnology to  
318catalyze challenging reactions.

### 319**METHODS**

320 Unless stated otherwise, all cultivation and biochemical processes were conducted under  
321strictly anaerobic conditions<sup>37</sup> in an anaerobic glove box (Type B, Coy Laboratory Products, Inc.,  
322Grass Lake, MI) with a nominal gas composition of 85% N<sub>2</sub> – 10% CO<sub>2</sub> – 5% H<sub>2</sub> (ultra-high

323purity, anaerobic mixture) maintained at ambient temperature (~22°C). Glass, plastic, and  
324stainless steel materials used to manipulate microbial cells, cell-free extracts, and purified  
325enzymes in the glove box were allowed to degas in the anaerobic glove box for at least one day  
326before use, as were heat-labile solids that could not be prepared in autoclaved and purged  
327solutions. Highly purified water (18 MΩ resistance) obtained from a Barnstead Nanopure  
328system (Thermo Scientific, Waltham, MA) was used to prepare all aqueous solutions described  
329in this article. Chemicals used in this study were of the highest purity available and were used as  
330received.

### 331 **Cultivation of anaerobic sewage and lake sediment cultures**

332 Anaerobic cultivation of sewage-derived cultures has been described previously<sup>8</sup>. In a  
333similar fashion, reducing sediments from a lake in Berkeley, California, were used to inoculate  
334cultures under anaerobic conditions using TP<sup>9</sup> or modified TP<sup>8</sup> growth medium in an anaerobic  
335glove box. Amended phenylacetate (typically 200 μM) and evolved toluene were monitored by  
336LC/MS and GC/MS, respectively, using methods described previously<sup>8</sup>.

### 337 **Partial purification of phenylacetate decarboxylase activity in sewage cultures with FPLC**

338 As described in detail elsewhere<sup>8</sup>, cell-free extracts from the sewage-derived culture were  
339generated under strictly anaerobic conditions with a French pressure cell<sup>19</sup> (138 MPa) and  
340clarified by ultracentrifugation, before being subjected to FPLC fractionation in an anaerobic  
341glove box with a Bio-Scale Mini CHT-II ceramic hydroxyapatite column (5-mL bed volume, 40-  
342μm particle diameter; Bio-Rad, Hercules, CA) and Bio-Rad Econo Gradient Pump. ~~Toluene~~  
343~~synthase~~Phenylacetate decarboxylase activity in FPLC fractions was determined with a GC/MS  
344static headspace assay that measured conversion of phenylacetic acid-2-<sup>13</sup>C (Icon Isotopes,  
345Summit, NJ; 99 atom% <sup>13</sup>C) to [*methyl*-<sup>13</sup>C]toluene<sup>8</sup>.

346

#### 347**Proteomic analysis of FPLC fractions by LC/MS/MS**

348 Details on proteomic analysis of selected FPLC fractions, including data processing, were  
349provided by Zargar et al.<sup>8</sup>. Briefly, proteomic LC/MS/MS analysis was performed with a Q  
350Exactive Orbitrap mass spectrometer (Thermo Scientific) in conjunction with a Proxeon Easy-  
351nLC II HPLC (Thermo Scientific) and Proxeon nanospray source.

#### 352**Characterization of sewage and lake cultures by next-generation sequencing of** 353**metagenomes and PCR-amplified 16S rRNA genes**

354 Extraction of genomic DNA from toluene-producing cultures was performed with a bead-  
355beating method involving hexadecyltrimethylammonium bromide (CTAB) extraction buffer  
356described elsewhere<sup>8</sup>. Genomic DNA was purified with Allprep DNA/RNA kits (Qiagen,  
357Valencia, CA). Methods used by the Joint Genome Institute (JGI) for metagenome library  
358construction, next-generation sequencing (Illumina and PacBio), and assembly for sewage- and  
359lake-derived cultures are summarized in Supplementary Data File 6, along with accession  
360numbers for NCBI's SRA (Sequence Read Archive;  
361<https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi>). The automated annotation pipeline for  
362metagenome sequences was described previously<sup>38</sup>.

363 Composition of the sewage-derived community was analyzed at the JGI by Illumina  
364sequencing of 16S rRNA genes amplified from the V4 region (primers 515F and 806R). Library  
365construction and sequencing methods are described in Supplementary Data File 6, and data  
366analysis with iTagger v. 1.1 was performed as described previously<sup>8</sup>.

367 Composition of the lake sediment-derived community was also assessed by Illumina  
368sequencing of 16S rRNA genes amplified from the V4 region (primers 515F and 806R). Library



369construction was performed according to the Earth Microbiome Project standard protocol  
370(<http://www.earthmicrobiome.org/protocols-and-standards/16s/>). Sequencing was conducted by  
371the QB3-Berkeley Core Research Facility at UC Berkeley on the Illumina MiSeq platform (San  
372Diego, CA) with paired-end, 300-bp reads (MiSeq Reagent Kit v3, 600 cycle). The UPARSE  
373method was used for sequence processing and operational taxonomic unit (OTU) clustering at  
37497% identity to process raw sequences (fastq\_maxdiffs=3, fastq\_truncLen=250,  
375fastq\_maxee=0.1). A set of 217 OTUs from a total of 108,041 filtered sequences were identified.  
376For each OTU, a representative sequence was selected as described by Edgar<sup>39</sup>. Taxonomic  
377assignments were made with a Naïve Bayes Classifier using the V4 region of the SILVA<sup>40</sup> SEED  
378sequences and their taxonomic identities as a training set.

#### 379Cloning, expression, *in vitro* reconstitution, and purification of PhdA and PhdB

380 Bacterial strains and plasmids used in this study are listed in Extended Data Table 1.  
381Strains and plasmids along with their associated information (annotated GenBank-format  
382sequence files) have been deposited in the public version of the JBEI Registry ([https://public-  
383registry.jbei.org](https://public-registry.jbei.org); entries JPUB\_XXX to JPUB\_XXX) and are physically available from the authors  
384and/or addgene (<http://www.addgene.org>) upon request [Note to editor: JPUB names will be  
385generated and made public upon publication of the manuscript]. Restriction enzymes were  
386purchased from Thermo Scientific (Waltham, MA), and Phusion DNA polymerase and T4 ligase  
387were from New England Biolabs (Ipswich, MA). Plasmid extractions were carried out using  
388Qiagen (Valencia, CA) miniprep kits. Oligonucleotide primers were designed using the web-  
389based PrimerBlast program ([http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?  
390LINK\\_LOC=BlastHomeAd](http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHomeAd)) and synthesized by Integrated DNA Technologies (IDT), Inc. (San  
391Diego, CA) or Eurofins MWG Operon (Huntsville, AL).

392 *phdA* and *phdB* were codon optimized (GenScript, Piscataway, NJ) for expression in  
393 *E.coli* BL21(DE3) (listed as Sequences 3 and 8 in Supplementary Data File 4). Each codon-  
394 optimized gene was individually cloned into plasmid pET28b (Novagen, Madison, WI). *phdA*  
395 was cloned between NdeI and BamHI restriction sites (primers listed in Extended Data Table 2),  
396 resulting in a construct that encodes an N-terminal His<sub>6</sub>-PhdA protein (pAS004; Extended Data  
397 Table 1). *phdB* was cloned between NdeI and XhoI restriction sites (primers listed in Extended  
398 Data Table 2). To enhance soluble PhdB yield, the construct also included the gene encoding  
399 maltose-binding protein (MBP) and a sequence encoding the tobacco etch virus (TEV) protease  
400 recognition site, which were inserted downstream of the N-terminal His<sub>6</sub> sequence and upstream  
401 of the *phdB* start codon, resulting in a construct that encodes a His<sub>6</sub>-MBP-PhdB fusion protein  
402 with a TEV protease-cleavable His<sub>6</sub>-MBP tag (pAS010, Extended Data Table 1). Plasmids were  
403 transformed into chemically competent *E.coli* DH10B cells grown on lysogeny broth (LB) agar  
404 plates under 50 µg/mL kanamycin selection (LB Kan-50 plates; Teknova, Hollister, CA).  
405 Plasmids were sequence-confirmed (Genewiz, South San Francisco, CA). Plasmids pAS006  
406 (with *phdA*) and pAS010 (with *phdB*) were separately transformed into chemically competent  
407 *E.coli* BL21(DE3) cells (New England Biolabs) on LB Kan-50 plates. Transformants were  
408 grown in LB broth and stored as 100 µL glycerol stock aliquots at -80°C.

409 For overexpression of PhdA, a frozen glycerol stock of strain AS013 (Extended Data  
410 Table 1) was used to inoculate 50 mL LB broth containing 50 µg/mL kanamycin (Teknova) in a  
411 250-mL shake flask. The starter culture was incubated overnight at 30°C with constant shaking at  
412 200 rpm. For larger scale growth, the starter culture was diluted 100-fold in a 2-L baffled shake  
413 flask containing 1L LB broth supplemented with 50 µg/mL kanamycin, and grown aerobically at  
414 37°C with constant shaking (190 rpm). At OD<sub>600</sub> ~0.7, the culture was induced with isopropyl β-

415D-1-thiogalactopyranoside (IPTG; IBI Scientific, Peosta, IA) to a final concentration of 0.5 mM  
416and supplemented with an aqueous solution of  $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$  (Sigma, St. Louis, MO;  
417prepared anaerobically) to a final concentration of 200  $\mu\text{M}$ . Following induction, the temperature  
418was decreased to 18°C and the culture was propagated overnight at this temperature for ~18  
419hours. Cells were then harvested by centrifugation and cell pellets were stored at -80°C until  
420lysis.

421 For overexpression of PhdB, strain AS019 (Extended Data Table 1) was cultivated in  
422autoinduction medium<sup>41</sup>. A frozen glycerol stock was used to inoculate 50 mL ZYP-0.8G  
423medium containing 100  $\mu\text{g}/\text{mL}$  kanamycin in a 250-mL shake flask incubated overnight at 30°C  
424with constant shaking (200 rpm). The starter culture was diluted 100-fold into a 2-L baffled  
425shake flask containing 1-L ZYP-5052 medium with 100  $\mu\text{g}/\text{mL}$  kanamycin and grown  
426aerobically at 37°C with constant shaking at 190 rpm. At  $\text{OD}_{600} \sim 1.5$ , the temperature was  
427decreased to 18°C and the culture was propagated overnight at this temperature for ~18 hours.  
428Cells were then harvested by centrifugation and cell pellets were stored at -80°C until lysis.

429 All purification steps were carried out under strictly anaerobic conditions. For lysis, cells  
430were passed three times through a French pressure cell (138 MPa) under anaerobic conditions.  
431Sealed lysates were centrifuged under anaerobic conditions at 19,000 rpm at 4°C for 40 min.  
432Clarified lysates were purified within an anaerobic glove box as described below using an  
433Econo-Gradient pump coupled with a model 2110 fraction collector (Bio-Rad).

434 For PhdA purification, strain AS013 cell pellets were resuspended in buffer A [50 mM  
435TRIS (pH 7.5; EMD Millipore, Billerica, MA), 300 mM NaCl (EMD Millipore), 10 mM  
436imidazole (Sigma), 0.1 mM DL-dithiothreitol (DTT; VWR, Visalia, CA)] and mixed with  
437powdered protease inhibitors (Pierce EDTA-free tablets, Thermo Scientific), chicken egg

438lysozyme (300  $\mu\text{g}/\text{mL}$ , Sigma) and DNaseI (10  $\mu\text{g}/\text{mL}$ , Sigma). This mixture was incubated for  
43920 min followed by cell lysis and clarification of the lysate as described above. The clarified  
440lysate was filtered through a 0.45- $\mu\text{m}$  filter (EMD Millipore) and loaded onto a 5-mL HisTrap  
441HP column (GE Healthcare, Chicago, IL) that was pre-equilibrated with buffer A. The column  
442was then washed with 3 column volumes (CV) of buffer A to remove unbound proteins and  
443eluted using a stepwise imidazole gradient made by mixing buffer A with buffer B [50 mM TRIS  
444(pH 7.5), 300 mM NaCl, 500 mM imidazole, 0.1 mM DTT] using stepwise concentrations of 20  
445mM, 50 mM, 250 mM, and 400 mM imidazole. Each step was set to 1.6 CV and 2-mL fractions  
446were collected. Fractions containing PhdA were dark red-brown and eluted at a concentration of  
447250 mM imidazole. The purity of PhdA fractions was confirmed by SDS-PAGE. Elution  
448fractions were pooled and DTT was added to a final concentration of 2 mM. To keep the protein  
449anoxic during concentration outside the glove box, a 10-kDa molecular weight cutoff (MWCO)  
450concentrator (EMD Millipore) was sealed inside a 250-mL centrifuge bottle (Nalgene, Rochester,  
451NY) with an O-ring-sealed cap. Concentrated protein was exchanged into buffer C [50 mM TRIS  
452(pH 7.5), 300 mM NaCl, 5 mM DTT] using a pre-equilibrated PD-10 desalting column (GE  
453Healthcare). Protein concentration was determined using the Bradford assay (Bio-Rad).  
454Collected UV-visible spectra (UV-2450; Shimadzu Scientific, Pleasanton, CA) indicated the  
455presence of [2Fe-2S] clusters bound to the protein (Extended Data Figure [118](#))<sup>42</sup>.

456 For reconstitution of [4Fe-4S] clusters in PhdA, which are required for activity, the  
457protein was diluted to 0.2 mM in buffer C in a stoppered serum bottle and cooled to 4°C. DTT  
458was then added to a final concentration of 10 mM and the solution was incubated at 4°C for ~1  
459hour. Aqueous  $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$  was added to a final concentration of 1 mM and incubated  
460at 4°C for ~3-4 hours. Aqueous  $\text{Na}_2\text{S} \cdot 9\text{H}_2\text{O}$  was then added to a final concentration of 0.9 mM

461 and the mixture was incubated at 4°C overnight (~18 hr). The protein mixture was then filtered  
462 through a 0.45-µm filter, concentrated, and diluted 15-fold in buffer D [50 mM TRIS (pH 7.5),  
463 20 mM NaCl, 2 mM DTT]. The diluted protein was then loaded onto a 5-mL Bioscale High Q  
464 column (Bio-Rad) that was pre-equilibrated with buffer D and eluted using buffer E [50 mM  
465 TRIS (pH 7.5), 1 M NaCl, 2 mM DTT] with a stepwise NaCl gradient of concentrations 40 mM,  
466 100 mM, 500 mM, and 800 mM NaCl. Each step was set to 1.6 CV and 2-mL fractions were  
467 collected. PhdA eluted at a concentration of ~500 mM NaCl and fractions were yellow-brown.  
468 Purity of eluted fractions was confirmed by SDS-PAGE. Pooled fractions were concentrated and  
469 exchanged into assay buffer [50 mM TRIS (pH 7.5), 150 mM NaCl, 1 mM MgCl<sub>2</sub> (Sigma), 5  
470 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (Sigma), 5 mM DTT] using a pre-equilibrated PD-10 column and stored at 4°C in  
471 a stoppered serum bottle. Protein concentration was determined using the Bradford assay. UV-  
472 visible spectra confirmed the presence of [4Fe-4S] clusters bound to the protein (Extended Data  
473 Figure [118](#))<sup>42</sup>.

474 For PhdB purification, strain AS019 (Extended Data Table 1) cell pellets were washed in  
475 buffer containing 50 mM TRIS (pH 7.5), 150 mM NaCl, and 0.5 mM dithionite. For purification,  
476 cell pellets were resuspended in buffer A [20 mM TRIS (pH 7.5), 200 mM NaCl, 1 mM EDTA  
477 (EMD Millipore), 1 mM DTT] and mixed with powdered protease inhibitors, chicken egg  
478 lysozyme (300 µg/mL) and DNaseI (10 µg/mL). This mixture was incubated for 20 minutes,  
479 followed by cell lysis with a French pressure cell under anaerobic conditions and clarification of  
480 the lysate as described for PhdA. The clarified lysate was filtered through a 0.45-µm filter  
481 (Millipore) and loaded on to a 5 mL-MBPTrap HP column (GE Healthcare) that was pre-  
482 equilibrated with buffer A. The column was then washed with 3 CV of buffer A to remove  
483 unbound proteins and eluted using a program consisting of a stepwise maltose gradient made by

484mixing buffer A with buffer B [20 mM TRIS (pH 7.5), 200 mM NaCl, 1 mM EDTA, 10 mM  
485maltose (Sigma), 1 mM DTT] using concentrations of 0.4 mM, 1 mM, 5 mM, and 8 mM  
486maltose. Each step was set to 1.6 CV and 1-mL fractions were collected. PhdB eluted at a  
487concentration of ~1 mM maltose and purity of fractions was confirmed by SDS-PAGE. Elution  
488fractions were pooled and DTT was added to a final concentration of 2 mM and the protein was  
489concentrated anaerobically as described for PhdA (except with a 50-kDa MWCO rather than 10-  
490kDa MWCO). Concentrated protein was exchanged into assay buffer [50 mM TRIS (pH 7.5),  
491150 mM NaCl, 1 mM MgCl<sub>2</sub>, 5 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 5 mM DTT] using a pre-equilibrated PD-10  
492desalting column (GE Healthcare). Protein concentration was determined using the Bradford  
493assay (Bio-Rad). During initial purifications, the identity of the protein was confirmed by  
494Western blotting using HRP-conjugated anti-MBP antibody (New England Biolabs).

#### 495**Anaerobic *in vitro* assays for PhdA activity with recombinant protein**

496 In an anaerobic chamber at ambient temperature, 0.7 mM reconstituted PhdA was  
497incubated in assay buffer [50 mM TRIS (pH 7.5), 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 5 mM  
498(NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 5 mM DTT] with 2 mM dithionite (Sigma) for 1 hour in 4-mL screw-capped glass  
499vials (Supelco). This was followed by the addition of 2 mM SAM [*S*-(5'-adenosyl)-L-methionine  
500chloride dihydrochloride; Sigma]. The reaction mixture (1.2 mL) was shaken at low speed on a  
501tabletop orbital shaker. After initiation of the PhdA reaction by SAM addition, sampling was  
502conducted from 0 to 180 min at 30-min intervals. Immediately upon sampling, 75 µL of reaction  
503mixture was quenched by addition of 75 µL LC/MS grade methanol (Honeywell Research  
504Chemicals, Muskegon, MI) and gentle bubbling of 0.5 mL of air (from a sealed serum bottle).  
505Control reaction mixtures excluding PhdA were assayed in an identical manner. Post quenching,  
506samples were centrifuged at 13,000 rpm for 15 min, then diluted in 50% (v/v) methanol in

507LC/MS grade water (J.T. Baker, Phillipsburg, NJ) in preparation for LC/MS measurement.

508Replicates involved separate assays rather than multiple analyses of a given assay sample.

509 For analysis of methionine produced by PhdA activity with SAM, external standard  
510quantification was performed with five-point calibration standards ranging from 0.25-10  $\mu$ M  
511methionine (Sigma) in 50/50 (v/v) methanol/water. Samples were run on an LC/MSD SL  
512(Agilent, Santa Clara, CA) equipped with a model 1260 Infinity Binary Pump and operated in  
513the electrospray ionization, positive-ion mode. The mobile phase initially flowed at 0.6 mL/min  
514(0 - 13 min), and later at 1 mL/min (13-15 min), through a Kinetex HILIC column (2.6- $\mu$ m  
515particle size, 4.6-mm inner diameter x 50-mm length; Phenomenex, Torrance, CA). The initial  
516mobile phase composition was 10 vol% A (20 mM ammonium acetate in water) and 90 vol% B  
517(10 mM ammonium acetate in 90% acetonitrile, 10% water), which was decreased linearly to  
51870% B at 4 minutes, then decreased linearly to 40% B from 6 - 11.5 minutes, and then increased  
519linearly to 90% B from 12 - 15 minutes to re-equilibrate the column to initial conditions. Sample  
520injection volume was 2  $\mu$ L. Source conditions included 3.5 kV capillary voltage, 250°C drying  
521gas temperature, 12 L/min drying gas flow, and 25 psig nebulizer pressure. Data acquisition for  
522methionine was in the selected ion monitoring (SIM) mode at  $m/z$  150.2. Peak areas were  
523integrated using Mass Hunter software (Agilent, version B.05.00).

#### 524**Anaerobic *in vitro* assays for phenylacetate decarboxylase activity with recombinant PhdA** 525**and PhdB**

526 Assays for phenylacetate decarboxylase activity were performed under strictly anaerobic  
527conditions within a glove box. Assays, which were performed in 4-mL glass vials sealed with 13-  
528mm diameter PTFE Mininert screw-cap valves (Sigma-Aldrich), contained 250  $\mu$ M PhdA in  
529assay buffer [50 mM TRIS (pH 7.5), 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 5 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 5 mM

530 DTT)], to which 2 mM dithionite was added and incubated for ~1 hour, followed by the addition  
531 of 2 mM SAM, 2.5  $\mu$ M PhdB in assay buffer, and 2.5 mM phenylacetic acid- $^{13}\text{C}$  in a final  
532 volume of 1 or 1.5 mL (depending on the specific experiment). Quantitative standards contained  
533 the same headspace/liquid ratios as assays and a dimensionless Henry's constant of 0.27<sup>43</sup> was  
534 used to calculate aqueous concentration. Negative controls were run concurrently and were  
535 identical except for the absence of SAM (Figure 3b) or other conditions specified in Extended  
536 Figure 4. The vials were shaken on a tabletop orbital shaker at low speed. Gaseous headspace  
537 samples (100  $\mu$ L) were taken within the glove box using a 500- $\mu$ L gastight syringe (Sample-Lok  
538 series A-2; Sigma-Aldrich) and analyzed immediately by GC/MS, as described previously<sup>8</sup>.  
539 Briefly, toluene was analyzed by static headspace, electron ionization (EI) GC/MS using a model  
540 7890A GC (Agilent, Santa Clara, CA) with a DB-5 fused silica capillary column (30-m length,  
541 0.25-mm inner diameter, 0.25- $\mu$ m film thickness; Agilent) coupled to an HP 5975C series  
542 quadrupole mass spectrometer. As described elsewhere<sup>8</sup>, the identity of [*methyl- $^{13}\text{C}$* ]toluene was  
543 confirmed with the expected  $m/z$  93/92 ratio of 0.6. Replicates involved separate assays rather  
544 than multiple analyses of a given assay sample. In assays testing whether PhdB could  
545 decarboxylate *p*-hydroxyphenylacetate to *p*-cresol, conditions were as described above except  
546 that equimolar amounts (2.5 mM) of *p*-hydroxyphenylacetic acid (Sigma) and phenylacetic acid-  
547  $^{13}\text{C}$  were added, and GC/MS analysis of *p*-cresol in 1- $\mu$ L liquid injections of concentrated  
548 hexane extracts were conducted as described previously<sup>8</sup>. The identity of *p*-cresol was assessed  
549 using retention time and the expected  $m/z$  108/107 ratio of 0.83 based on authentic standards.

#### 550 **PCR amplification of *phd* gene cluster from genomic DNA from lake sediment culture**

551 *phdA*, *phdB*, and an adjacent putative transcription factor were PCR-amplified from  
552 genomic DNA extracted from the lake sediment community using primers shown in Extended



553Data Table 2. Primer design was guided in part by partial gene sequences available from  
554metagenomes (IMG Taxon ID 2100351000 and 3300001865). Amplified and gel-purified DNA  
555was sequenced by Genewiz.

### 556**Construction of maximum likelihood tree of glycy radical enzymes in sewage-derived** 557**culture**

558 The maximum-likelihood tree in Fig. 1 encompasses protein sequences of putative glycy  
559radical enzymes (GREs) detected in the sewage culture metagenome (IMG Taxon ID  
5603300001865) based on BLASTP<sup>31</sup> searches against known GREs (> 30% sequence identity),  
561searches for the glycy radical motif (FIMO<sup>44</sup>), and a minimum length of 171 amino acids (not all  
562were full length). The following model sequences were also included in the tree to provide  
563context (accession numbers in parenthesis): PflB (GenBank: NP\_415423), HpdB (GenBank:  
564AJ543425.1), CsdB (GenBank: ABB05046.1), CutC (PDB: 5A0Z), NrdD (GenBank:  
565NP\_418659), and Gdh (PDB: 1R8W). The collected set of model and putative GRE sequences  
566( $n=81$ , mean =  $675\pm 194$  aa) were aligned using MUSCLE v. 3.8.31<sup>45</sup>. The resulting MSA was  
567screened for ambiguous C and N termini as well as columns with >97% gaps. The final  
568alignment spanned 1138 columns. A maximum likelihood phylogenetic tree was inferred with  
569RAxML v. 7.6.3<sup>46</sup>, under the LG plus Gamma model of evolution as follows:

```
570raxmlHPC-PTHREADS-SSE3 -# 50 -m PROTGAMMAGTR -p 777 -x 2000 -f a
```

571The tree was constructed with iTOL<sup>47</sup>.

### 572**Binning of sewage culture metagenomes and recovery of *Acidobacteria* strain Tolsyn** 573**genome**

574 For binning, two groups of sewage metagenomes (Group 1 from SRA accession numbers  
575SRP077640, SRP072654, and SRP099295 and Group 2 from SRA accession numbers

576SRP105442 and SRP105443) were separately co-assembled using metaSPAdes v3.6<sup>48</sup> with the  
577"--careful" option. The two co-assemblies were separately binned using MaxBin 2.0<sup>49</sup> with  
578default parameters (-min\_contig\_len 1000). The *Acidobacteria* strain Tolsyn bins were separately  
579identified within the two co-assemblies, and scaffolds that were shared (with >98% identity)  
580were selected to constitute the draft *Acidobacteria* genome. The scaffolds were further refined by  
581mapping against the hybrid assemblies of the sewage sludge samples (IMG Taxon ID  
5823300017643, 3300017642, and 3300017814) and extracting scaffolds that unambiguously  
583connected two or more sequences in the draft *Acidobacteria* genome. Genes were predicted from  
584the genome using Prodigal (parameter: -p meta)<sup>50</sup>. Amino acid sequence identity between the  
585draft Tolsyn genome and the *Ca. Koribacter versatilis* genome was carried out by comparing  
586predicted proteins from the two genomes using BLASTP<sup>31</sup> with an e-value cutoff of 1e-10 and  
587coverage cutoff 0.4. Annotation was performed by matching identical genes identified by the  
588IMG pipeline (IMG Taxon ID 3300001865) using BLASTP with minimum amino acid identity  
589set to 95% and minimum coverage set to 40%; the best matching IMG annotations were then  
590assigned for those genes. CheckM software<sup>51</sup> reported that the genome was 96.35% complete  
591with a contamination ratio of 1.69%. The circular genome plot (Fig. 6a) was made using  
592Circos<sup>52</sup>. The 16S rRNA gene was identified as follows. A partial 16S rRNA gene (756 bp) was  
593identified in a 1.7-kb scaffold and was 100% identical to a 16S rRNA gene identified from 16S  
594rRNA iTag analysis: *Acidobacteria* OTU (Operational Taxonomic Unit) #9 (Supplementary Data  
595File 3). When OTU9 was used as query sequence for BLASTN searches of the sewage culture  
596metagenome (IMG Taxon ID 3300001865), it had a 100% match with scaffold  
597JGI2065J20421\_1000212, which contained a 1382-bp 16S rRNA gene

598(JGI2065J20421\_10002126). As a result, the partial 16S rRNA gene in the *Acidobacteria* strain 599Tolsyn genome was replaced by the 1382-bp 16S rRNA gene.

#### 600**Construction of phylogenetic trees for *Acidobacteria* strain Tolsyn**

601 The 16S rRNA tree (Extended Data Fig. 96) was constructed by aligning selected 16S  
602rRNA gene sequences using MUSCLE<sup>45</sup> and then applying FastTree<sup>53</sup> to the alignment file. The  
603concatenated protein tree (Fig. 6b) was constructed with ezTree  
604(<https://github.com/yuwwu/ezTree>; manuscript under review), a pipeline for identifying single-  
605copy marker genes from a collection of complete or draft genomes and using the marker genes to  
606generate a concatenated protein tree.

#### 607Molecular modeling of PhdB in complex with its phenylacetate substrate

608 A molecular model of PhdB (Extended Figure 8) was created based on homology  
609modeling of three-dimensional protein structures implemented in the program SWISS-  
610MODEL<sup>54</sup>. The GRE 1,2-propanediol dehydratase from *Roseburia inulinivorans* (PDB ID:  
6115I2A), which shares 32% sequence identity with PhdB, was used as a template to generate the  
612molecular model of PhdB. Superposition of the CsdB in complex with *p*-hydroxyphenylacetate  
613(PDB ID: 2YAJ)<sup>28</sup> against the molecular model of PhdB with the program COOT<sup>55</sup> was used to  
614extract the binding position of phenylacetate. A structure idealization of PhdB-phenylacetate  
615using REFMAC<sup>56</sup> enabled the final molecular model of the complex. The overall stereochemical  
616quality of the final models was assessed using the program MolProbity<sup>57</sup>.

#### 617**Data Availability**

618 Data that support the findings of this study are available within the paper, its supplementary  
619information files, and data repositories cited therein [including JGI's IMG-M site  
620(<https://img.jgi.doe.gov/cgi-bin/mer/main.cgi>), NCBI's Sequence Read Archive

621(<https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi>), and the public version of the JBEI Registry  
622(<https://public-registry.jbei.org>), which contains strains, plasmids, and their associated  
623information].

## 624References

625

6261 Galperin, M. Y. & Koonin, E. V. From complete genome sequence to 'complete'

627 understanding? *Trends Biotechnol.* **28**, 398-406, doi:10.1016/j.tibtech.2010.05.006

628 (2010).

6292 Gerlt, J. A. *et al.* The Enzyme Function Initiative. *Biochemistry* **50**, 9950-9962,

630 doi:10.1021/bi201312u (2011).

6313 Anton, B. P. *et al.* The COMBREX project: design, methodology, and initial results.

632 *PLoS Biol.* **11**, e1001638, doi:10.1371/journal.pbio.1001638 (2013).

6334 Lespinet, O. & Labedan, B. Orphan enzymes? *Science* **307**, 42,

634 doi:10.1126/science.307.5706.42a (2005).

6355 Sorokina, M., Stam, M., Medigue, C., Lespinet, O. & Vallenet, D. Profiling the orphan

636 enzymes. *Biol. Direct* **9**, 10, doi:10.1186/1745-6150-9-10 (2014).

6376 McKenna, R. & Nielsen, D. R. Styrene biosynthesis from glucose by engineered *E. coli*.

638 *Metab. Eng.* **13**, 544-554, doi:10.1016/j.ymben.2011.06.005 (2011).

6397 Jüttner, F. & Henatsch, J. J. Anoxic hypolimnion is a significant source of biogenic

640 toluene. *Nature* **323**, 797-798 (1986).

6418 Zargar, K. *et al.* *In vitro* characterization of phenylacetate decarboxylase, a novel enzyme

642 catalyzing toluene biosynthesis in an anaerobic microbial community. *Scientific Reports*

643 **6**, 31362, doi:10.1038/srep31362 (2016).

6449 Fischer-Romero, C., Tindall, B. J. & Juttner, F. *Tolumonas auensis* gen. nov., sp. nov., a

645 toluene-producing bacterium from anoxic sediments of a freshwater lake. *Int. J. Syst.*

646 *Bacteriol.* **46**, 183-188, doi:10.1099/00207713-46-1-183 (1996).

64710 Pons, J. L., Rimbault, A., Darbord, J. C. & Leluan, G. [Biosynthesis of toluene in  
648 *Clostridium aerofetidum* strain WS]. *Ann. Microbiol. (Paris)* **135B**, 219-222 (1984).

64911 Akhtar, M. K., Turner, N. J. & Jones, P. R. Carboxylic acid reductase is a versatile  
650 enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proc.*  
651 *Natl. Acad. Sci. U S A* **110**, 87-92, doi:10.1073/pnas.1216516110 (2013).

65212 Schirmer, A., Rude, M. A., Li, X., Popova, E. & del Cardayre, S. B. Microbial  
653 biosynthesis of alkanes. *Science* **329**, 559-562, doi:10.1126/science.1187936 (2010).

65413 Selmer, T. & Andrei, P. I. *p*-Hydroxyphenylacetate decarboxylase from *Clostridium*  
655 *difficile*. A novel glycyl radical enzyme catalysing the formation of *p*-cresol. *Eur. J.*  
656 *Biochem.* **268**, 1363-1372 (2001).

65714 Yu, L., Blaser, M., Andrei, P. I., Pierik, A. J. & Selmer, T. 4-Hydroxyphenylacetate  
658 decarboxylases: properties of a novel subclass of glycyl radical enzyme systems.  
659 *Biochemistry* **45**, 9584-9592, doi:10.1021/bi060840b (2006).

66015 Selmer, T., Pierik, A. J. & Heider, J. New glycyl radical enzymes catalysing key  
661 metabolic steps in anaerobic bacteria. *Biol. Chem.* **386**, 981-988,  
662 doi:10.1515/BC.2005.114 (2005).

66316 Shisler, K. A. & Broderick, J. B. Glycyl radical activating enzymes: structure,  
664 mechanism, and substrate interactions. *Arch. Biochem. Biophys.* **546**, 64-71,  
665 doi:10.1016/j.abb.2014.01.020 (2014).

66617 Leuthner, B. *et al.* Biochemical and genetic characterization of benzylsuccinate synthase  
667 from *Thauera aromatica*: a new glycyl radical enzyme catalysing the first step in  
668 anaerobic toluene metabolism. *Mol. Microbiol.* **28**, 615-628 (1998).

66918 O'Brien, J. R. *et al.* Insight into the mechanism of the B12-independent glycerol  
670 dehydratase from *Clostridium butyricum*: preliminary biochemical and structural  
671 characterization. *Biochemistry* **43**, 4635-4645, doi:10.1021/bi035930k (2004).

67219 Beller, H. R. & Spormann, A. M. Substrate range of benzylsuccinate synthase from  
673 *Azoarcus* sp. strain T. *FEMS Microbiol. Lett.* **178**, 147-153 (1999).

67420 Becker, A. *et al.* Structure and mechanism of the glycyl radical enzyme pyruvate formate-  
675 lyase. *Nat. Struct. Biol.* **6**, 969-975, doi:10.1038/13341 (1999).

67621 Larsson, K. M., Andersson, J., Sjoberg, B. M., Nordlund, P. & Logan, D. T. Structural  
677 basis for allosteric substrate specificity regulation in anaerobic ribonucleotide reductases.  
678 *Structure* **9**, 739-750 (2001).

67922 Heider, J., Spormann, A. M., Beller, H. R. & Widdel, F. Anaerobic bacterial metabolism  
680 of hydrocarbons. *FEMS Microbiology Reviews* **22**, 459-473 (1998).

68123 Feliks, M., Martins, B. M. & Ullmann, G. M. Catalytic mechanism of the glycyl radical  
682 enzyme 4-hydroxyphenylacetate decarboxylase from continuum electrostatic and  
683 QC/MM calculations. *J. Am. Chem. Soc.* **135**, 14574-14585, doi:10.1021/ja402379q  
684 (2013).

68524 Kalnins, G. *et al.* Structure and function of CutC choline lyase from human microbiota  
686 bacterium *Klebsiella pneumoniae*. *J Biol Chem* **290**, 21732-21740,  
687 doi:10.1074/jbc.M115.670471 (2015).

68825 Craciun, S. & Balskus, E. P. Microbial conversion of choline to trimethylamine requires a  
689 glycyl radical enzyme. *Proc. Natl. Acad. Sci. U S A* **109**, 21307-21312,  
690 doi:10.1073/pnas.1215689109 (2012).

69126 Levin, B. J. *et al.* A prominent glycyl radical enzyme in human gut microbiomes  
692 metabolizes *trans*-4-hydroxy-L-proline. *Science* **355**, doi:10.1126/science.aai8386 (2017).

69327 Funk, M. A., Marsh, E. N. & Drennan, C. L. Substrate-bound structures of  
694 benzylsuccinate synthase reveal how toluene is activated in anaerobic hydrocarbon  
695 degradation. *J. Biol. Chem.* **290**, 22398-22408, doi:10.1074/jbc.M115.670737 (2015).

69628 Martins, B. M. *et al.* Structural basis for a Kolbe-type decarboxylation catalyzed by a  
697 glycyl radical enzyme. *J. Am. Chem. Soc.* **133**, 14666-14674, doi:10.1021/ja203344x  
698 (2011).

69929 Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., van Veen, J. A. & Kuramae, E. E. The  
700 Ecology of *Acidobacteria*: Moving beyond Genes and Genomes. *Front. Microbiol.* **7**,  
701 744, doi:10.3389/fmicb.2016.00744 (2016).

70230 Ward, N. L. *et al.* Three genomes from the phylum *Acidobacteria* provide insight into the  
703 lifestyles of these microorganisms in soils. *Appl. Environ. Microbiol.* **75**, 2046-2056,  
704 doi:10.1128/AEM.02294-08 (2009).

70531 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment  
706 search tool. *J. Mol. Biol.* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).

70732 Dawson, L. F., Stabler, R. A. & Wren, B. W. Assessing the role of *p*-cresol tolerance in  
708 *Clostridium difficile*. *J. Med. Microbiol.* **57**, 745-749, doi:10.1099/jmm.0.47744-0  
709 (2008).

71033 Schneider, S., Mohamed, M. E. S. & Fuchs, G. Anaerobic metabolism of L-phenylalanine  
711 via benzoyl-CoA in the denitrifying bacterium *Thauera aromatica*. *Arch. Microbiol.*  
712 **168**, 310-320 (1997).



71334 Carmona, M. *et al.* Anaerobic catabolism of aromatic compounds: a genetic and genomic  
714 view. *Microbiol. Mol. Biol. Rev.* **73**, 71-133, doi:10.1128/MMBR.00021-08 (2009).

71535 Molenaar, D., Bosscher, J. S., ten Brink, B., Driessen, A. J. & Konings, W. N. Generation  
716 of a proton motive force by histidine decarboxylation and electrogenic  
717 histidine/histamine antiport in *Lactobacillus buchneri*. *J Bacteriol* **175**, 2864-2870  
718 (1993).

71936 Pereira, C. I., Matos, D., San Romao, M. V. & Crespo, M. T. Dual role for the tyrosine  
720 decarboxylation pathway in *Enterococcus faecium* E17: response to an acid challenge and  
721 generation of a proton motive force. *Appl Environ Microbiol* **75**, 345-352,  
722 doi:10.1128/AEM.01958-08 (2009).

72337 Beller, H. R., Legler, T. C. & Kane, S. R. Genetic manipulation of the obligate  
724 chemolithoautotrophic bacterium *Thiobacillus denitrificans*. *Methods Mol. Biol.* **881**, 99-  
725 136, doi:10.1007/978-1-61779-827-6\_5 (2012).

72638 Huntemann, M. *et al.* The standard operating procedure of the DOE-JGI Microbial  
727 Genome Annotation Pipeline (MGAP v.4). *Stand. Genomic Sci.* **10**, 86,  
728 doi:10.1186/s40793-015-0077-y (2015).

72939 Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads.  
730 *Nat. Methods* **10**, 996-998, doi:10.1038/nmeth.2604 (2013).

73140 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data  
732 processing and web-based tools. *Nucleic Acids Res.* **41**, D590-596,  
733 doi:10.1093/nar/gks1219 (2013).

73441 Studier, F. W. Protein production by auto-induction in high density shaking cultures.  
735 *Protein Expr. Purif.* **41**, 207-234 (2005).

73642 Gao, H. *et al.* *Arabidopsis thaliana* Nfu2 accommodates [2Fe-2S] or [4Fe-4S] clusters  
737 and is competent for *in vitro* maturation of chloroplast [2Fe-2S] and [4Fe-4S] cluster-  
738 containing proteins. *Biochemistry* **52**, 6633-6645, doi:10.1021/bi4007622 (2013).

73943 Mackay, D. & Shiu, W. Y. A critical review of Henry's Law constants for chemicals of  
740 environmental interest. *Journal of Physical and Chemical Reference Data* **10**, 1175-1199  
741 (1981).

74244 Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given  
743 motif. *Bioinformatics* **27**, 1017-1018, doi:10.1093/bioinformatics/btr064 (2011).

74445 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high  
745 throughput. *Nucleic Acids Res.* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

74646 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of  
747 large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033  
748 (2014).

74947 Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and  
750 annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242-245,  
751 doi:10.1093/nar/gkw290 (2016).

75248 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to  
753 single-cell sequencing. *J. Comput. Biol.* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).

75449 Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm  
755 to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607,  
756 doi:10.1093/bioinformatics/btv638 (2016).

75750 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site  
758 identification. *BMC bioinformatics* **11**, 119, doi:10.1186/1471-2105-11-119 (2010).

75951 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:  
760 assessing the quality of microbial genomes recovered from isolates, single cells, and  
761 metagenomes. *Genome Res.* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).

76252 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics.  
763 *Genome Res.* **19**, 1639-1645, doi:10.1101/gr.092759.109 (2009).

76453 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution  
765 trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641-1650,  
766 doi:10.1093/molbev/msp077 (2009).

76754 Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure  
768 using evolutionary information. *Nucleic acids research* **42**, W252-258,  
769 doi:10.1093/nar/gku340 (2014).

77055 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta*  
771 *Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).

77256 Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and  
773 guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184-2195,  
774 doi:10.1107/S0907444904023510 (2004).

77557 Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and  
776 nucleic acids. *Nucleic acids research* **35**, W375-383, doi:10.1093/nar/gkm216 (2007).

77758 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence  
778 alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539, doi:10.1038/msb.2011.75  
779 (2011).

780

781 **Supplementary Information** is linked to the online version of the paper at

782 [www.nature.com/nature](http://www.nature.com/nature).

### 783 **Acknowledgments**

784 We thank the following people from JBEI, LBNL, and JGI for their valuable contributions to this  
785 work: Ulas Karaoz, Nathan Hillson, Andy DeGiovanni, ~~Paul Adams, Jose Henrique Pereira~~, Ee-  
786 Been Goh, Edward Baidoo, Xi Wang, Shi Wang, Patrick Sorensen, Suzan Yilmaz, Garima Goyal,  
787 Joshua Heazlewood, Tijana Glavina del Rio, Stephanie Malfatti, Emiley Eloie-Fadrosch, and  
788 Adam Rivers; we also thank Michelle Salemi (UC Davis Genome Center, Proteomics Core  
789 Facility).

790 This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported  
791 by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental  
792 Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National  
793 Laboratory and the U. S. Department of Energy. Work conducted by the Department of Energy  
794 Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of  
795 Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

### 796 **Author Contributions**

797 H.R.B., A.V.R., K.Z. and R.M.S. conceived of and designed the experiments. A.V.R., K.Z.,  
798 H.R.B., A.K.S., and R.S. performed the experiments. H.R.B., Y.W.W., and A.V.R. analyzed the  
799 data. S.G.T. oversaw metagenomic data production and C.J.P. oversaw metaproteomic data  
800 production. The manuscript was written by H.R.B. (primarily) and all authors, including J.D.K.,  
801 contributed to refining the text.

802

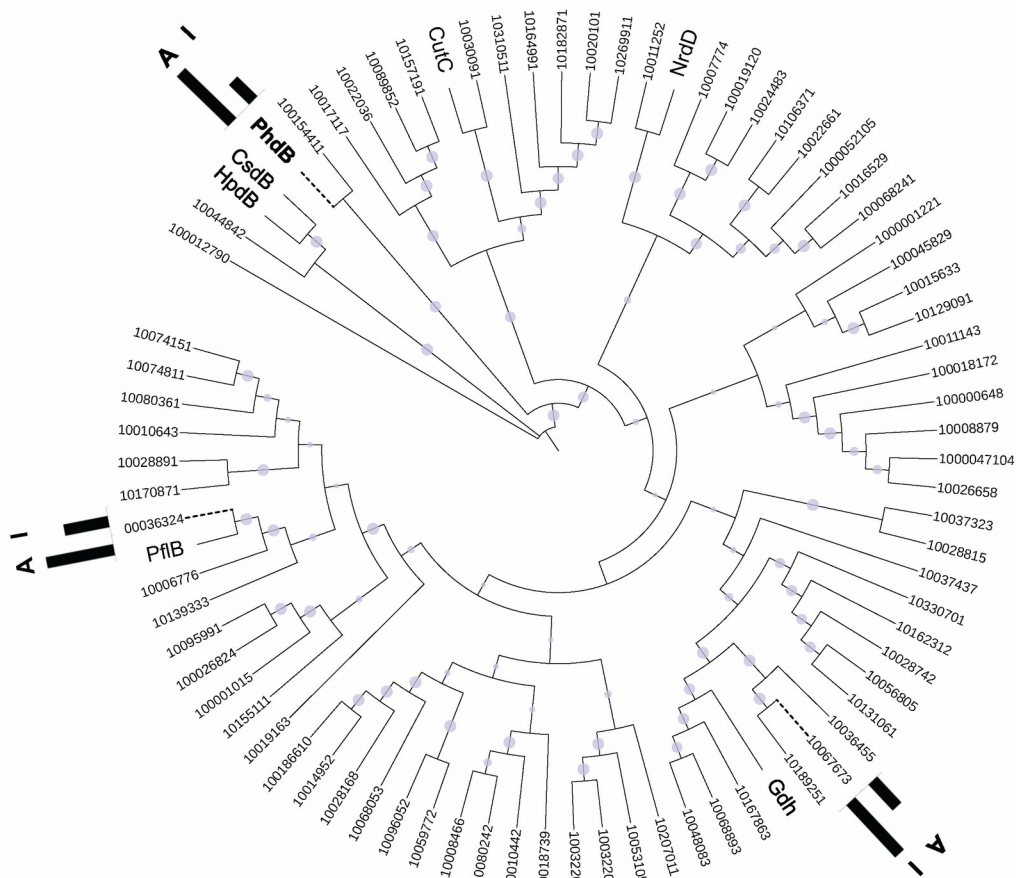
803

804

805 **Author Information**

806 Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). J.D.K. has a  
807 financial interest in Amyris and Lygos. Correspondence and requests for materials should be  
808 addressed to H.R.B ([HRBeller@lbl.gov](mailto:HRBeller@lbl.gov)).

809

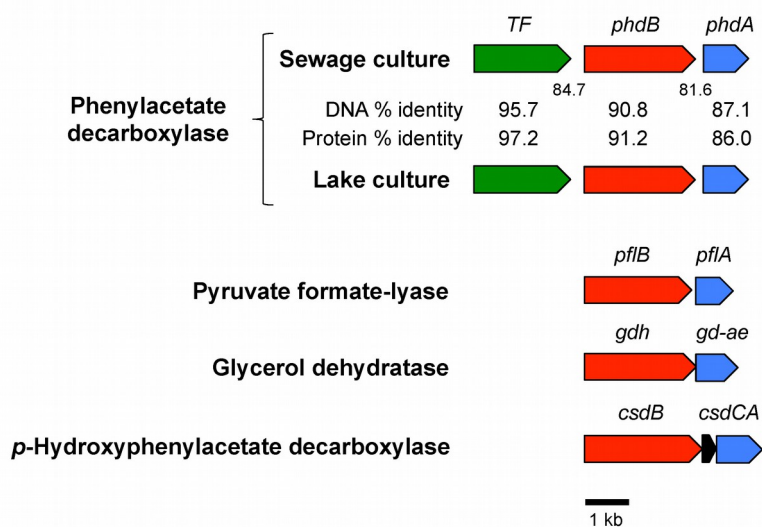


810  
811

**812 Figure 1 | Glycyl radical enzymes encoded in the toluene-producing sewage culture**  
**813 metagenome and their association with *in vitro* toluene synthase activity.** This maximum-  
**814 likelihood tree is based on protein sequences of putative glycy radical enzymes (GREs) detected**  
**815 in the sewage-derived metagenome [IMG Taxon ID 3300001865 on JGI's IMG-M site**  
**816 (<https://img.jgi.doe.gov/cgi-bin/mer/main.cgi>)]. Numerical values on the leaves represent locus**  
**817 tags in the metagenome from which the prefix “JGI2065J20421\_” has been truncated for brevity.**  
**818 Leaves with protein names rather than locus tags are known GREs provided for context (see**  
**819 Methods for details). The leaf marked PhdB represents the GRE characterized in this study.**  
**820 Leaves with dashed lines represent proteins detected by LC/MS/MS in active FPLC fractions,**  
**821 and the histograms on these leaves represent the maximum abundance of this protein in (A) the**  
**822 two most active fractions and (I) the two flanking inactive or less active fractions**  
**823 (Supplementary Data File 1); histograms are normalized to the greatest of the A and I values.**  
**824 Purple circles on leaves represent bootstrap support values for each node (largest symbols are**  
**825 100).**

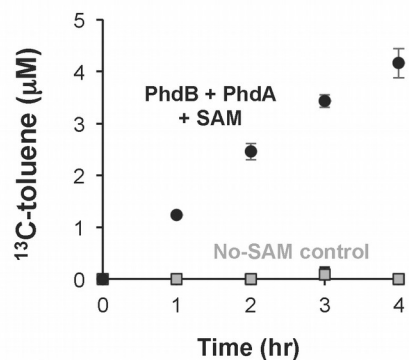
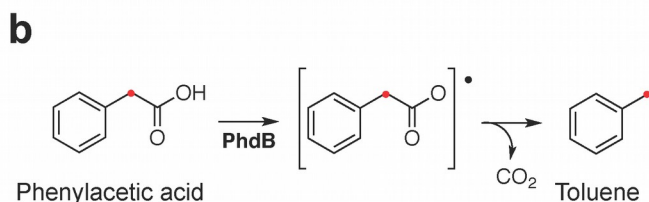
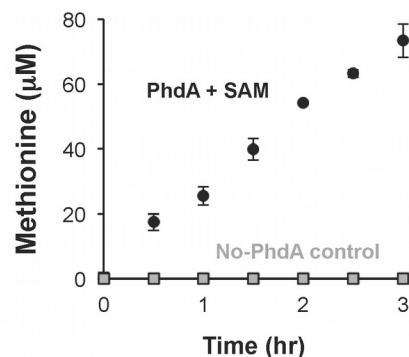
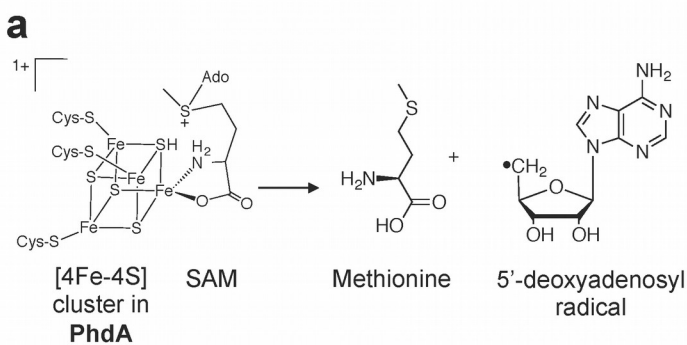
826  
827  
828  
829

830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856



**Figure 2 | Homologous *toluene-synthasphenylacetate decarboxylase* gene clusters from sewage and lake sediment cultures.** *phdB*, phenylacetate decarboxylase (a glycy radical enzyme); *phdA*, a cognate activating enzyme for *phdB*; *TF*, putative transcription factor. Sequence identity is shown for the coding sequences as well as the two intergenic regions. Gene clusters for selected GREs (in red) and their cognate activating enzymes (in blue) are shown for comparison, including pyruvate formate-lyase (*pflB*, *pflA*), glycerol dehydratase (*gdh*, *gd-ae*), and *p*-hydroxyphenylacetate decarboxylase (*csdB*, *csdC*, *csdA*). A 1-kb scale bar is included.

857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879

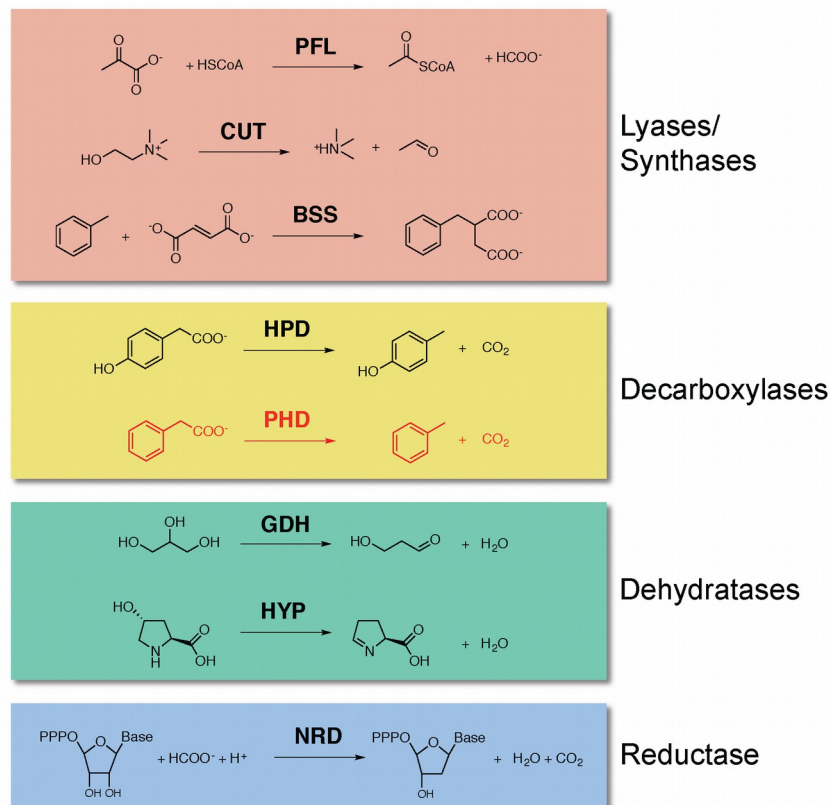


**Figure 3 | Reactions catalyzed by PhdA and PhdB.** **a**, proposed reaction of PhdA with SAM, as supported *in vitro* by methionine production by re-constituted and purified recombinant PhdA (black circles). Controls without PhdA are also shown (gray squares). **b**, proposed reaction of PhdB with phenylacetic acid-2-<sup>13</sup>C, as supported *in vitro* by [*methyl*-<sup>13</sup>C]toluene production by partially purified PhdB in combination with PhdA and SAM (black circles). Controls without SAM are also shown (gray squares). <sup>13</sup>C-labeled C atoms in the proposed reaction are highlighted with a red circle. Data points represent means and error bars represent one standard deviation (*n*=3). Experiments demonstrating PhdA-catalyzed production of methionine from SAM were replicated twice-three times and experiments demonstrating labeled toluene production from labeled phenylacetate in the presence of PhdB and PhdA were performed 6 times (four times with no-SAM negative controls).

891  
892  
893  
894  
895  
896  
897



898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921

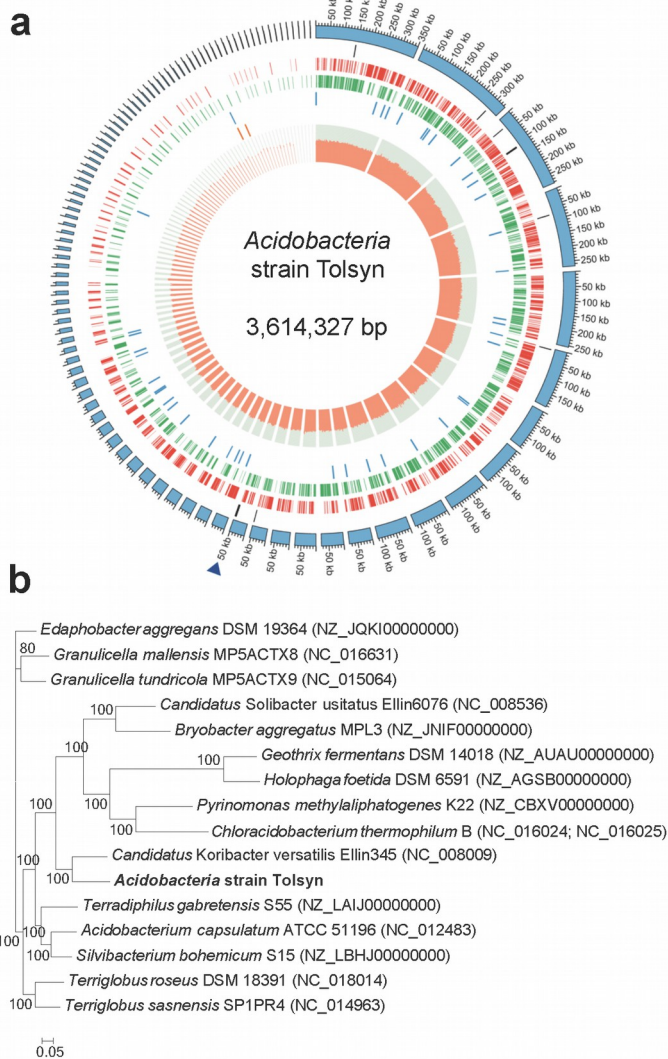


922**Figure 4 | Reactions catalyzed by characterized GREs.** PFL, pyruvate formate-lyase; CUT,  
923choline trimethylamine-lyase; BSS, benzylsuccinate synthase; HPD, *p*-hydroxyphenylacetate  
924decarboxylase; PHD, phenylacetate decarboxylase (this study); GDH, glycerol dehydratase;  
925HYP, *trans*-4-hydroxy-L-proline dehydratase; and NRD, anaerobic ribonucleotide reductase.  
926  
927  
928

929	<b>a</b>			
930	PflB	MLLDAMENPEKYPQLTIRV <b>S</b> <sup>*</sup> GYAVRFNSLTKEQQQDVITRFTQSM	760	
931	CsdB	TLRDAQLTPEKYRELMVRV <b>A</b> GF <b>T</b> QYWCEIGKPIQDEVIYRTEYDK	897	
932	BssA	EMRAAQREPEKHHDLIVR <b>V</b> S <b>G</b> YSARFVDIPTYGQNTI IARQEQDFSASDL	857	
933	Gdh	ILLAAQKNPEKYQDLIVR <b>V</b> A <b>G</b> YSAQFISLDKSIQNDI IARTEHVM	787	
934	CutC	VLKKAQQEPEKYRDLIVR <b>V</b> A <b>G</b> YSAYFVELCKEVQDEIISRTEVIEKF	1128	
935	HypD	VLLEAQKNPQDYKDLIVR <b>V</b> A <b>G</b> YSDHFNNLSRTLQDEIIGRTEQTF	789	
936	PhdB-s	TLRAAQKDPDSYRDLIVR <b>V</b> A <b>G</b> FSAYFITLCPEVQDEIVSRTECQTW	839	
937	PhdB-l	TLRAAQKDPDSFRDLIVR <b>V</b> A <b>G</b> FSAYFITLCPEVQNEIVSRTECQW	839	
938	<b>b</b>			
939	PflB	DDYAIA <b>CC</b> VSPMIVG-----KQM-----QF-----FGARANLAKTML	444	
940	CsdB	RAWCLGG <b>CLE</b> SAPGCFLPLEYNGKVTMIPGGASPT <b>CG</b> TGV <b>H</b> FIGMPKVLE	545	
941	BssA	HNWVNV <b>LCMS</b> PGI <b>H</b> G-----RRK-----TQ <b>KTR</b> SEGGGSIFPAKLE	521	
942	Gdh	RDYGI <b>IIG</b> <b>CV</b> EPQKPG-----KTE-----GWH-----DSAFFNLARIVE	458	
943	CutC	RDYCLMG <b>CV</b> EPQKSG-----RIY-----QWT-----STGYTQWPPIAIE	796	
944	HypD	RLGGTSG <b>CV</b> ETGCFG-----K-E-----AYV-----LTGYMNIPIKILE	458	
945	PhdB-s	RDQAVAG <b>CV</b> QSIIGG-----KTD-----GT-----WEARFNMTKMME	506	
946	PhdB-l	RDQAVAG <b>CV</b> QSIIGG-----KTD-----GT-----WEARFNMCKMIE	506	
947	<b>c</b>			
948	NrdG			
949	BssD			
950	PflA			
951	CutD			
952	CsdA			
953	GD-AE			
954	HypD-AE			
955	PhdA-s			
956	PhdA-l			

957**Figure 5 | Multiple sequence alignments comparing PhdB and PhdA with other glycy**  
958**radical enzymes and glycy radical activating enzymes, respectively.** **a**, C-terminal region of  
959GREs containing the conserved glycy radical motif, with the glycy radical site highlighted in  
960red with an asterisk and other conserved residues in bold. **b**, mid-sequence region of GREs  
961containing conserved, active-site cysteine residue (which bears the thyl radical that interacts  
962with the substrate), highlighted in red with an asterisk, along with other conserved residues  
963shown in blue. **c**, N-Terminal region of activating enzymes highlighting the CxxxCxxC motif  
964(highlighted with asterisks) coordinating with the [4Fe-4S] cluster. Sequences used in these  
965alignment comparisons include the following GREs and AEs [PDB (Protein Data Bank) or  
966GenBank accession number]: PflB (GenBank: NP\_415423), PflA (GenBank: NP\_415422), CsdB  
967(GenBank: ABB05046.1), CsdA (GanBank: 2580384209), BssA (PDB: 4PKC:A), BssD  
968(GenBank: CAA05050.2), Gdh (PDB: 1R8W), GD-AE (GenBank: AAM54729), CutC (PDB:  
9695A0Z), CutD (GenBank: EPO20361.1), HypD (UniProt: A0A031WDE4), HypD-AE (UniProt:  
970A0A069AMK2), NrdG (GenBank: NP\_418658). The “s” and “l” suffixes for PhdB and PhdA  
971stand for sewage and lake, respectively. Alignment was performed with Clustal Omega<sup>58</sup>.  
972

973  
 974  
 975  
 976  
 977  
 978  
 979  
 980  
 981  
 982  
 983  
 984  
 985  
 986  
 987  
 988  
 989  
 990  
 991  
 992  
 993  
 994  
 995  
 996  
 997  
 998  
 999  
 1000  
 1001  
 1002  
 1003  
 1004



1005 **Figure 6 | Characterization of the putatively toluene-producing *Acidobacterium* strain**  
 1006 **Tolsyn based on its recovered genome. a**, schematic circular diagram of the genome, with  
 1007 contigs in size order, displaying contigs and their corresponding lengths (outer ring), genes  
 1008 encoding radical-related enzymes (second ring; the contig containing *phdA* and *phdB* is indicated  
 1009 with a filled triangle), genes on the forward strand (third ring), genes on the reverse strand  
 1010 (fourth ring), tRNA genes (fifth ring), rRNA genes (sixth ring), and GC content (seventh ring;  
 1011 GC is averaged every 1000 bp and is represented as orange, whereas AT is light green). **b**,  
 1012 Phylogenetic relationships among *Acidobacterium* strain Tolsyn and the most closely related  
 1013 *Acidobacteria* sequenced isolates based upon 129 concatenated marker proteins (GenBank  
 1014 accession numbers for species are shown in the tree). Numbers at nodes represent bootstrap  
 1015 support values. The scale bar represents substitution rate per site.

1016  
 1017