

UC Berkeley

UC Berkeley Previously Published Works

Title

Oxidative Pathways of Deoxyribose and Deoxyribonate Catabolism

Permalink

<https://escholarship.org/uc/item/3vn4b417>

Journal

mSystems, 4(1)

ISSN

2379-5077

Authors

Price, Morgan N

Ray, Jayashree

Iavarone, Anthony T

et al.

Publication Date

2019-02-26


DOI

10.1128/msystems.00297-18

Peer reviewed



# Oxidative Pathways of Deoxyribose and Deoxyribonate Catabolism

 Morgan N. Price,<sup>a</sup> Jayashree Ray,<sup>a</sup> Anthony T. Iavarone,<sup>b</sup> Hans K. Carlson,<sup>a</sup> Elizabeth M. Ryan,<sup>c</sup> Rex R. Malmstrom,<sup>c</sup> Adam P. Arkin,<sup>a,d</sup> Adam M. Deutschbauer<sup>a,e</sup>

<sup>a</sup>Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, California, USA

<sup>b</sup>QB3/Chemistry Mass Spectrometry Facility, University of California, Berkeley, California, USA

<sup>c</sup>DOE Joint Genome Institute, Walnut Creek, California, USA

<sup>d</sup>Department of Bioengineering, University of California, Berkeley, California, USA

<sup>e</sup>Department of Plant and Microbial Biology, University of California, Berkeley, California, USA

**ABSTRACT** Using genome-wide mutant fitness assays in diverse bacteria, we identified novel oxidative pathways for the catabolism of 2-deoxy-D-ribose and 2-deoxy-D-ribonate. We propose that deoxyribose is oxidized to deoxyribonate, oxidized to ketodeoxyribonate, and cleaved to acetyl coenzyme A (acetyl-CoA) and glycerol-CoA. We have genetic evidence for this pathway in three genera of bacteria, and we confirmed the oxidation of deoxyribose to ketodeoxyribonate *in vitro*. In *Pseudomonas simiae*, the expression of enzymes in the pathway is induced by deoxyribose or deoxyribonate, while in *Paraburkholderia bryophila* and in *Burkholderia phytofirmans*, the pathway proceeds in parallel with the known deoxyribose 5-phosphate aldolase pathway. We identified another oxidative pathway for the catabolism of deoxyribonate, with acyl-CoA intermediates, in *Klebsiella michiganensis*. Of these four bacteria, only *P. simiae* relies entirely on an oxidative pathway to consume deoxyribose. The deoxyribose dehydrogenase of *P. simiae* is either nonspecific or evolved recently, as this enzyme is very similar to a novel vanillin dehydrogenase from *Pseudomonas putida* that we identified. So, we propose that these oxidative pathways evolved primarily to consume deoxyribonate, which is a waste product of metabolism.

**IMPORTANCE** Deoxyribose is one of the building blocks of DNA and is released when cells die and their DNA degrades. We identified a bacterium that can grow with deoxyribose as its sole source of carbon even though its genome does not contain any of the known genes for breaking down deoxyribose. By growing many mutants of this bacterium together on deoxyribose and using DNA sequencing to measure the change in the mutants' abundance, we identified multiple protein-coding genes that are required for growth on deoxyribose. Based on the similarity of these proteins to enzymes of known function, we propose a 6-step pathway in which deoxyribose is oxidized and then cleaved. Diverse bacteria use a portion of this pathway to break down a related compound, deoxyribonate, which is a waste product of metabolism. Our study illustrates the utility of large-scale bacterial genetics to identify previously unknown metabolic pathways.

**KEYWORDS** deoxyribonate catabolism, deoxyribose catabolism, high-throughput genetics


DNA is present in every cell, and once released into the environment by cell lysis, it is hydrolyzed by a variety of secreted or extracellular enzymes (1). DNA is hydrolyzed to oligonucleotides and then to individual deoxynucleotides or deoxynucleosides (1). Deoxynucleotides can be hydrolyzed extracellularly to deoxynucleosides (2), and deoxynucleosides can be cleaved to release 2-deoxy-D-ribose and nucleobases. Some

**Citation** Price MN, Ray J, Iavarone AT, Carlson HK, Ryan EM, Malmstrom RR, Arkin AP, Deutschbauer AM. 2019. Oxidative pathways of deoxyribose and deoxyribonate catabolism. *mSystems* 4:e00297-18. <https://doi.org/10.1128/mSystems.00297-18>.

**Editor** Korneel Rabaey, Ghent University

**Copyright** © 2019 Price et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

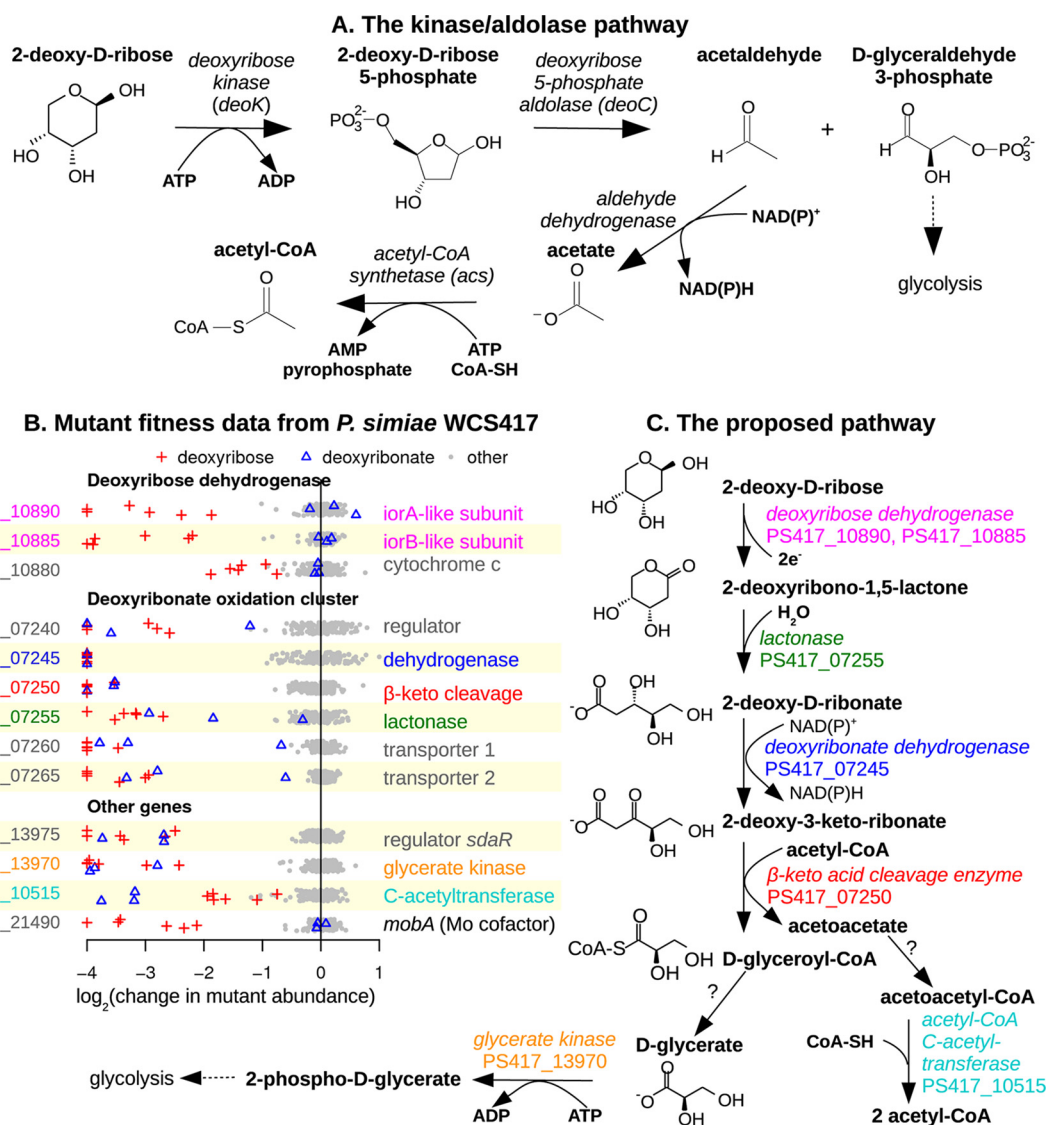
Address correspondence to Morgan N. Price, [morgannprice@yahoo.com](mailto:morgannprice@yahoo.com), or Adam P. Arkin, [aparkin@lbl.gov](mailto:aparkin@lbl.gov).

 Deoxyribonate is a waste product of metabolism and is present in urine, but nothing was known about its metabolism. Using high-throughput genetics, we show that diverse bacteria grow on deoxyribonate by an oxidative pathway

**Received** 19 November 2018

**Accepted** 12 January 2019

**Published** 5 February 2019



**FIG 1** An oxidative pathway for deoxyribose utilization in *Pseudomonas simiae* WCS417. (A) The previously described kinase/aldolase pathway. (B) Gene fitness in *P. simiae*. Each point shows the gene's fitness value (x axis) in a different experiment. Fitness values under  $-1$  indicate a defect in growth. Values outside of the plot's range are shown at  $-4$  or  $+1$ . Within each gene's section, the y axis is random. Experiments with deoxyribose or deoxyribonate as the carbon source are indicated by symbols. (C) The proposed pathway for deoxyribose catabolism in *P. simiae*. Putative enzymes are color coded to match the gene descriptions in panel B.

nucleosidases are specific for 2'-deoxyribonucleosides over ribonucleosides (3), which suggests that the cleavage of deoxynucleosides is their physiological role.

We are studying the catabolism of the end products of DNA hydrolysis, including deoxyribose. The only previously known pathway for catabolizing deoxyribose (of which we are aware) involves deoxyribose kinase (*deoK*) (4) and deoxyribose 5-phosphate aldolase (*deoC*) (5), which yield acetaldehyde and D-glyceraldehyde 3-phosphate (Fig. 1A). However we recently screened 28 bacteria for growth on deoxyribose and found that the plant-root-associated bacterium *Pseudomonas simiae* WCS417 grows with deoxyribose as the sole source of carbon (6), even though its genome does not contain either *deoK* or *deoC*. We also tested which genes are important for the growth of *P. simiae* under various conditions, including growth with deoxyribose as the sole source of carbon (6), by using a pool of over 100,000 randomly barcoded transposon mutants of *P. simiae*. If a gene is important for growth under a condition, then the abundance of its mutants will decrease during growth, and the

abundance of each mutant in the pool can be measured by amplifying and sequencing the DNA barcodes (7).

Here, we use these genetic data to identify a novel oxidative pathway for deoxyribose catabolism. An intermediate in the proposed pathway is 2-deoxy-D-ribonate, which may also be widely available in the environment, as deoxyribonate has been detected in human urine (8). Deoxyribonate might form by the oxidation of deoxyribose, by the repair of oxidatively damaged DNA (9), or from a spontaneous reaction of hydrogen peroxide with 2-keto-3-deoxy-D-gluconate (which is an intermediate in hexuronate metabolism and in variants of the Entner-Doudoroff pathway of glycolysis) (10).

We used the pool of mutants to study the growth of *P. simiae* on deoxyribonate and found that *P. simiae* uses many of the same genes to catabolize deoxyribonate as deoxyribose. Furthermore, these genes are induced during growth on either deoxyribose or deoxyribonate. We also reconstituted the enzymatic conversion of deoxyribose to deoxyribonate to ketodeoxyribonate *in vitro*.

We then used pooled mutant fitness assays to study the catabolism of deoxyribose and deoxyribonate in other bacteria. In *Paraburkholderia bryophila* 376MFSha3.1 and in *Burkholderia phytofirmans* PsJN, similar genes to those in *P. simiae* are involved in the catabolism of both deoxyribose and deoxyribonate. *Klebsiella michiganensis* M5a1 uses an analogous pathway with acyl coenzyme A (acyl-CoA) intermediates to consume deoxyribonate (but not deoxyribose). Thus, diverse bacteria use oxidative pathways to consume deoxyribose and deoxyribonate.

## RESULTS

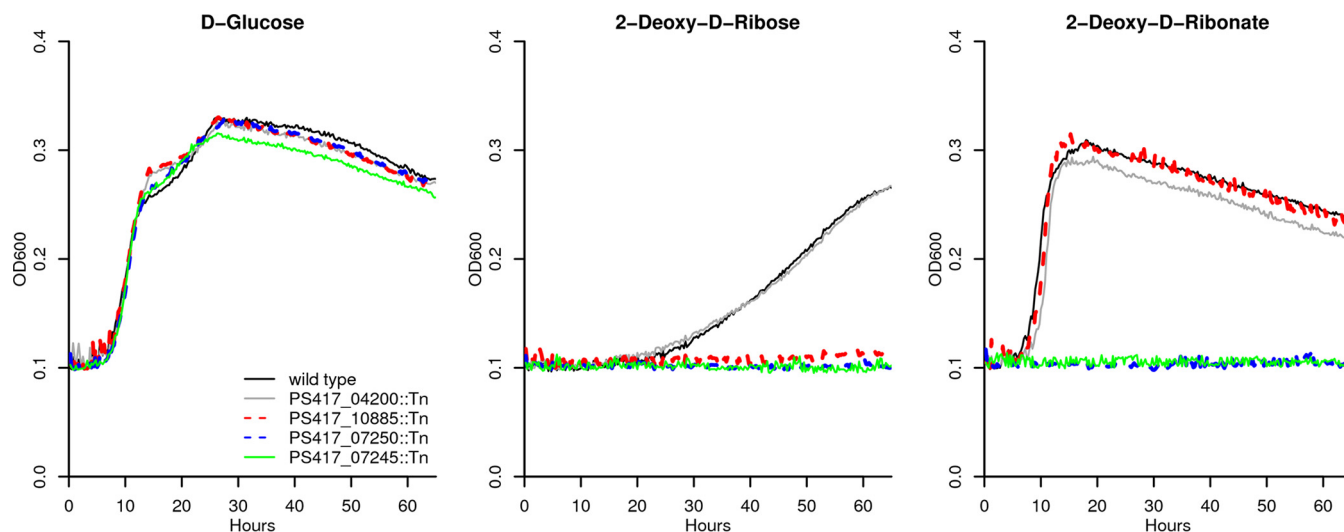
**The genetic basis of growth on deoxyribose and deoxyribonate in *Pseudomonas simiae*.** To supplement the previously described mutant fitness data for *P. simiae* WCS417 (6), we tested deoxyribose as the carbon source on 2 additional days, and we also tested deoxyribonate as the carbon source. For each experiment, we grew a library of over 100,000 barcoded mutant strains together and used DNA sequencing to measure the change in each strain's abundance. Overall, our data set for *P. simiae* included 6 genome-wide assays of fitness in deoxyribose, 3 assays of fitness in deoxyribonate, and 153 other fitness assays from diverse conditions, including 49 other carbon sources.

From these data, we identified 13 genes that were consistently important for utilizing deoxyribose and were not important for fitness under most other conditions (see Materials and Methods). The fitness data for these genes are shown in Fig. 1B: each fitness value is the log<sub>2</sub> change of the relative abundance of mutants in that gene during the experiment. In the genome of *P. simiae*, 11 of the 13 genes cluster into three groups of nearby genes that are on the same strand and are probably operons.

Of the 13 genes, eight are likely to code for catabolic enzymes:

- two subunits of a molybdenum-dependent dehydrogenase (PS417\_10890 and PS417\_10885) that are distantly related to isoquinoline 1-oxidoreductase (*iorAB* [49% or 29% amino acid identity]) (11)
- a cytochrome *c* that is in the same putative operon (PS417\_10880) and might accept electrons from the *iorAB*-like dehydrogenase
- a dehydrogenase of the short-chain dehydrogenase/reductase (SDR) superfamily (PS417\_07245)
- a  $\beta$ -keto acid cleavage enzyme (PS417\_07250) (12)
- a lactonase (PS417\_07255)
- glycerate kinase (PS417\_13970)
- and acetyl-CoA C-acetyltransferase (PS417\_10515).

The remaining genes that are specifically important for deoxyribose utilization are two putative transporters, two regulatory genes, and *mobA* (PS417\_21490), which is involved in the biosynthesis of molybdenum cofactor and is probably required for the activity of the molybdenum-dependent dehydrogenase (PS417\_



**FIG 2** Growth curves for individual transposon mutants. We compared the growth of wild-type *P. simiae* WCS417 and of individual transposon mutant strains in defined media with three carbon sources: D-glucose, deoxyribose, and deoxyribonate. The mutant of PS417\_04200 (2-ketoglutaric semialdehyde dehydrogenase) is included as a control. Each curve is the median of six replicates. Each carbon source was provided at 10 mM.

10890 to PS417\_10885). Most of these genes are also important for growth on deoxyribonate (Fig. 1B). The exceptions are the genes coding for the two subunits of the molybdenum-dependent dehydrogenase, the associated cytochrome *c* gene, *mobA*, and perhaps the lactonase gene (Fig. 1B); we therefore propose that these genes act upstream of deoxyribonate formation (Fig. 1C).

To confirm the genome-wide fitness data, we obtained individual mutant strains for some of the key genes and grew them with either D-glucose, deoxyribose, or deoxyribonate as the sole source of carbon. We found that a transposon mutant with mutation of the *iorB*-like subunit of the molybdenum-dependent dehydrogenase (PS417\_10885) grows on glucose or deoxyribonate but not on deoxyribose (Fig. 2). Mutants of the SDR dehydrogenase (PS417\_07245) or the  $\beta$ -keto acid cleavage enzyme (PS417\_07250) grow on glucose but not on deoxyribose or deoxyribonate (Fig. 2).

**The proposed pathway for deoxyribose catabolism in *P. simiae*.** The genetic data suggest that *P. simiae* WCS417 catabolizes deoxyribose by an oxidative pathway (Fig. 1C). Because the molybdenum-dependent dehydrogenase was not required for growth on deoxyribonate, and because the lactonase was more important for growth on deoxyribose than on deoxyribonate (Fig. 1B), it appears that the molybdenum-dependent dehydrogenase oxidizes deoxyribose to a lactone, which is hydrolyzed to deoxyribonate by the lactonase. We tentatively predict that the lactone intermediate is 2-deoxy-D-ribo-1,5-lactone instead of a 1,4-lactone because the pyranose form of deoxyribose is more stable in solution at 30°C (13).

Oxidation of the deoxyribonate by the SDR enzyme could yield 2-deoxy-3-keto-D-ribonate, which is a  $\beta$ -keto acid and could be cleaved by the  $\beta$ -keto acid cleavage enzyme (PS417\_07250).  $\beta$ -Keto acid cleavage enzymes remove an acetyl group from a  $\beta$ -keto acid and transfer the acetyl group to acetyl-CoA (12). In the proposed pathway, the  $\beta$ -keto acid cleavage enzyme is expected to yield D-glyceryl-CoA and acetoacetate. The requirement for the glycerate kinase suggests that the D-glyceryl-CoA is converted to glycerate. We looked for a thiolase or CoA-transferase that might act on D-glyceryl-CoA and that is important for deoxyribose utilization in the mutant fitness data. There are many candidate genes in the genome, but none with that phenotype. We suspect that the conversion of D-glyceryl-CoA to D-glycerate is genetically redundant and therefore cannot be identified by assaying mutant strains with only one transposon insertion. The glycerate kinase of *P. simiae* (PS417\_13970) is 51% identical to glycerate kinase 1 from *Escherichia coli* (*gark*), which forms 2-phosphoglycerate (14, 15).

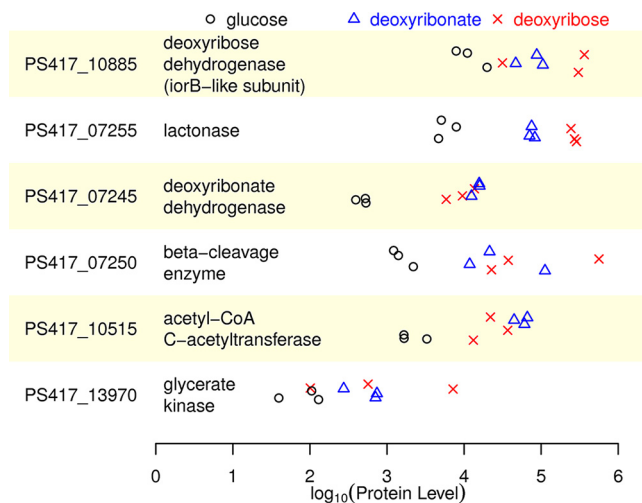
2-Phosphoglycerate is an intermediate in glycolysis and thus links the proposed pathway to central metabolism.

In the preceding analysis, we proposed reactions for all of the putative enzymes with strong negative phenotypes on deoxyribose (relative to other conditions), except for the acetyl-CoA C-acetyltransferase (PS417\_10515). That enzyme converts acetoacetate (which is produced by the  $\beta$ -keto acid cleavage enzyme) back to acetyl-CoA, but acetoacetate must first be activated to acetoacetyl-CoA. Acetoacetate might be activated by a CoA-transferase (PS417\_10525 to PS417\_10520) or a CoA-synthetase (there are several candidates in the genome). None of these genes is important for the utilization of deoxyribose, so we propose that this activity is genetically redundant. Overall, the proposed pathway accounts for all of the putative enzymes that are specifically important for deoxyribose utilization in *P. simiae*.

**Oxidation of deoxyribose to ketodeoxyribonate *in vitro*.** We tested the activity of the deoxyribose dehydrogenase and the deoxyribonate dehydrogenase *in vitro*. (For details, see the Appendix.) First, we overexpressed the *iorA*- and *iorB*-like subunits of deoxyribose dehydrogenase (PS417\_10890 and PS417\_10885) in *E. coli* and tested the resulting cell lysate. We found that this preparation oxidized deoxyribose to deoxyribonate with phenazine methosulfate as the electron acceptor. The lactonase (PS417\_07255) was not necessary: deoxyribonolactone may hydrolyze spontaneously in water, or it might be hydrolyzed by enzymes in the cell lysate, or the deoxyribose dehydrogenase may form deoxyribonate instead of deoxyribonolactone. Second, we purified the deoxyribonate dehydrogenase (PS417\_07245) and showed that it oxidized deoxyribonate with NADH as the electron acceptor. Tandem mass spectrometry (MS/MS) showed that the oxidized product was a ketodeoxyribonate, but we were not able to determine which position was oxidized. We also tested the two dehydrogenases on each others' substrates (deoxyribose dehydrogenase on deoxyribonate or deoxyribonate dehydrogenase on deoxyribose) and did not observe any activity. Thus, biochemical assays confirmed the proposed activities of deoxyribose dehydrogenase and deoxyribonate dehydrogenase.

**Accessory genes for deoxyribose catabolism in *P. simiae*.** Based on the proposed pathway, we can explain the roles of the other genes with specific phenotypes on deoxyribose (Fig. 1B). First, the putative operon that includes the lactonase, the SDR deoxyribonate dehydrogenase, and the  $\beta$ -keto acid cleavage enzyme also includes two genes from the major facilitator superfamily (MFS) of transporters (PS417\_07260 and PS417\_07265). We propose that these two genes are required for the uptake of deoxyribonate or deoxyribonolactone. We suspect that deoxyribose is oxidized in the periplasm, because the *iorB*-like subunit has a potential signal for Tat-dependent export (RRRFLA) near its N terminus, and the associated cytochrome *c* protein has a membrane anchor and may be exported. In contrast, the lactonase and the deoxyribonate dehydrogenase are predicted to be in the cytoplasm (16). If deoxyribose is oxidized in the periplasm, then deoxyribonate could be the substrate of the transporter when either deoxyribose or deoxyribonate is the carbon source, which would explain why these transporter genes were important under both conditions. However we do not understand why two different transporter genes appear to be required. These transporter genes are at the end of the putative operon, so polar effects are not likely.

Second, a *gntR*-like transcription factor (PS417\_07240) is present at the beginning of the deoxyribonate oxidation cluster. By comparing the sequences upstream of this cluster and upstream of similar clusters from other strains of *Pseudomonas* or from *Burkholderia graminis*, we identified the conserved palindrome GTGATCAC. This sequence occurs at positions  $-47$  to  $-40$  relative to the predicted start codon of PS417\_07240 and at similar positions relative to its homologs. The same motif was identified among the binding sites of the transcriptional activator AkgR from *Rhodobacter sphaeroides* (17), which is 32% identical to PS417\_07240. We propose that PS417\_07240 regulates the expression of the cluster by binding to this motif.



**FIG 3** Levels of protein expression in *Pseudomonas simiae* WCS417. We show the normalized abundance of enzymes in the deoxyribose utilization pathway, as estimated from the integrated ion abundances of the detected peptide ions. There are three biological replicates of each condition, and the y axis within each panel is random. Note the log scale for the x axis.

Third, the transcription factor PS417\_13975 (*sdaR*) is probably an activator for the expression of the glycerate kinase (PS417\_13970) gene, which is downstream of *sdaR*. Similarly, a close homolog of PS417\_13975 from *Pseudomonas fluorescens* Pf-5 (PFL\_3379 [88% identical]) is predicted in RegPrecise (18) to regulate a downstream glycerate kinase.

Finally, *mobA* (PS417\_21490) is expected to be necessary for the attachment of a nucleotide to molybdenum cofactor to form a molybdenum dinucleotide cofactor. This cofactor is probably required for the activity of deoxyribose dehydrogenase, which is homologous to the *coxLS* subunits of a CO dehydrogenase that has a molybdenum-cytosine dinucleotide cofactor (19). While *mobA* is required for the activity of deoxyribose dehydrogenase, it is not important for oxidation of deoxyribonate (Fig. 1B), which is expected because none of the other enzymes in the pathway uses a molybdenum cofactor. Similarly, genes for the biosynthesis of molybdopterin (a precursor to molybdenum cofactor) are important for utilizing deoxyribose but not deoxyribonate (*moaABCE*: PS417\_21395, PS417\_08670, PS417\_04865, and PS417\_04875). These genes' phenotypes are not specific to deoxyribose (and hence are not shown in Fig. 1B) because they are also important for fitness under other conditions such as inosine utilization and nitrate stress (6).

**Induction of the oxidative pathway in *P. simiae*.** Next we asked if the enzymes in the oxidative pathway are induced when *P. simiae* grows on deoxyribose or deoxyribonate. As shown in Fig. 3, all of these proteins seem to be expressed more highly during growth on deoxyribose or deoxyribonate than during growth on glucose. (The *iorA*-like subunit of deoxyribose dehydrogenase and the associated cytochrome were not detected; they may be membrane associated.) The differences in expression levels were statistically significant for the lactonase, the deoxyribonate dehydrogenase, and acetyl-CoA C-acetyltransferase (false-discovery rate of <5% by analysis of variance [ANOVA]).

**Comparative genetics of the oxidative pathway.** To determine if other bacteria use the oxidative pathway, we examined the genomic organization of the putative enzymes and their homologs. We focused on homologs from bacteria for which mutant libraries are available, and we obtained genome-wide mutant fitness data on deoxyribose, deoxyribonate, or deoxynucleosides for three additional bacteria: *Burkholderia phytofirmans* PsJN, *Paraburkholderia bryophila* 376MFSHa3.1, and *Klebsiella michiganensis* M5a1. This section gives an overview of the genes that were important for fitness

under these conditions, the chromosomal organization of these genes, and the catabolic pathways that we propose.

Although these three additional bacteria, as well as *P. simiae*, grow on deoxyribonate by oxidative pathways, their other catabolic capabilities vary. Figure 4A shows the relationships between the proposed pathways for the catabolism of deoxyribose, deoxyribonate, or deoxynucleosides in the four bacteria based on our analysis of the genetic data. The full rationale for the assignment of each bacterium to a particular pathway is contained in the following sections.

First, the deoxyribonate oxidation genes are clustered together, with some variation, in diverse bacteria (Fig. 4B). The clusters from two related bacteria, *Paraburkholderia bryophila* 376MFSha3.1 and *Burkholderia phytofirmans* PsJN, contain the deoxyribonate dehydrogenase and  $\beta$ -keto acid cleavage enzyme genes, and these genes were important for deoxyribonate utilization. The genomes of both of these bacteria also encode the kinase/aldolase pathway for deoxyribose utilization (Fig. 4C). Genes from both pathways were important for deoxyribose utilization, so we propose that these bacteria consume deoxyribose by the kinase/aldolase and oxidative pathways in parallel.

Second, *Klebsiella michiganensis* M5al has a gene cluster that shares some genes with the deoxyribonate oxidation cluster, but *K. michiganensis* does not encode deoxyribonate dehydrogenase or the  $\beta$ -keto acid cleavage enzyme (Fig. 4B). Nevertheless, this gene cluster was important for deoxyribonate utilization by *K. michiganensis*. We propose that *K. michiganensis* consumes deoxyribonate by another oxidative pathway with acyl-CoA intermediates. Also, *K. michiganensis* does not grow on deoxyribose, but it can grow on deoxynucleosides via DeoABD, which can convert them to deoxyribose-5-phosphate and deoxyribose 5-phosphate aldolase (Fig. 4C).

Third, the deoxyribose dehydrogenase gene of *P. simiae* is part of a conserved gene cluster, but its biological role does not seem to be conserved. The conserved cluster includes *iorA*- and *iorB*-like subunits and a putative cytochrome *c* (Fig. 4D). (In contrast, the *iorAB* genes of isoquinoline 1-oxidoreductase are more distantly related, and no gene for a cytochrome *c* was found near *iorB* [11] or near its closest homologs in complete genomes [WP\_083944220.1 or WP\_113935277.1].) The deoxyribose dehydrogenase-like gene cluster is found in some other *Pseudomonas* species and in some genera of *Proteobacteria* such as *Bordetella* and *Agrobacterium*. Although the cluster is found in *Pseudomonas putida* KT2440, *P. putida* does not grow on deoxyribose or deoxyribonate, and its genome does not contain the deoxyribonate oxidation genes. While studying the seemingly unrelated process of vanillin (4-hydroxy-3-methoxybenzaldehyde) catabolism by *P. putida*, we identified mutant phenotypes for this cluster. We propose that the cluster from *P. putida* (PP\_3621 to PP\_3623) encodes a novel vanillin dehydrogenase.

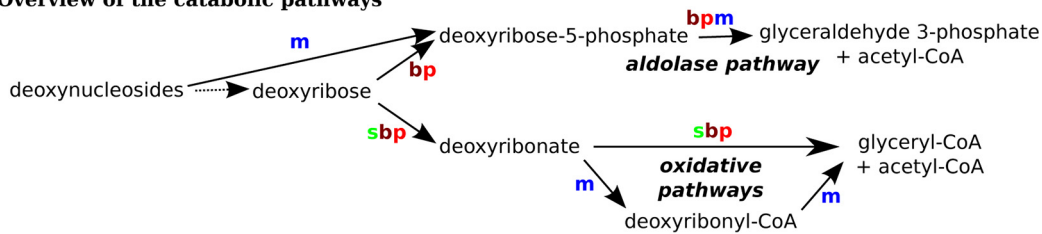
We will now describe the genetic data for each of the bacteria in more detail.

**Oxidation of deoxyribose and deoxyribonate in *Paraburkholderia bryophila* and *Burkholderia phytofirmans*.** *Paraburkholderia bryophila* 376MFSha3.1 and *Burkholderia phytofirmans* PsJN contain close homologs of the deoxyribonate dehydrogenase and the  $\beta$ -keto acid cleavage enzyme (50 to 60% identical to the proteins in *P. simiae*), but they do not contain close homologs of the deoxyribose dehydrogenase or the lactonase. In *P. bryophila*, the standard kinase/aldolase pathway was important for growth on deoxyribose, but not on deoxyribonate or glucose (Fig. 5). Enzymes for the oxidative catabolism of deoxyribonate—the deoxyribonate dehydrogenase (H281DRAFT\_00644) and the  $\beta$ -keto acid cleavage enzyme (H218DRAFT\_00641)—were important for growth on both deoxyribose and deoxyribonate. Also important for growth on both substrates were glycerate kinase (H281DRAFT\_00894) and genes for the conversion of acetoacetate to acetoacetyl-CoA to acetyl-CoA (H281DRAFT\_04495 and H281DRAFT\_00852). These genes were not important for glucose utilization. These data strongly suggest that the oxidative pathway in *P. bryophila* is the same as in *P. simiae*.

For most of the genes in the oxidative pathway, the mutant phenotypes were stronger (the fitness values were more negative) during growth on deoxyribonate than on deoxyribose (Fig. 5). It appears that *P. bryophila* uses the oxidative pathway to grow

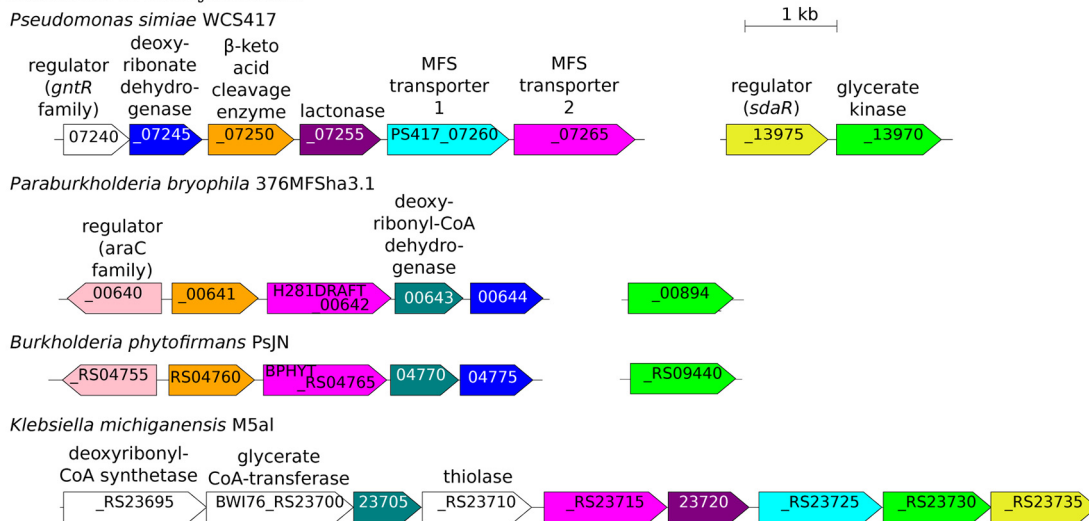


### A. Overview of the catabolic pathways

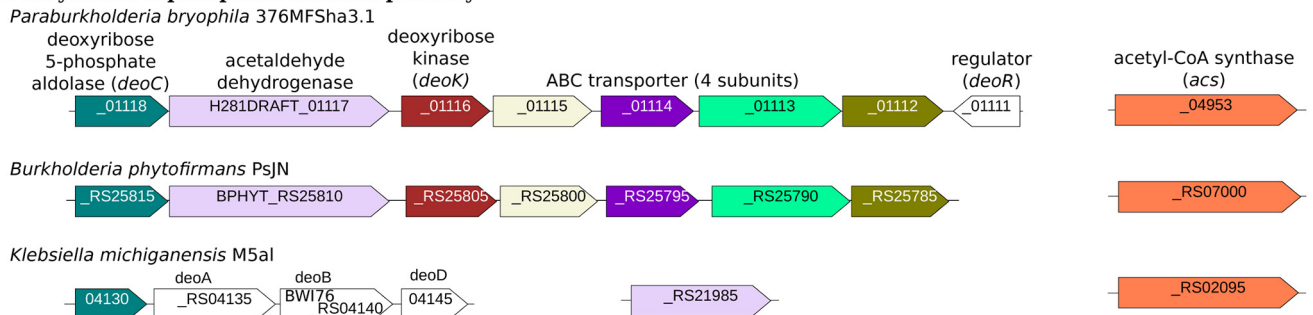


**s** *P. simiae*  
**b** *P. bryophila*  
**p** *B. phytofirmans*  
**m** *K. michiganensis*

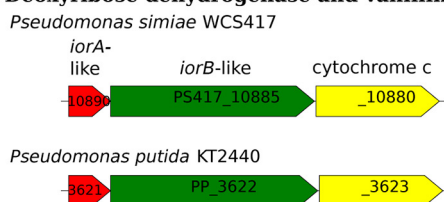
### B. Oxidation of deoxyribonate



### C. Deoxyribose-5-phosphate aldolase pathway



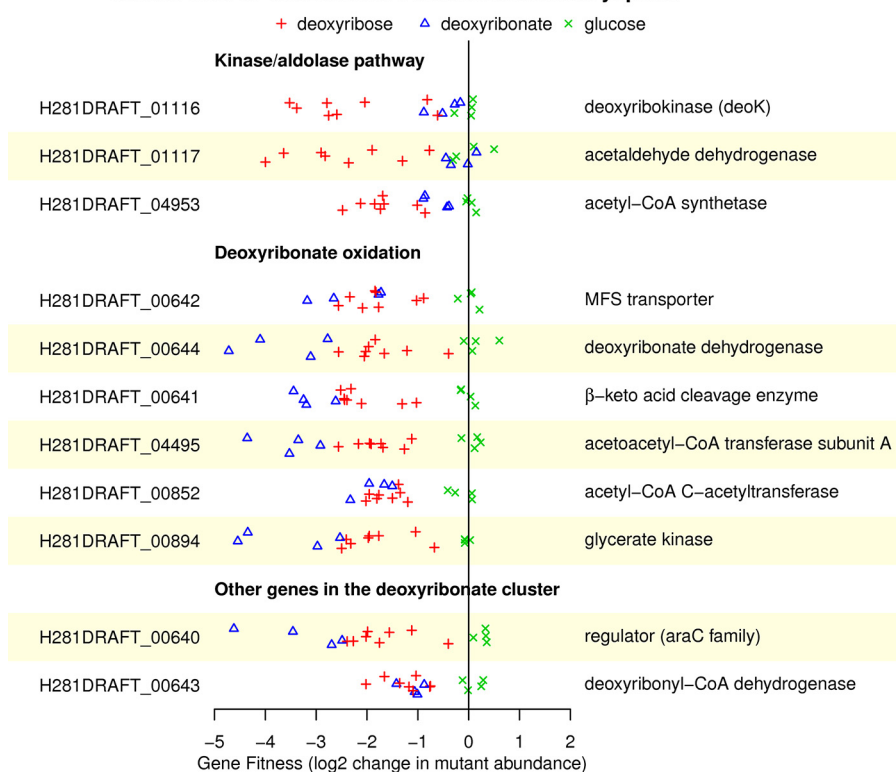
### D. Deoxyribose dehydrogenase and vanillin dehydrogenase



**FIG 4** Genome organization of the deoxyribose utilization pathways. (A) Overview of the catabolism of deoxyribose and related compounds by 4 bacteria. For each transformation, we show which of the bacteria performs it. (B) Deoxyribonate oxidation in four bacteria. (C) The deoxyribose-phosphate aldolase pathway in three bacteria. (D) Deoxyribose dehydrogenase from *P. simiae* and vanillin dehydrogenase from *P. putida*. In panels B to D, adjacent genes are connected by horizontal lines, genes that are potential orthologs are the same color, and the scale is the same. All of the putative enzymes shown are important for the utilization of deoxyribose or related compounds, except for the *iorAB*-like cluster in *P. putida* and the cleavage enzyme and deoxyribonyl-CoA dehydrogenase in *B. phytofirmans*. We lack fitness data for *deoC* from *P. bryophila*.

on deoxyribonate, but deoxyribose is metabolized by the oxidative pathway in parallel with the kinase/aldolase pathway.

We did not identify any candidates for deoxyribose dehydrogenase in the fitness data from *P. bryophila*. This activity may be genetically redundant or nonspecific, or the kinase/aldolase pathway may suffice for efficient growth on deoxyribose.

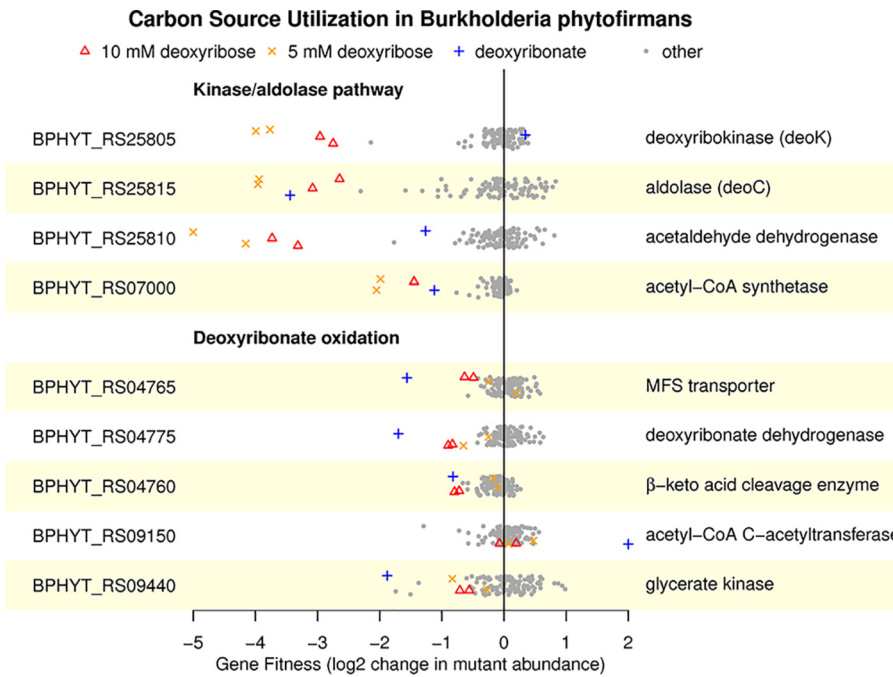
Carbon Source Utilization in *Paraburkholderia bryophila*

**FIG 5** Mutant fitness data for the kinase/aldolase and oxidative pathways in *Paraburkholderia bryophila* 376MF5ha3.1. Fitness values for the deoxyribose 5-phosphate aldolase (*deoC*) and for subunit B of acetoacetyl-CoA transferase were not estimated due to inadequate coverage. The data are plotted as in Fig. 1B.

In the genome of *P. bryophila*, the oxidative enzymes are clustered with a transporter (H281DRAFT\_00642), a regulator of the *araC* family (H281DRAFT\_00640), and a putative β-hydroxyacyl-CoA dehydrogenase (H281DRAFT\_00643) (Fig. 4B). These genes are also important for the utilization of both deoxyribose and deoxyribonate, although the β-hydroxyacyl-CoA dehydrogenase mutants have a milder phenotype (Fig. 5). Based on the phenotype of a homolog of this protein in *K. michiganensis* (see below), we suspect that the β-hydroxyacyl-CoA dehydrogenase acts on deoxyribonyl-CoA. So, we speculate that *P. bryophila* may use a third parallel pathway for deoxyribose catabolism.

When we performed fitness assays with *B. phytofirmans*, we found similar results to *P. bryophila* for deoxyribonate utilization (Fig. 6). In particular, deoxyribonate dehydrogenase (BPHYT\_RS04775), the β-keto acid cleavage enzyme (BPHYT\_RS04760), and glycerate kinase (BPHYT\_RS09440) are all important during growth on deoxyribonate, although β-keto acid cleavage enzyme mutants have a milder phenotype. One unexplained result is that the acetyl-CoA C-acetyltransferase (BPHYT\_RS09150) is detrimental to fitness on deoxyribonate (mutants in this gene increased more than 4-fold in abundance relative to other mutants in the pooled assay). During growth of *B. phytofirmans* on deoxyribose, we observe strong phenotypes for the enzymes in the kinase/aldolase pathway (Fig. 6). Genes in the oxidative pathway are somewhat important for fitness during growth on deoxyribose at 10 mM (but not at 5 mM). Although these defects are modest, the defects are larger during growth on 10 mM deoxyribose than in most of 80 other conditions that we previously studied (data from reference 6) (Fig. 6).

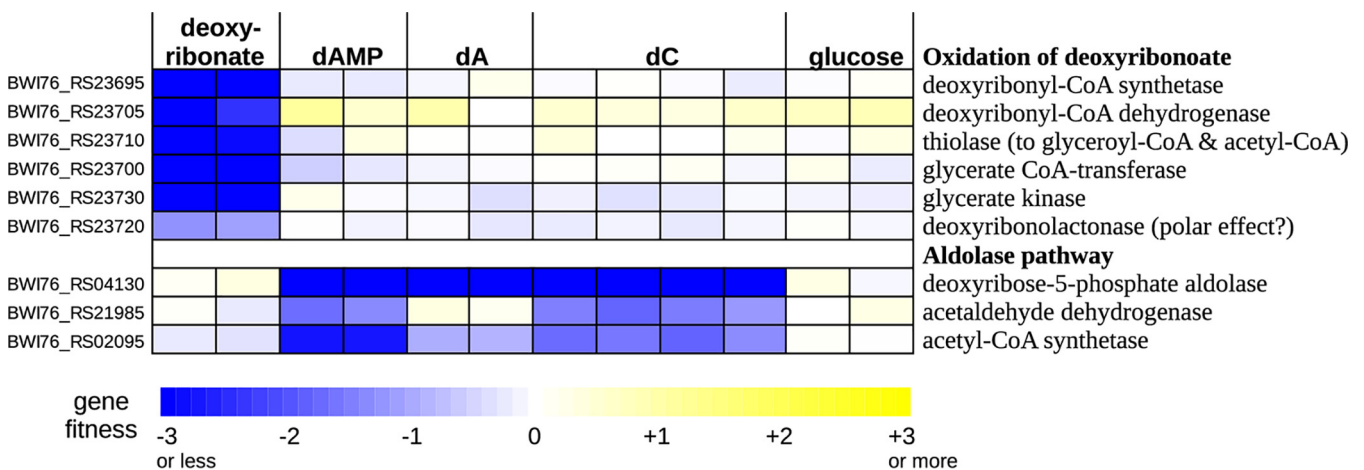
Overall, we conclude that both *Paraburkholderia bryophila* and *Burkholderia phytofirmans* catabolize deoxyribonate by the same oxidative pathway that *P. simiae* uses.



**FIG 6** Mutant fitness data for the aldolase and oxidative pathways in *Burkholderia phytofirmans* PsJN. Fitness values outside of the plot’s range are shown at -5 or +2. The data are plotted as in Fig. 1B.

Furthermore, the oxidative pathway contributes to deoxyribose utilization in *P. bryophila* and probably in *B. phytofirmans* as well. However, in contrast to *P. simiae*, both *P. bryophila* and *B. phytofirmans* also use the aldolase-kinase pathway for deoxyribose catabolism.

**Oxidation of deoxyribonate in *Klebsiella michiganensis*.** As mentioned above, the genome of *Klebsiella michiganensis* contains a gene cluster that shares some genes with the deoxyribonate oxidation cluster of *P. simiae* (Fig. 4B). In *K. michiganensis*, the cluster is important for the utilization of deoxyribonate but not for the utilization of deoxynucleosides or deoxynucleotides (Fig. 7) or under any of 92 other conditions that we previously profiled (6). The cluster includes a putative CoA-synthetase (BWI76\_RS23695), a β-hydroxyacyl-CoA dehydrogenase (BWI76\_RS23705), a thiolase (BWI76\_RS23710), a CoA-transferase (BWI76\_23700), and glyc-



**FIG 7** Mutant fitness data for the utilization of deoxyribonate and various deoxynucleotides/deoxynucleosides in *Klebsiella michiganensis* M5aI. dAMP is 2'-deoxyadenosine 5'-phosphate, dA is 2'-deoxyadenosine, and dC is 2'-deoxycytidine.

erate kinase (BWI76\_RS23730). We propose that the CoA-synthetase converts deoxyribonate to deoxyribonyl-CoA, the dehydrogenase oxidizes this to 3-ketodeoxyribonyl-CoA, and the thiolase cleaves this to glyceryl-CoA and acetyl-CoA (see Fig. S1 in the supplemental material). These transformations are analogous to the oxidative pathway of the other three bacteria, but with CoA bound intermediates. In all four bacteria, deoxyribonate is oxidized and cleaved to yield glyceryl-CoA and acetyl-CoA, but the point at which the CoA thioesterification occurs varies. In *K. michiganensis*, the cluster also includes a CoA-transferase that may convert the glyceryl-CoA to glycerate and a glycerate kinase that forms 2-phosphoglycerate, which feeds into glycolysis (as in the other bacteria).

The deoxyribonate utilization cluster of *K. michiganensis* also includes a close homolog (70% amino acid identity) of the putative deoxyribonolactonase from *P. simiae*. This gene (BWI76\_RS23720) was also somewhat important for deoxyribonate utilization (fitnesses of  $-1.3$  and  $-1.1$ ). It is possible that the latter phenotype is due to polar effects that inhibit the expression of downstream genes (which include glycerate kinase). Or, because deoxyribonolactone and deoxyribonate interconvert in solution, the lactonase's activity may be useful. We also observed a similar mild phenotype for the lactonase during deoxyribonate utilization in *P. simiae* (Fig. 1B).

As in *P. simiae*, the cluster in *K. michiganensis* contains two MFS transporter genes (Fig. 4B). These are probably orthologs of the two transporters in *P. simiae*'s cluster. (Each pair of proteins is over 50% identical between the two organisms.) As in *P. simiae*, both of the transporters in *K. michiganensis* are important for growth on deoxyribonate, although BWI76\_RS23715 mutants have a stronger phenotype than BWI76\_RS23725 mutants (average fitness of  $-4.0$  versus  $-1.4$ ).

Although the genome of *K. michiganensis* contains a deoxyribose-phosphate aldolase (BWI76\_RS04130), it does not seem to encode deoxyribose kinase. This might explain why it did not grow with deoxyribose as the carbon source. We did obtain fitness data with three deoxynucleotides or deoxynucleosides as carbon sources: 2'-deoxyadenosine 5'-phosphate, 2'-deoxyadenosine, and 2'-deoxycytidine (Fig. 7). These data confirmed that *K. michiganensis* uses the aldolase pathway to break down deoxynucleosides. Deoxynucleosides are probably converted to deoxyribose 5-phosphate by the combination of a phosphorylase (either DeoA or DeoD; BWI76\_RS04135 or BWI76\_RS04145) which forms deoxyribose 1-phosphate and a phosphopentomutase (DeoB, BWI76\_RS04140) which converts it to deoxyribose 5-phosphate. *deoABD* were important for growth on deoxynucleotides and deoxynucleosides (all fitness values were under  $-1$ , except for *deoD* on 2'-deoxycytidine). Based on these data, we propose that *K. michiganensis* uses an oxidative pathway with acyl-CoA intermediates to metabolize deoxyribonate, it uses the aldolase pathway to metabolize deoxynucleosides, and it cannot metabolize deoxyribose.

**Deoxyribose dehydrogenase is closely related to a novel vanillin dehydrogenase from *Pseudomonas putida* KT2440.** We found that all three genes from the deoxyribose dehydrogenase-like cluster of *Pseudomonas putida* KT2440 (PP\_3621 to PP\_3623) were specifically important for growth on 5 mM vanillin (all fitness values of  $\leq -2$  in two replicate experiments). These genes were not important for growth with other aromatic compounds as carbon sources (vanillic acid, benzoic acid, ferulic acid, or *p*-coumaric acid) (see Fig. S2 in the supplemental material).

Because PP\_3621 to PP\_3623 are important for growth on vanillin but not on vanillic acid (which is the oxidized form), we propose that it is the major vanillin dehydrogenase in *P. putida*. Although another vanillin dehydrogenase from *P. putida* (*vdh*, or PP\_3357) has been studied biochemically, *vdh* is not required for vanillin utilization (20). (Our data instead suggest a role for *vdh* as a *p*-hydroxybenzaldehyde dehydrogenase during the catabolism of *p*-coumaric acid [Fig. S2].) Vanillin utilization by *P. putida* requires the uptake of molybdate (21), which we can now explain because molybdate is a precursor for the molybdenum cofactor of PP\_3621 to PP\_3623. The vanillin dehydrogenase is expected to produce vanillate, which is demethylated (by PP\_3736 to PP\_3737) to protocatechuic acid, which feeds into the ortho-cleavage pathway that

begins with protocatechuate 3,4-dioxygenase (PP\_4655 to PP\_4656). The demethylase and dioxygenase genes are also specifically important during growth on vanillin (all fitness values under  $-2$  [Fig. S2]).

The genome of *P. putida* does not contain the deoxyribonate oxidation genes, and *P. putida* does not grow with either deoxyribose or deoxyribonate as the sole carbon source. Conversely, none of the four other bacteria that we studied grew with vanillin as the sole carbon source, and their genomes do not encode orthologs of the vanillin *O*-demethylase. However all three components of *P. simiae*'s deoxyribose dehydrogenase are very similar to the components of *P. putida*'s vanillin dehydrogenase (84 to 92% amino acid identity). We speculate that both enzymes are active on a broader range of substrates.

## DISCUSSION

**Oxidation and cleavage of deoxyribose and deoxyribonate.** Using a genetic approach, we identified pathways for the oxidation of deoxyribose and/or deoxyribonate in four genera of bacteria. Although we have biochemical evidence for the conversion of deoxyribose to a ketodeoxyribonate in *Pseudomonas simiae*, we do not have direct evidence for the cleavage to  $D$ -glyceryl-CoA and acetyl-CoA. However the importance of glycerate kinase for deoxyribonate utilization by four different bacteria is difficult to explain otherwise. Also, for the three bacteria that use a  $\beta$ -keto cleavage enzyme during growth on deoxyribonate, experimental data from a homologous enzyme is consistent with our proposal. Specifically, *in vitro* assays of the  $\beta$ -keto cleavage enzyme BKACE\_178 (UniProt accession no. [A6X2V8](#)), which is 61% identical to PS417\_07250, found that it had some activity with 5-hydroxy- $\beta$ -ketoheptanoate,  $\beta$ -ketopentanoate,  $\beta$ -ketoisocaproate, and  $\beta$ -ketoheptanoate as substrates (12). These substrates are similar to 2-deoxy-3-keto-ribonate (which is 4,5-dihydroxy- $\beta$ -keto-pentanoate). Furthermore, the oxidative cluster in *Klebsiella michiganensis* contains the glycerate kinase gene along with close homologs of three genes from the oxidative cluster of *P. simiae* (the lactonase gene and both MFS transporter genes [Fig. 4B]). This suggests that these gene clusters evolved to catabolize deoxyribonate via glycerate and 2-phosphoglycerate.

**What are the natural substrates of the oxidative pathways?** We identified deoxyribose dehydrogenase in just one of the four bacteria that grow on deoxyribonate. Furthermore, the natural role of the deoxyribose dehydrogenase of *P. simiae* remains uncertain. Its close similarity to the major vanillin dehydrogenase of *Pseudomonas putida* suggests that both enzymes might act on a broader range of substrates. In addition, *P. simiae* grows much more slowly on deoxyribose than on deoxyribonate, with a doubling time of around 12 h instead of around 2 h (Fig. 2), which suggests that deoxyribose might not be the natural substrate. Overall, it appears that deoxyribose utilization is not the main reason that these bacteria maintain a cluster for deoxyribonate oxidation.

We also wondered if 2-deoxyglucose might be a substrate for the oxidative pathway. A strain of *Pseudomonas* was reported to grow on 2-deoxyglucose by oxidizing it to 2-deoxy-3-ketogluconate (22). This transformation is chemically analogous to the conversion of 2-deoxyribose to 2-deoxy-3-ketoribonate. We tested the five bacteria considered here with 2.5 to 20 mM 2-deoxyglucose as a carbon source and found that none of them grew. Also, 2-deoxyglucose is not, as far as we know, a naturally occurring compound.

Instead, we propose that deoxyribonate is the natural substrate for the oxidative pathways. There are several plausible pathways by which deoxyribonate may form. First, nonspecific oxidation of deoxyribose, which may occur in *Paraburkholderia bryophila* and in *Burkholderia phytofirmans*, could yield deoxyribonate. Although we did not identify the genes responsible for this nonspecific activity, other sugar 1-dehydrogenase enzymes are known to be promiscuous, such as bifunctional  $L$ -arabinose/ $D$ -galactose 1-dehydrogenases (6, 23, 24). Also, because the deoxyribose dehydrogenase from *P. simiae* probably acts on substrates in the periplasm,

**TABLE 1** Primers used in this study

Primer name	Sequence (5' to 3')
confirm_mariner	GAGCTTAGTACGTAATC
confirm_04200::Tn	ATGGCGATAGTGCCCGGTC
confirm_07245::Tn	GCGGGCGTCGGCAATTCCTG
confirm_07250::Tn	GTCGGTGCAGCGCAGGATG
confirm_10885::Tn	CTGGGTGAGCATCAGGTGG
Barseq_P1 (5N)	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTNNNNNGTCGACCTGCAGCGTACG
Barseq_P1_2N	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTNNNGTCGACCTGCAGCGTACG
Barseq_P1_3N	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTNNNGTCGACCTGCAGCGTACG
Barseq_P1_4N	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTNNNNNGTCGACCTGCAGCGTACG
BarSeq_P2_noindex	CAAGCAGAAGACGGCATACGAGATNNNNNNNGTACTGGAGITTCAGACGTGTGCTCTCCGATCTGATGTCCACGAGGTCTCT
BarSeq_P1_SangerSeq	GAGATCTACACTCTTTCCCTACAC
amplify_10885_forward	AACCTCATCAACGAAGTCTCG
amplify_10885_reverse	TACGCTGAAAGTGTGTGCGGCTC
amplify_10890_forward	GAATTACGAATCAACCAAAAGGC
amplify_10890_reverse	TACAGCGCCCCCTGCTTG
amplify_07255_forward	CGTATCGAAGTCTTGTGCGA
amplify_07255_reverse	CTAGCTGGCGAACCGTGC
amplify_07245_forward	AGTGTCTACAACGGCTG
amplify_07245_reverse	TTACAGGTACTCGACGTTACC
amplify_pET32_rev	ATGATGATGATGATGGTGCATATGGCC
amplify_pET32_for	TCTTCTGGTCTGGTCCACGCGGTTCC

some bacteria might obtain energy (but not carbon) by oxidizing deoxyribose in the periplasm without metabolizing it further. Deoxyribose in DNA is abundant, so even a low rate of nonspecific oxidation could select for deoxyribonate catabolism. All four of the deoxyribonate-catabolizing bacteria in this study are associated with plant roots. Also, the deoxyribonate utilization cluster is present in about 20% of the *Pseudomonas* genomes in PATRIC (25) and in 34% of plant-associated *Pseudomonas*, but it is rarely found in *Pseudomonas* isolates from humans (under 1%). DNA is estimated to be ~3% of the dry weight of a typical plant cell (26), which implies that deoxyribose in DNA is around 1% of biomass. Second, deoxyribonate or deoxyribonolactone could be released after oxidative damage of DNA (9). Third, deoxyribonate is proposed to form spontaneously from the metabolic intermediate 2-keto-3-deoxy-D-gluconate and hydrogen peroxide (10). Although we do not know of any measurement of the concentration of deoxyribonate in plants or in soils, 2-deoxyribonate is present in human urine at about 40-fold less than the concentration of creatinine (8), or around 25 mg/day excreted by the typical adult.

## MATERIALS AND METHODS

**Strains and growth media.** *Pseudomonas simiae* WCS417 and *Paraburkholderia bryophila* 376MF-Sha3.1 were provided by Jeff Dangl (University of North Carolina). Pools of barcoded transposon mutants of *Pseudomonas simiae* WCS417, *Pseudomonas putida* KT2440, *Burkholderia phytofirmans* PsJN, and *Klebsiella michiganensis* M5al were described previously (6, 27). Each mutant has a transposon insertion that includes *kanR* (providing kanamycin resistance), as well as a 20-nucleotide barcode flanked by common priming sites.

Individual strains from the library of mutants for *P. simiae* were obtained from an arrayed collection (28). Briefly, the library was spread on agar plates containing LB (Luria-Bertani) plus kanamycin, and colonies were picked and arrayed into 384-well plates containing LB with 7.5% glycerol and kanamycin and allowed to grow overnight. Next, 25 nl of each well was collected as part of a multiplexing strategy involving pooling of rows, columns, and plates, and these pools were subjected to PCR amplification and sequencing of the insertion barcodes. The location of each mutant was determined using an in-house script to identify the row, column, and plate pools sharing the same barcode. The insertion barcodes of the specific mutants used in this study were confirmed by Sanger sequencing of colonies recovered from glycerol stocks, using PCR with BarSeq\_P1 and BarSeq\_P2\_noindex primers to amplify the region and BarSeq\_P1\_SangerSeq as the sequencing primer (Table 1). Furthermore, the presence of an insertion within the expected gene was confirmed by PCR amplification of the transposon junction using a mariner transposon-specific primer and a gene-specific primer (see the “confirm” primers in Table 1).

A transposon mutant library of *Paraburkholderia bryophila* 376MFSha3.1 (Burk376\_ML3) was constructed via conjugation with an *E. coli* donor strain (APA752) carrying the randomly barcoded mariner transposon delivery vector pKMW3, as described previously (7). Briefly, we conjugated a 1:1 ratio of mid-log-phase *P. bryophila* to mid-log-phase APA752 for 12 h at 30°C on LB plates supplemented with diaminoipimelic acid (DAP). The *E. coli* donor strain is derived from WM3064, a DAP auxotroph. After

conjugation, we selected for mutants by plating on LB plates supplemented with 100  $\mu\text{g/ml}$  kanamycin at 30°C. After pooling about 100,000 colonies, we subinoculated the cells into 50 ml of LB with 100  $\mu\text{g/ml}$  kanamycin at a starting optical density at 600 nm ( $\text{OD}_{600}$ ) of 0.2, grew the library to saturation at 30°C, and made multiple freezer stocks in glycerol. The mapping of transposon insertion locations and the identification of their associated DNA barcodes were performed as described previously (7). Burk376\_ML3 contains 106,195 mapped strains with unique DNA barcodes and contains at least one central insertion (in 10 to 90% of the gene's length) for 5,624 of 6,555 predicted protein-coding genes. Of 63,242 strains with insertions in the central parts of genes, 59,045 were at sufficient abundance in the recovered library and were used to compute gene fitness.

All bacteria were cultured in LB medium during recovery from glycerol stocks. For the recovery of mutant libraries or individual mutant strains, 50  $\mu\text{g/ml}$  kanamycin was added to the LB. All mutant libraries except *K. michiganensis* were recovered until the cells reached mid-log phase ( $\text{OD}_{600}$  of around 1.0). For *K. michiganensis*, we recovered the mutant library until the culture reached saturation ( $\text{OD}_{600}$  of around 4.0). All bacterial growth described in this study was conducted at 30°C. For cloning work, LB was used as the rich medium. All chemicals for medium preparations, mutant fitness assays, growth assays, and biochemical reactions were purchased from Sigma unless otherwise mentioned. For deoxyribonate, we used the lithium salt.

To assay fitness or growth on a specific carbon source, we generally used a defined medium with 5 to 20 mM of the carbon source as well as 30 mM PIPES [piperazine-*N,N'*-bis(2-ethanesulfonic acid)] sesquisodium salt, 0.25 g/liter ammonium chloride, 0.1 g/liter potassium chloride, 0.6 g/liter sodium phosphate monobasic monohydrate, Wolfe's vitamins, and Wolfe's minerals (<http://fit.genomics.lbl.gov>). For some fitness assays of *P. putida*, we instead used a minimal MOPS medium [40 mM 3-(*N*-morpholino)propanesulfonic acid, 4 mM Tricine, 1.32 mM potassium phosphate dibasic, 10  $\mu\text{M}$  iron(II) sulfate heptahydrate, 9.5 mM ammonium chloride, 0.276 mM aluminum potassium sulfate dodecahydrate, 0.5  $\mu\text{M}$  calcium chloride, 0.525 mM magnesium chloride hexahydrate, 50 mM sodium chloride, 3 nM ammonium heptamolybdate tetrahydrate, 0.4  $\mu\text{M}$  boric acid, 30 nM cobalt chloride hexahydrate, 10 nM copper(II) sulfate pentahydrate, 80 nM manganese(II) chloride tetrahydrate, and 10 nM zinc sulfate heptahydrate].

For the growth assays with individual mutants of *P. simiae* (Fig. 2), we washed the cells 3 times with defined media before starting the growth experiments. (The cells were pregrown in LB.) We grew these individual strains in a 96-well microplate within a Tecan Sunrise instrument, with readings every 15 min, as described previously (6).

**Genome-wide mutant fitness assays.** For pooled fitness assays (RB-TnSeq), cells from the recovered mutant library were washed 3 times with defined media without a carbon source and then inoculated into the medium of choice at an  $\text{OD}_{600}$  of 0.02 and grown until saturation in a transparent 24-well microplate plate in a Multitron shaker. Total genomic DNA was extracted and barcodes were PCR amplified as described previously (7), but with some modifications to the primers. For most of the experiments, the P1 primer included variable spacing between the Illumina adapter and the sequence that flanks the barcode (2 to 5 N's instead of always 5 N's [Table 1]). Barcodes were sequenced using an Illumina HiSeq 4000, which can lead to some bleed-through between samples (in our experience, 0.1% to 0.3% of reads from BarSeq; for the likely mechanism behind bleed-through, see reference 29). To eliminate this bleed-through, some experiments used a P1 primer that contained an additional sequence to verify that it came from the expected sample. (The P2 primer contains the tag that is used for demultiplexing by Illumina software; in our earlier design, the P1 primer was the same mix for all the samples.) Specifically, the new P1 primers contain the same sequence as BarSeq\_P1 before the N's, 1 to 4 N's (which varies with the index), the reverse (not the reverse complement) of the 6-nucleotide (nt) index sequence, and the sequence from BarSeq\_P1 after the N's.

**Analysis of the mutant fitness data.** The fitness data were analyzed as described previously (7) using publicly available scripts (<https://bitbucket.org/berkeleylab/feba>). Briefly, the fitness of a strain is the normalized  $\log_2$  ratio of the number of reads for its barcode in the sample after growth versus the control sample (taken after the cells recover from the freezer). The fitness of a gene is the weighted average of the fitness of strains with insertions in the central 10 to 90% of the gene. For example, fitness values were computed for 4,414 of the 5,033 predicted nonessential proteins in *P. simiae*, and for the typical (median) protein, the fitness values of five independent mutant strains were averaged. For each experiment, the gene fitness values are normalized so that the mode of the distribution is at zero.

To assess the statistical significance of each fitness value, we used a *t*-like test statistic of the form  $\text{fitness}/\sqrt{\text{estimated variance}}$  (7). For most of the bacteria, a gene was considered to have a specific phenotype in an experiment if  $\text{fitness} < -1$ ,  $t < -5$ , 95% of fitness values are between  $-1$  and 1, and  $|\text{fitness}| > 95\text{th percentile}(|\text{fitness}|) + 0.5$ , as described previously (6). However, for *P. simiae*, 9/162 experiments (6%) are for deoxyribose or deoxyribonate, so the 95th percentile threshold for a specific phenotype was not appropriate. We used the 90th percentile instead.

In *P. simiae*, we found 19 genes with a specific phenotype in at least one of six deoxyribose experiments. We manually excluded 6 of these genes from Fig. 1B because they did not have a consistent fitness defect. These genes were as follows: PS417\_08675, or *modA*, which is involved in molybdenum cofactor biosynthesis; PS417\_23250, or *tsaA*, which is involved in tRNA modification; and four regulatory or signaling genes, PS417\_00445, PS417\_07615, PS417\_13295, and PS417\_22670.

**Sequence analysis.** Protein sequences were analyzed by using PaperBLAST (30) to search for characterized homologs, with the Fitness Browser (<http://fit.genomics.lbl.gov> [6]), which incorporates results from Pfam (31), TIGRFAMs (32), KEGG (33), and SEED/RAST (34), by using the Conserved Domain

Database (35) to search for similarity to gene families and by using PSORTb (16) to predict protein localization.

To identify a motif for the putative binding sites of the transcription factor PS417\_07240, we first used the MicrobesOnline tree-browser (36) to identify orthologs with conserved synteny. We identified similar proteins from *P. fluorescens* SBW25, *P. fluorescens* Pf-5, *Pseudomonas syringae* B728a, and *B. graminis* C4D1M (44 to 99% amino acid identity) that were also upstream of an SDR dehydrogenase and a  $\beta$ -keto acid cleavage enzyme. We ran MEME (37) on the 200 nucleotides upstream of these five genes.

**Cloning and protein purification.** We used the expression vector pET32a for cloning and protein purification as described previously (38). We PCR amplified four genes from *P. simiae* WCS417 (PS417\_10885, PS417\_10890, PS417\_07255, and PS417\_07245) using the amplify primers listed in Table 1 and cloned the PCR products into pET32a using Gibson assembly (39). For PS417\_07255, PS417\_07245, and PS417\_10890, we transformed the Gibson assembly products directly into the *E. coli* expression strain BL21(DE3). However, for PS417\_10885, we first transformed into *E. coli* TOP10 to identify the correct plasmid. We then transformed this plasmid into the *E. coli* BL21(DE3) expression strain. Protein was purified using a cobalt affinity column (Clontech Talon kit) and quantified using a Nanodrop spectrophotometer (Thermo Fisher). Purified PS417\_07245 and PS417\_07255 had the expected molecular weights (29 kDa and 39 kDa, respectively). Although we were not able to purify PS417\_10885 or PS417\_10890, protein gels of cell lysates from *E. coli* strains that expressed these proteins showed predominant bands at the expected molecular weights (80 and 17 kDa, respectively).

**Enzymatic assays.** Enzymatic assays were performed with purified enzymes (2 to 5  $\mu$ M), except that clarified cell lysates of *E. coli* expressing PS417\_10885 or PS417\_10890 were mixed together to obtain deoxyribose dehydrogenase. Each enzymatic assay was done in 50  $\mu$ l of 100 mM ammonium bicarbonate ( $\text{NH}_4\text{HCO}_3$ ) buffer with three independent replicates, run for 1 h at room temperature, and then diluted 1,000-fold in ammonium bicarbonate buffer and frozen at  $-20^\circ\text{C}$ . NADH oxidase was purchased from Nacalai USA (San Diego, CA).

**Liquid chromatography and mass spectrometry of end products.** Samples were analyzed using a Thermo-Dionex UltiMate3000 RSLCnano liquid chromatograph that was connected in-line with an LTQ-Orbitrap-XL mass spectrometer equipped with a nano-electrospray ionization (nano-ESI) source (Thermo Fisher Scientific, Waltham, MA). The liquid chromatograph was equipped with a  $\text{C}_{18}$  analytical column (length, 150 mm; inner diameter, 0.075 mm; particle size, 3  $\mu$ m; pore size, 100  $\text{\AA}$ ; Thermo Acclaim) and a 1- $\mu$ l sample loop. Acetonitrile, formic acid (Optima grade, 99.9%; Fisher), and water purified to a resistivity of 18.2  $\text{M}\Omega\text{-cm}$  (at  $25^\circ\text{C}$ ) using a Milli-Q Gradient ultrapure water purification system (Millipore, Billerica, MA) were used to prepare mobile-phase solvents. Solvent A was 99.9% water–0.1% formic acid, and solvent B was 99.9% acetonitrile–0.1% formic acid (vol/vol). The elution program consisted of isocratic flow at 1% B for 3 min, a linear gradient to 95% B over 1 min, isocratic flow at 95% B for 3 min, and isocratic flow at 1% B for 13 min, at a flow rate of 300 nl/min. Full-scan mass spectra were acquired in the positive-ion mode over the range  $m/z = 70$  to 1,500 using the Orbitrap mass analyzer, in profile format, with a mass resolution setting of 60,000 (at  $m/z = 400$ , measured at full width at half-maximum peak height). For tandem mass spectrometry (MS/MS) analysis, precursor ions were fragmented using collision-induced dissociation (CID) under the following conditions: MS/MS spectra acquired using the linear ion trap, centroid format, isolation width of 2  $m/z$  units, normalized collision energy of 35%, default charge state of 1+, activation Q of 0.25, and activation time of 30 ms. Data acquisition and analysis were performed using Xcalibur software (version 2.0.7; Thermo).

**Proteomics.** *P. simiae* was inoculated at an initial  $\text{OD}_{600}$  of 0.04 into 1 ml of defined medium that contained 10 mM either glucose, deoxyribose, or lithium deoxyribonate. At late log phase ( $\text{OD}_{600}$  of 0.4 to 0.7), cells were pelleted by centrifugation at  $4,000 \times g$  and resuspended in 100 mM  $\text{NH}_4\text{HCO}_3$  (pH 7.5) buffer (Sigma-Aldrich, St. Louis, MO). Fifty microliters of a 1- $\mu\text{g}/\mu\text{l}$  sample of each cell suspension was mixed with 10  $\mu\text{l}$  100 mM  $\text{NH}_4\text{HCO}_3$  (pH 7.5) and 25  $\mu\text{l}$  of a 0.2% aqueous solution of Rapigest SF (Waters, Milford, MA), and the sample was placed in a block heater at  $80^\circ\text{C}$  for 15 min and vortexed briefly to lyse cells and solubilize proteins. Five microliters of a 1- $\mu\text{g}/\mu\text{l}$  solution of Trypsin Gold porcine protease (Promega, Madison, WI) was added, and the sample was incubated overnight at  $37^\circ\text{C}$  for the tryptic digest of proteins. The next morning, 10  $\mu\text{l}$  of trifluoroacetic acid (Sigma-Aldrich) was added, and the sample was incubated for 90 min at  $37^\circ\text{C}$  to hydrolyze the Rapigest SF. The sample was centrifuged at  $21,130 \times g$  for 30 min at  $6^\circ\text{C}$ , and the supernatant was transferred to an autosampler vial for mass spectrometry analysis.

Trypsin-digested proteins were analyzed using a Synapt G2-Si high-definition ion mobility mass spectrometer that was equipped with a nano-electrospray ionization source and connected in line with an Acquity M-class ultraperformance liquid chromatography (UPLC) system (Waters, Milford, MA). Data-independent, ion mobility-enabled mass spectra and tandem mass spectra (40–42) were acquired in the positive-ion mode. Data acquisition was controlled using MassLynx software (version 4.1). Tryptic peptide identification, relative protein quantification using a label-free approach (43, 44), and calculation of statistical analysis of variance (ANOVA) *P* values were performed using Progenesis Q1 for Proteomics software (version 4.0; Waters). The abundance of 1,277 proteins was quantified.

**Data availability.** The fitness data are available from the Fitness Browser (<http://fit.genomics.lbl.gov>). The fitness data and the proteomics data are archived at figshare (<https://doi.org/10.6084/m9.figshare.7304159.v1>). The 96 primer sequences described in the section “Genome-wide mutant fitness assays” are available with the analysis code (<https://bitbucket.org/berkeleylab/feba>) in the source file named primers/barseq3.index2.



## APPENDIX

**Oxidation of deoxyribose to a ketodeoxyribonate *in vitro*.** To test the activity of deoxyribose dehydrogenase from *P. simiae in vitro*, we recombinantly expressed the *iorA*- and *iorB*-like components (PS417\_10890 and PS147\_10885) with polyhistidine (6× His) tags in *E. coli*. We tried to use a cobalt affinity column to purify each subunit, but did not succeed. Instead, we combined cell lysates from two strains of *E. coli* that overexpressed each subunit.

We incubated these cell lysates with 1 mM deoxyribose and with 100 μM phenazine methosulfate as the electron acceptor. The *iorAB*-like dehydrogenase converted deoxyribose to a compound with the same mass-to-charge ratio (*m/z*) as deoxyribonate (see Fig. S3 in the supplemental material). Tandem mass spectrometry confirmed that this product had a fragmentation pattern very similar to that of a 2-deoxy-D-ribonate standard (see Fig. S4 in the supplemental material). In contrast, addition of an equivalent amount of cell lysate from *E. coli* with an empty vector yielded only a small amount of deoxyribonate (less than a tenth as much [Fig. S3]). This shows that the deoxyribose dehydrogenase activity was primarily due to the *iorAB*-like dehydrogenase from *P. simiae* and not to an enzyme that is native to *E. coli*. No deoxyribonate formed if no lysate was added, so the conversion does not occur spontaneously in solution. Using polyhistidine tags, we were able to purify the lactonase (PS417\_07255). The conversion of deoxyribose to deoxyribonate occurred with or without the added lactonase, perhaps because deoxyribonolactone in water can spontaneously hydrolyze to deoxyribonate or because lactonases are present in the cell lysate. In both reactions, and in the starting sample of deoxyribose, we detected a small amount of a compound with an *m/z* of 155.0 that could be deoxyribonolactone, although the abundance of this ion was too low for tandem mass spectrometry.

We also purified the deoxyribonate dehydrogenase (PS417\_07245) using a polyhistidine tag and a cobalt affinity column. We incubated the purified enzyme with 1 mM deoxyribonate and with 1 mM NAD<sup>+</sup> as the electron acceptor. We observed a product with the molecular weight of 2-deoxy-3-keto-D-ribonate (a putative sodium adduct with an *m/z* of 171.02). The putative ketodeoxyribonate was not detected in the absence of the deoxyribonate dehydrogenase or when either deoxyribose or deoxyribonate was incubated with the *iorAB*-like dehydrogenase. Tandem mass spectrometry showed the loss of CO<sub>2</sub> or H<sub>2</sub>O, as expected for a sugar acid (see Fig. S5 in the supplemental material). The amount of product was small and difficult to quantify (perhaps 3 to 4% of the ion intensity of the input deoxyribonate), but in the absence of the enzyme, the product was below the limit of detection. We also tested the deoxyribonate dehydrogenase for activity with deoxyribose as the substrate, but we did not observe any formation of deoxyribonate or ketodeoxyribonate.

Because the deoxyribonate dehydrogenase reaction is similar to the 2-deoxygluconate 3-dehydrogenase reaction, which is thermodynamically unfavorable (22), we tried to reoxidize the NADH in order to drive the reaction forward. Specifically, we incorporated 0.002 U/μl of NADH oxidase, which converts NADH + O<sub>2</sub> to NAD<sup>+</sup> + H<sub>2</sub>O<sub>2</sub>. This change increased the yield of the putative ketodeoxyribonate product to 12 to 14%. The retention time of the product by high-performance liquid chromatography (HPLC) was identical to that of deoxyribonate, so we did not try to separate them.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00297-18>.

**FIG S1**, PDF file, 0.2 MB.

**FIG S2**, PDF file, 0.3 MB.

**FIG S3**, PDF file, 0.1 MB.

**FIG S4**, PDF file, 0.1 MB.

**FIG S5**, PDF file, 0.1 MB.

## ACKNOWLEDGMENTS

We thank Mitchell Thompson for growing the *P. putida* mutant library on several carbon sources.

This material by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies; <http://enigma.lbl.gov>), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under contract no. DE-AC02-05CH11231. The QB3/Chemistry Mass Spectrometry Facility at UC Berkeley receives support from NIH (grant no. 1S10OD020062-01). This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH Instrumentation Grant S10 OD018174.

## REFERENCES

- Levy-Booth DJ, Campbell RG, Gulden RH, Hart MM, Powell JR, Klironomos JN, Pauls KP, Swanton CJ, Trevors JT, Dunfield KE. 2007. Cycling of extracellular DNA in the soil environment. *Soil Biol Biochem* 39: 2977–2991. <https://doi.org/10.1016/j.soilbio.2007.06.020>.
- Rittmann D, Sorger-Herrmann U, Wendisch VF. 2005. Phosphate starvation-inducible gene *ushA* encodes a 5' nucleotidase required for growth of *Corynebacterium glutamicum* on media with nucleotides as the phosphorus source. *Appl Environ Microbiol* 71:4339–4344. <https://doi.org/10.1128/AEM.71.8.4339-4344.2005>.
- Koszalka GW, Krenitsky TA. 1979. Nucleosidases from *Leishmania donovani*. Pyrimidine ribonucleosidase, purine ribonucleosidase, and a novel purine 2'-deoxyribonucleosidase. *J Biol Chem* 254:8185–8193.
- Tourneux L, Bucurenci N, Saveanu C, Kaminski PA, Bouzon M, Pistotnik E, Namane A, Marlière P, Bârzu O, Li De La Sierra I, Neuhard J, Gilles AM. 2000. Genetic and biochemical characterization of *Salmonella enterica* serovar Typhi deoxyribokinase. *J Bacteriol* 182:869–873.
- Valentin-Hansen P, Boëtius F, Hammer-Jespersen K, Svendsen I. 1982. The primary structure of *Escherichia coli* K12 2-deoxyribose 5-phosphate aldolase. Nucleotide sequence of the *deoC* gene and the amino acid sequence of the enzyme. *Eur J Biochem* 125:561–566.
- Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, Kuehl JV, Melnyk RA, Lamson JS, Suh Y, Carlson HK, Esquivel Z, Sadeeshkumar H, Chakraborty R, Zane GM, Rubin BE, Wall JD, Visel A, Bristow J, Blow MJ, Arkin AP, Deutschbauer AM. 2018. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* 557:503–509. <https://doi.org/10.1038/s41586-018-0124-0>.
- Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, Deutschbauer A. 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio* 6:e00306-15. <https://doi.org/10.1128/mBio.00306-15>.
- Chamberlin BA, Sweeley CC. 1987. Metabolic profiles of urinary organic acids recovered from absorbent filter paper. *Clin Chem* 33:572–576.
- Kappen LS, Goldberg IH. 1989. Identification of 2-deoxyribonolactone at the site of neocarzinostatin-induced cytosine release in the sequence d(AGC). *Biochemistry* 28:1027–1032.
- Lerma-Ortiz C, Jeffries JG, Cooper AJL, Niehaus TD, Thamm AMK, Frelin O, Aunins T, Fiehn O, de Crecy-Lagard V, Henry CS, Hanson AD. 2016. "Nothing of chemistry disappears in biology": the top 30 damage-prone endogenous metabolites. *Biochem Soc Trans* 44:961–971. <https://doi.org/10.1042/BST20160073>.
- Lehmann M, Tshisuaka B, Fetzner S, Lingens F. 1995. Molecular cloning of the isoquinoline 1-oxidoreductase genes from *Pseudomonas diminuta* 7, structural analysis of *iorA* and *iorB*, and sequence comparisons with other molybdenum-containing hydroxylases. *J Biol Chem* 270:14420–14429.
- Bastard K, Smith AA, Vergne-Vaxelaire C, Perret A, Zaparucha A, De Melo-Minardi R, Mariage A, Boutard M, Debarb A, Lechaplais C, Pelle C, Pellouin V, Perchat N, Petit JL, Kreimeyer A, Medigue C, Weissenbach J, Artiguenave F, De Berardinis V, Vallenet D, Salanoubat M. 2014. Revealing the hidden functional diversity of an enzyme family. *Nat Chem Biol* 10:42–49. <https://doi.org/10.1038/nchembio.1387>.
- Lemieux RU, Anderson L, Conner AH. 1971. The mutarotation of 2-deoxy-D-erythro-pentose ("2-deoxy-D-ribose"). Conformations, kinetics, and equilibria. *Carbohydr Res* 20:59–72.
- Bartsch O, Hagemann M, Bauwe H. 2008. Only plant-type (GLYK) glycerate kinases produce D-glycerate 3-phosphate. *FEBS Lett* 582: 3025–3028. <https://doi.org/10.1016/j.febslet.2008.07.038>.
- Zelbuch L, Razo-Mejia M, Herz E, Yahav S, Antonovsky N, Kroytoro H, Milo R, Bar-Even A. 2015. An in vivo metabolic approach for deciphering the product specificity of glycerate kinase proves that both *E. coli*'s glycerate kinases generate 2-phosphoglycerate. *PLoS One* 10:e0122957. <https://doi.org/10.1371/journal.pone.0122957>.
- Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26:1608–1615. <https://doi.org/10.1093/bioinformatics/btq249>.
- Imam S, Noguera DR, Donohue TJ. 2015. CceR and AkgR regulate central carbon and energy metabolism in alphaproteobacteria. *mBio* 6:e02461-14. <https://doi.org/10.1128/mBio.02461-14>.
- Novichkov PS, Kazakov AE, Ravcheev DA, Leyn SA, Kovaleva GY, Sutormin RA, Kazanov MD, Riehl W, Arkin AP, Dubchak I, Rodionov DA. 2013. RegPrecise 3.0—a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* 14:745. <https://doi.org/10.1186/1471-2164-14-745>.
- Dobbek H, Gremer L, Meyer O, Huber R. 1999. Crystal structure and mechanism of CO dehydrogenase, a molybdo iron-sulfur flavoprotein containing S-selenylcysteine. *Proc Natl Acad Sci U S A* 96:8884–8889.
- Simon O, Kläiber I, Huber A, Pfannstiel J. 2014. Comprehensive proteome analysis of the response of *Pseudomonas putida* KT2440 to the flavor compound vanillin. *J Proteomics* 109:212–227. <https://doi.org/10.1016/j.jprot.2014.07.006>.
- Graf N, Altenbuchner J. 2014. Genetic engineering of *Pseudomonas putida* KT2440 for rapid and high-yield production of vanillin from ferulic acid. *Appl Microbiol Biotechnol* 98:137–149. <https://doi.org/10.1007/s00253-013-5303-1>.
- Eichhorn MM, Cynkin MA. 1965. Microbial metabolism of 2-deoxyglucose; 2-deoxygluconic acid dehydrogenase. *Biochemistry* 4:159–165.
- Watanabe S, Shimada N, Tajima K, Kodaki T, Makino K. 2006. Identification and characterization of L-arabinonate dehydratase, L-2-keto-3-deoxyarabinonate dehydratase, and L-arabinolactonase involved in an alternative pathway of L-arabinose metabolism. Novel evolutionary insight into sugar metabolism. *J Biol Chem* 281:33521–33536. <https://doi.org/10.1074/jbc.M606727200>.
- Aro-Kärkkäinen N, Toivari M, Maaheimo H, Ylilauri M, Pentikäinen OT, Andberg M, Oja M, Penttilä M, Wiebe MG, Ruohonen L, Koivula A. 2014. L-Arabinose/D-galactose 1-dehydrogenase of *Rhizobium leguminosarum* bv. trifolii characterised and applied for bioconversion of L-arabinose to L-arabinonate with *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol* 98:9653–9665. <https://doi.org/10.1007/s00253-014-6039-2>.
- Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res* 45: D535–D542. <https://doi.org/10.1093/nar/gkw1017>.
- Landenmark HKE, Forgan DH, Cockell CS. 2015. An estimate of the total

- DNA in the biosphere. *PLoS Biol* 13:e1002168. <https://doi.org/10.1371/journal.pbio.1002168>.
27. Rand JM, Pisithkul T, Clark RL, Thiede JM, Mehrer CR, Agnew DE, Campbell CE, Markley AL, Price MN, Ray J, Wetmore KM, Suh Y, Arkin AP, Deutschbauer AM, Amador-Noguez D, Pfleger BF. 2017. A metabolic pathway for catabolizing levulinic acid in bacteria. *Nat Microbiol* 2:1624–1634. <https://doi.org/10.1038/s41564-017-0028-z>.
  28. Cole BJ, Feltcher ME, Waters RJ, Wetmore KM, Mucyn TS, Ryan EM, Wang G, Ul-Hasan S, McDonald M, Yoshikuni Y, Malmstrom RR, Deutschbauer AM, Dangel JL, Visel A. 2017. Genome-wide identification of bacterial plant colonization genes. *PLoS Biol* 15:e2002860. <https://doi.org/10.1371/journal.pbio.2002860>.
  29. Sinha R, Stanley G, Gulati GS, Ezran C, Travaglini KJ, Wei E, Chan CKF, Nabhan AN, Su T, Morganti RM, Conley SD, Chaib H, Red-Horse K, Longaker MT, Snyder MP, Krasnow MA, Weissman IL. 2017. Index switching causes “spreading-of-signal” among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. *bioRxiv* <https://doi.org/10.1101/125724>.
  30. Price MN, Arkin AP. 2017. PaperBLAST: text mining papers for information about homologs. *mSystems* 2:e00039-17. <https://doi.org/10.1128/mSystems.00039-17>.
  31. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heeger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the Protein Families Database. *Nucleic Acids Res* 42:D222–D230. <https://doi.org/10.1093/nar/gkt1223>.
  32. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. 2012. Tigrfams and genome properties in 2013. *Nucleic Acids Res* 41:D387–D395. <https://doi.org/10.1093/nar/gks1234>.
  33. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462. <https://doi.org/10.1093/nar/gkv1070>.
  34. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>.
  35. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI’s conserved domain database. *Nucleic Acids Res* 43:D222–D226. <https://doi.org/10.1093/nar/gku1221>.
  36. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, Dubchak IL, Alm EJ, Arkin AP. 2010. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res* 38:D396–D400. <https://doi.org/10.1093/nar/gkp919>.
  37. Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28–36.
  38. Baran R, Bowen BP, Price MN, Arkin AP, Deutschbauer AM, Northen TR. 2013. Metabolic footprinting of mutant libraries to map metabolite utilization to genotype. *ACS Chem Biol* 8:189–199. <https://doi.org/10.1021/cb300477w>.
  39. Gibson DG. 2011. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol* 498:349–361. <https://doi.org/10.1016/B978-0-12-385120-8.00015-2>.
  40. Plumb RS, Johnson KA, Rainville P, Smith BW, Wilson ID, Castro-Perez JM, Nicholson JK. 2006. UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun Mass Spectrom* 20:1989–1994. <https://doi.org/10.1002/rcm.2550>.
  41. Shliha PV, Bond NJ, Gatto L, Lilley KS. 2013. Effects of traveling wave ion mobility separation on data independent acquisition in proteomics studies. *J Proteome Res* 12:2323–2339. <https://doi.org/10.1021/pr300775k>.
  42. Distler U, Kuharev J, Navarro P, Levin Y, Schild H, Tenzer S. 2014. Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat Methods* 11:167–170. <https://doi.org/10.1038/nmeth.2767>.
  43. Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, Lee A, van Sluyter SC, Haynes PA. 2011. Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* 11:535–553. <https://doi.org/10.1002/pmic.201000553>.
  44. Nahnsen S, Bielow C, Reinert K, Kohlbacher O. 2013. Tools for label-free peptide quantification. *Mol Cell Proteomics* 12:549–556. <https://doi.org/10.1074/mcp.R112.025163>.