An Oxidative Pathway of Deoxyribose Catabolism

Authors: Jayashree Ray,[1] Morgan N. Price,[1*] Anthony T. Iavarone,[2] Hans K. Carlson,[1] Elizabeth M. Ryan,[3] Rex R. Malmstrom,[3] Adam P. Arkin,[1*] and Adam M. Deutschbauer[1*]

[1] Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley CA, USA
[2] QB3/Chemistry Mass Spectrometry Facility, University of California Berkeley, Berkeley CA, USA
[3] DOE Joint Genome Institute, Walnut Creek CA, USA
[*] morgannprice@yahoo.com, aparkin@lbl.gov, amdeutschbauer@lbl.gov

Abstract

High-throughput genetics is a powerful approach to identify new genes involved in bacterial carbon catabolism. Here, we used genome-wide fitness assays to identify a novel pathway for 2-deoxy-D-ribose catabolism in *Pseudomonas simiae* WCS417. The genes that were important for deoxyribose utilization but not in other conditions included two putative dehydrogenases, a lactonase, a β-keto acid cleavage enzyme, and a glycerate kinase. We propose that deoxyribose is oxidized twice to 2-deoxy-3-keto-D-ribonoate (a β-keto acid) before cleavage to D-glycerate, which is phosphorylated and enters lower glycolysis. We purified the two dehydrogenases and reconstituted the enzymatic pathway for the conversion of deoxyribose to 2-deoxy-3-keto-D-ribonoate *in vitro*.

Introduction

DNA is present in every cell, and once released into the environment by cell lysis, some of the DNA is probably hydrolyzed to 2-deoxy-D-ribose (deoxyribose) (Levy-Booth et al. 2007). Indeed, some nucleosidases are specific for 2'-deoxyribonucleosides over ribonucleosides and release deoxyribose (Koszalka and Krenitsky 1979). The only previously-published pathway for catabolizing deoxyribose of which we are aware involves deoxyribose kinase (*deoK) (Tourneux et al. 2000)* and deoxyribose 5-phosphate aldolase (*deoC) (Valentin-Hansen et al. 1982)*, which yields acetaldehyde and D-glyceraldehyde 3-phosphate.

During a large-scale screen for the growth of *Pseudomonas simiae* WCS417 on various carbon sources (Price et al. 2016), we found that it grows on deoxyribose even though its genome does not contain either *deoK* or *deoC*. Using a pool of over 100,000 randomly-barcoded transposon mutants of *P. simiae*, we tested which genes were important for growth in 89 different conditions, including growth on deoxyribose as the sole source of carbon (Price et al. 2016). Here, we use this genetic data to identify a novel oxidative pathway for deoxyribose catabolism

in *P. simiae*. To test our prediction of this novel pathway, we assayed the pool of mutants during growth on 2-deoxy-D-ribonoate (deoxyribonoate), which is an intermediate in the pathway; we measured the growth of individual mutants in pathway enzymes; and we reconstituted the enzymatic conversion of deoxyribose to deoxyribonoate to ketodeoxyribonoate *in vitro*.

Results

The genetic basis of growth on deoxyribose and deoxyribonoate in *P. simiae* WCS417

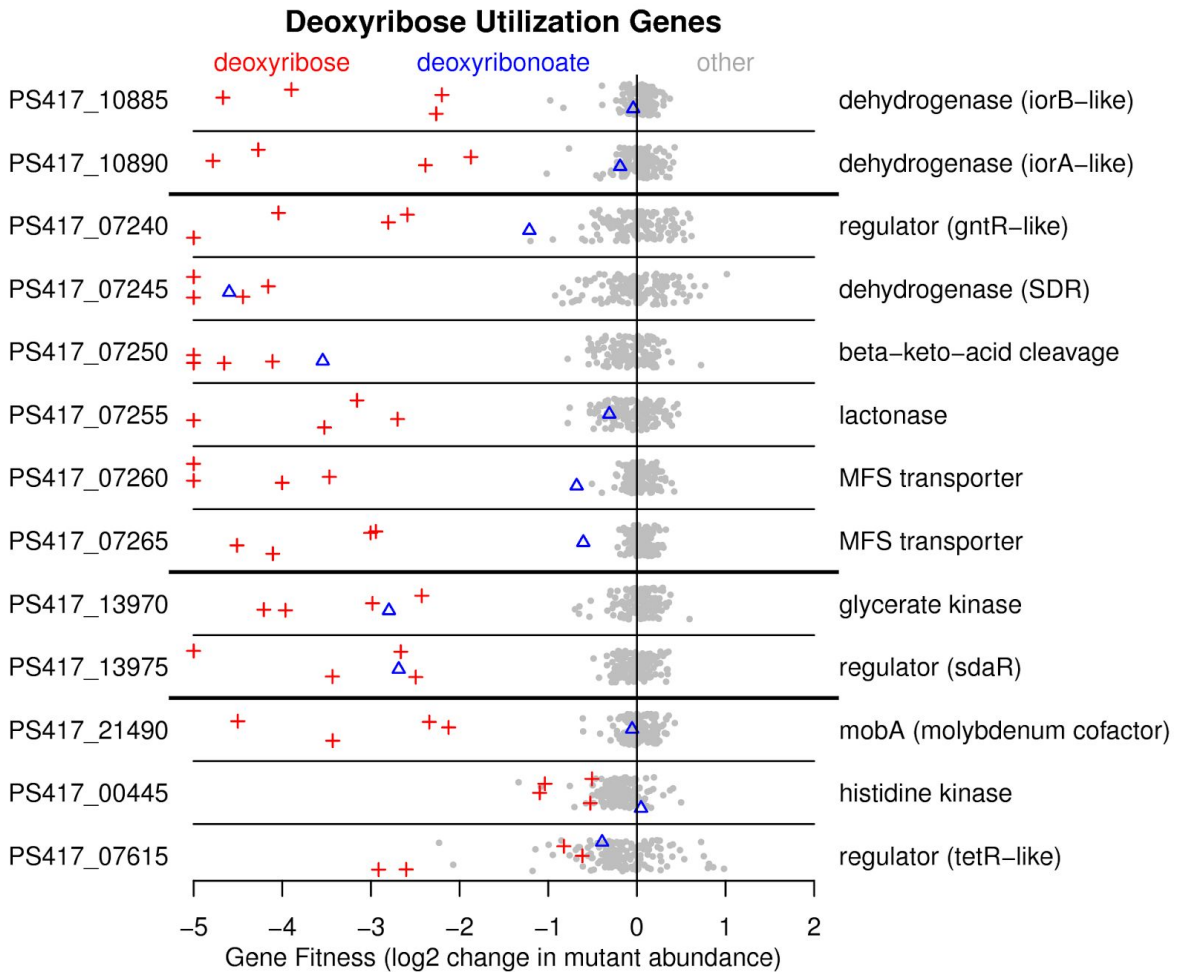

Figure 1: Mutant fitness data for the 13 genes that are specifically important for utilizing deoxyribose. In every panel, each point shows the gene's fitness value (*x* axis) in a different experiment, and the *y* axis is random. Fitness values under -1 indicate a defect in growth. Fitness values under -5 are shown at -5. Fitness experiments with deoxyribose or deoxyribonoate as the carbon source are highlighted. The genes are grouped together if they cluster on the chromosome. (For adjacent genes, the locus numbers differ by five.)

To supplement the previously-described fitness data for WCS417 (Price et al. 2016), we tested deoxyribose as the carbon source on another day, and we also tested deoxyribonoate as the carbon source. Overall we had 4 biological replicates for fitness in deoxyribose, 1 experiment for fitness in deoxyribonoate, and 146 other fitness assays in diverse conditions, including 47 additional carbon sources. From these data, we identified 13 genes that were important for utilizing deoxyribose, had significant phenotypes in at least 2 of the 4 replicates, and were not important for fitness in most other conditions (see Methods). The fitness data for these genes is shown in Figure 1. In the genome of WCS417, 10 of the 13 genes cluster into three groups of adjacent genes that are on the same strand and are probably operons.

The 13 genes include two putative transporters, four regulatory genes, and *mobA* (PS417_21490), which is involved in the biosynthesis of molybdenum cofactor. The remaining six genes are likely catabolic enzymes: two subunits of a molybdenum-dependent dehydrogenase (PS417_10890, PS417_10885) that is related to isoquinoline 1-oxidoreductase *iorAB* (Lehmann et al. 1995); a dehydrogenase of the short chain dehydrogenase/reductase (SDR) superfamily (PS417_07245); a β-keto acid cleavage enzyme (PS417_07250) (Bastard et al. 2014); a lactonase (PS417_07255); and glycerate kinase (PS417_13970).

Most of the enzymatic genes were also important for growth on deoxyribonoate (Figure 1). The exceptions were the two subunits of the molybdenum-dependent dehydrogenase and the lactonase.

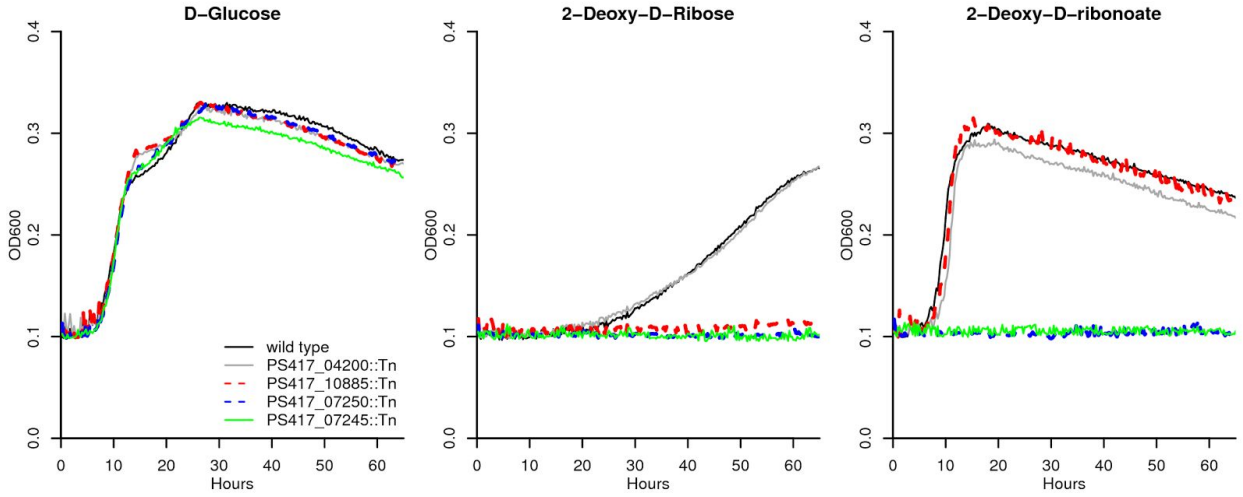

Figure 2: Growth curves for individual transposon mutants. We compared the growth of wild-type *P. simiae* WCS417 and of individual transposon mutant strains in defined media with three carbon sources: D-glucose, deoxyribose, and deoxyribonoate. The mutant of PS417_04200 (2-ketoglutaric semialdehyde dehydrogenase) is included as a control. Each curve is the median of six replicates. Each carbon source was provided at 10 mM.

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 deoxyribonoate as the sole

source of carbon. We found that a transposon mutant of the *iorB*-like subunit of the molybdenum-dependent dehydrogenase (PS417_10885) grew on glucose or deoxyribonoate but not on deoxyribose (Figure 2). Mutants of the SDR dehydrogenase (PS417_07245) or the β-keto acid cleavage enzyme (PS417_07250) grew on glucose but not on deoxyribose or deoxyribonoate (Figure 2).

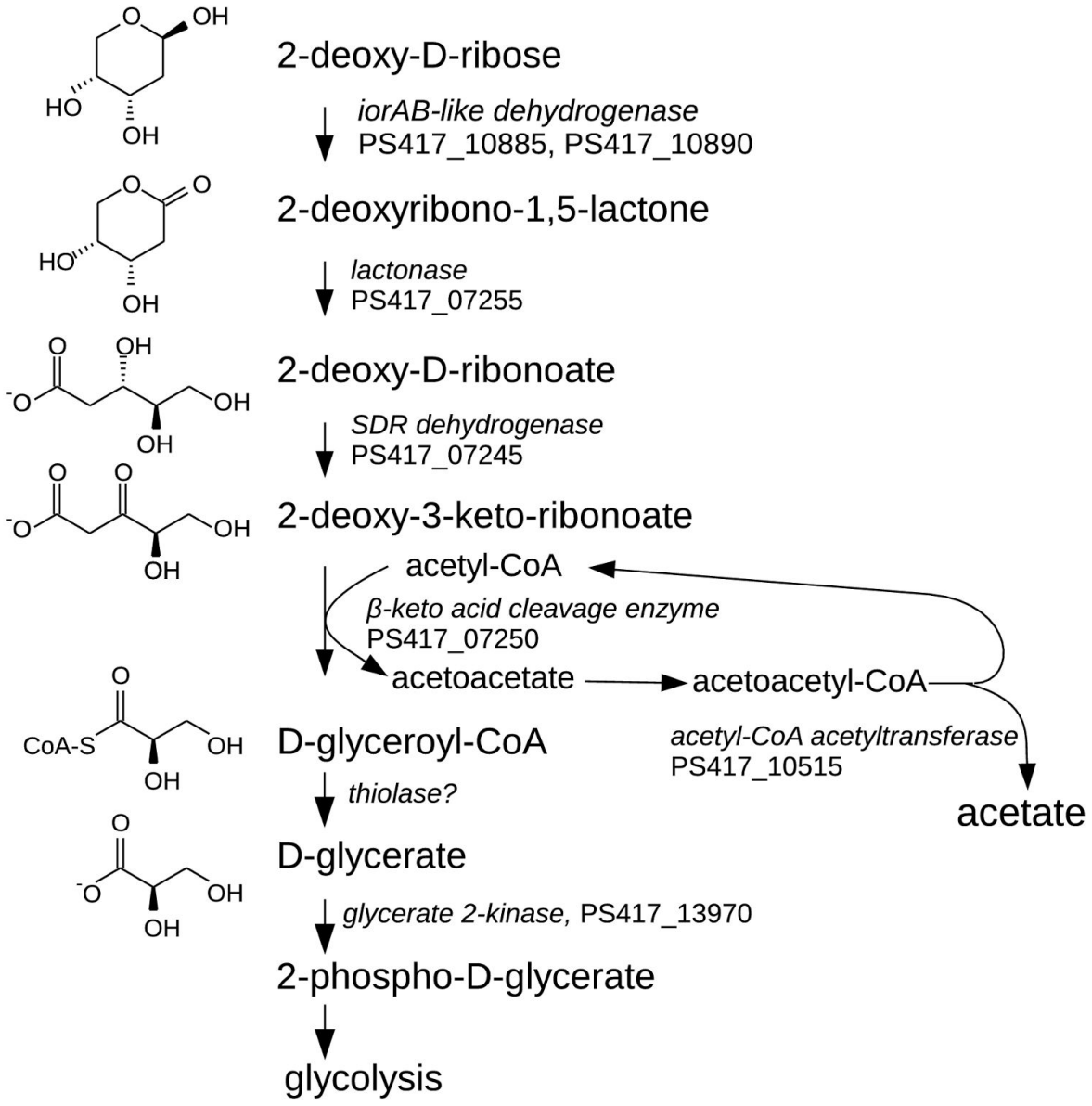

Figure 3: The proposed pathway of deoxyribose catabolism in *P. simiae* WCS417.

Proposed pathway for deoxyribose catabolism in *P. simiae* WCS417

The genetic data suggested that deoxyribose is catabolized by an oxidative pathway (Figure 3). Because the molybdenum-dependent dehydrogenase and the lactonase were not required for growth on deoxyribonoate, it appears that the molybdenum-dependent dehydrogenase oxidizes deoxyribose to a lactone, which is hydrolyzed to deoxyribonoate by the lactonase. We predict that the lactone intermediate is 2-deoxy-D-ribono-1,5-lactone instead of the 1,4-lactone because the pyranose form of deoxyribose is more stable in solution at 30℃ (Lemieux et al. 1971), but we have no relevant data.

Oxidation of the deoxyribonoate by the SDR enzyme could yield 2-deoxy-3-keto-D-ribonoate, which is a β-keto acid and could be cleaved by the β-keto acid cleavage enzyme encoded by PS417_07250. These cleavage enzymes remove an acetyl group from a β-keto acid and transfer it to acetyl-CoA (Bastard et al. 2014). In this case, the β-keto acid cleavage enzyme is expected to yield D-glyceroyl-CoA and acetoacetate. The requirement for the glycerate kinase suggests that the D-glyceroyl-CoA is converted to glycerate. The glycerate kinase is 51% identical to glycerate kinase 1 from *Escherichia coli (garK)*, which forms 2-phosphoglycerate (Bartsch et al. 2008; Zelcbuch et al. 2015). 2-phosphoglycerate is an intermediate in glycolysis and thus links the catabolic pathway to central metabolism.

This proposed pathway accounts for all of the putative enzymes that are specifically important for growth on deoxyribose. The two remaining issues are the conversion of D-glyceroyl-CoA to D-glycerate and the fate of the acetoacetate that would be produced by the β-keto acid cleavage enzyme. First, we looked for a thiolase or CoA-transferase that might act on D-glyceroyl-CoA and that was important for deoxyribose utilization in the mutant fitness data. There were many candidate genes in the genome, but none with that phenotype. This activity could be genetically redundant and might not be identified by assaying mutants with one transposon insertion. Second, we expect that the acetoacetate would be activated to acetoacetyl-CoA by a CoA-transferase (PS417_10525:PS417_10520) or a CoA-synthetase (there are several candidates) and then cleaved by acetyl-CoA acetyltransferase to give acetyl-CoA and acetate. None of the genes for the activation step was important for the utilization of deoxyribose; they might be genetically redundant. But one of the acetyl-CoA acetyltransferases (PS417_10515) was important for fitness (fitness < -1) in three of the four deoxyribose experiments as well as on deoxyribonoate. This gene did not appear in the list of specifically-important genes because it met the threshold for statistical significance in just 1 of the 4 deoxyribose experiments. Nevertheless, this supports our proposal that acetoacetate is an intermediate in deoxyribose utilization.

Predicted roles of other genes in deoxyribose utilization

Given the proposed pathway, we can explain the roles of several of the other genes with specific phenotypes on deoxyribose (Figure 1). First, the putative operon that includes the lactonase, the SDR deoxyribonoate dehydrogenase, and the β-keto acid cleavage enzyme also

includes two genes from the major facilitator superfamily of transporters (MFS) (PS417_07260, PS417_07265). These are presumably involved in the uptake of deoxyribose, although we cannot explain why there are two transporter genes instead of just one. MFS transporters usually have 12 transmembrane helices, but PS417_07260 is predicted (Sonnhammer et al. 1998) to have just 11 transmembrane helices, so it may have another role. The clustering of the lactonase and subsequent enzymes with the putative deoxyribose transporter suggests that deoxyribose is the natural substrate for this pathway.

Second, a *gntR*-like transcription factor (PS417_07240) is also present in this operon. By comparing the regions upstream of this operon and of similar operons from other strains of *Pseudomonas* or from *Burkholderia graminis*, we identified the conserved palindrome GTGATCAC. This sequence occurs at -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* (Imam et al. 2015), which is 32% identical to PS417_07240. We propose that PS417_07240 regulates the expression of the operon by binding to this motif.

Third, the transcription factor PS417_13975 (*sdaR*) is probably an activator for the expression of the glycerate kinase (PS417_13970), which is downstream of *sdaR*. Similarly, a close homolog of PS417_13975 from *P. fluorescens* Pf-5 (PFL_3379, 88% identical) is predicted in RegPrecise (Novichkov et al. 2013) to regulate a downstream glycerate kinase.

Fourth, *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 the deoxyribose dehydrogenase, which is homologous to the *coxLS* subunits of a CO dehydrogenase that has a molybdenum-cytosine dinucleotide cofactor (Dobbek et al. 1999). If *mobA* is required for the activity of deoxyribose dehydrogenase, then this also explains why it is not important for utilizing deoxyribonoate (Figure 1). Similarly, genes for the biosynthesis of molybdopterin (a precursor to molybdenum cofactor) are important for utilizing deoxyribose but not deoxyribonoate (*moaABCE*: PS417_21395, PS417_08670, PS417_04865, PS417_04875). These genes' phenotypes are not specific to deoxyribose (and hence are not shown in Figure 1) because they are also important for fitness in other conditions such as inosine utilization and nitrate stress (data of (Price et al. 2016)).

The two remaining genes encode regulatory proteins (the histidine kinase PS417_00445 and the *tetR*-like regulator PS417_07615) and their phenotypes on deoxyribose are less consistent (Figure 1). These genes might be indirectly involved in the regulation of deoxyribose catabolism.

Conversion of deoxyribose to ketodeoxyribonoate *in vitro*

We recombinantly expressed four proteins with polyhistidine (6X His) tags in *E. coli*: both subunits of the *iorAB*-like dehydrogenase (PS417_10885 and PS417_10890), the lactonase (PS417_07255), and the SDR dehydrogenase (PS417_07245). Using a cobalt affinity column,

we were able to purify the lactonase and the SDR deoxyribonoate dehydrogenase. We were not able to purify the subunits of the *iorAB*-like dehydrogenase, so to assay the activity of this enzyme, we instead combined lysates from strains of *E. coli* that overexpressed each subunit.

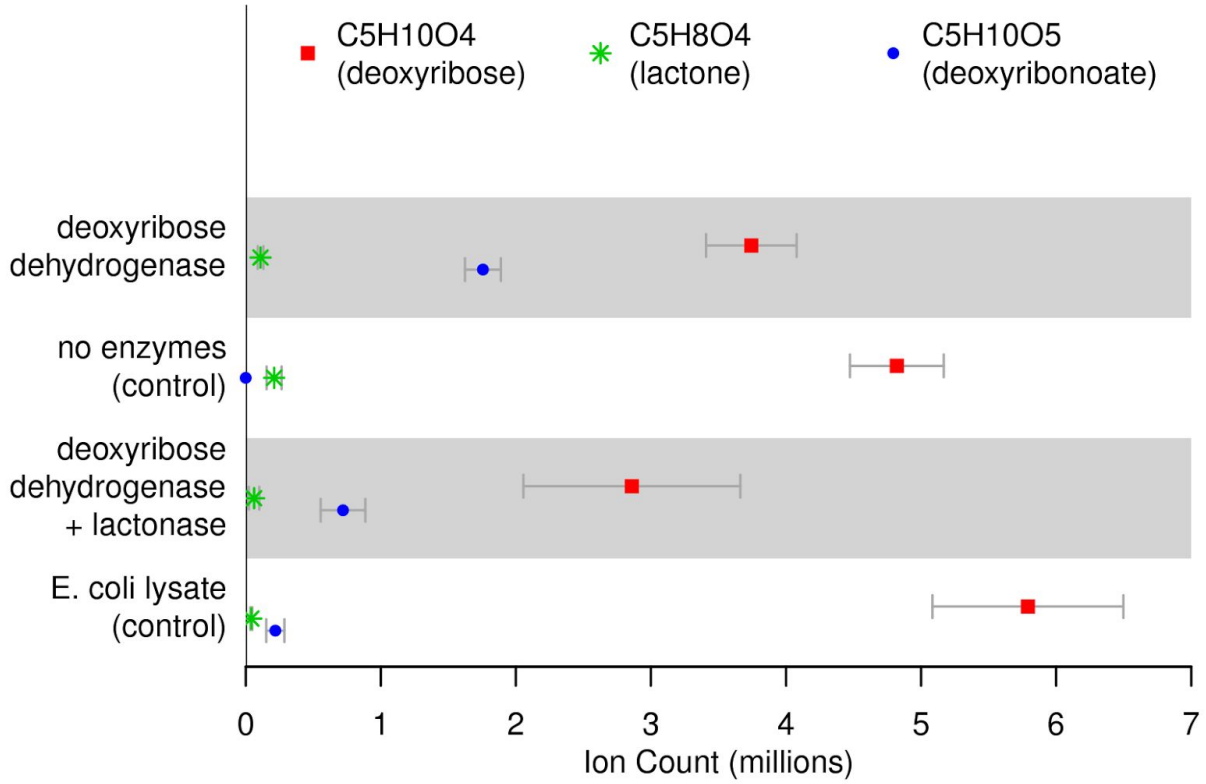

Figure 4: Reconstitution of the oxidation of deoxyribose to deoxyribonoate *in vitro*. Deoxyribose, a potential deoxyribonolactone, and deoxyribonoate were detected as sodium adducts (*m/z* = 157.05, 155.0, and 173.04, respectively). Each point is the average of 3 replicates and the error bar is the standard deviation.

To test the conversion of deoxyribose to deoxyribonoate *in vitro*, we incubated 1 mM deoxyribose with cell lysates containing both subunits of the *iorAB*-like dehydrogenase and 100 μM phenazine methosulfate as the electron acceptor. We optionally added the purified lactonase as well. As shown in Figure 4, the *iorAB*-like dehydrogenase converted deoxyribose to a compound with the same mass-to-charge ratio (*m/z*) as deoxyribonoate. Tandem mass spectrometry confirmed that this product had a fragmentation pattern very similar to that of a 2-deoxy-D-ribonoate standard (Figure 5). In contrast, addition of an equivalent amount of cell lysate from *E. coli* with an empty vector yielded only a small amount of deoxyribonoate (Figure 4). This shows that the deoxyribose dehydrogenase activity was due to the *iorAB*-like dehydrogenase and not to an enzyme that is native to *E. coli.* No deoxyribonoate formed if no enzymes were added (Figure 4), so the conversion does not occur spontaneously. The conversion of deoxyribose to deoxyribonoate occurred with or without the added lactonase,

perhaps because deoxyribonolactone in water can spontaneously hydrolyze to deoxyribonoate. In both reactions, and in the starting sample of deoxyribose, we detected a small amount of a compound with *m/z* = 155.0 that could be the deoxyribonolactone (Figure 4) although the abundance of this ion was too low for tandem mass spectrometry.

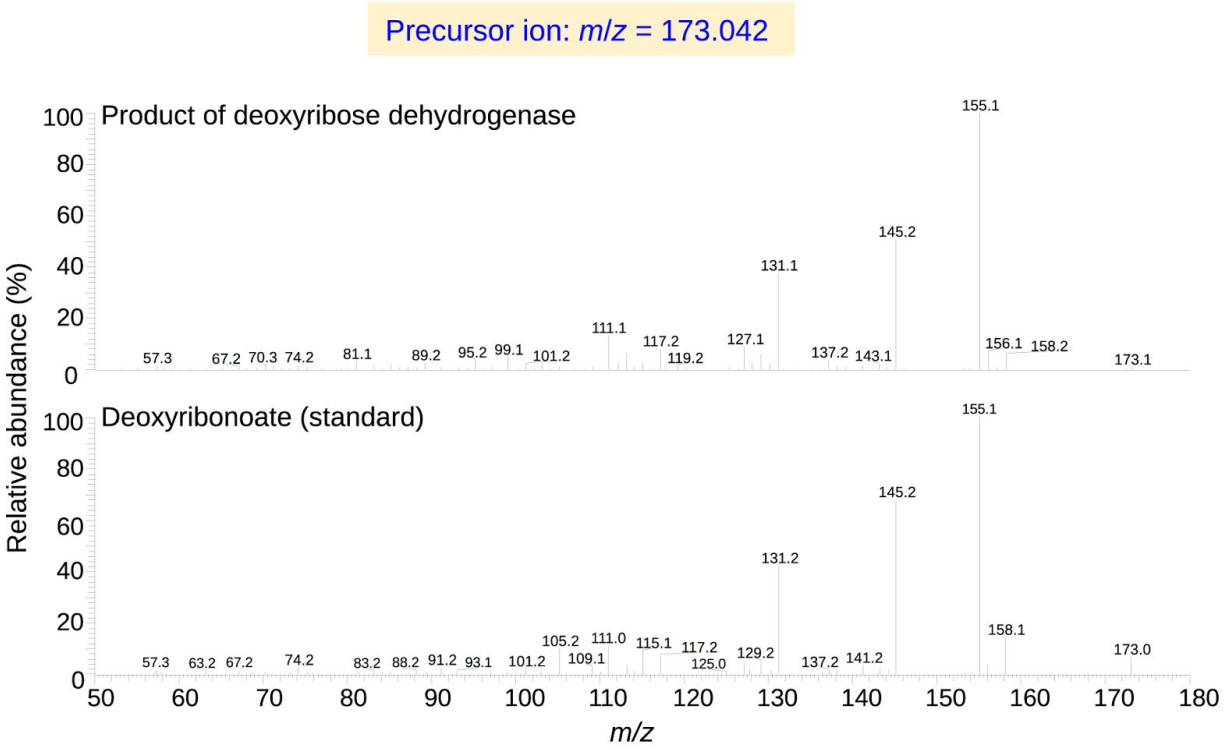

Figure 5: Comparison of the tandem mass spectra of the product of the *iorAB*-like dehydrogenase and deoxyribonoate.

We also incubated 1 mM deoxyribonoate with purified SDR dehydrogenase (PS417_07245) and with 1 mM NAD^+ as the electron acceptor. We observed a product with the molecular weight of 2-deoxy-3-keto-D-ribonoate (a putative sodium adduct with *m/z* = 171.02). The putative ketodeoxyribonoate was not detected in any of the other samples (in particular, it was not observed when deoxyribose was incubated with the *iorAB*-like dehydrogenase). Tandem mass spectrometry showed the loss of CO_2 or H_2O, as expected for a sugar acid (Figure 6). The amount of product was small and difficult to quantify (perhaps 3-4% of the ion intensity of the input deoxyribonoate), but in the absence of the enzyme, the product was below the limit of detection. These data confirm that the SDR dehydrogenase encoded by PS417_07245 is a deoxyribonoate dehydrogenase, but does not prove that the oxidation occurred at the third carbon of deoxyribonoate.

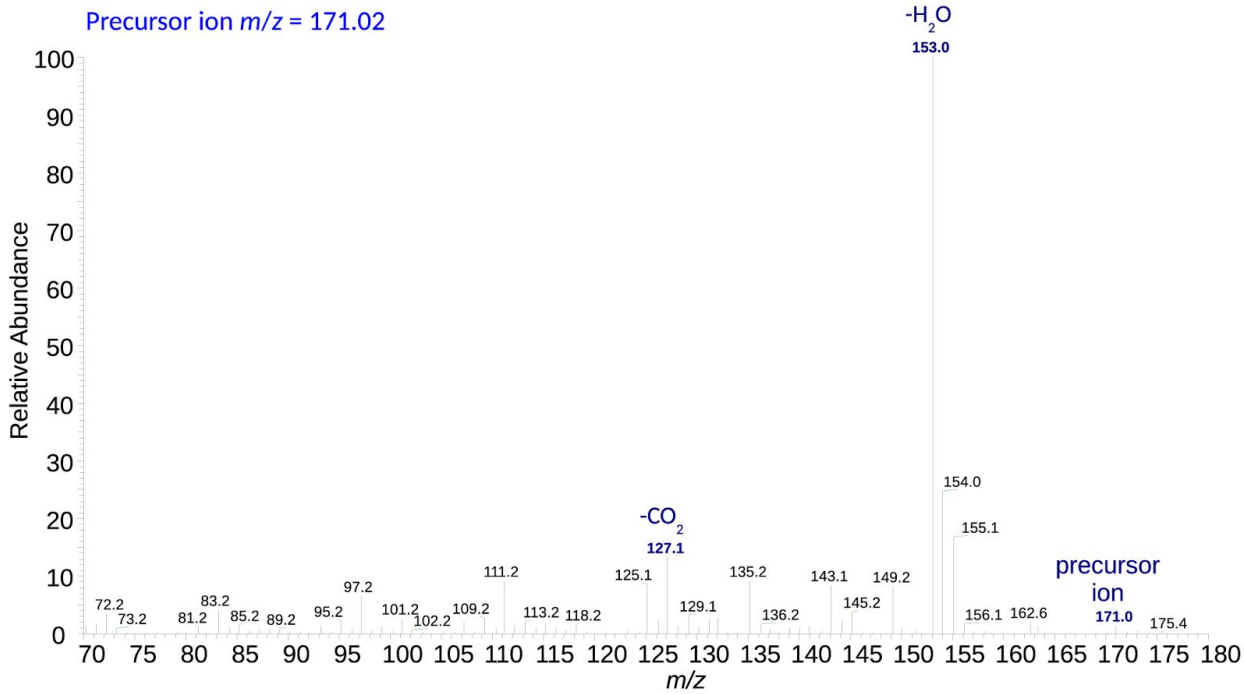

Figure 6: Tandem mass spectrum (MS/MS) of the putative 2-deoxy-3-keto-D-ribonoate from deoxyribonoate incubated with purified SDR dehydrogenase.

Discussion

Although we have direct evidence for the conversion of deoxyribose to a ketodeoxyribonoate, we do not have biochemical evidence for the conversion to D-glycerate. But the role of the putative glycerate kinase (PS417_13970) in growth on deoxyribose and deoxyribonoate is difficult to explain otherwise. Also, there is some experimental data from a homologous cleavage enzyme that is consistent with our proposal. Specifically, *in vitro* assays of the β-keto cleavage enzyme BKACE_178 (UniProt A6X2V8), which is 61% identical to PS417_07250, found that it had some activity with 5-hydroxy-β-ketohexanoate, β-ketopentanoate, β-ketoisocaproate, and β-ketohexanoate as substrates (Bastard et al. 2014). These substrates are similar to 2-deoxy-3-keto-ribonoate (which is 4,5-dihydroxy-β-ketopentanoate).

In summary, we used genetic data to identify an oxidative pathway for the catabolism of deoxyribose by *Pseudomonas simiae* WCS417. The proposed pathway includes four biochemical reactions that have not (as far as we know) been reported before: deoxyribose 1-dehydrogenase, deoxyribonolactonase, deoxyribonoate 3-dehydrogenase, and the transfer of an acetyl group from 2-deoxy-3-keto-D-ribonoate to acetyl-CoA by a β-keto-acid cleavage enzyme. Biochemical assays confirmed the oxidation of deoxyribose to deoxyribonoate by a two-subunit molybdenum-dependent dehydrogenase and the oxidation of deoxyribonoate to a ketodeoxyribonoate by a member of the short chain dehydrogenase/reductase superfamily.

Materials and Methods

Strains and growth media

P. simiae WCS417 was provided by Jeff Dangl (University of North Carolina). The pool of randomly-barcoded transposon mutants of *P. simiae* WCS417 was previously described (Price et al. 2016). 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 this library were obtained from an arrayed collection (Cole et al. 2017). Briefly, the library was spread on LB + kanamycin agar plates, 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).

WCS417 and mutants thereof were recovered from the freezer in LB (Luria-Bertani) medium. For the recovery of the mutant pool or individual mutants, 50 µg/mL kanamycin was added. WCS417 was grown at 30℃ for all experiments. For cloning work, LB was used as the rich medium. All chemicals for media preparations, mutant fitness assays, growth assays and biochemical reactions were purchased from Sigma unless otherwise mentioned.

For assaying fitness or growth on a specific carbon source, we used a defined medium with 10 mM of the carbon source as well as 30 mM PIPES sesquisodium salt, 0.25 g/L ammonium chloride, 0.1 g/L potassium chloride, 0.6 g/L sodium phosphate monobasic monohydrate, Wolfe's vitamins, and Wolfe's minerals (http://fit.genomics.lbl.gov; (Price et al. 2016)).

Genome-wide mutant fitness assays

For fitness assays, the recovered mutant library was inoculated into the medium of choice at OD_{600} = 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 (Wetmore et al. 2015), except that the P1 primer included variable spacing between the Illumina adapter and the sequence that flanks the barcode (2-5 Ns instead of always 5 Ns; see Table 1). Barcodes were sequenced using an Illumina HiSeq 4000.

Table 1: Primers used in this study.

Primer Name	Sequence (5' to 3')
confirm_mariner	GAGCTTAGTACGTACTATC
confirm_04200::Tn	ATGGCGATAGTGGCCCGGTC
confirm_07245::Tn	GGCGGCGTCGGCAATTTCCTG
confirm_07250::Tn	GTCGGTGCAGCGCAGGATG
confirm_10885::Tn	CTGGGTCGAGCATCAGGTGG
Barseq_P1 (5N)	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGTCGACCTGCAGCGTACG
Barseq_P1_2N	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNGTCGACCTGCAGCGTACG
Barseq_P1_3N	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNGTCGACCTGCAGCGTACG
Barseq_P1_4N	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCGACCTGCAGCGTACG
BarSeq_P2_noindex	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT
BarSeq_P1_SangerSeq	GAGATCTACACTCTTTCCCTACAC
amplify_10885_forward	AACCTCATCAACGAACTGCTCG
amplify_10885_reverse	TCAGCTGAAAGTGTGTCGGCTC
amplify_10890_forward	GAATTACGAATCAACCAAAAGGC
amplify_10890_reverse	TCAGACGCCCCCCTGCTTG
amplify_07255_forward	CGTATCGAAGTGCTTGTCGA
amplify_07255_reverse	CTAGCTGGCGAACCGTCGC
amplify_07245_forward	AGTGTCCTACAACGGCTG
amplify_07245_reverse	TTACAGGTACTCGACGTTACC
amplify_pET32_rev	ATGATGATGATGATGGTGCATATGGCC
amplify_pET32_for	TCTTCTGGTCTGGTGCCACGCGGTTC

Analysis of the mutant fitness data

The fitness data was analyzed as described previously (Wetmore et al. 2015) using publically 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 initial sample (after recovery from the freezer). The fitness of a gene is the weighted average of the fitness of strains with insertions in the central 10-90% of the gene. Fitness values were computed for 4,414 of the 5,033 predicted non-essential 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 fitness / sqrt(estimated variance) (Wetmore et al. 2015). A gene was considered to have a specific phenotype in an experiment (Price et al. 2016) if fitness < -1 and *t* < -5; the gene had little phenotype (|fitness| < 1) in 95% of experiments, and the fitness value was at least 0.5 more extreme than the 95th percentile of |fitness| for that gene.

Sequence analysis

Protein sequences were analyzed using the Fitness Browser (http://fit.genomics.lbl.gov) (Price et al. 2016), which incorporates results from PFam (Finn et al. 2014), TIGRFam (Haft et al. 2013), KEGG (Kanehisa et al. 2016), and SEED/RAST (Overbeek et al. 2014).

To identify a motif for the putative binding sites of the transcription factor PS417_07240, we first used the MicrobesOnline tree-browser (Dehal et al. 2010) to identify orthologs with conserved synteny. We identified similar proteins from *P. fluorescens* SBW25, *P. fluorescens* Pf-5, *P. syringae* B728a, and *B. graminis* C4D1M (44-99% amino acid identity) that were also upstream of an SDR dehydrogenase and a β-keto acid cleavage enzyme. We ran MEME (Bailey and Elkan 1994) 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 (Baran et al. 2013). 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 (Gibson 2011). 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 (Clonetech TALON kit) and quantified using a Nanodrop spectrophotometer (Thermofisher). 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 kDa and 17 kDa, respectively).

Enzymatic assays

Enzymatic assays were performed with purified enzymes (2-5 μM) 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 μL of 100 mM ammonium

bicarbonate (NH_4HCO_3) buffer with three independent replicates, run for one hour at room temperature, and then diluted 1000-fold in ammonium bicarbonate buffer and frozen at -20°C.

Liquid chromatography and mass spectrometry

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 nanoelectrospray ionization (nanoESI) source (Thermo Fisher Scientific, Waltham, MA). The LC was equipped with a C18 analytical column (length: 150 mm; inner diameter: 0.075 mm; particle size: 3 µm; pore size: 100 Å; Thermo Acclaim®) and a 1-µL sample loop. Acetonitrile, formic acid (Optima grade, 99.9%, Fisher), and water purified to a resistivity of 18.2 MΩ·cm (at 25°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 (v/v). 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 1500 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: 2 m/z units, normalized collision energy: 35%, default charge state: 1+, activation Q: 0.25, and activation time: 30 ms. Data acquisition and analysis were performed using Xcalibur software (version 2.0.7, Thermo).

Data availability

The fitness data is available from the Fitness Browser (http://fit.genomics.lbl.gov) or at figshare (https://doi.org/10.6084/m9.figshare.5134840.v1).

Acknowledgements

This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

Funding: 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 & Environmental Research under contract number DE-AC02-05CH11231. The QB3/Chemistry Mass Spectrometry Facility at UC Berkeley receives support from NIH (grant number 1S10OD020062-01).

Bibliography

Bailey, T.L. and Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology* 2, pp. 28–36.

Baran, R., Bowen, B.P., Price, M.N., Arkin, A.P., Deutschbauer, A.M. and Northen, T.R. 2013. Metabolic footprinting of mutant libraries to map metabolite utilization to genotype. *ACS Chemical Biology* 8(1), pp. 189–199.

Bartsch, O., Hagemann, M. and Bauwe, H. 2008. Only plant-type (GLYK) glycerate kinases produce d-glycerate 3-phosphate. *FEBS Letters* 582(20), pp. 3025–3028.

Bastard, K., Smith, A.A.T., Vergne-Vaxelaire, C., Perret, A., Zaparucha, A., De Melo-Minardi, R., Mariage, A., Boutard, M., Debard, A., Lechaplais, C., Pelle, C., Pellouin, V., Perchat, N., Petit, J.-L., Kreimeyer, A., Medigue, C., Weissenbach, J., Artiguenave, F., De Berardinis, V., Vallenet, D. and Salanoubat, M. 2014. Revealing the hidden functional diversity of an enzyme family. *Nature Chemical Biology* 10(1), pp. 42–49.

Cole, B.J., Feltcher, M.E., Waters, R.J., Wetmore, K.M., Mucyn, T.S., Ryan, E.M., Wang, G., Ul-Hasan, S., McDonald, M., Yoshikuni, Y., Malmstrom, R.R., Deutschbauer, A.M., Dangl, J.L. and Visel, A. 2017. Genome-wide identification of bacterial plant colonization genes. *PLoS Biology* 15(9), p. e2002860.

Dehal, P.S., Joachimiak, M.P., Price, M.N., Bates, J.T., Baumohl, J.K., Chivian, D., Friedland, G.D., Huang, K.H., Keller, K., Novichkov, P.S., Dubchak, I.L., Alm, E.J. and Arkin, A.P. 2010. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Research* 38(Database issue), pp. D396-400.

Dobbek, H., Gremer, L., Meyer, O. and Huber, R. 1999. Crystal structure and mechanism of CO dehydrogenase, a molybdo iron-sulfur flavoprotein containing S-selanylcysteine. *Proceedings of the National Academy of Sciences of the United States of America* 96(16), pp. 8884–8889.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J. and Punta, M. 2014. Pfam: the protein families database. *Nucleic Acids Research* 42(Database issue), pp. D222-30.

Gibson, D.G. 2011. Enzymatic assembly of overlapping DNA fragments. *Methods in Enzymology* 498, pp. 349–361.

Haft, D.H., Selengut, J.D., Richter, R.A., Harkins, D., Basu, M.K. and Beck, E. 2013. Tigrfams and genome properties in 2013. *Nucleic Acids Research* 41(Database issue), pp. D387-95.

Imam, S., Noguera, D.R. and Donohue, T.J. 2015. CceR and AkgR regulate central carbon and energy metabolism in alphaproteobacteria. *mBio* 6(1).

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. 2016. KEGG as a

reference resource for gene and protein annotation. *Nucleic Acids Research* 44(D1), pp. D457-62.

Koszalka, G.W. and Krenitsky, T.A. 1979. Nucleosidases from Leishmania donovani. Pyrimidine ribonucleosidase, purine ribonucleosidase, and a novel purine 2'-deoxyribonucleosidase. *The Journal of Biological Chemistry* 254(17), pp. 8185–8193.

Lehmann, M., Tshisuaka, B., Fetzner, S. and 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. *The Journal of Biological Chemistry* 270(24), pp. 14420–14429.

Lemieux, R.U., Anderson, L. and Conner, A.H. 1971. The mutarotation of 2-deoxy- -D-erythro-pentose ("2-deoxy- -D-ribose"). Conformations, kinetics, and equilibria. *Carbohydrate Research* 20(1), pp. 59–72.

Levy-Booth, D.J., Campbell, R.G., Gulden, R.H., Hart, M.M., Powell, J.R., Klironomos, J.N., Pauls, K.P., Swanton, C.J., Trevors, J.T. and Dunfield, K.E. 2007. Cycling of extracellular DNA in the soil environment. *Soil Biology and Biochemistry* 39, pp. 2977–2991.

Novichkov, P.S., Kazakov, A.E., Ravcheev, D.A., Leyn, S.A., Kovaleva, G.Y., Sutormin, R.A., Kazanov, M.D., Riehl, W., Arkin, A.P., Dubchak, I. and Rodionov, D.A. 2013. RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* 14, p. 745.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F. and Stevens, R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research* 42(Database issue), pp. D206-14.

Price, M.N., Wetmore, K.M., Waters, R.J., Callaghan, M., Ray, J., Kuehl, J.V., Melnyk, R.A., Lamson, J.S., Suh, Y., Esquivel, Z., Sadeeshkumar, H., Chakraborty, R., Rubin, B.E., Bristow, J., Blow, M.J., Arkin, A.P. and Deutschbauer, A.M. 2016. Deep Annotation of Protein Function across Diverse Bacteria from Mutant Phenotypes. *bioRxiv*. Available at: http://dx.doi.org/10.1101/072470.

Sonnhammer, E.L., von Heijne, G. and Krogh, A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology* 6, pp. 175–182.

Tourneux, L., Bucurenci, N., Saveanu, C., Kaminski, P.A., Bouzon, M., Pistotnik, E., Namane, A., Marlière, P., Bârzu, O., Li De La Sierra, I., Neuhard, J. and Gilles, A.M. 2000. Genetic and biochemical characterization of Salmonella enterica serovar typhi deoxyribokinase. *Journal of Bacteriology* 182(4), pp. 869–873.

Valentin-Hansen, P., Boëtius, F., Hammer-Jespersen, K. and 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. *European Journal of Biochemistry*

/ *FEBS* 125(3), pp. 561–566.

Wetmore, K.M., Price, M.N., Waters, R.J., Lamson, J.S., He, J., Hoover, C.A., Blow, M.J., Bristow, J., Butland, G., Arkin, A.P. and Deutschbauer, A. 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio* 6(3), pp. e00306-15.

Zelcbuch, L., Razo-Mejia, M., Herz, E., Yahav, S., Antonovsky, N., Kroytoro, H., Milo, R. and 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(3), p. e0122957.