

Essential Genes Predicted in the Genome of *Rubrivivax gelatinosus*

Patrick D. Curtis

Department of Biology, University of Mississippi, University, Mississippi, USA

ABSTRACT

Rubrivivax gelatinosus is a betaproteobacterium with impressive metabolic diversity. It is capable of phototrophy, chemotrophy, two different mechanisms of sugar metabolism, fermentation, and H₂ gas production. To identify core essential genes, *R. gelatinosus* was subjected to saturating transposon mutagenesis and high-throughput sequencing (TnSeq) analysis using nutrient-rich, aerobic conditions. Results revealed that virtually no primary metabolic genes are essential to the organism and that genomic redundancy only explains a portion of the nonessentiality, but some biosynthetic pathways are still essential under nutrient-rich conditions. Different essentialities of different portions of the Pho regulatory pathway suggest that overexpression of the regulon is toxic and hint at a larger connection between phosphate regulation and cellular health. Lastly, various essentialities of different tRNAs hint at a more complex situation than would be expected for such a core process. These results expand upon research regarding cross-organism gene essentiality and further enrich the study of purple nonsulfur bacteria.

IMPORTANCE

Microbial genomic data are increasing at a tremendous rate, but physiological characterization of those data lags far behind. One mechanism of high-throughput physiological characterization is TnSeq, which uses high-volume transposon mutagenesis and high-throughput sequencing to identify all of the essential genes in a given organism's genome. Here TnSeq was used to identify essential genes in the metabolically versatile betaproteobacterium *Rubrivivax gelatinosus*. The results presented here add to the growing TnSeq field and also reveal important aspects of *R. gelatinosus* physiology, which are applicable to researchers working on metabolically flexible organisms.

Rubrivivax gelatinosus (formerly *Rhodocyclus gelatinosus*) is a purple nonsulfur bacterium in the *Betaproteobacteria* clade. Its physiology has principally been studied for its photosynthetic capability (1–8), but the organism is very metabolically versatile; it is capable of growing phototrophically (including photoheterotrophically) in anaerobic light conditions and chemoheterotrophically under aerobic conditions in light or dark (9, 10). It has the necessary components for the Embden-Meyerhof-Parnas pathway and the Entner-Doudoroff pathway as well as for the pentose phosphate pathway (11). It is capable of carbon and nitrogen fixation, fermentation, and H₂ gas production, making it of interest in biofuel production. Interestingly, while 16S rRNA gene sequencing places the organism in the *Betaproteobacteria*, the phylogenetic history of the photosynthesis genes strongly suggests that they were horizontally transferred from purple photosynthetic *Alphaproteobacteria* (12).

To identify a minimal set of essential genes in *R. gelatinosus* IL-144, saturating transposon mutagenesis and high-throughput sequencing (TnSeq) was performed on the organism using aerobic, dark, nutrient-rich conditions. TnSeq is a method of identifying essential genes in a high-throughput fashion by mutagenizing a target strain with a transposon, generating upwards of over one million transformants (13). Mutants are pooled, and high-throughput sequencing is used to map hundreds of thousands of unique insertions; these data are then used to find genes that are essential under mutagenesis conditions by identifying genes that cannot tolerate disruptive insertions. This technique has been used on a number of model bacterial systems (13–17). Here, TnSeq was applied to *R. gelatinosus*, not only to investigate its metabolism but also to reveal other unusual features of the organism.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *R. gelatinosus* was grown using a peptone-yeast extract (PY) medium (peptone, 10 g/liter; yeast extract, 10

g/liter; 1.5% agar for solid medium) at 30°C. Agar plates with 12.5 μg/ml kanamycin were used to select for transposon insertions.

***R. gelatinosus* mutagenesis.** EZ-Tn5 Transposome (Epicentre) was electroporated into *R. gelatinosus* using conditions previously described for the organism *Brevundimonas subvibrioides* (15). Briefly, a 25-ml *R. gelatinosus* culture was grown to an optical density at 600 nm (OD₆₀₀) of ~1.0 and harvested by centrifugation (12,000 × g for 5 min at 4°C). Cells were resuspended in 25 ml ice-cold water and centrifuged. This process was repeated, and cells were finally resuspended in 250 μl water. One microliter of EZ-Tn5 Transposome was added to the cells and mixed by gentle pipetting. Cells were then aliquoted to 5 chilled electroporation cuvettes (0.1 mm) and electroporated (1.5 kV, 25 μF, 400 Ω). Each of the aliquots was resuspended in 1.0 ml PY medium and was grown by shaking at 200 rpm at 30°C for 3 h. The cultures were plated 3 by 350 μl on 150-mm-diameter PY plus kanamycin plates. This process was repeated for a total of 100 transformations. Plates were incubated at 30°C for 2 days and stored at 4°C until colonies were harvested. Cells were harvested from plates with a sterile spreader, combined into a single pool of transformants, and frozen at –80°C.

DNA extraction, library preparation, and high-throughput sequencing. Genomic DNA was extracted and transposon mapping was performed as described previously (15). Briefly, genomic DNA was harvested from transformant pools using the Qiagen Maxiprep kit. Cell pel-

Received 25 April 2016 Accepted 30 May 2016

Accepted manuscript posted online 6 June 2016

Citation Curtis PD. 2016. Essential genes predicted in the genome of *Rubrivivax gelatinosus*. *J Bacteriol* 198:2244–2250. doi:10.1128/JB.00344-16.

Editor: I. B. Zhulin, University of Tennessee

Address correspondence to pdcurtis@olemiss.edu.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/JB.00344-16>.

Copyright © 2016, American Society for Microbiology. All Rights Reserved.

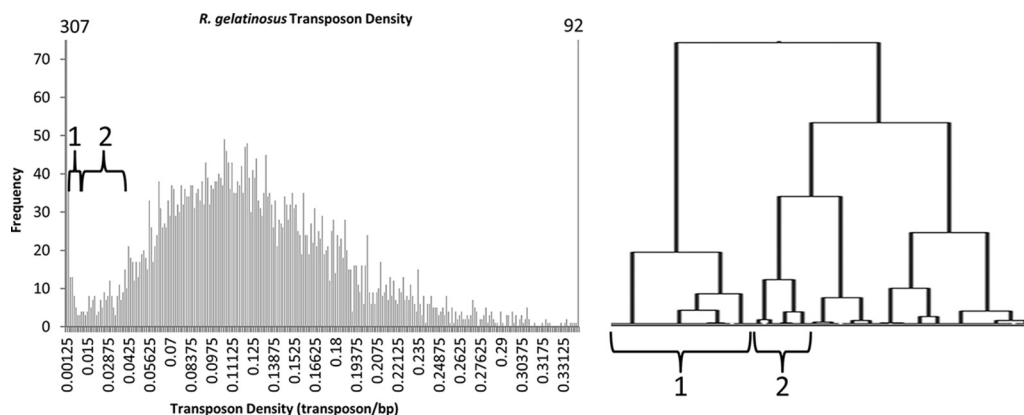


FIG 1 Transposon density distribution. The transposon density of the internal 60% of each ORF was plotted on a histogram (left). The distribution shows a small population of densities close to the y axis (essential genes) and a larger population with higher densities (nonessential genes). Genes in the trough area between peaks likely contain essential genes and genes that have a significant growth defect when disrupted. Numbers in parenthesis represent the height of bars exceeding the y axis (307 for the 0 to 0.0025 range and 92 for the above 0.34 range). A portion of the Ward's cluster analysis covering the essential gene peak, trough, and early nonessential gene peak is depicted (right). Brackets indicate major clusters and how those portions of the clustering analysis correspond to portions of the histogram.

lets were thawed with buffer B1 until the supernatant had an OD_{600} of ~ 2.0 . This supernatant was then used for the Maxiprep kit according to the manufacturer's recommendations.

Library preparation, sequencing, and sequence analysis were performed by the Indiana University Center for Genomics and Bioinformatics. Ten micrograms of DNA was resuspended in a final volume of 100 μ l of Tris-EDTA (TE; pH 8.0) buffer and nebulized at 45 lb/in² for 1 min and 15 s. Nebulized DNA was then purified and assayed on a DNA7500 chip (Agilent Technologies). DNA was subjected to end repair as recommended by Illumina and cleanup using AMPure XP (Beckman Coulter). Ligation reactions involving the Illumina PE adaptor were then performed, and following purification with a 1 \times ratio of AMPure XP, the samples were assayed on the TapeStation using D1K high sensitivity tapes to verify that adaptor ligation was successful. Amplification reactions were performed with the following conditions: 28 μ l DNA template, 1 μ l multiplex primer, 10.3 μ l water, 5 μ l 10 \times AccuStart buffer, 2 μ l 50 mM MgSO₄, 1 μ l 10 mM deoxynucleoside triphosphate (dNTP) mix, 1 μ l Transleft primer (see reference 15), 1 μ l IN2.0 primer (0.5 μ M), and 0.7 μ l AccuStart Hifi *Taq*. The following thermocycler conditions were used: 94°C for 1 min and 30 s; 5 cycles of 94°C for 15 s, 58°C for 45 s, and 68°C for 30 s; 15 cycles of 94°C for 15 s, 65°C for 45 s, and 68°C for 30 s; and 68°C for 10 min. Following amplification, the reactions were purified using a 1 \times ratio of AMPure XP beads and eluted in 28 μ l elution buffer. Samples were then assayed on the TapeStation using D1K high sensitivity tapes.

Sequencing was performed on an Illumina GA2-PEM2X instrument. Sequencing runs were 146-bp paired ends with 7-base index reads. Sequencing runs were performed using v8.3px and v4 Illumina Cluster kits (120 tiles per lane) along with a v5 Illumina SBS run reagent kit. Sequencing out of the transposon was accomplished by using a custom primer (see reference 15). This primer annealed over the junction of the 3' end of the inner transposon and the 3' mosaic sequence; thus, all sequencing reads outward of the 3' end of the transposon. Exposure time for the lasers was 1,225 ms (A = 500 ms, C = 350 ms, G = 200 ms, T = 175 ms). Image analysis was performed using Illumina RTA v1.9.35.

Sequence analysis. Sequence analysis was performed as described previously (15). Briefly, sequence reads that lacked the 10-bp transposon sequence at the start were removed, and the transposon sequences were trimmed from reads of interest prior to mapping. Sequences were mapped to the *R. gelatinosus* genome (GenBank accession number NC_017075.1). The BWA mapping tool was used, allowing an n of 3 (max number of mismatches per read) and an o of 1 (max number of gap openings) (18, 19). From the alignment file, the insertion site for each read was listed,

where the insertion site was defined as the coordinate to which the 5' end of the read mapped. The insertion sites were annotated based on open reading frame (ORF) coordinates.

To identify essential genes, transposons in the 5' and 3' 20% of each ORF were discarded to avoid complications in misannotated start sites or nondisruptive insertions at the 3' terminus (as performed previously [15]), and the transposon density of the internal 60% of each ORF was calculated and plotted on a histogram. To draw tentative cutoffs for essential and nonessential gene designations, transposon densities were subjected to Ward's clustering analysis as was done previously (15). This analysis is an agglomerative hierarchical clustering mechanism where each value is initially in its own cluster, and pairs of clusters are then merged into larger and larger clusters such that the total increase of in-cluster variance is minimal. In this case, each "value" is the transposon density of a given ORF, and similar densities are merged into larger and larger clusters. A graphical representation of clusters containing low transposon densities is shown as a tree in Fig. 1 (right). In this tree, branch length has no bearing on the relationships between clusters, and the horizontal size of the clusters is dependent on the number of densities (i.e., ORFs) in each cluster. This method was a way of creating groups of ORFs with similar transposon densities so that they could be categorized for essentiality; as shown in Fig. 1, one group of clusters contained ORFs in the histogram peak of abnormally low densities (categorized essential), while other groups of clusters contained ORFs with densities in the area between histogram peaks (categorized unresolved). However, all essentiality designations are putative until verified by more direct genetic analysis.

RESULTS AND DISCUSSION

Analysis of insertion bias and identification of essential genes.

An estimated 850,275 mutant colonies were harvested into the transposon pool, and sequencing identified 567,436 unique insertions, leading to an estimated library recovery of 66.7%, which is similar to previous recoveries (15). This resulted in an average of 119 transposons per ORF. A comparison of the GC content of 100-bp windows around each transposon insertion to a random sampling of 100-bp genomic windows shows that there was a slight bias of insertion into more AT-rich regions (see the supplemental material), but in total, most transposons inserted into regions with similar GC contents as would be expected by random chance, suggesting that GC content had little impact on trans-

poson insertion. Similarly, gene orientation was not a significant contributor to transposon insertion. For each ORF, the number of insertions oriented in the positive strand was divided by the number of insertions oriented in the negative strand, and these numbers were \log_2 transformed. Therefore, an ORF with equal numbers of insertions in the positive and negative strands will have a \log_2 strand analysis number of 0. The strand analysis number for each ORF was plotted and shown in the supplemental material. The majority of ORFs cluster around the 0 point, suggesting that ORF direction did not significantly impact insertion direction.

Gene essentiality designation was assigned as previously performed for *Agrobacterium tumefaciens* and *B. subvibrioides* (15). Transposons in the 5' and 3' 20% of each ORF were discarded to avoid complications in misannotated start sites or nondisruptive insertions at the 3' terminus, and the transposon density of the internal 60% of each ORF was calculated and plotted on a histogram (Fig. 1, left). Transposon densities were also subjected to Ward's clustering analysis. As described in the Materials and Methods, this clustering mechanism is an agglomerative mechanism that joins data values (in this case transposon densities) such that the least variance is introduced in clusters. This analysis therefore joined transposon densities into clusters, and smaller clusters into larger cluster groups, based on density similarity. A tree showing cluster groups that contain low transposon densities is shown in Fig. 1 (right). Brackets indicate which cluster group covers which portion of the transposon densities in the histogram (Fig. 1, left). The group of clusters indicated by bracket 1 corresponds to the histogram peak containing abnormally low densities; these 388 genes have been designated essential. The group of clusters indicated by bracket 2 covers genes in the trough between the two peaks, and these 103 genes were categorized as unresolved. The remaining 4,276 genes were categorized as nonessential. It should be emphasized that essential gene designations are putative until verified by more traditional genetic analysis. Lists of genes and accompanying essentiality are found in the supplemental material.

As a percentage of total genomic content, *R. gelatinosus* is on the low end for essentiality. Only 8.1% of its ORFs are essential; more than 10% is typical. This is likely a result of a combination of factors. The first factor is the larger size of the *R. gelatinosus* genome (4,767 total ORFs). Organisms with more genomic content tend to have lower essential gene content by percentage. For example, *Salmonella typhi* has 356 essential ORFs of its total 4,537 (7.8%) (13), *Agrobacterium tumefaciens* has 372 of 5,460 (6.8%) (15), and *Pseudomonas aeruginosa* has 352 of 5,678 (6.2%) (16); conversely, *Caulobacter crescentus* has 480 of 3,876 (12.4%) (14), *Brevundimonas subvibrioides* has 447 of 3,393 (13.2%) (15), and *Streptococcus pyogenes* has 241 of 1,785 (13.5%) (20). Another contributing factor to lower essential gene content is the general lack of metabolic genes. As noted in the original genome publication (11), *R. gelatinosus* has a number of redundant genes, particularly for metabolic pathways. This is reflected in the low number of essential metabolic genes, as exemplified by those of the tricarboxylic acid (TCA) cycle (see below). Similarly, the metabolic flexibility of the organism can contribute to lower essential gene content, as a blockage in one type of metabolic pathway may be bypassed by adopting a different type of metabolism. However, *Rhodospseudomonas palustris*, a purple photosynthetic and metabolically versatile alphaproteobacterium, has a similar genome

size to *R. gelatinosus* with 4,838 protein-encoding ORFs, and TnSeq analysis, using a rich medium that was also supplemented with light (thereby providing conditions for more types of metabolism than were provided for *R. gelatinosus*), found 552 essential genes, which covers 11.2% of the genome (21). Therefore, genome size and metabolic flexibility cannot be the sole determinants of essential gene content.

Analysis and comparison of essential genes by COG category. Table 1 summarizes the *R. gelatinosus* essential gene content by clusters of orthologous group (COG) category. Nearly all essential genes had a predicted function, and most of them belonged to systems one would expect to be essential. By far, the COG category with the most essential genes is J (translation, ribosomal structure, and biogenesis) with 90 essential genes. This large number is mostly composed of the many ribosomal proteins and amino acyl-tRNA synthetases. While the vast majority of ribosomal proteins were categorized as essential, the rRNAs were not. As noted in the original genome announcement, *R. gelatinosus* has a large number of redundant genes, including those for the rRNAs (11). The lack of essentiality suggests that at least two of the rRNA operons can be functionally expressed. Other COG categories with large numbers of essential genes include H (coenzyme transport and metabolism) (36 genes), I (lipid transport and metabolism) (20 genes), and M (cell wall/membrane/envelope biogenesis) (35 genes). It is unsurprising that there are many essential genes in the last category, as it contains genes for peptidoglycan synthesis as well as assembly, which tend to be essential in other organisms (14, 15, 21). Essential genes in category I not only include genes for fatty acid and phospholipid biosynthesis but several genes involved in isoprenoid biosynthesis as well. It is likely that this pathway is needed for the synthesis of undecaprenol, which is an essential component of extracellular sugar polymerization pathways, including peptidoglycan and lipopolysaccharide synthesis. Of particular interest is the large number of genes found in category H (coenzyme transport and metabolism); the vast majority of genes in this category do not encode cofactor transport enzymes but instead encode cofactor biosynthetic enzymes. Given that the organism was grown in a relatively nutrient-rich/nutrient-diverse peptone-yeast extract medium, these results suggest that the medium does not have enough cofactors or the appropriate cofactors to support the growth of the organism. Instead, it appears that *R. gelatinosus* must synthesize a significant portion of its cofactor complement itself.

Included in Table 1 is a COG category breakdown of the essential genes of *R. palustris* (21), the most comparable organism to *R. gelatinosus* on which TnSeq analysis has been performed. Similar to *R. gelatinosus*, the COG category with the most genes in *R. palustris* was J (translation, ribosomal structure, and biogenesis) with 91 essential genes. Also similar were larger numbers of essential genes in H (coenzyme transport and metabolism), I (lipid transport and metabolism), and M (cell wall/membrane/envelope biogenesis), the same categories that are populous in *R. gelatinosus* essential genes. Clearly, these processes are critical to cellular growth, so their essentiality is to be expected. Some categories have larger representation in *R. palustris* essential genes than in *R. gelatinosus* essential genes. There are more genes in categories C (energy production and conversion), E (amino acid transport and metabolism), and F (nucleotide transport and metabolism) in *R. palustris*. The larger number in C is likely due to more TCA cycle enzyme representation (see below) as well as more electron trans-

TABLE 1 Summary of *R. gelatinosus* and *R. palustris* essential genes by COG category

COG category	Function(s)	No. of <i>R. gelatinosus</i> genes	No. of <i>R. palustris</i> genes ^a
A	RNA processing and modification	0	0
B	Chromatin structure and dynamics	0	0
C	Energy production and conversion	28	43
D	Cell cycle control, cell division, and chromosome partitioning	9	10
E	Amino acid transport and metabolism	12	53
F	Nucleotide transport and metabolism	11	29
G	Carbohydrate transport and metabolism	9	12
H	Coenzyme transport and metabolism	36	49
I	Lipid transport and metabolism	20	22
J	Translation, ribosomal structure, and biogenesis	90	91
K	Transcription	15	23
L	Replication, recombination, and repair	20	22
M	Cell wall/membrane/envelope biogenesis	35	48
N	Cell motility	0	0
O	Posttranslational modification, protein turnover, and chaperones	14	21
P	Inorganic ion transport and metabolism	4	7
Q	Secondary metabolites biosynthesis, transport, and catabolism	1	3
R	General function prediction only	16	25
S	Function unknown	10	25
T	Signal transduction mechanism	1	13
U	Intracellular trafficking, secretion, and vesicular transport	11	16
V	Defense mechanisms	3	6
Z	Cytoskeleton	0	0
Not in COG		4	34

^a Based on reference 21.

port proteins. In the latter case, it may be because light was used to supplement the growth of *R. palustris*, as purely heterotrophic conditions proved insufficient to allow efficient growth in the presence of kanamycin. As noted in the *R. palustris* paper (21), there are more amino acid biosynthesis (as well as nucleotide biosynthesis) enzymes categorized essential than would be expected given a Casamino Acids-yeast extract growth medium; certainly, there are significantly more of these enzymes categorized essential in *R. palustris* than in *R. gelatinosus*. As the authors suggested, these differences are likely due to the particular needs of the organisms on their different growth media. Supplementation of the media with amino acids and nucleotides (and cofactors for that matter) would likely reduce the number of essential biosynthesis enzymes.

Also noted in the *R. palustris* paper is the surprisingly larger number of essential genes with unknown function. *R. gelatinosus* has a total of 30 essential genes between R (general function prediction only), S (function unknown), and genes with no COG assignment. In these same three categories, *R. palustris* has 84 essential genes. It is possible that *R. palustris* has an abnormally high unknown essential gene content, but it is also just as feasible that *R. gelatinosus* has an abnormally low unknown essential gene content. Not enough of these kinds of studies have been performed to predict an average unknown essential gene content. In those three categories, *B. subtilis* has 35 essential genes and *A. tumefaciens* has 39 essential genes (15). *C. crescentus* TnSeq found 49 essential genes of unknown function (14), and *P. aeruginosa* has 72 essential genes listed as hypothetical alone (16). While this unknown essential gene content may represent biology yet to be discovered, the number of genes in these categories is likely also dependent on the age or effectiveness of genome annotation. In-depth manual curation and characterization of these genes are needed.

Lastly, *R. palustris* has 13 essential genes in category T (signal transduction mechanisms) while *R. gelatinosus* only has one. This result likely reflects the alphaproteobacterial heritage of *R. palustris*, where many organisms use two-component signal transduction proteins in complex networks to regulate the cell cycle as exemplified by the developmental cell cycle of *C. crescentus* (reviewed in reference 22). Indeed, many of the essential signal transduction genes in *R. palustris* encode two-component proteins homologous to known *C. crescentus* developmental regulators (21). The lone essential signal transduction gene in *R. gelatinosus* is a putative homolog of *dksA*. DksA has been found to participate in the regulation of rRNA operons, tRNA genes, and some amino acid biosynthesis operons, serving to reduce their expression under nutrient-limited or stationary-phase conditions (23, 24). DksA is nonessential in *Escherichia coli* (25); the essential nature of this putative *dksA* homolog in *R. gelatinosus* is unknown.

Essentiality of primary metabolic pathways. Contributing to the low essential gene content of *R. gelatinosus* is the virtual complete absence of primary metabolism genes. There were very few essential genes from any of the glycolytic pathways. This result is not unusual; few glycolytic genes are essential in other organisms (13–17). TnSeq is usually performed on rich media containing combinations of tryptone, peptone, and yeast extract; therefore, the principal carbon and energy sources in these media are amino acids and small peptides. These molecules are entered into the primary metabolism typically through the TCA cycle, making glycolytic pathways nonessential. Theoretically, glycolytic pathways may be necessary for anabolic reactions to generate needed structural sugars (and, indeed, two enzymes involved in glucosamine metabolism are essential in *R. gelatinosus*), but the general lack of glycolytic genes suggests that residual sugars in the growth medium are sufficient for growth. A gluconolactonase (RGE_02330) and a phosphogluconate dehydratase (RGE_34470) likely partic-

TABLE 2 TCA cycle enzyme essentiality

TCA cycle enzyme	Essentiality ^a in:					
	<i>R. gelatinosus</i>	<i>R. palustris</i> ^b	<i>C. crescentus</i> ^c	<i>B. subvibrioides</i> ^d	<i>A. tumefaciens</i> ^d	<i>P. aeruginosa</i> ^e
Citrate synthase	U	E	N	N	E	E
Aconitase	E	E	E	E	N	N
Isocitrate dehydrogenase	N	E	E	N	E	N
α-Ketoglutarate dehydrogenase	E	E	E	E	E	E
Succinyl-CoA synthetase	N	N	E	E	E	N
Succinate dehydrogenase	U	E	E	E	E	E
Fumarase	E	N	E	E	E	E
Malate dehydrogenase	N	E	E	E	N	N
Total no. of essential enzymes	3	6	7	6	6	4

^a Essentiality is reported as essential (E), nonessential (N), or unresolved/high fitness (U).

^b Designations are based on reference 21.

^c Designations are based on reference 14.

^d Designations are based on reference 15.

^e Designations are based on reference 16.

ipating in the Entner-Doudoroff pathway were found to be essential. It is not clear why these enzymes are essential while the rest of the enzymes in the pathway are not.

While the TCA cycle is probably the primary input for carbon and energy in many TnSeq studies, including this one, *R. gelatinosus* has a remarkably low number of essential genes encoding TCA cycle enzymes. A comparison of results from this study to other TnSeq studies of primarily environmental (i.e., principally nonpathogenic) organisms is shown in Table 2. In organisms other than *R. gelatinosus*, genes encoding 6 or 7 of the 8 TCA cycle processes are essential, except in *P. aeruginosa*. In the majority of cases where a process is not essential, redundant enzymes encoded in the genome are the most likely explanation. At least two, if not more, genes can be found encoding a *R. palustris* fumarase, *C. crescentus* citrate synthase, *B. subvibrioides* citrate synthase, *A. tumefaciens* aconitase and malate dehydrogenase, and *P. aeruginosa* aconitase, isocitrate dehydrogenase, and malate dehydrogenase. There is no redundancy to explain the nonessentiality of succinyl coenzyme A (succinyl-CoA) synthetase in *R. palustris*, but the authors proposed that succinate may be generated by the glyoxylate bypass and excess succinyl-CoA may be shuttled into heme synthesis (21). Similarly, there is no redundancy to explain the succinyl-CoA synthetase nonessentiality in *P. aeruginosa*, but that organism does possess a succinyl-CoA:acetate CoA transferase. In this case, instead of directly hydrolyzing succinyl-CoA to succinate and CoA, the CoA moiety is transferred to acetate to make acetyl-CoA (which can be fed into the TCA cycle or other metabolic pathways, such as fatty acid synthesis), thereby freeing up succinate to complete the cycle. Strangely, there is no redundancy for the nonessential *B. subvibrioides* isocitrate dehydrogenase. This is a reaction that is bypassed by the glyoxylate shunt, and the organism does possess the enzymes for that process. The glyoxylate shunt may figure into the survival of *B. subvibrioides* without isocitrate dehydrogenase activity.

R. gelatinosus has the fewest essential TCA cycle processes of those listed in Table 2 with only 3 essential processes. There is redundancy for some of these processes but not all of them. *R. gelatinosus* has two apparent citrate synthases: RGE_12310, which is nonessential, and RGE_29810, which is unresolved. While there are two aconitases, RGE_12220 and RGE_38310, the former is

essential while the latter is not, which suggests that RGE_38310 is either nonfunctional or not expressed. Similar to *B. subvibrioides*, there is only one isocitrate dehydrogenase (RGE_31820), and it is nonessential. There are single copies of the E1 (RGE_27750) and E2 (RGE_27740) components of the α-ketoglutarate dehydrogenase complex, and both are essential. There are two copies for both the α and β subunits of the succinyl-CoA synthetase (RGE_12330 and RGE_12340, RGE_46560 and RGE_46570), none of which are essential. There are five components to the succinate dehydrogenase: subunits A, B, C, D, and a membrane anchor protein. In *R. gelatinosus* A (RGE_12280), B (RGE_12290), C (RGE_12260), and the membrane anchor protein (RGE_12270) are all in the same operon, while D (RGE_21250) is in a different operon with another copy of A (RGE_21280). While all of the components in the first operon were categorized as unresolved, the components in the second operon, including the second copy of A, were found to be nonessential. *R. gelatinosus* has two apparent fumarases, but RGE_18620 is essential while RGE_28550 is not. Lastly, there is only one apparent malate dehydrogenase (RGE_12240), but it is nonessential. In total, genetic redundancy can explain some of the low essential gene content for the TCA cycle, but it does not explain the cases where there are multiple genes and only one is essential, and more importantly, there are three cases where there are only single copies of genes encoding TCA cycle activities and they are not essential: isocitrate dehydrogenase, succinate dehydrogenase, and malate dehydrogenase. This is in stark contrast to other organisms where TCA cycle processes are either essential or the loss of those activities can be compensated for by redundant enzymes or bypass pathways. It is not clear how *R. gelatinosus* achieves this flexibility in its metabolism, but it may be that the organism can live with an incomplete or altered TCA cycle by adopting a strictly fermentative metabolism, where the TCA cycle is typically used more in a biosynthetic capacity or at least not an integral part of primary metabolism. However, this is speculation as the fermentation capability of the organism has not been extensively studied.

Most components of the electron transport chain were also missing from the *R. gelatinosus* essential gene set. Only the ATP synthase and, interestingly, the large NADH quinone oxidoreductase complex were essential. The NADH quinone oxidoreductase

complex was also essential in *R. palustris*, though a second homolog system was not essential (21); the authors proposed that the second system may function exclusively for reverse electron transport. *R. gelatinosus* does not appear to have a second NADH quinone oxidoreductase system.

Essentiality of cell cycle genes. Most cell division genes were categorized as essential, as well as the chromosome partitioning protein ParB, but the closest *parA* homolog was not essential (RGE_00580). However, a second potential *parA* homolog also exists in the chromosome (RGE_38650, nonessential); it is possible that they are functionally redundant. The septum site selection *min* system also had various essentiality; *minC* and *minD* were nonessential, while *minE* was categorized as unresolved, suggesting that it is either essential or there is a fitness cost to its disruption. Most cell wall synthesis genes were essential, including a PBP2 homolog (RGE_46640). Neither PBP1A homolog (RGE_08160 or RGE_12210) was essential, suggesting redundancy. The cell shape-determining genes *rodA*, *mreB*, *mreC*, and *mreD* were all categorized as essential, as they have been in other organisms (26–28).

The Pho regulon. An unusual discovery in the essential gene set was two genes involved in phosphate uptake: *phoU* and *pstB*. PhoU functions with the PstSCAB phosphate uptake system to control the Pho regulon. In the presence of excess phosphate, PhoU inhibits the activity of the histidine kinase PhoR, which prevents phosphorylation and thus activation of the DNA-binding response regulator PhoB, keeping the phosphate starvation regulon repressed (29). Not only were *phoU* and *pstB* essential, but the *pstS*, *pstC*, and *pstA* genes were all unresolved, suggesting their disruption is either lethal or physiologically costly. However, the *phoR* (histidine kinase) and *phoB* (response regulator) genes were nonessential. In other systems, disruption of *phoU* or any of the *pst* components leads to constitutive activation of the Pho regulon, while disruption of *phoB* or *phoR* leads to constitutive repression (30). The fact that *pstSCAB* and *phoU* are unresolved/essential while *phoR* and *phoB* are nonessential in *R. gelatinosus* suggests that the overactivation of the regulon has a toxic effect on the cell. This is an unexpected result. Overactivation of this pathway is nonfatal in other species and typically leads to accumulation of inorganic polyphosphate (31, 32). However, *phoU* was categorized as essential in *C. crescentus*, *B. subvibrioides*, and *R. palustris*, suggesting this phenotype may be common (14, 15, 21). There is a growing link between Pho regulation and stress response, potentially by the action of ppGpp (31, 33). The results of this and other TnSeq studies support the idea that the Pho regulon has larger impacts on cellular physiology than just regulating phosphate levels.

Unusual essentiality is found in the *R. gelatinosus* tRNA complement. The published genome annotation found 50 tRNA genes in the *R. gelatinosus* genome. However, several of these tRNAs are functionally redundant, and in actuality, only 43 of 61 codons are accounted for by identified tRNAs. As noted in the original genome paper, three rRNA loci are found in the genome containing the 23S, 16S, and 5S rRNAs (11). Given that each locus also has the same Ala and Ile tRNAs, it is likely that the extra loci are the result of duplication; in fact, two of the loci are immediately adjacent to each other. The tRNAs in all three loci are nonessential, suggesting at least two sets are functional if not all three. Redundancy also explains the nonessentiality of four other tRNAs. Two Glu and two Gly tRNAs encode redundant codons,

respectively; none of them are essential. However, redundancy does not explain all of the nonessential tRNAs. One Gly, one Leu, and one Val tRNA are nonessential, and no redundant tRNAs are apparent for these codons. General third codon position wobble rules can compensate for the loss of the Gly tRNA but not for the Leu or Val tRNAs. On the other hand, a lack of redundancy does not explain all of the essentiality results. Three Met tRNAs are found in the genome (with the same anticodon), and all three are essential. Initiator fMet-tRNAs are characterized by (i) a non-Watson-Crick base pair at the top of the acceptor stem (C-A in *E. coli*), (ii) a purine-11–pyrimidine-24 base pair instead of a pyrimidine-11–purine-24 base pair, and (iii) a GGG-CCC base pair set in the anticodon stem (34–36). Of the *R. gelatinosus* Met tRNAs, RGE_t0060 satisfies all three conditions, while RGE_t0080 and RGE_t0450 satisfy none of them, indicating that there is one initiator Met tRNA and two elongation Met tRNAs. The lone initiator tRNA being essential is expected, but it is not clear why both redundant elongation Met tRNAs are essential. RGE_t0080 is in the middle of an apparent four- or five-gene operon, but no other genes in the operon are essential, and RGE_t0450 is at the beginning of an apparent three-gene operon where again no other genes are essential. Therefore, it is unlikely that polar effects explain the essentiality. Clearly, the tRNA dynamics are more complex in this organism than would be expected. Lastly, *R. gelatinosus* contains a selenocysteine tRNA; however, this tRNA is nonessential, suggesting that selenocysteine is not a necessary participant of *R. gelatinosus* physiology.

The essential genome data presented here gives a glimpse into this metabolically versatile organism and reveals physiological and genomic complexity. With further TnSeq studies of versatile organisms, it will be possible to identify the core processes necessary for these complicated organisms and perhaps allow researchers to manipulate them for biotechnological purposes.

ACKNOWLEDGMENTS

I thank Doug Rusch, Ram Podicheti, and James Ford of the Indiana University Center for Genomics and Bioinformatics for the assistance in library construction, sequencing, and data analysis.

Early stages of this work were conducted in Yves Brun's laboratory at Indiana University and were funded by NIH grant GM51986.

REFERENCES

1. Agalidis H, Reiss-Husson F. 1991. Resolution of *Rhodocyclus gelatinosus* photoreceptor unit components by temperature-induced phase separation in the presence of decyltetraoxyethylene. *Biochem Biophys Res Commun* 177:1107–1112. [http://dx.doi.org/10.1016/0006-291X\(91\)90653-O](http://dx.doi.org/10.1016/0006-291X(91)90653-O).
2. Agalidis I, Rivas E, Reiss-Husson F. 1990. Reaction center light harvesting B875 complexes from *Rhodocyclus gelatinosus*: characterization and identification of quinones. *Photosynth Res* 23:249–255. <http://dx.doi.org/10.1007/BF00034855>.
3. Banci L, Bertini I, Cambria MT, Capozzi F, Dikuy A. 1994. 1H one-dimensional and two-dimensional NMR studies of the ferricytochrome c 551 from *Rhodocyclus gelatinosus*. *Eur J Biochem* 219:663–669. <http://dx.doi.org/10.1111/j.1432-1033.1994.tb19982.x>.
4. Bertini I, Gori G, Luchinat C, Vila AJ. 1993. One- and two-dimensional NMR characterization of oxidized and reduced cytochrome c' from *Rhodocyclus gelatinosus*. *Biochemistry* 32:776–783. <http://dx.doi.org/10.1021/bi00054a006>.
5. Brunisholz RA, Suter F, Zuber H. 1994. Structural and spectral characterisation of the antenna complexes of *Rhodocyclus gelatinosus*. Indications of a hairpin-like-arranged antenna apoprotein with an unusually high alanine content. *Eur J Biochem* 222:667–675.
6. Nagashima KV, Matsuura K, Shimada K, Vermeglio A. 2002. High-potential iron-sulfur protein (HiPIP) is the major electron donor to the

- reaction center complex in photosynthetically growing cells of the purple bacterium *Rubrivivax gelatinosus*. *Biochemistry* 41:14028–14032. <http://dx.doi.org/10.1021/bi026511a>.
7. Nagashima KV, Shimada K, Matsuura K. 1996. Shortcut of the photosynthetic electron transfer in a mutant lacking the reaction center-bound cytochrome subunit by gene disruption in a purple bacterium, *Rubrivivax gelatinosus*. *FEBS Lett* 385:209–213. [http://dx.doi.org/10.1016/0014-5793\(96\)00382-1](http://dx.doi.org/10.1016/0014-5793(96)00382-1).
 8. Nagashima S, Shimada K, Vermeleglio A, Nagashima KV. 2011. The cytochrome c_8 involved in the nitrite reduction pathway acts also as electron donor to the photosynthetic reaction center in *Rubrivivax gelatinosus*. *Biochim Biophys Acta* 1807:189–196. <http://dx.doi.org/10.1016/j.bbabi.2010.10.020>.
 9. Prasertsan P, Choorit W, Suwanno S. 1993. Isolation, identification and growth conditions of photosynthetic bacteria found in seafood processing wastewater. *World J Microbiol Biotechnol* 9:590–592. <http://dx.doi.org/10.1007/BF00386301>.
 10. Prasertsan P, Choorit W, Suwanno S. 1993. Optimization for growth of *Rhodocyclus gelatinosus* in seafood processing effluents. *World J Microbiol Biotechnol* 9:593–596. <http://dx.doi.org/10.1007/BF00386302>.
 11. Nagashima S, Kamimura A, Shimizu T, Nakamura-Isaki S, Aono E, Sakamoto K, Ichikawa N, Nakazawa H, Sekine M, Yamazaki S, Fujita N, Shimada K, Hanada S, Nagashima KV. 2012. Complete genome sequence of phototrophic betaproteobacterium *Rubrivivax gelatinosus* IL144. *J Bacteriol* 194:3541–3542. <http://dx.doi.org/10.1128/JB.00511-12>.
 12. Nagashima KV, Shimada K, Matsuura K. 1993. Phylogenetic analysis of photosynthetic genes of *Rhodocyclus gelatinosus*: possibility of horizontal gene transfer in purple bacteria. *Photosynth Res* 36:185–191. <http://dx.doi.org/10.1007/BF00033037>.
 13. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK. 2009. Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Res* 19:2308–2316. <http://dx.doi.org/10.1101/gr.097097.109>.
 14. Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Collier JA, Fero MJ, McAdams HH, Shapiro L. 2011. The essential genome of a bacterium. *Mol Syst Biol* 7:528.
 15. Curtis PD, Brun YV. 2014. Identification of essential alphaproteobacterial genes reveals operational variability in conserved developmental and cell cycle systems. *Mol Microbiol* 93:713–735. <http://dx.doi.org/10.1111/mmi.12686>.
 16. Lee SA, Gallagher LA, Thongdee M, Staudinger BJ, Lippman S, Singh PK, Manoil C. 2015. General and condition-specific essential functions of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 112:5189–5194. <http://dx.doi.org/10.1073/pnas.1422186112>.
 17. van Opijnen T, Camilli A. 2012. A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res* 22:2541–2551. <http://dx.doi.org/10.1101/gr.137430.112>.
 18. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595. <http://dx.doi.org/10.1093/bioinformatics/btp698>.
 19. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <http://dx.doi.org/10.1093/bioinformatics/btp324>.
 20. Le Breton Y, Belew AT, Valdes KM, Islam E, Curry P, Tettelin H, Shirliff ME, El-Sayed NM, McIver KS. 2015. Essential genes in the core genome of the human pathogen *Streptococcus pyogenes*. *Sci Rep* 5:9838. <http://dx.doi.org/10.1038/srep09838>.
 21. Pechter KB, Gallagher L, Pyles H, Manoil CS, Harwood CS. 2015. Essential genome of the metabolically versatile alphaproteobacterium *Rhodopseudomonas palustris*. *J Bacteriol* 198:867–876.
 22. Curtis PD, Brun YV. 2010. Getting in the loop: regulation of development in *Caulobacter crescentus*. *Microbiol Mol Biol Rev* 74:13–41. <http://dx.doi.org/10.1128/MMBR.00040-09>.
 23. Mallik P, Paul BJ, Rutherford ST, Gourse RL, Osuna R. 2006. DksA is required for growth phase-dependent regulation, growth rate-dependent control, and stringent control of *fis* expression in *Escherichia coli*. *J Bacteriol* 188:5775–5782. <http://dx.doi.org/10.1128/JB.00276-06>.
 24. Paul BJ, Barker MM, Ross W, Schneider DA, Webb C, Foster JW, Gourse RL. 2004. DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* 118:311–322. <http://dx.doi.org/10.1016/j.cell.2004.07.009>.
 25. Kang PJ, Craig EA. 1990. Identification and characterization of a new *Escherichia coli* gene that is a dosage-dependent suppressor of a *dnaK* deletion mutation. *J Bacteriol* 172:2055–2064.
 26. Figue RM, Divakaruni AV, Gober JW. 2004. MreB, the cell shape-determining bacterial actin homologue, co-ordinates cell wall morphogenesis in *Caulobacter crescentus*. *Mol Microbiol* 51:1321–1332. <http://dx.doi.org/10.1111/j.1365-2958.2003.03936.x>.
 27. Jones LJ, Carballido-Lopez R, Errington J. 2001. Control of cell shape in bacteria: helical, actin-like filaments in *Bacillus subtilis*. *Cell* 104:913–922. [http://dx.doi.org/10.1016/S0092-8674\(01\)00287-2](http://dx.doi.org/10.1016/S0092-8674(01)00287-2).
 28. Kruse T, Bork-Jensen J, Gerdes K. 2005. The morphogenetic MreBCD proteins of *Escherichia coli* form an essential membrane-bound complex. *Mol Microbiol* 55:78–89.
 29. Wanner BL. 1993. Gene regulation by phosphate in enteric bacteria. *J Cell Biochem* 51:47–54. <http://dx.doi.org/10.1002/jcb.240510110>.
 30. Wanner BL. 1996. Phosphorus assimilation and control of the phosphate regulon, p 1357–1381. In Neidhardt FC, Curtiss R, III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umberger HE (ed), *Escherichia coli* and *Salmonella*, 2 ed, vol 1. ASM Press, Washington, DC.
 31. de Almeida LG, Ortiz JH, Schneider RP, Spira B. 2015. *phoU* inactivation in *Pseudomonas aeruginosa* enhances accumulation of ppGpp and polyphosphate. *Appl Environ Microbiol* 81:3006–3015. <http://dx.doi.org/10.1128/AEM.04168-14>.
 32. Morohoshi T, Maruo T, Shirai Y, Kato J, Ikeda T, Takiguchi N, Ohtake H, Kuroda A. 2002. Accumulation of inorganic polyphosphate in *phoU* mutants of *Escherichia coli* and *Synechocystis* sp. strain PCC6803. *Appl Environ Microbiol* 68:4107–4110. <http://dx.doi.org/10.1128/AEM.68.8.4107-4110.2002>.
 33. Santos-Beneit F. 2015. The Pho regulon: a huge regulatory network in bacteria. *Front Microbiol* 6:402.
 34. Lee CP, Seong BL, RajBhandary UL. 1991. Structural and sequence elements important for recognition of *Escherichia coli* formylmethionine tRNA by methionyl-tRNA transformylase are clustered in the acceptor stem. *J Biol Chem* 266:18012–18017.
 35. Seong BL, RajBhandary UL. 1987. *Escherichia coli* formylmethionine tRNA: mutations in GGGCCC sequence conserved in anticodon stem of initiator tRNAs affect initiation of protein synthesis and conformation of anticodon loop. *Proc Natl Acad Sci U S A* 84:334–338. <http://dx.doi.org/10.1073/pnas.84.2.334>.
 36. Sprinzl M, Gauss DH. 1982. Compilation of sequences of tRNA genes. *Nucleic Acids Res* 10:r57–r81.