LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-TR-215218

Characterization of the Cobalamin and Fep Operons in Methylobium petrolphilum PM1

J. Ewing

September 9, 2005

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Characterization of the Cobalamin and Fep Operons in *Methylobium petrolphilum* PM1

Jane Ewing

Department of Energy

Office of Science, CCL

Merced College

Lawrence Livermore National Laboratory

Livermore, California

August 23, 2005

Table of Contents

Abstract

The bacterium *Methylobium petroleophilum* PM1 is economically important due to its ability to degrade methyl tert-butyl ether (MTBE), a fuel additive. Because PM1 is a representative of all MTBE degraders, it is important to understand the transport pathways critical for the organism to survive in its particular environment. In this study, the cobalamin pathway and select iron transport genes will be characterized to help further understand all metabolic pathways in PM1. PM1 contains a total of four cobalamin operons. A single operon is located on the chromosome. Located on the megaplasmid are two tandem repeats of cob operons and a very close representative of the cob operon located on the chromosome. The fep operon, an iron transport mechanism, lies within the multiple copies of the cob operon. The cob operon and the fep operon appear to be unrelated except for a shared need for the TonB-dependent energy transduction complex to assist the operons in moving large molecules across the outer membrane of the cell. A genomic study of the cob and the fep operons with that of phylogenetically related organisms helped to confirm the identity of the cob and fep operons and to represent the pathways. More study of the pathways should be done to find the relationship that positions the two seemingly unrelated cob and fep genes together in what appears to be one operon.

Introduction

Methylobium petroleophilum PM1

Methylobium petroleophilum PM1 is a gram-negative bacterium that exhibits the ability to degrade a fuel additive, methyl tert-butyl ether (MTBE), and displays many metal resistance genes. Along with other metal resistant species, PM1 is a member of the family Comamonadaceae. Economically, the understanding of the metabolic pathways of PM1 including that of MTBE degradation and metal resistance may assist in the restoration of gasoline-impacted aquifers [1]. Analysis at the genome level will assist in the understanding of the metabolic processes such as the transport and absorption of vitamin B12 and metal resistance in PM1.

The Cobalamin Operon in PM1

The production and metabolism of vitamin-B_{12} in animals is of evolutionarily and economical importance. It has been shown that vitamin-B_{12} is an essential early evolutionary cofactor. As an essential nutrient of almost all animals, it is important economically to understand the production, transportation and utilization of this nutrient. Bacteria such as PM1 are the main producers of vitamin-B_{12} [2]. Many organisms can synthesize cobalamin by either anaerobic or aerobic pathways [8]. The presence of multiple, extensive cobalamin operons combined with iron siderophore transport genes in PM1 has led to curiosity as to whether or not the genes are mislabeled or the possibility of an association between the transportation of vitamin-B_{12} and iron siderophores.

Description of the fep operon

A fully functional fep operon consists of fepABC and D. A diagrammatic example of the fep operon can be seen in figure 3. The fep operon inside two of the cob operons of PM1

consists of fepBD and C. The fep operon functions in the synthesis of polypeptides required for the uptake of ferric enterobactin. Located in the outer membrane, fepA, functions in the initial recognition and reception of ferric enterobactin. FepB which is located in the periplasm functions in the internalization of ferric enterobactin into the cell. FepC is located in the cytoplasm and also functions in the production of the polypeptides or ATPase components required for the active transport if iron. FepD functions as an inner membrane permease component. The fep operon is dependent on the presence of tonB and exbB which are involved in other transport mechanisms including the cob operon [3]. TonB and exbB function in the transduction of the energy required for transport.

The Importance of the TonB-dependent energy transduction complex

Because the outer membrane of gram-negative bacteria such as PM1 prevents large molecules such as iron siderophores and cobalamin from entering the organism, the TonB-dependent energy transduction complex is necessary for the transportation of large molecules through the outer membrane of the cell. Both the fep operon and the cob operon are associated with the TonB-dependent energy transduction complex which provides the mechanism for the active transport of iron siderophores and cobalamin across the outer membrane of the cell [6].

The TonB-dependent energy transduction complex consists minimally of TonB, exbB, and exbD. TonB is not functional without the exb operon which consists of exbB and exbD. If either exbB or exbD are mutated in *E. coli*, research has shown that TonB functions have decreased by 90% [5]. The genes exbB and exbD are homologous to tolQ and tolR, and sometimes TolQ and TolR have been seen to function in the place of exbB and exbD. ExbB and exbD function together as signal transducers [5].

Focus of Study

In this report, specific regions of the genomic sequence of PM1 will be analyzed to help understand the vitamin metabolism and metal homeostasis in this organism. To ascertain operon functionality, the presence and location of the necessary proteins for the operons to be functional will be analyzed. To characterize the cob and fep operons, a genomic study of PM1 and phylogenetically related organisms will be used. Also, the cobalamin biosynthetic pathway will be analyzed in PM1. For confirmation of the identity of the fep operons in PM1, a comparison of the genes to one another and to the fep genes in other organisms such as *E. coli* will be used.

Materials and Methods

BLAST Searches

All blast searches were performed by protein-to-protein sequence alignments using the Basic Local Alignment Search Tool (BLAST) located on the internet at the National Center for Biotechnology Information (NCBI). Only the top hits were taken into consideration in this study. Gene functions were also established using protein-to-protein sequence alignments. The Integrated Microbial Genome database (IMG) was used to provide the functions and gene numbers of all genes used in this study. The information provided about the selected organisms in this study at the Integrated Microbial Genome database (IMG), was then supplemented with the functions of the top hits and the closest putative family. The genome of PM1 was represented in this study by the Sanger Institute's Artemis (release 7) [10].

Biosynthetic Pathways

Biosynthetic pathways were either created from combined reading and functional assignments or taken from the Pathways Tool Software [9]. The images taken from the Pathways Tool Software were supplemented with data used in this article regarding protein function in order to give a more complete picture of the pathway.

Phylogenetic and Gene Order Comparison

For the phylogenetic and gene order comparison, all operon sequences were found and represented using the Joint Genome Institute (JGI). All figures representing genomes were found using JGI's view phylogenetic neighbors option. JGI contains all of the sequences used in the Integrated Microbial Genome database (IMG) except for the genes of *E. coli* which are represented by genbank accession numbers.

Results

Similarity of Fep Operon Repetitions

When the protein sequences of the two supposed fep operons and the surrounding cobalamin clusters located on the megaplasmid and the chromosome were aligned, the proteins displayed significant similarity to one another. The percent identities were 61% for fepB, 64% for fepD, and 61% for fepC (see fig. 1). All of the genes on the two clusters aligned with greater than 25% identity but most were above 50%.

Comparison of PM1 fep Genes to E. coli

Figure 2 displays the results of the protein-to-protein sequence alignments of PM1 fepCD and B to that of *E. coli*. When the fepB genes were aligned with that of *E. coli*, they displayed no significant similarity. However, when the fepC and the fepD genes were aligned with those of *E. coli*, they did in fact show significant homology.

Search for fepA and the TonB-dependent Energy Transduction Complex

Using fepA genes from *Methylobacillus flagellatus* a protein-to-protein BLAST search was performed along the entire sequence of PM1. Figure 3 displays the data collected when two fepA genes were used from *M. flagellatus* as models for the fepA sequence. Five genes were collected with some overlapping from the search. CirA, btuB, fecA, and a single fepA showed significant similarity to the fepA gene of *M. flagellatus.* All of the genes displaying similarity to fepA function as outer membrane receptors for iron and are located on the chromosome of PM1.

Because exbB or tolQ genes are required for the cob and fep operons to be functional, a broad search for their presence was performed using Artemis. In this search, many tolQ genes were found present in more than one operon. The tolQ genes were ORF 653, 1303, 1729, 2426, 2737, and 3642. All of the tolQ genes were present on the chromosome. ExbD and TonB were

also present only on the chromosome. The exbD genes were ORF 654, 655, 4243, 1304, 1730, 2736, 3643, and 3644. The tonB genes were ORF 1956, 2425, and 3489. The tolQ operons generally consisted of tolQ, exbD, and btuB.

Gene Sequence Comparison of Phylogenetically Related Organisms to PM1

To compare the order of the cobalamin genes of PM1 and four phylogenetically related organisms, *Polaranomas sp.*, *Rhodoferax ferrireducens,* and *E. coli* all of the gene sequences were analyzed. PM1 and *Polaromonas* seem to display very similar gene order as seen in the comparison of figures 8-11 with figure 12. *Polaromonas* has a gene order that is relatively similar to all four clusters. Figure 13 displays the gene numbers of *Polaromonas* as found in the Joint Genome Institute's database of sequenced genomes. It also displays the functions of the proteins encoded by the genes.

The cobalamin sequence found in *R. ferrireducens* is found in figure 14 and appears to display similarity to the cobalamin genes in PM1. As with the megaplasmid cluster resembling the chromosomal cluster in PM1, the cobalamin operon in *R. ferrireducens* also contains a protein functioning in histidinol-phosphate transfer. The sequence also contains three genes related to iron transport which is similar to that of the chromosomal cluster of cobalamin genes found in PM1. Figure 15 displays the functions of the cobalamin genes along with their numbers as found in the JGI database.

Methylobacillus flagellatus displays an extensive grouping of cobalamin genes (figure 16). Although not part of a single, concise operon, the genes are grouped very closely together. There are also some genes in this grouping that have been seen in the other organisms in this study such as a protein functioning as a nitroreductase. There are many genes which complicate this section of the sequence of *M. flagellatus* that are unrelated to the transport of cobalamin or

iron-siderophores. Figure 17 shows the gene numbers as found in JGI, the gene names and/or the protein functions of all the genes under consideration.

In figure 18, one can see the chromosomal cluster of cobalamin genes in *E. coli* chosen for consideration in this study. The functions of the cobalamin genes and the surrounding genes are shown in figure 19. The genes are grouped in a single operon together with erfK, nac, cbl, and yeeO which are genes functioning in the transport of nitrogen.

Phylogenetic Comparison of the Cobalamin Genes of Closely Related Organisms to PM1

Figure 20 shows the phylogenetic relationships of the four organisms chosen for a phylogenetic comparison with PM1. *Polaramonas sp., Rhodoferax ferrireducens,* and PM1 are all in family Comamonadaceae. *Methylobacillus flagellatus* and *E. coli* are in different families and orders than PM1. In figure 21, a phylogenetic comparison based on the protein-to-protein sequence alignments of the cobalamin genes of PM1 to that of the chosen four phylogenetically related organisms. Generally, the closest sequence match found in *Polaromonas sp.* to PM1 was the same protein or a protein with a similar function. The percent identity of the hits ranged from 22.02 to 83.26 percent except in the case of open reading frame (orf) 428 for which there were no hits.

R. ferrireducens displays many hits that share similar functions to that of PM1, but very few direct hits. The percent identity of PM1 against the cobalamin genes of *R. ferrireducens* ranged from 22.45 to 72.49 percent. The results of the alignment of PM1 with *M. flagellatus* displayed some direct hits to PM1. The percent identity ranged from 25.71 to 55.98 percent except for orf428 which displayed no significant similarity to any of the genes. The small cobalamin operon found in *E. coli* displayed homology to PM1 not only in its cobalamin genes but also in the other genes in the operon. The percent identity of the top hits in PM1 to *E. coli* extended from 22.73 to 66.67 percent.

Discussion and Conclusion

The Identity and Functionality of the Fep Operons

Because the fepC and D genes found in PM1 display significant homology with the fep operon in *E. coli*, it can be concluded from this study that the genes are correctly labeled. FepB didn't produce significant homology. However, this may be due to the fact that in previous studies, it was found that the fepB in *E.* coli displayed no significant homologies to the Genbank database [4]. Therefore, the discrepancies may lie with the fepB in *E. coli* and not with that of PM1. As expected, the repeating fep operons on the megaplasmid and the chromosome display significant protein homology to each other that is very close to the similarity of the overall clusters in which they are located. This implies that they are comparable and inserted almost identically into both the megaplasmid and the chromosome clusters.

PM1 has many genes including cirA, btuB, fecA, and fepA which due to their significant homology with fepA may function as the necessary outer membrane receptor for iron siderophores. However, none of the genes are present in operons consisting of fepA or possible equivalent, and fepBD and C. More research should be done to conclude whether or not fepA needs to be located in the same operon with the other fep genes. However, it should be noted that cirA (orf1131) has 54% identity with the fepA of *M. flagellatus,* and is located extremely close to the fep operon on the chromosome. The fact that *M. flagellatus* is only functionally related to PM1 and not very close phylogenetically indicates a very high probability that cirA does perhaps function as fepA and is perhaps incorrectly annotated. However, due to the broadness of the search for fepA it is reasonable to assume that there may be more fepA genes or fepA-like genes.

The presence of fepABD and C together would, most likely, indicate a fully functional operon. However, it is possible due to protein homology that the fep operon is also using the btuB genes present in the cobalamin operons for iron assimilation. In PM1, there are many sets of the TonB-dependent energy transduction complexes. In these sets, exbB is replaced by tolQ according to current annotation. The presence of many tolQ, exbD, btuB, and tonB genes in PM1 suggests that these operons may be working together with the cob and fep operons to assimilate cobalamin and iron into the organism. Although questionable as to whether or not the fep operon and cobalamin operons are functional in PM1, it is reasonable to conclude from the results in this study that all of the genes necessary for functionality are present. However, it is unclear as to whether or not the fep genes need to be grouped together.

The Cobalamin Biosynthetic Pathway in PM1

Figures 5 and 6 illustrate the biosynthetic pathways of PM1. There are two biosynthetic pathways responsible for cobalamin production in PM1: biosynthesis I and biosynthesis II. Depending on the point of oxygen insertion, cobalamin biosynthesis can be either aerobic or anaerobic. Both biosynthesis I and II represent aerobic pathways [8]. Although PM1 appears to synthesize cobalamin by an aerobic pathway, more research should be done to see the exact conditions under which PM1 synthesizes cobalamin.

Similarities Between the Fep and Cob Operons to Each Other and Across Species

It is unclear why the cobalamin and fep operons are located together in the same clusters. The cobalamin and fep operons both require the functionality of the TonB-dependent mechanism in order to transport their large molecules through the gram-negative cell wall. Another similarity between the cobalamin and fep operons is that both operons share cysG, which in PM1, is present within the tandem repeat clusters of cobalamin genes in PM1. CysG functions in

the methylation of uroporphyrinogen III into precorrin-2 which is a necessary step for the synthesis of vitamin-B_{12} in the biosynthesis II pathway (figure 6). CysG is also needed to catalyze the ring oxidation and iron insertion in the synthesis of siroheme which is a necessary precursor for the production of iron siderophores [7]. It should also be noted that PM1 is not the only organism in this study containing iron transport related genes within its cobalamin operons. *Polaromonas sp.*, *R. ferrireducens,* and *M. flagellatus* all have iron transport genes either within or in the near proximity, as in the case of *M. flagellatus,* to their cobalamin operons. Also, although *Polaromonas sp.,* and *R. ferireducens* are in the same family as PM1, *M. capsulatus* is not even in the same order and yet displays similar combinations of iron transport genes and cobalamin clusters.

Gene Order Comparison of PM1 to Phylogenetically Related Organisms

The gene order of the tandem repeat clusters is comparable to the gene order of the cob operon seen in *Polaromonas sp.*. In comparison with the tandem repeats, *Polaromonas sp.* appears to have a very similar gene order in respect to the cobalamin genes. However, there are many iron-transport related genes inserted in the sequence of *Polaromonas* which are not present in the tandem repeats of PM1. The third cluster on the megaplasmid of PM1 appears to resemble the gene order of *Polaromonas sp.* more than the tandem repeats. However, *Polaromonas sp.* is lacking hisC, btuB and gymB which are present in the third cluster. The cluster on the chromosome of PM1 partially resembles that of *Polaromonas sp.* but *Polaromonas sp.* lacks cobQ, cobT, cobS, and gpmB. The most significant similarity from this comparison is that the operons, for the most part, seem to begin with btuB, cobB, and cobU and then seem to diverge from there.

R. ferrireducens also begins with btuB and cobB and then enters into a fairly similar sequence consisting of a periplasmic binding protein, and a cob gene followed by genes related to iron transport. *R. ferrireducens* also contains a cbiB and a possible hisC which are present in third cluster on the megaplasmid. *M. flagellatus* didn't have a simple cobalamin operon, but rather had an extensive grouping of cob genes. Most of the operons which are grouped together seem to begin with btuB which is seen in PM1. However, there appears to be very little similarity in the gene order of *M. flagellatus* to PM1 which implies that there is little functional similarity. The cluster of cob genes in *E. coli* also appears to have little similarity in its gene order to that of PM1.

Because *Polaromonas sp.* and *R. ferrireducens* are in the same phylogenetic family as PM1, they are expected to have better protein-to-protein sequence alignments and more similar gene orders with PM1 than that of *M. flagellatus* and *E. coli.* The alignments between *Polaromonas sp.* and *R. ferrireducens* do confirm this assumption in this study. All three genes in the family Comamonadaceae have relatively similar gene orders and protein sequence similarity. However, *M. flagellatus* and *E. coli* display very different operon sequence order and protein similarity to PM1 which implies that *M. flagellatus, E. coli*, and PM1 operate slightly different cobalamin operons than those of the family Comamonadaceae.

Sequence Alignment of Phylogenetically Related Organisms to PM1

The cobalamin operons of PM1 show significant similarity to near phylogenetic neighbors both in gene order and protein homology. The results of a protein-to-protein alignment of PM1 to *Polaromonas sp.* demonstrate similarity between both the order of the cob genes and the order of fep genes inside of the cobalamin operon. The presence of the fep operon in the cobalamin operons of PM1 and *Polaromonas sp.* may illustrate a need for a greater

understanding of the transport of cobalamin and iron. The relationship between PM1 and *R. ferrireducens* is also very close. There are both cobalamin and iron transport genes present in the studied operon in *R. ferrireducens* which, again, illustrates similar results to that seen in *Polaromonas sp.*. The comparison of PM1 to *M. flagellatus* displayed ambiguous results. The best sequence alignments were generally of the genes surrounding the cobalamin genes which were placed in the study for broadness. Also, the fep genes on the chromosome of PM1 seem to have very good percent identities to cobalamin genes in both *M. flagellatus* and *R. ferrireducens.* For example, fepB and fepD have very good sequence alignment to cobU and cobQ, respectively, in both *R. ferrireducens* and *M. flagellatus.* FepB also seems to resemble cobT in *E. coli.* It should be noted that fepB also produced ambiguous results with the protein-to-protein sequence alignment of fepB in PM1 to fepB in *E. coli* which may or may not be due to the *E. coli* protein used in this study as discussed previously.

Conclusion

To conclude, although there are discrepancies about the identity and functionality of the fep operons, it may be reasonable to conclude from this study that the operon is in fact operational and correctly annotated due to its similarity to the operons found in other organisms in family Comamonadaceae. The fep operons of other organisms besides PM1 can be used to compare the sequences and help confirm that in a functional fep operon, fepA may be located outside of the fep cluster. The annotation of the cobalamin operon also appears to the correct. Further research could be performed to see how the multiple cobalamin operons compliment each other in their functions. Perhaps the shared needs of the cobalamin and fep operons have placed them together in the same operon as seen in the family Comamonadaceae. The gene order of the cobalamin genes in PM1 appears to be unique to this organism with only slight

overlapping in closely relating organisms. However, there is a lot of protein homology indicating similar functions of corresponding genes.

Acknowledgments

The research in this article was performed at the Lawrence Livermore National Laboratory. Thank you to the Department of Energy and the Office of Science for creating and funding the CCL program. Thanks to my mentor, Anu Chakicherla, for sharing her advice, support, and expertise in her field with me. I would also like to thank Daniel Barsky for his willingness to always give support to my project when needed.

References

[1] Finished Genome: *Methylobium petroleophilum. Doe Joint Genome Institute.* [Online]. Available: http://genome.jgi-psf.org/finished_microbes/metpe/metpe.home.html

[2] John R. Roth, Jeffrey G. Lawrence, Marc Rubenfield, Stephen Kieffer-Higgins, and George M. Church. "Characterization of the Cobalamin (Vitamin$_{B12}$) Biosynthesis Genes of *Salmonella typhimurium.*" *Journal of Bacteriology*, vol. 175, pp. 3303-3316, Mar. 1993.

[3] Bradley A. Ozenberger, Mary Shrodt Nahlik, and Mark A. Mcintosh. "Genetic Organization of Multiple *fep* Genes Encoding Ferric Enterobactin Transport Functions in *Escherichia coli."* *Journal of Bacteriology*, vol. 169, pp. 3638-3646, Aug. 1987.

[4] Margaret F. Elkins and Charles F. Earhart, "Nucleotide Sequence and Regulation of the *Escherichia coli* Gene for Ferrienterobactin Transport Protein FepB," Journal of Bacteriology, vol. 171, pp. 5443-5451, Oct. 1989.

[5] Penelope I. Higgs, Paul S. Myers and Kathleen Postle, "Interaction in the TonB-Dependent Energy Transduction Complex: ExbB and ExbD Form Homomultimers," Journal of Bacteriology, vol. 180, pp. 6031-6038, Nov. 1998.

[6] Pierre Germon, Thierry Clavel, Anne Vianney, Raymond Portalier, and Jean Claude Lazzaroni. "Mutational Analysis of the Escherichia coli K-12 TolA N-Terminal Region and Characterization of Its TolQ-Interacting Domain by Genetic Suppresion," Journal of Bacteriology, vol. 180, pp. 6433-6439, Sep. 1998.

[7] John R. Roth, Jeffrey G. Lawrence, Marc Rubenfield, Stephen Kieffer-Higgens, and George M. Church. "Characterization of the Cobalamin (Vitamin B_{12}) Biosynthetic Genes of *Salmonella typhimurium,*" Journal of Bacteriology, vol. 175, pp. 3303-3316, Mar. 1993.

[8] Martin J. Warren, Evelyne Raux, Heidi L. Schubert, and Jorge C. Escalante-Semerena. (2002, June). The Biosynthesis of Adenosylcobalamin (Vitamin B12). *Natural Product Reports*. [Online]. pp. 390-412. Available: http://www.rsc.org/delivery/_ArticleLinking/DisplayArticleForFree.cfm?doi=b108967f.

[9] Peter D. Karp, Suzanne Paley, and Pedro Romero, "The Pathway Tools Software," Bioinformatics, vol. 18, pp. s1-s8, Mar. 2002.

[10] K. Rutherford, J. Parkhill, J. Crook, T. Horsnell, P. Rice, M-A. Rajandream and B. Barrell, "Artemis: sequence visualization and annotation." Bioinformatics, vol. 16 (10), pp. 944-945, May 2000.

Figures

Protein Sequence Alignment of FepC, FepD, FepB and Surrounding Clusters in PM1

PM1 Gene on megaplasmid		PM1 Gene on Chromosome		% Identity
447	BtuB	1144	BtuB	25.24
448	CobB	no comparable gene		
449	CobU	1143	CobU	55.43
450	**FepB**	**1142**	**FepB**	**61.4**
451	**FepD**	**1141**	**FepD**	**64.52**
452	**FepC**	**1140**	**FepC**	**61.45**
453	BtuR	no comparable gene		
454	CbiB	no comparable gene		
616	HisC	no comparable gene		
455	CobQ	1139	CobQ	58.19
456	CobT	1138	CobT	61.38
617	CobS	4241	CobS	42.48
457	GpmB	4240	GpmB	58.62

Figure 1. Results of the protein-to-protein sequence alignments of the fep genes located on the megaplasmid and the chromosome that appear to repeat. As expected, the fep genes display relatively the same similarity as the corresponding genes in the cluster. This table illustrates very well how similar the megaplasmid and chromosome operons are. On the megaplasmid, are extra genes which are lacking on the chromosome.

Protein Sequence Alignment of the Fep Operons in PM1 Against that of *E. coli*

Gene in PM1 Against Same Gene in *E. coli*	Percent Identity	Percent Positives	Gaps (percent)	Gene Index for *E. coli*
FepB 450 against *E. coli*	No Signifigant Similarity	No Signifigant Similarity	No Signifigant Similarity	581195
FebB 1142 against *E. coli*	No Signifigant Similarity	No Signifigant Similarity	No Signifigant Similarity	581195
FepC 450 against *E. coli*	33%	47%	4%	41432
FepC 1140 against *E. coli*	39%	52%	1%	41432
FepD 451 against *E. coli*	47%	55%	1%	41430
FepD 1141 against *E. coli*	39%	52%	1%	41430

Figure 2. Results of a protein to protein sequence alignment of the fep genes in PM1 to that of *E. coli.* The percent identity represents the similarity of the entire protein sequences. The percent positive displays how many hits were exact, and the gap percentage represents any discrepancies in the sequences. The gene index number can be used to find the protein on NCBI.

Matches to fepA in *M. flagellatus* Found in PM1

***M. flagellatus* fepA gene 685**

Gene in PM1	%id	gene name	Function
			outer membrane receptor protein for mostly Fe transport
1131	54%	cirA	~ TonB dependent receptor (fepA?)
			outer membrane receptor protein for mostly Fe transport
1409	40%	cirA	~ TonB dependent receptor (fepA?)
			***M. flagellatus* fepA gene 1152**
Gene in PM1	**%id**	**gene name**	**Function**
			outer membrane receptor protein for mostly Fe transport
1131	40%	cirA	~ TonB dependent receptor (fepA?)
			outer membrane receptor protein for mostly Fe transport
1409	38%	cirA	~ TonB dependent receptor (fepA?)
			outer membrane receptor protein for mostly Fe transport
			~ TonB dependent
1144	28%	btuB	~ putative receptor for hemin and siderophores
			outer membrane receptor protein for Fe+3 dicitrate
3901	21%	fecA	~ TonB dependent
			outer membrane receptor protein for mostly Fe transport
1586	24%	fepA	~ TonB dependent ~ receptor for ferrienterochelins and colicins

Figure 3. Protein to protein alignment search of the entire genome of PM1 looking for matches to fepA in *M. flagellatus. M. flagellatus* proteins 685 and 1152 can be found on the JGI database.

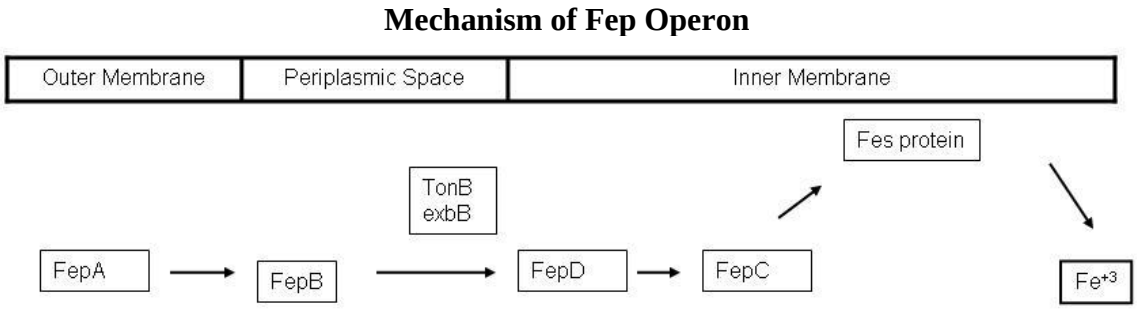

Figure 4. This figure shows the mechanism by which the fep operon collects ferric enterobactin, moves it into the cell, and then converts it to Fe^{+3}. FepA functions as an outer membrane receptor for ferric enterobactin. FepB transports the molecule through the periplasmic space with the help of the TonB-dependent energy transduction mechanism. Next, the ferric enterobactin is transportated through the inner membrane by fep D and fepC then and converted into Fe^{+3} by fes proteins [4].

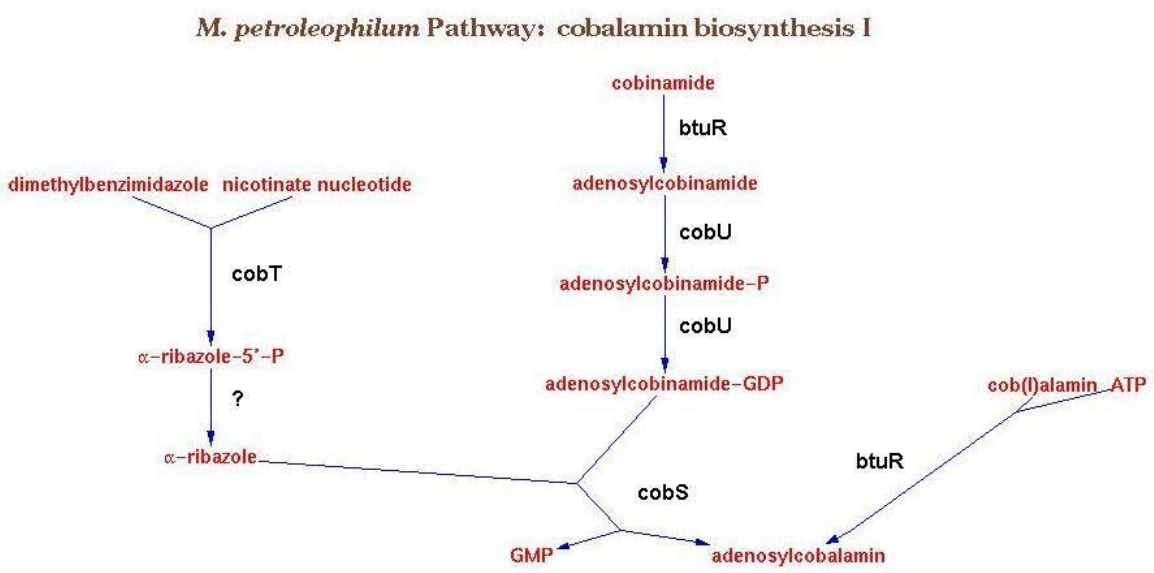

Figure 5. This figure displays the cobalamin biosynthesis I pathway in PM1. It is an aerobic pathway that converts dimethylbnezimdizole and nicotinate nucleotide, cobalamide, and cob(I)alamin to adenosylcobalamin. Figure is adapted from the Pathways Tools Software [9].

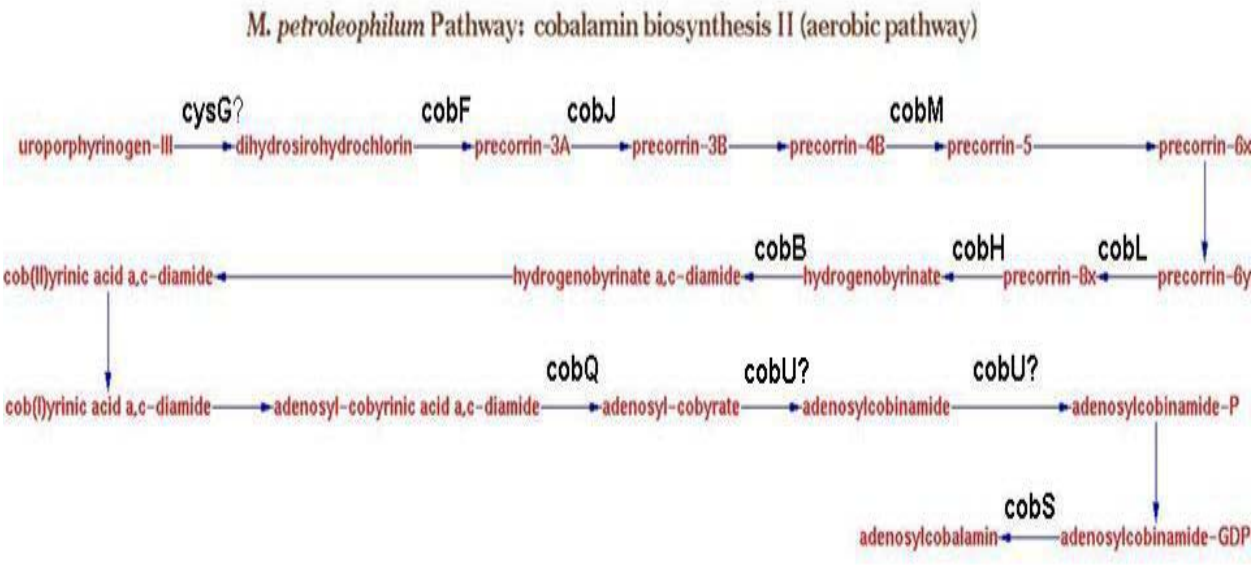

Figure 6. This figure displays the cobalamin biosynthesis II pathway in PM1. This is a aerobic pathway which converts uroporphyrinogen-III to adenosylcobalamin. Figure is adapted from the Pathways Tools Software [9].

Functional Assignments of Cob and Fep Operons in PM1

COBALAMIN GENES	GENE	FUNCTIONAL ASSIGN

CLUSTER 1 (380-396)		**TANDEM REPEAT 1**	
	380	CobB	cobrinic acid a,c-diamide synthase
	381	CobM	cobalt-precorrin-4 methyltransferase
	382	orf382	conserved hypothetical protein ~ putative sirohydrochlorin cobaltochelata
	383	CobH	precorrin-8X methylmutase ~ precorrin isomerase ???
	384	CbiD	Cobalamin (vitamin B12) biosynthesis protein
	385	CobL	precorrin -6Y C5-methyltransferase ~decarboxylation
	386	CobF	precorrin-2 C20-methyltransferase
	387	CbiG	Cobalamin (vitamin B12) biosynthesis protein ~precorrin methylase
	388	CbiG	precorrin methylase
	389	or389	conserved hypothetical integral membrane protein
	390	CobJ	precorrin-3 C-17 methylase
	391	BtuB	conserved hypothetical protein ~ outer membrane cobalamin (ferrientero
	392	FepC	putative branched-chain amino acid transporter ATP-binding protein
	393	FepC	ABC-type cobalamin and Fe+3 siderophores transport system, ATPase c
	394	or394	putative ferredoxin protein ~ energy production and conversion
	395	CysG	uroporphyrinogen-III methylase ~ nitroreductase family protein
	396	or396	uncharacterized enzyme of phosphonate metabolism ~ transposase and
CLUSTER 2 (412 - 428)		**TANDEM REPEAT 2**	
	412	CobB	cobrinic acid a,c-diamide synthase
	413	CobM	cobalt-precorrin-4 methyltransferase
	414	or414	conserved hypothetical protein ~ putative sirohydrochlorin cobaltochelata
	415	CobH	precorrin-8X methylmutase ~ precorrin isomerase
	416	CbiD	Cobalamin (vitamin B12) biosynthesis protein
	417	CobL	precorrin -6Y C5-methyltransferase ~ decarboxylation
	418	CobF	precorrin-2 C20-methyltransferase
	419	CbiG	Cobalamin (vitamin B12) biosynthesis protein~ precorrin methylase
	420	CbiG	precorrin methylase
	421	or421	conserved hypothetical integral membrane protein
	422	CobJ	precorrin-3 C-17 methylase
	423	BtuB	TonB-dependent receptor ~ conserved hypothetical protein ~ outer mem
	424	FepC	putative branched-chain amino acid transporter ATP-binding protein
	425	FepC	ABC-type cobalamin and Fe+3 siderophores transport system, ATPase c
	426	or426	putative ferredoxin protein ~ energy production and conversion
	427	CysG	Uroporphyrinogen-III methylase
	428	or428	Uncharacterized enzyme of phosphonate metabolism ~ transposase and
CLUSTER 3 (447 - 458)		**SIMILAR TO CHR CLUSTER**	
	447	BtuB	TonB-dependent receptor ~ conserved hypothetical protein ~ outer mem
	448	CobB	cobrinic acid a,c-diamide synthase
	449	CobU	Adenosyl cobinamide kinase
	450	FepB	Iron(III) dicitrate-binding protein ~ ABC-type Fe3+hydroxamate transport
	451	FepD	Iron(III) dicitrate-binding protein ~ ABC-type Fe3+hydroxamate transport
	452	FepC	Iron ABC transporter ATP-binding protein ~ ABC-type cobalamin/Fe3+si
	453	BtuR	Cob(I)alamin adenosyltransferase
	454	CbiB	cobalamin biosynthesis
	616	HisC	Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarb
	455	CobQ	Cobyric acid synthase
	456	CobT	Nicotinate-nucleotide—dimethylbenzimidazole phosphoribosyltransferas

617	CobS	cobalamin-5-phosphate synthase
457	GpmB	fructose-2,6-biphosphatase ~ phosphoglycerate mutase enzyme
CLUSTER 4 (4240 - 1144)	**ON CHROMOSOME**	
		TonB-dependent receptor ~ conserved hypothetical protein ~ outer mem
1144	BtuB	
1143	CobU	Adenosyl cobinamide kinase
1142	FepB	Iron(III) dicitrate-binding protein ~ ABC-type Fe3+hydroxamate transport
1141	FepD	Iron(III) dicitrate-binding protein ~ ABC-type Fe3+hydroxamate transport
1140	FepC	Iron ABC transporter ATP-binding protein ~ ABC-type cobalamin/Fe3+si
1139	CobQ	Cobyric acid synthase
1138	CobT	Nicotinate-nucleotide—dimethylbenzimidazole phosphoribosyltransferas
4241	CobS	cobalamin-5-phosphate synthatse ~
4240	GmpB	fructose-2,6-biphosphatase ~ phosphoglycerate mutase enzyme

Figure 7. Displays gene identity and function of cob and fep operons in PM1.

First Tandem Repeat of Cobalamin Operon on Megaplasmid in PM1

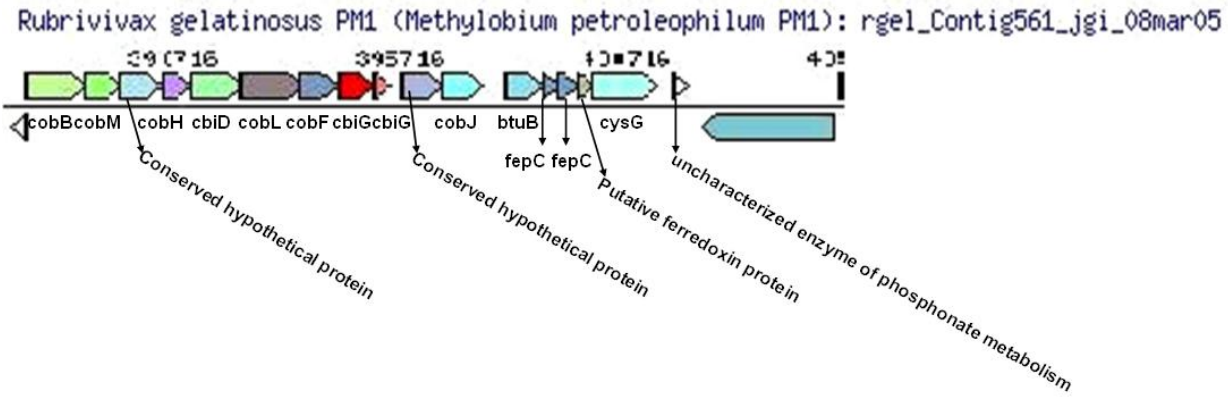

Figure 8. Displays the first tandem repeat of the cobalamin genes found on the megaplasmid of PM1. Figure includes genes 380 through 396 as displayed in fig. 4.

Second Tandem Repeat of Cobalamin Operon on Megaplasmid in PM1

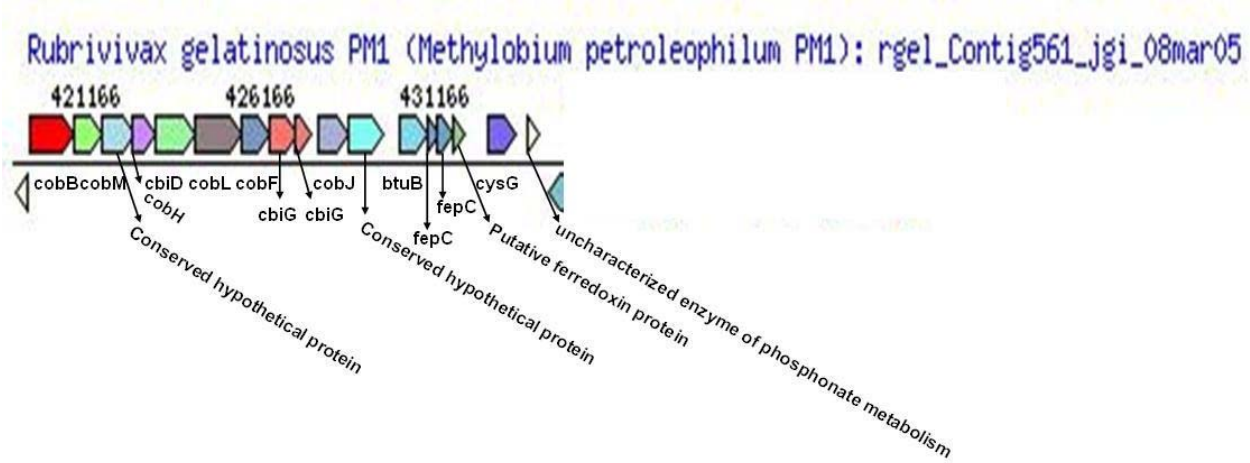

Figure 9. Displays the second tandem repeat of the cobalamin genes found on the megaplasmid of PM1. Figure includes genes 412 through 428 as displayed in fig. 6.

Cobalamin Cluster on Megaplasmid of PM1 that is Similar to the Chromosome Cluster

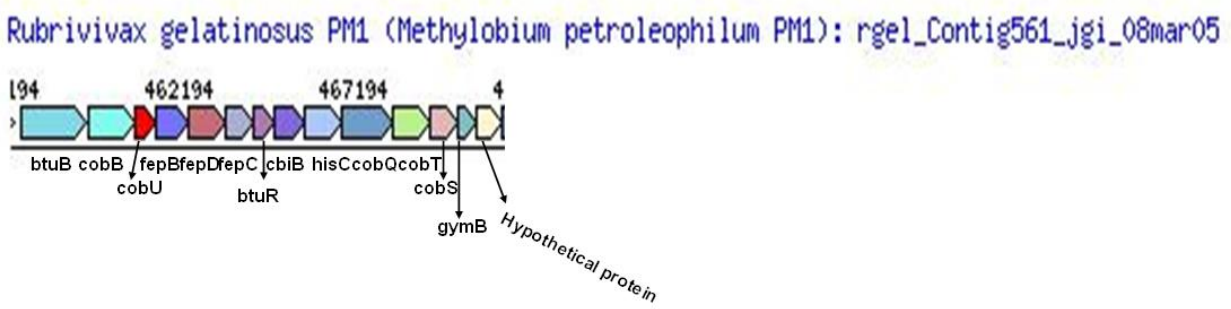

Figure 10. Displays the cluster of cobalamin genes on the megaplasmid that are similar to the cluster found on the chromosome of PM1. Figure includes genes 447 through 458.

Cluster of Cobalamin Genes Found on the Chromosome in PM1

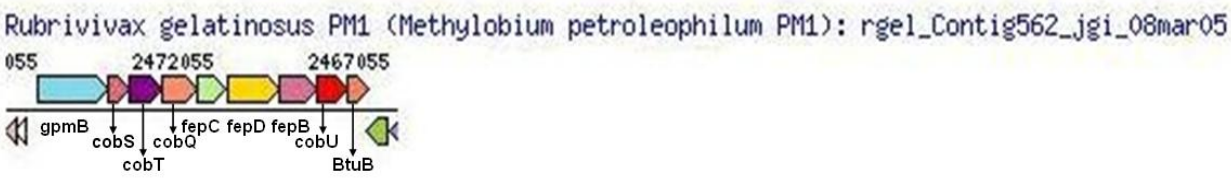

Figure 11. Displays the cluster of cobalamin genes found on the chromosome of PM1. Includes genes 4240-1144. This cluster contains outer membrane receptors for both ferric enterobactin and cobalamin

Polaronomonas sp. Cobalamin Protein Sequence

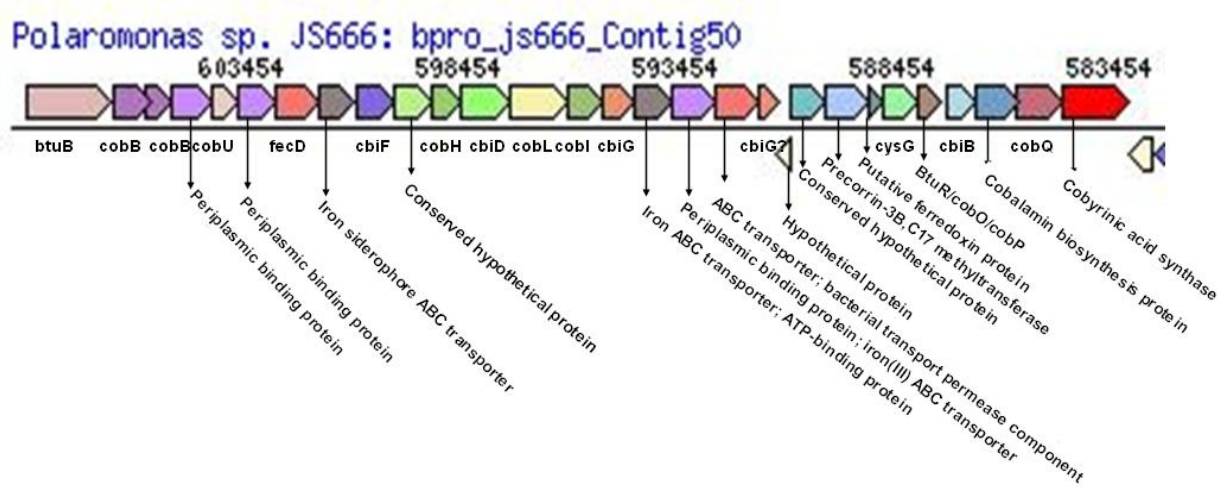

Figure 12. Displays the largest cobalamin operon in *Polaronomas sp.* As with PM1, the cobalamin operon in *Polaromonas sp.* contains many iron-siderophore transport genes. The proteins are labeled by their names or lacking the name are labeled by their function.

Functions of Cobalamin Genes Found in *Polaromonas*

***Polaromonas sp.* Cob Operon**

Gene	Name	Function
5389	btuB	TonB-dependent receptor ~ putative outer membrane hemin/siderophore receptor protein
5388	cobB	cobyrinic acid a,c diamide synthase
5387	cobB	cobyrinic acid a,c diamide synthase
5386		periplasmic binding protein ~ possible substrate-binding protein
5385	cobU	cobalamin biosynthesis enzyme

5384		periplasmic binding protein
5383	fecD	bacterial transport system permease protein ~ iron III dicitrate transport protein
5382		ABC transporter ~ iron-siderophore
5381	cbiF	precorrin-4 C11 methyl transferase
5380		conserved hypothetical protein
5379	cobH	precorrin-8X methylmutase
5378	cbiD	cobalamin (vitamin B12) biosynthesis protein
5377	cobL?	precorrin-6Y C5, 15 methyltransferase
5376	cobI	precorrin -2 C20 methyltransferase
5375	cbiG	cobalamin (vitamin B12) biosynthesis protein
5374		iron ABC transporter, ATP binding protein
5373		periplasmic binding protein ~ iron III ABC transporter
5372		ABC transporterbacterial transport system permease component
5371	cbiG?	cobalamin (vitamin B12) biosynthesis protein ~ precorrin methylase
5342		hypothetical protein
5370		conserved hypothetical protein (integral membrane transport?)
5369		precorrin-3B C17 methyltransferase
5368		putative ferredoxin protein
5367	cysG?	uroporphyrin-III C-methyltransferase C-terminal
5366		ATP corrinoid adenosyltransferase BtuR/cobO/cobP
5365	cbiB	Nitroreductase
5341		cobalamin biosynthesis protein
5364		histidinol-phosphate aminotransferase
5362	cobQ	cobrynic acid synthase

Figure 13. This figure displays the cobalamin operon found in *Polaromonas sp*. and the functions of the proteins. All genes numbers are from JGI.

Rhodoferax ferrireducens Cobalamin Protein Sequence

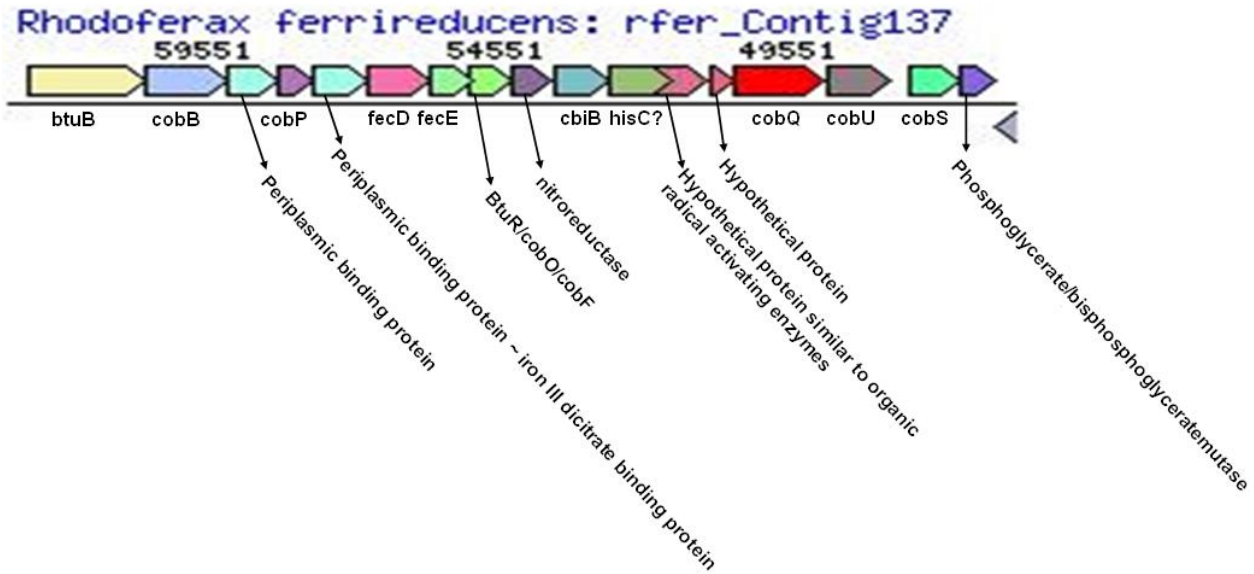

Figure 14. This figure displays the gene sequence of *R. ferrireducens*. The proteins are labeled by their names or lacking the name are labeled by their function.

Functions of Cobalamin Genes Found in *Rhodoferax ferrireducens*

Rhodoferax ferrireducens Cobalamin Operon		
gene	**name**	**Function**
2307	btuB	TonB-dependent receptor ~ outer membrane hemin/siderophore receptor
2306	cobB	cobyrinic acid a,c diamide
2305		periplasmic binding protein
2304	cobP	cobalbumin biosynthesis enzyme
2303		periplasmic binding protein ~ iron III dicitrate binding protein
2302	fecD	bacterial transport system permease protein ~ iron III dicitrate transport system permease protein
2301	fecE	ABC transporter ~ putative iron III dicitrate ABC transporter ~ ATP-binding component
2300		ATP corrinoid adenosyltransferase BtuR/CobO/CobF Nitroreductase
2299		
2298	cbiB	cobalamin biosynthesis protein
2415		histidinol-phosphate aminotransferase
2297		hypothetical protein similar to organic radical activating enzymes
2296		6-pyruvoyl tetrahydropterin synthase and hypothetical protein
2295	cobQ	cobyrinic acid synthase
2294	cobU	nicotinate-nucleotide-dimethylbenzimadazole phosphoribosyltransferase
2293	cobS	cobalamin-5-phosphate synthase
2292		phosphoglycerate/bisphosphoglyceratemutase

Figure 15. This figure displays the functions of the cobalamin proteins found in the same operon in *R. ferrireducens*. All genes numbers are from JGI.

Methylobacillus flagellatus Cobalamin Protein Sequence

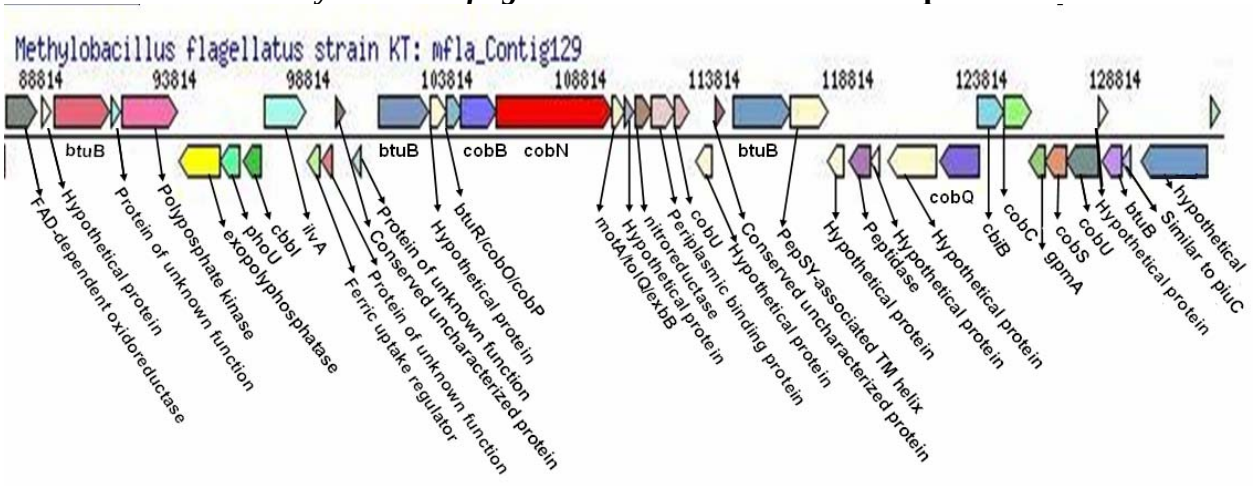

Figure 16. This figure displays a section of the genomic sequence of *M. flagellatus* containing many cobalamin genes. Although not grouped together in a single neatly arranged operon, they are closely located which may indicate a relationship in the functionality of the genes and their corresponding proteins. Also in the near proximity of the cobalamin genes are many proteins coding for the transport of iron and iron siderophores.

Functions of Cobalamin Genes Found in *M. flagellatus*

M. flagellatus Cobalamin Operon		
Gene	**Gene Name**	**Function**
1987		FAD dependent oxidoreductase ~ NAD binding site: D-amino acid oxidase

1988		hypothetical protein
1989	btuB	TonB-dependent receptor ~ outer membrane siderophore receptor
1990		protein of unknown function
1991		Polyphosphate kinase
2064		Exopolyphophatase
2063	PhoU	
2062	cbbI, ppi, rpiA	Ribose 5-phosphate isomerase
1992	ilvA	Threonine dehydratase I
2061	Fur1?	Ferric uptake regulator ~ probably ferric uptake transcriptional repressor
2060		protein of unknown function
1993		similar to uncharacterized protein conserved in bacteria
2058		protein of unknown function
1994	btuB	TonB-dependent receptor ~ outer membrane siderophore receptor
1995		hypothetical protein
1996	BtuR/cobO/cobP	ATP-corrinoid adenosyltransferase ~ cob(I)alamin adenosyltransferase
1997	cobB	cobyrinic acid a,c diamide synthase
1998	cobN	magnesium chelatase
1999	motA/tolQ/exbB	proton channel
2000		hypothetical
2001		Nitroreductase
2002		periplasmic binding protein ~ substrate binding
2003	cobU	cobalbumin biosynthesis enzyme ~ adenosyl cobinamide kinase/adenosyl cobinamide phosphate guanylyltransferase
2055		hypothetical
2004		similar to uncharacterized protein conserved in bacteria
2005	btuB	TonB-dependent receptor ~ outer membrane siderophore receptor
2006		PepSY-associated TM helix
2054		hypothetical
2053		peptidase M24A methionine aminopeptidase subfamily
2052		hypothetical
2051		Hypothetical
2050	cobQ	cobyric acid synthase
2008	cbiB	cobalamin biosynthesis protein
2009	cobC	cobalamin biosynthetic protein ~ aminotransferase class I and II
2049	gpmA	phophoglycerate/bisphosoglycerate mutase
2048	cobS	cobalamin -5-phophate synthase
2047	cobU	nicotinate-nucletide dimethylbenzimidazole
2046		hypothetical protein ~ sel1-like repeat
2045		similar to uncharacterized iron regulated protein such as PKHD-type hydroxylase piuC
2044	btuB	TonB-dependent receptor ~ outer membrane siderophore receptor
2010		rhodanese-like (hypothetical)

Figure 17. The above table represents the cobalamin genes and the surrounding operons around the genes which are closely grouped together in this organism. There are also a few iron and iron siderophore transport genes in the close proximity of the grouped cobalamin genes. All genes numbers are from JGI.

E. coli Cobalamin Protein Sequence

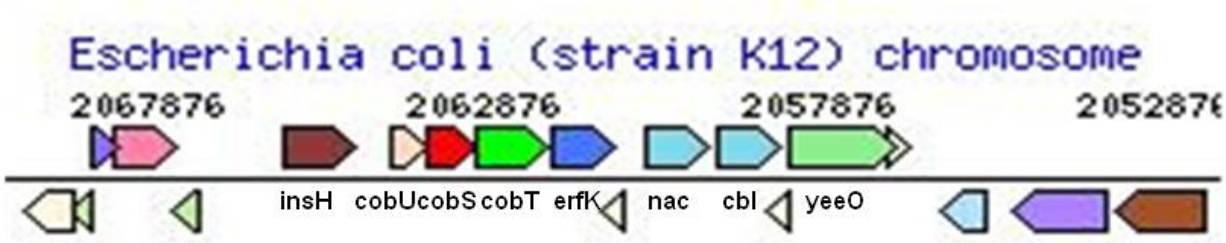

Figure 18. Figure 16 displays one of the most extensive cobalamin operons found in *E. coli.* As with the cobalamin operons found in PM1 and the other organisms, *E. coli* has proteins associated with nitrogen fixation (nac) and hypothetical proteins such as yeeO.

Functions of Cobalamin Genes Found in *E. coli*

***E. coli* Cob Operon**			
Gene	**Gene Accession #**	**Gene Name**	**Function**
6007720	P03837	insH	transposase insH for insertion sequence element
6007710	P46886	cobU	bifunctional adenosyl cobalamin biosynthesis protein
6007700	Q8X8U9	cobs	cobalamin synthesis
6007690	Q8X9U9	cobT	nicotinate-nucleotide-dimethyl benzimidazole phophoribosyltransferase
6007680	P39176	erfK	protein erfK/srfK precursor
6007660	Q47005	Nac	nitrogen assimilation regulatory protein
6007650	Q47083	Cbl	HTH-type transcriptional regulator cbl
6007630	P76352	yeeO	hypothetical protein

Figure 19. This figure displays the gene names and functions of the small cobalamin operon found in *E. coli* used in this study. The nac, cbl and yeeO proteins are all related to nitrogen assimilation.

Phylogenetic Tree

Proteobacteria
- Burkholderiales
  - Comamonadaceae
    - *Polaramonas sp*. JS666
    - *Rhodoferax ferrireducens*
    - *Methylobium petrolphilum* PM1
- Methylophilales
  - Methylophilacea
    - *Methylobacillus flagellatus* KT
- Enterobacteriacea
  - Enterobacteriaceae
    - *Escherichia coli* K12

Figure 20. This figure shows the phylogenetic tree relating the four species chosen for a phylogenetic comparison with PM1. *Polaramonas sp., Rhodoferax ferrireducens,* and PM1 are all in the same family called Comamonadaceae. *Methylobacillus flagellatus* and *E. coli* are in different families and orders in comparison with PM1.

PM1 Cobalamin Genes Against Those of Four Phylogenetically Related Organisms

PM1	gene	Polaromonas Gene or Function	% ID to PM1	R. ferrireducens Gene or Function	%ID to PM1	M. flagellatus Gene or function	%ID to PM1	E. coli Gene or Function	%ID to PM1
380	CobB	cobB	59.27	cobB	56.25	cobB	53.5	nac	57.14
381	CobM	cbiF	77.99	periplasmic binding protein ~ iron III dicitrate binding	52.63	PepSY-associated TM	41.67	nac	50

				protein		helix			
382	orf382	conserved hypothetical protein	77.07	hypothetical protein similar to organic radical activating enzymes	41.18	hypothetical	31.25	yeeO	28
383	CobH	cobH	83.26	hypothetical protein similar to organic radical activating enzymes	36.11	polyphosphate kinase	40	cobS	26.79
384	CbiD	cbiD	73.87	periplasmic binding protein ~ iron III dicitrate binding protein	41.18	BtuR/cobO/cobP	27.03	cobT	61.54
385	CobL	cobL	69.92	btuB	22.64	cobN	35.14	cobS	33.33
386	CobF	cobI or cobF?	63.49	btuB	47.62	fur1?	41.67	yeeO	32.26
387	CbiG	cbiG	80.49	periplasmic binding protein	22.45	btuB	31.37	cbl	27.27
388	CbiG	cbiG	65.19	cobB	33.9	hypothetical	37.93	cobT	44
389	or389	conserved hypothetical protein	60.31	cobs	36.36	PepSY-associated TM helix	29.17	yeeO	39.02
390	CobJ	cobJ	83.17	periplasmic binding protein	30.49	cobC	25.71	cobT	46.67
391	BtuB	btuB	25.23	btuB	23.77	btuB	27.37	cobT	44.44
392	FepC	iron ABC transporter, ATP binding protein	36.49	fecE	42.37	similar to uncharacterized protein conserved in bacteria	50	nac	34.38
393	FepC	iron ABC transporter, ATP binding protein	45.12	fecE	54.55	cobs	54.55	yeeO	36.11
394	or394	putative ferredoxin protein	83.04	periplasmic binding protein	30.43	ilvA	29.41	yeeO	30
395	CysG	cysG?	78.1	nitroreductase	62.39	nitroreductase	55.98	yeeO	53.33
396	or396	no hits		fecD	50	no hits		nac	30
412	CobB	cobB	59.27	cobB	56.25	cobB	53.5	nac	57.14
413	CobM	cbiF	77.99	periplasmic binding protein ~ iron III dicitrate binding protein	52.63	PepSY-associated TM helix	41.67	nac	50
414	or414	conserved hypothetical protein	77.07	hypothetical protein similar to organic radical activating enzymes	41.18	hypothetical	31.25	yeeO	28
415	CobH	cobH	83.26	hypothetical protein similar to organic radical activating enzymes	36.11	polyphosphate kinase	40	cobS	26.79
416	CbiD	cbiD	73.87	periplasmic binding protein ~ iron III dicitrate binding protein	41.18	BtuR/cobO/cobP	27.03	cobT	61.54
417	CobL	cobL	69.92	btuB	22.64	cobN	35.14	cobS	33.33
418	CobF	cobI or cobF?	63.49	btuB	47.62	fur1?	41.67	yeeO	32.26
419	CbiG	cbiG	80.49	periplasmic binding protein	22.45	btuB	31.37	cbl	27.27
420	CbiG	cbiG	65.19	cobB	33.9	hypothetical	37.93	cobT	44
421	or421	conserved hypothetical protein	60.31	cobs	36.36	PepSY-associated TM helix	29.17	yeeO	39.02
422	CobJ	cobJ	83.17	periplasmic binding protein	30.49	cobC	25.71	cobT	46.67
423	BtuB	btuB	25.23	btuB	23.77	btuB	27.37	cobT	44.44
424	FepC	iron ABC transporter, ATP binding protein	36.49	fecE	42.37	similar to uncharacterized protein conserved in bacteria	50	nac	34.38
425	FepC	iron ABC transporter, ATP binding protein	45.12	fecE	54.55	cobs	54.55	yeeO	36.11

426	or426	putative ferredoxin protein	83.04	periplasmic binding protein	30.43	ilvA	29.41	yeeO	30
427	CysG	cysG?	78.1	nitroreductase	62.39	nitroreductase	55.98	yeeO	53.33
428	or428	no hits		fecD	50	no hits		nac	30
447	BtuB	btuB	22.02	btuB	26.64	btuB	46.12	yeeO	26.04
448	CobB	cobB	64.44	cobB	58.57	cobB	37.23	yeeO	22.73
449	CobU	cobU	46.7	cobP	50.24	cobU	27.2	yeeO	27.6
450	FepB	periplasmic binding protein	54.89	periplasmic binding protein ~ iron III dicitrate binding protein	55.64	periplasmic binding protein ~ substrate binding	32.56	cbl	21.21
451	FepD	bacterial transport system permease protein	55.48	fecD	54.95	cobs	35.71	yeeO	52.63
452	FepC	ABC transporter	44.49	fecE	44.94	hypothetical	44.38	yeeO	66.67
453	BtuR	BtuR/cobO/cobF	72.54	BtuR/CobO/CobF	72.49	BtuR/cobO/cobP	33.6	yeeO	50
454	CbiB	cobI	40	cbiB	50	cbiB	32.08	cbl	34.62
616	HisC	hisC?	45.77	histidinol-phosphate aminotransferase	46.61	cobB	44.11	yeeO	21.62
455	CobQ	cobQ	53.26	cobQ	53.54	cobQ	50.29	yeeO	58.33
456	CobT	cbiD	27.44	cobU	62.61	cobU	38.19	cobT	33.33
617	CobS	bacterial transport system permease protein	33.73	cobs	35.32	cobs	26.92	cobS	33.06
457	GpmB	cobB	29.36	phosphoglycerate/ bisphosphoglyceratemutase	29.41	gpmA		cobS	46.15
4240	GmpB	conserved hypothetical protein	70	phosphoglycerate/ bisphosphoglyceratemutase	33.89	gpmA	28.07	cobT	35
4241	CobU	conserved hypothetical protein	27.59	Cobs	46.21	cobs	43.41	cobS	35.69
1138	FepB	cbiD	29.67	cobU	57.89	cobU	55.59	cobT	34.62
1139	FepD	cobQ	63.14	cobQ	60.87	cobQ	39.31	cbl	29.58
1140	FepC	ABC transporter	48.43	fecE	51.45	hypothetical protein	52.63	cbl	33.33
1141	CobQ	fecD	60.63	fecD	61.02	no hits		erfK	33.33
1142	CobT	periplasmic binding protein	55.69	periplasmic binding protein ~ iron III dicitrate binding protein	55.24	periplasmic binding protein ~ substrate binding	27.35	cbl	32.26
1143	CobS	cobU	54.55	cobP	52.02	cobU	40.45	yeeO	32.8
1144	BtuB	btuB	59.27	btuB	56.69	btuB	28.08	erfK	38.46

Figure 21. The table above displays the best protein to protein sequence alignments when the cobalamin genes of PM1 are aligned with that of the other four individual organisms. All genes from *Polaromonas sp., R. ferrireducens, M. flagellatus,* and *E. coli* can be found in figures 11, 13, 15, and 17, respectively.