

Augustana College Augustana Digital Commons

Meiothermus ruber Genome Analysis Project

Biology

Winter 2-2016

E. coli b3639 and b3634 are orthologs of Mrub_2047 and Mrub_1372


Rong Zheng

Augustana College, Rock Island Illinois

Dr. Lori Scott

Augustana College, Rock Island Illinois

Follow this and additional works at: <http://digitalcommons.augustana.edu/biolmruber>

 Part of the [Bioinformatics Commons](#), [Biology Commons](#), [Genetics Commons](#), [Molecular Biology Commons](#), and the [Molecular Genetics Commons](#)

Recommended Citation

Zheng, Rong and Scott, Dr. Lori. "*E. coli* b3639 and b3634 are orthologs of Mrub_2047 and Mrub_1372" (2016). *Meiothermus ruber Genome Analysis Project*.

<http://digitalcommons.augustana.edu/biolmruber/2>

This Student Paper is brought to you for free and open access by the Biology at Augustana Digital Commons. It has been accepted for inclusion in Meiothermus ruber Genome Analysis Project by an authorized administrator of Augustana Digital Commons. For more information, please contact digitalcommons@augustana.edu.

E. coli b3934, b3634 are Orthologs of Mrub_2047, Mrub_1372

Introduction

Coenzyme A (CoA) is an essential cofactor that plays an important role in numerous biosynthetic, degradative and energy yielding metabolic pathways (Geerlof *et al.*, 1999). It is the principle acyl carrier in all living cells (Kupke *et al.*, 2002). In fatty acid synthesis, CoA delivers the 4'-phosphopantetheine moiety with acetyl group to the acyl carrier protein (ACP) and malonyl group to the growing hydrocarbon tail of the fatty acid molecules (Geerlof *et al.*, 1999). The product of glycolysis- pyruvate- combines with CoA to form acetyl CoA, which enters citric cycle. In citric acid cycle, succinyl CoA is formed as an intermediate.

In *E. coli*, CoA synthesis is a four-step process as shown in Figure 1 below. The first step is catalyzed by enzyme phosphopantothencysteine synthase, where (R)-4'-phosphopentothenate is combined with L-cysteine to form (R)-4'-phosphopentothenoil-L-cysteine (PPC) with CTP as the cofactor for the reaction. Then, (R)-4'-phosphopentothenoil-L-cysteine is decarboxylated by 4'-phosphopentothenoilcysteine decarboxylase, forming 4'-phosphopantetheine (PP). In next two steps, PP is first phosphorylated by phosphopantetheine adenyltransferase (PPAT) to form 3-dephosphate-CoA (dPCoA) followed by the second phosphorylation mediated by dephosphate-CoA kinase. Both phosphorylation steps are energy consuming mediated by high-energy carrier ATP, and coenzyme A is produced as the final product.

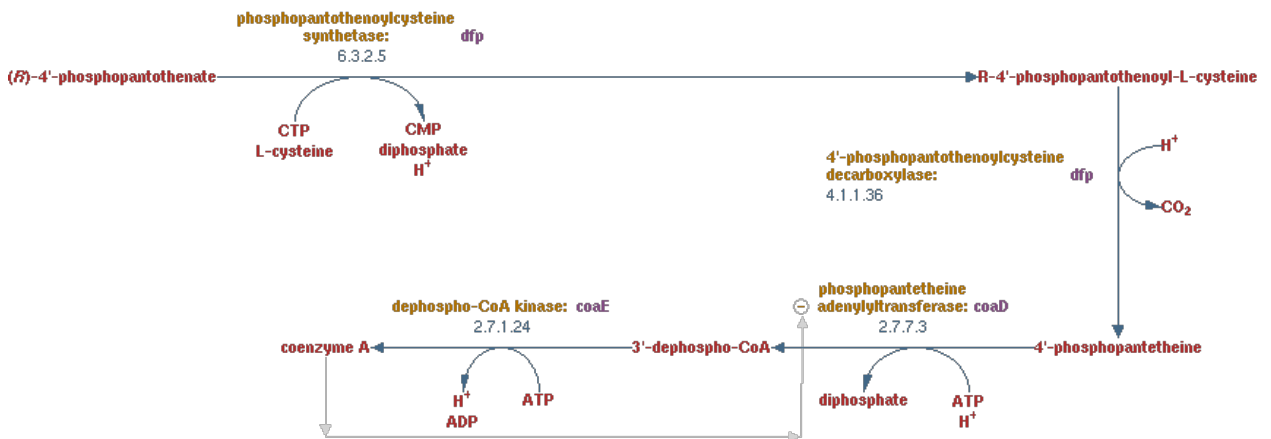


Figure 1. Gene product of dfp and coaD are involved in *E. coli* CoA biosynthesis. EcoCyc (<http://ecocyc.org/ECOLI/NEW-IMAGE?type=PATHWAY&object=COA-PWY>) created this coenzyme A biosynthesis pathway.

There are actually only three enzymes involving in CoA synthesis pathway of *E. coli* as the enzyme that catalyzes the first two sequential steps phosphopantothencysteine synthase and 4'-phosphopentothenoilcysteine

decarboxylase are two parts of a bi-functional enzyme fused 4'-phosphopantothienoylcysteine decarboxylase/phosphopantothienoylcysteine synthetase, FMN-binding (Strauss *et al.*, 2001). This bifunctional enzyme was encoded by *dfp* or *coaBC* (Strauss *et al.*, 2001). *Dfp* was first identified from a temperature sensitive mutant *dfp707* (Spitser & Weiss, 1985). Spitzer and Weiss found out that conditional lethality of *dfp 707* could be complemented by a plasmid carrying a gene segment of chromosomal DNA that is adjacent and independent to *dut* gene. No operon has been identified containing *dfp*. Although Spitzer and Weiss did not know the function of the gene was at that time, they characterized protein product of *dfp* and determined that *dfp* protein is a 45-kilodalton flavoprotein containing flavin mononucleotide (FMN) (Spitser & Weiss, 1985).

Dfp of *E. coli* is 1221bp long encoding for 406 amino acids. *Dfp* lies between nucleotide 3810754 and 3811974. It is a homodecamer protein (Kupke *et al.*, 2000). The C-terminal (CoaB) domain of the enzyme confers phosphopantothienoylcysteine synthase activity (Strauss *et al.*, 2001) The N-terminal of *dfp* is homologous to EpiD, a flavoenzyme that carries out the decarboxylation of cysteine residue of epidermin precursor (Kupke *et al.*, 2000). Kupke *et al.* (2000) had showed that N-terminal domain (CoaC) of *dfp* catalyzed the decarboxylation of 4'-phosphopantothienoylcystein (PPC) to 4'-phosphopantetheine (PP). Since the C-terminals and N-terminals of *dfp* catalyze two different reactions, *dfp* is also referred as CoaBC.

Phosphopantetheine adenylyltransferase (PPAT), in the third step of CoA synthesis is a hexameric protein, behaving as a dimer of trimers (Geerlof *et al.*, 1999). PPAT is made up of 159 amino acids. In eukaryotic cells, PPAT forms multi-enzyme complex with dephosphate-CoA kinase, while the PPAT in *E. coli* does not form complex with dephosphate-CoA kinase. PPAT is encoded by gene *coaD* with 480 nucleotides (Geerlof *et al.*, 1999). In *E. coli*, *coaD* covers nucleotide from 3807848 to 3808327.

E. coli is a well studied model microorganism, yet many other microorganisms are understudied. The focus of this project is to study the *coaBC* (Mrub_2047) gene and *coaD* (Mrub_1372) gene in *M. ruber* and its functional role in CoA synthesis by applying various bioinformatics tools. We hypothesize that *coaBC* (Mrub_2047) and *coaD* (Mrub_1372) in *M. ruber* are orthologs of *coaBC* (b3639) and *coaD* (b3634) in *E. coli*; they shares both structural and functional similarities.

Method

Bioinformatics programs with the GENI-ACT lab notebook are used for the analysis of *E. coli* b3639, b3634 and Mrub_2047 and Mrub_1372 (http://geniact.org/student/view_assignment/isolate_genome_gene/6fa2ae446a0244ad/88178c02e87e4060/). In addition, Blast searches of b3639 and b3934 in *Meiothermus ruber* genome were performed (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). More than ten species (15-20) were selected for T-coffee and Web logo analysis.

Results

b3639 and Mrub_2047

Table 1. shows the comparison between *E. coli* b3639 and Mrub_2047. The predicted location of both b3639 and Mrub_2047 are cytoplasm. BLAST search demonstrates the similarity in sequence between *E. coli* b3639 and Mrub_2047. Both *E. coli* b3639 and Mrub_2047 involve in pentothenate and CoA synthesis. There are structural similarities between *E. coli* b3639 and Mrub_2047 as they are in the same TIGRfam family and Pfam family. *E. coli* b3639 and Mrub_2047 are in the same COG category CoaBC. Although Mrub_2047 is also identified in another CDD category Ubix 3-polyprenyl-4-hydroxybenzoate decarboxylase, it is less significant than CoaBC since the E-value ($E=2.2e-10$) is much larger than the E-value of CoaBC ($E=1.29e-137$). PDB pulls out the same the protein for both *E. coli* b3639 and Mrub_2047 search. Both *E. coli* b3639 and Mrub_2047 have higher GC content in comparison with organisms' genome.

Table 1. *E. coli* b3639 and Mrub_2047 are orthologs

Description of evidence collected	<i>E. coli</i> b3639	<i>M. ruber</i> Mrub_2047
Cellular localization	Cytoplasm	Cytoplasm
Blast <i>E. coli</i> against <i>M. ruber</i>	Score: 270 bits; E- value: 1e-86	
KEGG pathway	Pentothenate and CoA biosynthesis E.C. 6.3.2.5/4.1.1.36	
Pfam-protein family	PF02441 Flavoprotein family ($E=2.2e-31$) PF04127 DFP ($E=1.9e-73$)	PF04127 Flavoprotein family ($E=5.8e-32$) PF04127 DFP ($E=2.7e-68$)
CDD (COG category)	COG0452 ($E=2.23e-166$) CoaBC Phosphopantothenoylecysteine synthetase/decarboxylase	COG0452 ($E=1.29e-137$) CoaBC Phosphopantothenoylecysteine synthetase/decarboxylase COG0163 ($E=2.02e-10$) Ubix 3-polyprenyl-4-hydroxybenzoate decarboxylase
TIGRfam-protein family	TIGR00521 ($E=1.3e-205$) coaBC_dfp: phosphopantothenoylecysteine deca	TIGR00521 ($E=2.8e-137$) coaBC_dfp: phosphopantothenoylecysteine deca
PDB	1U7U ($E=1.28351E-104$) Phosphopantothenoylecysteine synthetase	1U7U ($E=3.49585E-23$) Phosphopantothenoylecysteine synthetase
GC content of the gene	54%	71%

Sequence based similarities

The gene product of Mrub_2047 is phosphopantothenoylcysteine decarboxylase/phosphopantothenate/cysteine ligase. When using BLAST to search b3639 protein sequence in *M. ruber* genome, phosphopantothenoylcysteine decarboxylase/phosphopantothenate/cysteine ligase was pulled out from the database. Comparing b3639 and Mrub_2047, protein Blast produced the bit score of 270 (Fig. 2.). Approximately half of the amino acids (44%) of b3639 and Mrub_2047 are identical with 4% of gap (Fig. 2.). It is unlikely that similarities in amino acids between b3639 and Mrub_2047 is due to chance as the E value is relatively small ($E=1e-86$). Therefore, b3639 and Mrub_2047 have very similar amino acid sequence.

Score	Expect	Method	Identities	Positives	Gaps
270 bits(689)	1e-86	Compositional matrix adjust.	168/382(44%)	232/382(60%)	16/382(4%)
Query 3		LAGKKIVLGVSGGIAAYKTPELVRRRLDRGADVVRVAMTEAAKAFITPLSLQAVSGYPVSD			62
Sbjct 4		L GK+IVLGVSG IAAYK +L +L GA V V ++E A+ FITPL+ +++G			63
Query 63		SLLDPAEEAAMGHIELGKWADLVILAPATADLIARVAAGMANDLVSTICLATPAPVAVLP			122
Sbjct 64		SL + EA + H+ LG+ ADL+++AP TA +A++A G A++L++ LA PV V P			121
Query 123		AMNQMYRAAATQHNLEVLASRGLLIWGPDSGSGQACGDIGPGRMLDPLTI---VDMVAHA			179
Sbjct 122		AM+ M+ ATQ NLE L RG+ I GP G A G +G GRML+P I + +A+			181
Query 180		FSPVNDLKHNLNIMITAGPTREPLDPVRYISNHSSGKMGFAIAAAAARRGANVTLVSGPV-			238
Sbjct 182		P+ KH ++++AG T+EPLDPVRY++N SSGK GFA+A AA GA V+LV G			238
Query 239		-SLPTPPFVKRVDVMTALEMEAAVNASVQQQNIFIGCAAVADYRAATVAPEKIKKQATQG			297
Sbjct 239		+LPTP +R DV TA +M AV ++ ++ I AAVAD+R V KIKK+			295
Query 298		DELTIKMKVKNPDIAGVAALKDH--RP-YVVGFAAETNNVEEYARQKRIRKNDLICAND			354
Sbjct 296		+ +++ PDI+ VA + H RP VVGFAAE+ + E A+ K RK+L +I AND			355
Query 355		VSQPTQGFNSDNNALHLFWQDG 376			
Sbjct 356		+S P GF D N + L G 377			

Figure 2. b3639 and Mrub_2047 have highly similar amino acid sequence. Blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) created this amino acid sequence alignment.

Cellular localization data

The locations for Gene products of b3639 and Mrub_2047 are predicted to be cytoplasm. Based on TMHMM analysis, b3639 and Mrub_2047 proteins do not contain any trans-membrane region, meaning that they are unlikely to be found in cell membrane (Fig. 4.) Signal peptide D score for b3639 is 0.2 while the cutoff score is 0.57. Signal peptide D score for Mrub_2047 is 0.306, which is also lower than the cutoff score 0.57 (Fig. 5). Signal peptide is the N-terminal sequence that will be cleaved when trans-membrane proteins are inserted to the cell membrane. Therefore the absence of signal peptide also suggests that b3639 and Mrub_2047 are not located on cell membrane. Phobius analysis shows that b3639 and Mrub_2047 are noncytoplasmic, which is inconsistent with the results obtained from TMHMM and Signal P (Fig. 6). It is possible that both b3639 and Mrub_2047 have regions of hydrophobic amino acids, but not be membrane-bound. The preponderance of evidence indicates that the predicated location for b3636 and Mrub_2047 is cytoplasm, and in this case, it is not wise to interpret the result of Phobius as an accurate prediction.

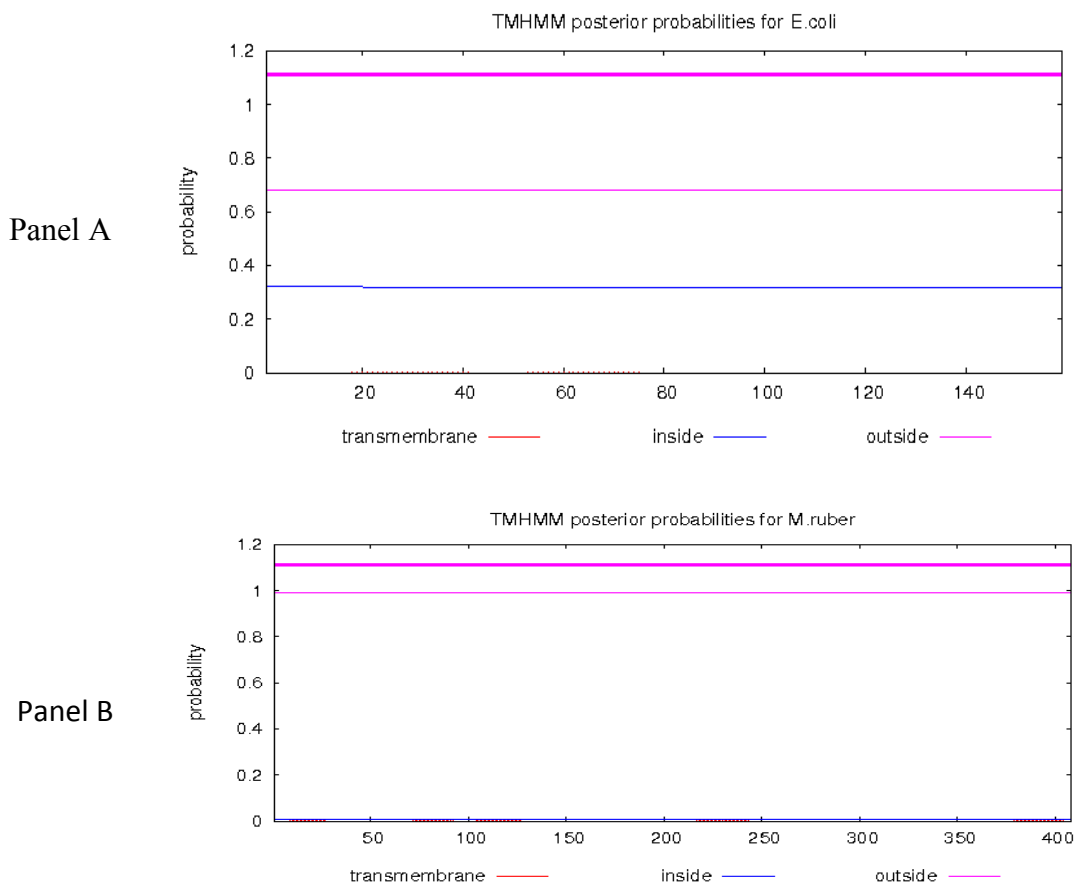


Figure 4. b3639 and Mrub_2047 do not contain TMH (trans-membrane helix) regions; cytoplasmic location is predicted. Panel A=*E. coli* *coaBC*/b3639; Panel B=Mrub_2047. TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM>) created this hydropathy plot.

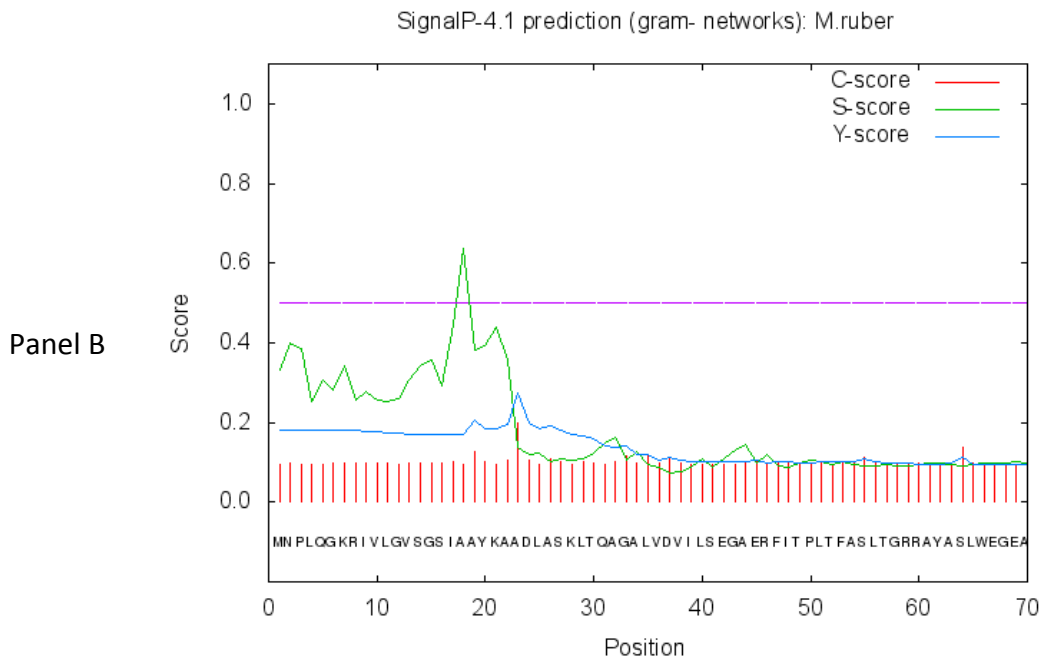
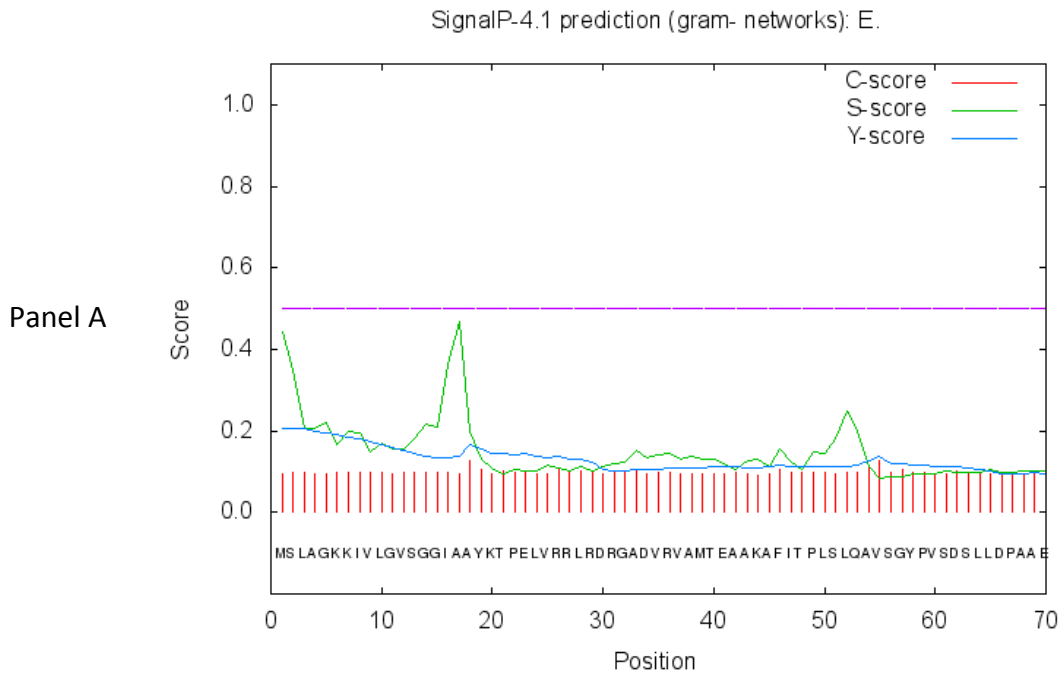
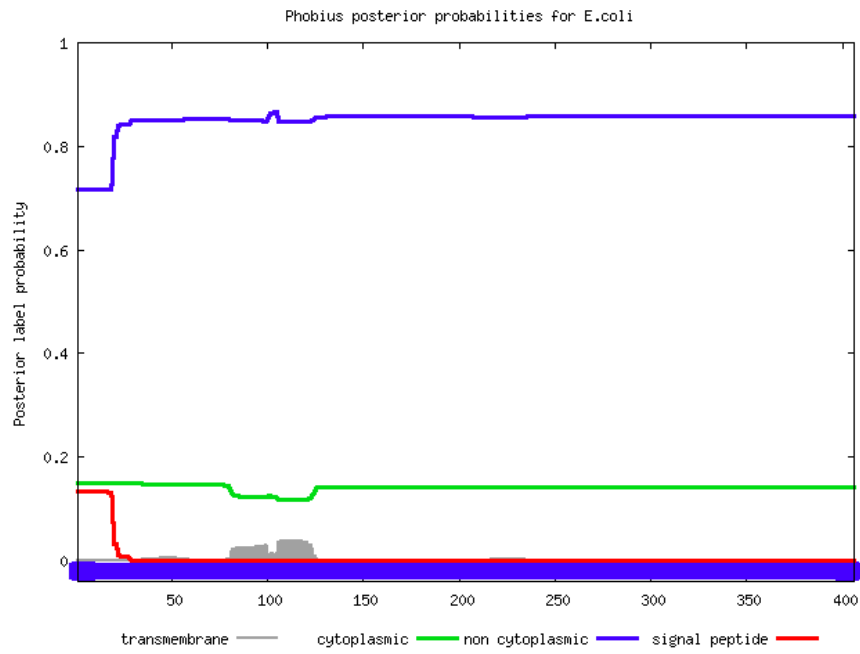


Figure 5. b3639 and Mrub_2047 do not contain signal peptide; cytoplasmic location is predicted. Panel A=*E. coli* *coaBC*/b3639; Panel B=Mrub_2047. Signal IP (<http://www.cbs.dtu.dk/services/SignalP>) created this hydropathy plot.

Panel A



Panel B

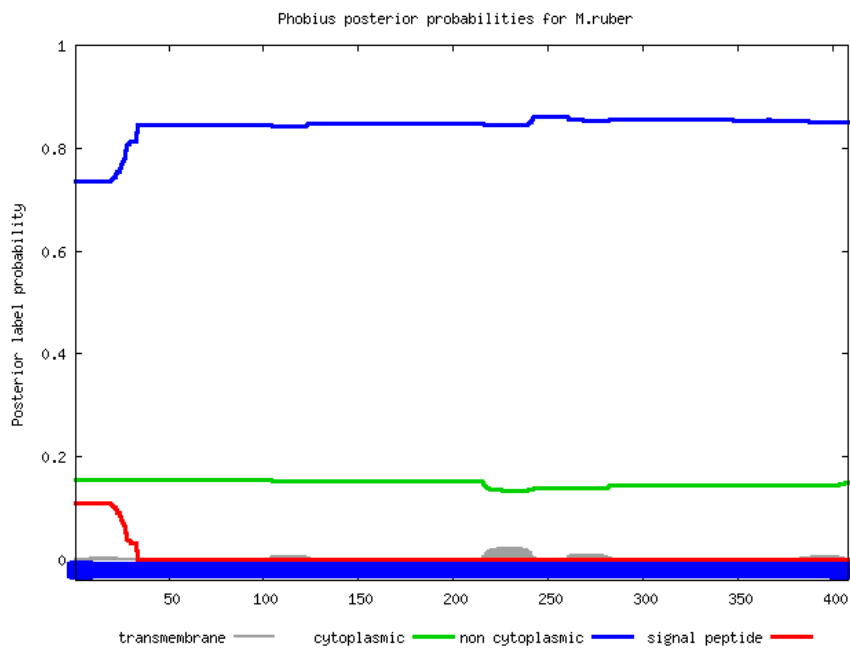


Figure 6. Both b3639 and Mrub_2047 is predicted to be noncytoplasmic. Panel A=*E. coli* *coaBC*/b3639; Panel B=Mrub_2047. Phobius (<http://phobius.sbc.su.se/>) created this hydropathy plot.

Evidence in structure

Flavoprotein domain is found in both *E.coli* b3639 and Mrub_2047. Domain is a tertiary structure and functional unit of a protein (Table 1.). Since *E.coli* 3639 and Mrub_2047 have flavoprotein domain in common, it is likely that they have similar structure and function. The E values for *E.coli* b3639 and Mrub_2047 are 2.2e-30 and 5.8e-32 (Table 1). With very small E values, it is unlikely that flavoprotein domain is found in both proteins by chance. The pairwise alignment shows that *E. coli* b3639 has conserved amino acids that are important to the function of flavoprotein domain, and those amino acids are G14, A43, P93, A100, G105 (Fig. 7). Mrub_2047 also has similar conserved amino acids G12, A44, A101, G106 (Fig. 7). The positions of conserved amino acids in Mrub_2047 protein are relatively closed to their positions in *E.coli* b3639.

In addition to flavoprotein, DNA/Pantothenate flavoprotein (DFP) domain is also found in both *E. coli* b3639 and Mrub_2047 proteins. The E values for both proteins are very low (E=1.9e-73, E=2.7e-68), suggesting that it is unlikely that presence of DFP domain in both *E. coli* b3639 and Mrub_2047 is due to chance. Pairwise alignment shows that amino acids G196, T197, E199, D202, V204, R205, N209, S211, S212, G213, G216, A220, G227, V230, A274, A275, D278, K288, G325, F326, N352 are the conserved amino acids of DFP domains in *E. coli* b3639, and G195, T197, E199, D202, V204, R205, N209, S211, S212, G213, G216, A220, G227, V230, A276, A277, D280, K290, G329, F330, N356 are conserved amino acids of DFP domains in Mrub_2047 (Fig. 8). Both *E.coli* b3639 and Mrub_2047 have exact same conserved amino acids in their proteins, and the positions of those amino acids in both proteins are either the same or relatively near to each other. The presence of both favoprotein and DFP domains in *E.coli* b3639 and Mrub_2047 suggests that both proteins share structurally and functional similarities.

Both b3639 and Mrub_2047 pull Phosphopantothenoylcysteine synthetase protein from database PDB, suggesting that b3639 and Mrub_2047 proteins share structural similarities with Phosphopantothenoylcysteine synthetase and with each other. The E values for b3639 is 1.28351E-104, and the E value for Mrub_2047 is 3.49585E-23. Low E values indicate that it is unlikely that similarities in structure are due to chance.

Panel A

```

#HMM      lkillgiTGsvaaikalelvekIkkegaeqvvtkaAekflskeksqkqsksekvltkdewdwtswq
#MATCH    +ki+lg++G +aa+k++elv++l+++ga+v+v +t+aA+ f+++ ++q+ s  v +
#PP       69*****977776666665...45678
#SEQ      KKIVLGVSGGIAAYKTPELVRRLRDRGADVVRVAMTEAAKAFITPLSLQAVSGYPVSDS----LDPA
          rtdpilhielrrwadllvvaPlsantlAkianGladnlltnaarawlkekkpliiapamntlmynspi
          ++ + hiel +wadl++ aP+++a +A+++ G+a++l+++++ a+  p+ + pamn+ my  +
          g*****...79*88*****9776
          AEAAMGHIELGKWADLVILAPATADLIARVAAGMANDLVSTICLAT---PAPVAVLPAAMNQMYRAAA

```

Panel B

```

#HMM      lkillgiTGsvaaikalelvekIkkegaeqvvtkaAekflskeksqkqsksekvltkdewdwtswq
#MATCH    ++i+lg++Gs+aa+ka++l++kl ++ga v+v+l++ Ae+f+++ ++++ + ++ +
#PP       69*****97777777775...558
#SEQ      KRIVLGVSGSIAAYKAADLASKLTQAGALVDVILSEGAERFITPLTFASLTGRRAYAS-----LWE
          rtdpilhielrrwadllvvaPlsantlAkianGladnlltnaarawlkekkpliiapamntlmynspt
          - + +lh+ l + adllv+aP+++a tIAk+a+G adnllt +a a+  + p+++apam  m+++p+t
          j99*****...9*99*****87
          GEAHVLVHVLGEEAADLLVIAPCTAHTLAKLAQGQADNLLTLTALAA---RCPVLVAPAMDGGMFEHPAT

```

Figure 7. One of the predicted domain in b3639 and Mrub_2047 is flavoprotein domain, and both b3639 and Mrub_2047 contain similar conserved amino acids that are important for the function of proteins. Panel A=*E. coli coaBC*/b3639; Panel B=Mrub_2047. Pfam (<http://pfam.sanger.ac.uk/search>) created this pairwise alignment.

Panel A

```

#HMM      egkkvlvtaGpTrEaiDpVrfisNrSSGkmGyalAeaaaarGaeVtlvsgpvsletpe
#MATCH    ++ ++++taGpTrE++DpVR+isN+SSGkmG+a+A+aaa+rGa+vtlvsgpvsletpe+
#PP       6789*****
#SEQ      HLNIMITAGPTREPLDPVRYISNHSSGKMGFIAAAAAARRGANVTLVSGPVSLETPP
          gvevvkvesaeemleaveealkeadivimaAAvaDyrpkeaaeeKikke.eeeeeitteleknpdilaelaek
          v++v+v +a em++av++ ++++i+i +AAvaDyr +++a eKikk+ +++elt++++knpdi+a +a
          *****98567899*****99
          FVKRVDVMTALEHEAAVNASVQQQIFIGCAAVADYRAATVAPEKIKKQATQGDLETIKMWKNPDIVAGVAAL
          kkqllvGFaaeteellenareklerkgldlivaNdvskeeagfgsdenevtl
          k++++vGFaaet++++e+ar+k+ rk+ldli+aNdvs++++gf+sd+n+++l
          99999*****9987
          KDHRPVVVGFAAEETNVVEEYARQKRIRKNLDLICANDVVSQPTQGFNSDNALHL

```

Panel B

```

#HMM      legkkvlvtaGpTrEaiDpVrfisNrSSGkmGyalAeaaaarGaeVtlvsgpv..
#MATCH    l+gk+v+v+aG+T+E++DpVR+++NrSSGk+G+a+lA+aaa+ Ga+V lv g +
#PP       689*****9744
#SEQ      LAGKHVVVSAGGTOEPLDPVRYLTNRSSGKGFALAOAALDLGARVSLVVGATaa
          sletpegvevvkvesaeemleaveealkeadivimaAAvaDyrpkeaaeeKikkeeeeeitteleknpdilaelaekkk...
          +l+tp+g+e+++ve+a+m+eav + ++ead++imaAAvaD+rp+++ ++Kikk e+e ++lele++pdil ++ae+++
          688999*****9.45679*****87777565
          ALPTPTGAERTDVTAAQMAEAVLRLCREADVLTMAAAVADFRPROVRAHKIKK--FEAPOLELEPTDILLVAEERRHLe
          kkqllvGFaaeteellenareklerkgldlivaNdvskeeagfgsdenevtl
          + ++vGFaae+e+l ena++klerk+l +ivaNd+s ++agf+ d+n+v+l
          4569*****998
          RPAVVVGFAAESEKLRENAQDKLERKSLSMIVANDISAPDAGFAVDTRNVLL

```

Figure 8. One of the predicted domains in b3639 and Mrub_2047 is DFP domain, and both b3639 and Mrub_2047 contain similar conserved amino acids that are important for the function of proteins. Panel A=*E. coli coaBC*/b3639; Panel B=Mrub_2047. Pfam (<http://pfam.sanger.ac.uk/search>) created this pairwise alignment.

Enzyme function

Figure 9 shows the biosynthesis pathway of CoA in *E. coli* and *M. ruber*. b3639 and Mrub_2047 are both involved in CoA synthesis pathway. Both b3639 and Mrub_2047 are assigned the same enzyme commission number E.C. 6.3.2.5/4.1.1.36, which means that b3639 and Mrub_2047 proteins catalyze the same type of reaction in CoA synthesis. Therefore, b3639 and Mrub_2047 proteins have same function.

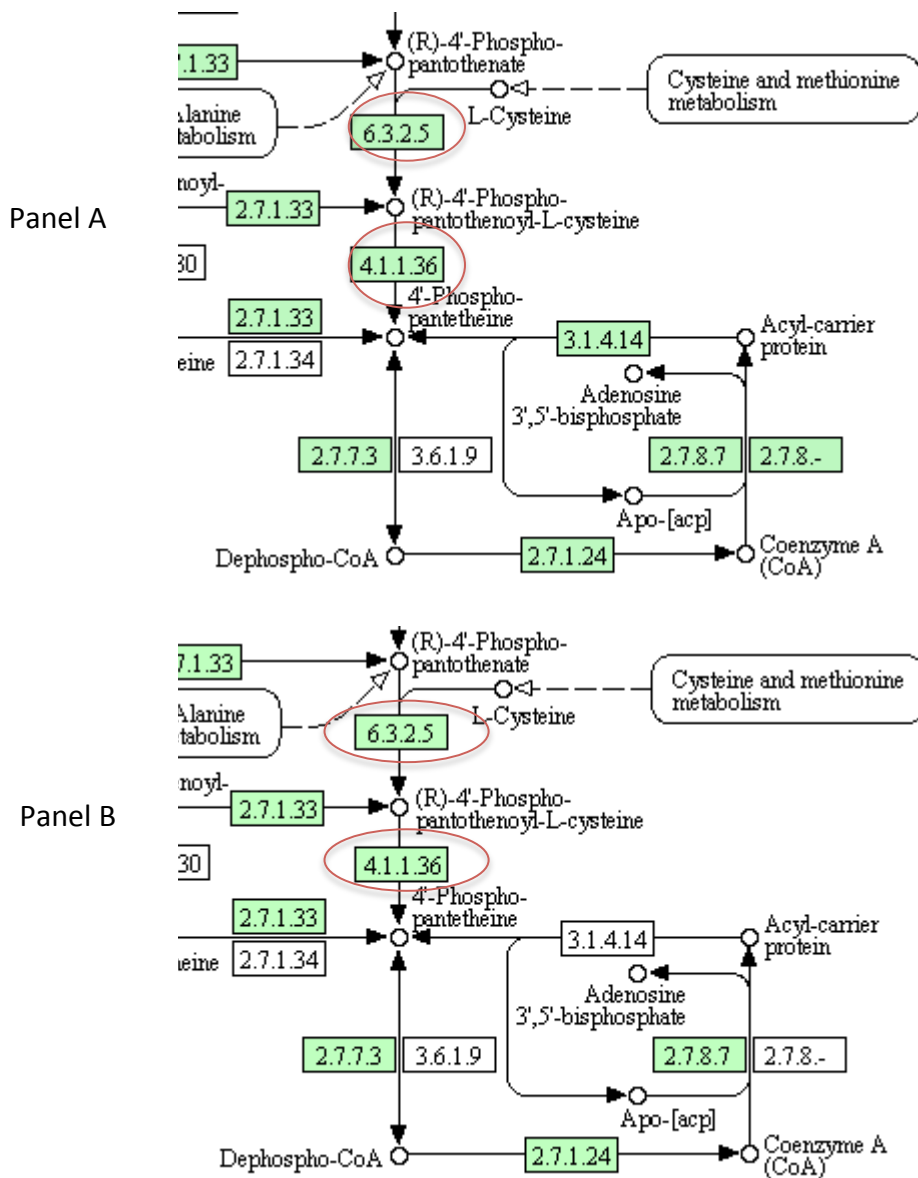


Figure 9. b3639 and Mrub_2047 both involve in CoaA biosynthesis, and they are assigned same enzyme commission number. Panel A=*E. coli* *coaBC*/b3639; Panel B=Mrub_2047. KEGG (<http://www.genome.jp/kegg/pathway.html>) created this pathway map.

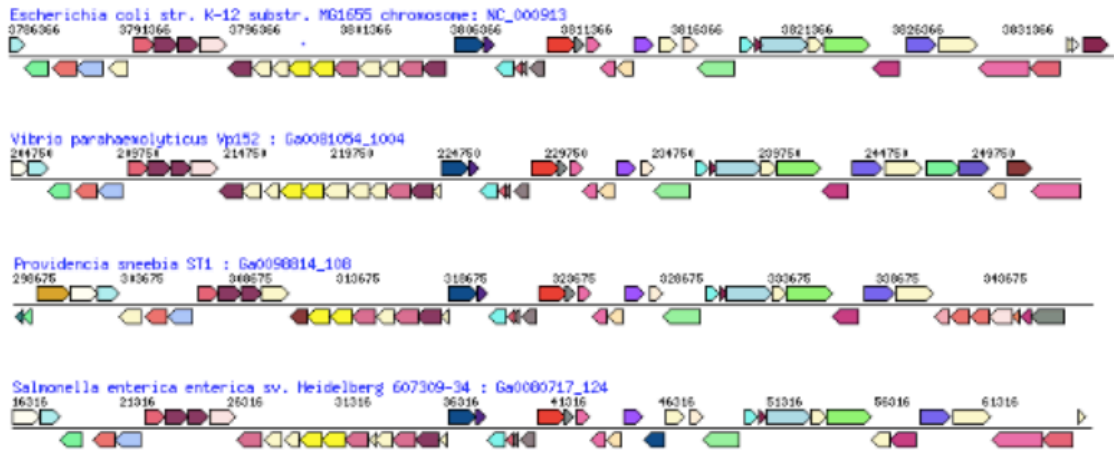
Horizontal gene transfer

E. coli b3639 (*coaBC*) is denoted by red color, and it is next to the gene *dut* (deoxyuridinetriphosphatase), facing the same direction. Orthologs of *coaBC* are seen in other organisms, and they are all positioned next to *dut* gene as well (Fig. 10. Panel A). However, *coaBC* and *dut* are not operon based on the gene context denoted by KEGG pathway (Fig 11. Panel A). Two different green colors for *coaBC* and *dut* indicates that *coaBC* and *dut* do not involve in the same metabolic pathway. Therefore, it is unlikely that *coaBC* and *dut*, which belong to two different metabolic pathways, would form an operon. The Average GC content of *coaBC* in *E. coli* is 54%, which is about 3% higher than the average GC content of its genome, suggesting that *coaBC* might possibly acquire from horizontal gene transfer (Table 1). Phylogenetic tree (Fig. 12. Panel A) shows that organisms containing most closely related b3639 (*coaBC*) are closed relatives of *E. coli* as they are from the same phylum as *E. coli* which is Proteobacteria. Therefore, it is unlikely that b3639 has recently undergone horizontal gene transfer.

Mrub_2047 (*coaBC*) is denoted by red color, and it is next to *coaX* in purple color (putative Baf family transcriptional activator), facing the same direction (Fig. 10. Panel B). Orthologs of *coaBC* are seen in other organisms, and they are all positioned next to *coaX* except for *Bacterium* sp. JAD2 whose *coaBC* is next to the gene codes for xanthine phosphoribosyltransferase (Fig. 10. Panel A). In addition, *coaX* and *coaBC* are involving in the same metabolic pathway as they are denoted by the same green color by KEGG (Fig. 11 Panel B). It is likely that Mrub_2047 and *coaX* works together as an operon in those species. The Average GC content of Mrub_2047 is 71%, which is 8% higher than the average GC content of its genome (63%), strongly suggesting that Mrub_2047 might be acquired from horizontal gene transfer as well (Table 1). If Mrub_2047 results from horizontal gene transfer (HGT), then that HGT might occur long time ago as the phylogenetic tree shows that organisms containing most closely related Mrub_2047 (*coaBC*) are not closed relatives as they are not in the same phylum of *M. ruber* which is Dienococcus-Thermus (Fig. 12. Panel B).

In conclusion, b3639 does not belong to an operon while Mrub_2047 form operon with *coaX*. It is unlikely that b3639 and Mrub_2047 have recently undergone horizontal gene transfer.

Panel A



Panel B



Figure 10. Both gene contexts of b3639 and Mrub_2047 are conserved among species. b3639 and *dut* gene order is conserved among several species; Mrub_2047 and *coaX* are conserved among several species as well. Panel A=*E. coli* *coaBC*/b3639; Panel B=Mrub_2047. IMG/ER (<https://img.jgi.doe.gov/cgi-bin/er/main.cgi>) created the image of gene context.

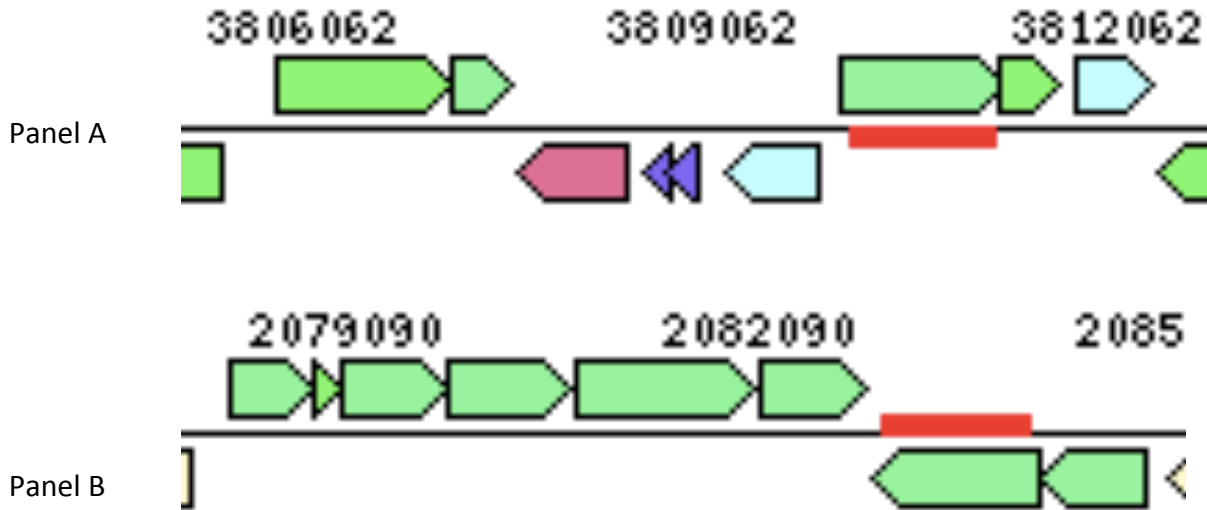
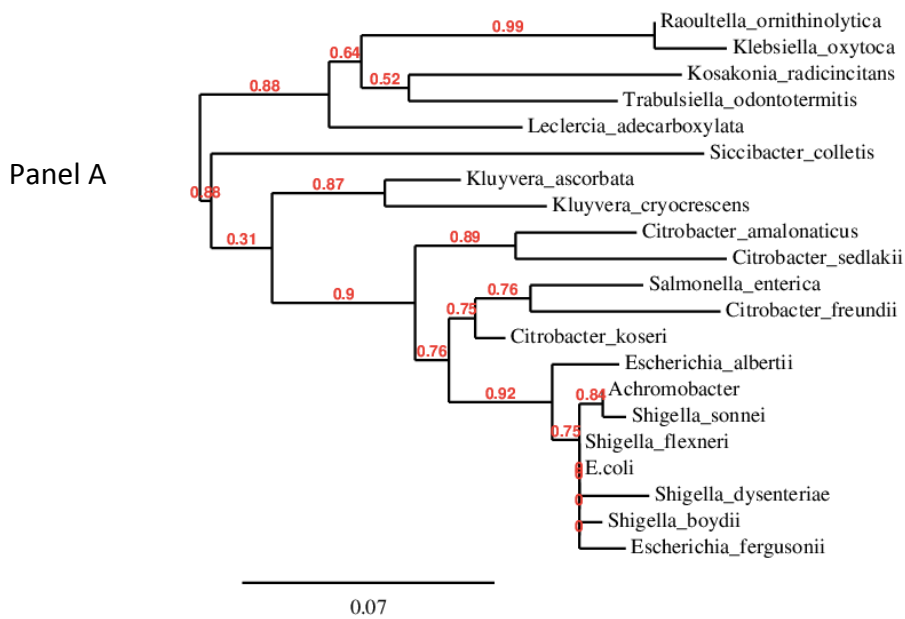


Figure 11. *b3639* and *dut* do not form an operon, and *Mrub_2047* and *coaX* are likely to form an operon. Panel A=*E. coli* *coaBC/b3639*; Panel B=*Mrub_2047*. IMG/ER (<https://img.jgi.doe.gov/cgi-bin/er/main.cgi>) created this image of gene context.



Panel B

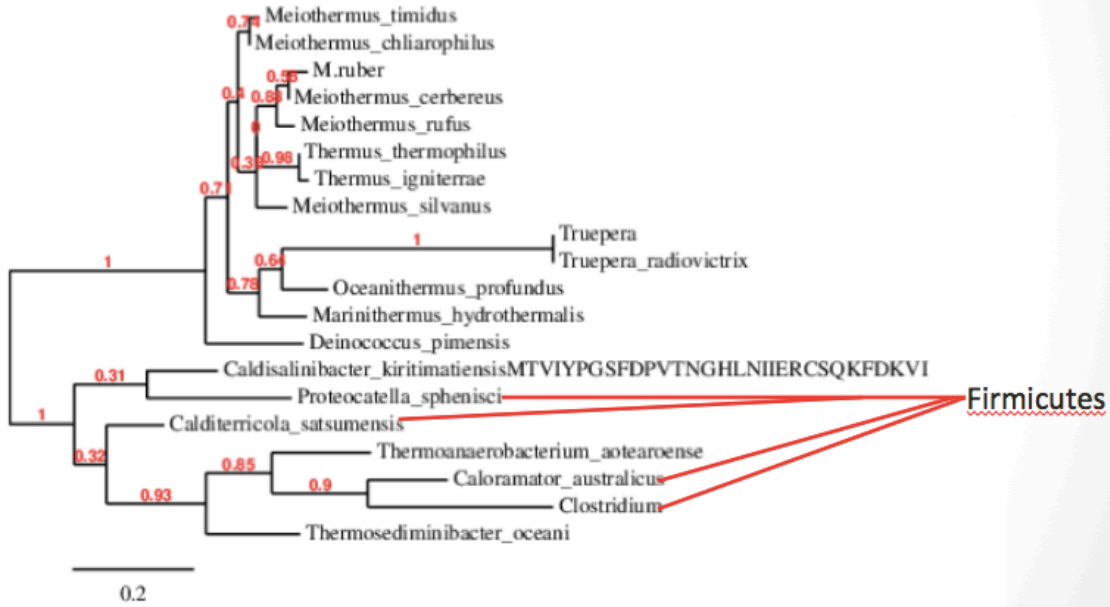


Figure 12. It is unlikely for both b3639 and Mrub_2047 have recently undergone horizontal gene transfer. Panel A=*E. coli* *coaBC*/b3639; Panel B=Mrub_2047. Phylogeny. fr (<http://www.phylogeny.fr>) created this phylogenetic tree.

B3634 and Mrub_1372

Table 2. shows the comparison between *E. coli* b3634 and Mrub_1372. The predicted location of both b3634 and Mrub_1372 is cytoplasm. BLAST search demonstrates the similarity in sequence between *E. coli* b3634 and Mrub_1372. Both *E. coli* b3634 and Mrub_1372 involve in pentothenate and CoA synthesis. There are structural similarities between *E. coli* b3634 and Mrub_1372 as they are in the same TIGRfam family and Pfam family. *E. coli* b3634 and Mrub_1372 are in the same COG category CoaBC. *E. coli* b3634 is also identified in COG category of CitC Citrate lyase synthetase. However CitC Citrate lyase synthetase hit in CDD is less significant than CcoaD as the E-value for CitC Citrate lyase synthetase is very large ($E=2.35e-05$). PDB pulls out the same the protein for both *E. coli* b3634 and Mrub_1372. GC content of b3634 is slightly higher than organism's genome, while Mrub_1372 has lower GC content than the average GC content in its genome.

Table 2. *E. coli* b3634 and Mrub_1372 are orthologs

Description of evidence collected	<i>E. coli</i> b3634	<i>M. ruber</i> Mrub_1372
Cellular localization	Cytoplasmic	Cytoplasmic
Blast <i>E. coli</i> against <i>M. ruber</i>	Score: 143 bits; E- value: 6e-44	
KEGG pathway	Pentothenate and CoA biosynthesis E. C. 2.7.7.3	
Pfam-protein family	PF01467 CTP transf like family (E=4.9e-23)	PF01467 CTP transf like family (E=7.9e-18)
CDD (COG category)	COG0669 (E=4.08e-101) coaD Phosphopantetheine adenylyltransferase COG3053 (E=2.35e-05) CitC Citrate lyase synthetase	COG0669 (E=6.35e-77) coaD Phosphopantetheine adenylyltransferase
TIGRfam-protein family	TIGR01510 (E=1.2e-84) coaD_prev_kdtB: pantetheine- phosphate adeny	TIGR01510 (E=4.2e-66) coaD_prev_kdtB: pantetheine- phosphate adeny
PDB	1B6T (E=8.93925E-91) Phosphpantetheine adenylyltransferase	1OD6 (E=5.55299E-63) Phosphpantetheine adenylyltransferase
GC content of the gene	53%	59%

Sequence based similarities

When using BLAST to search for b3634 protein sequence in *M. ruber* genome, Pentetheine-phosphate adenylyltrasferase was pulled out from the database, which is the gene product of Mrub_1372. Comparing b3634 and Mrub_1372, protein Blast produced the bit score of 143 (Fig. 13). Approximately half of the amino acids (43%) of b3634 and Mrub_1372 are identical with 2% of gap (Fig. 13). It is unlikely that similarities in amino acids between b3634 and Mrub_1372 is simply due to chance as the E value is relatively small (E=6e-44). Therefore, b3634 and Mrub_1372 have very similar amino acid sequence.

Score	Expect	Method	Identities	Positives	Gaps
143 bits(360)	6e-44	Compositional matrix adjust.	68/158(43%)	109/158(68%)	4/158(2%)
Query 6		IYPGTFDPITNGHIDIVTRATQMFHDHILAIASPSKKPM--FTLEERVALAQQA--TAH			61
Sbjct 4		+YPG+FDP+ NGH D++ RA++ F V +A+ +PSK+ + FT ERV + ++A +A			63
Query 62		LGNVEVVGFSDLMANFARNQHATVLRGLRAVADFEYEMQLAHMNRHLMPELESVFLMPS			121
Sbjct 64		L NVEV F+ L+A + + + V+++GLRAV+D+E E+Q+AH+NR E+ F+M +			123
Query 122		KEWSFISSSLVKEVARHQGDVTHFLPENVHQALMAKLA	159		
Sbjct 124		WSF+SS++VKE+AR+ GDV+ +P +AL KL+	161		

Figure 13. The amino acid sequences of b3634 and Mrub_1372 are highly similar. Blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) created this sequence alignment between b3634 and Mrub_1372.

Cellular localization

The predicated locations for b3634 and Mrub_1372 are cytoplasm. Based on TMHMM analysis, both b3634 and Mrub_1372 don not contain trans-membrane regions in their proteins, meaning that both gene products are unlikely to be found in cell membrane (Fig. 14). The signal peptide D score for b3634 is 0.124 while the cutoff score is 0.57. The signal peptide D score for Mrub_1372 is 0.098, which is also lower than the cutoff score 0.57 (Fig. 15). Signal peptide is the N-terminal sequence that will be cleaved when trans-membrane proteins are inserted to the cell membrane. Therefore, the absence of signal peptide is another evidence suggesting that b3634 and Mrub_1372 are not located on cell membrane as well. Phobius analysis predicts that both b3634 and Mrub_1372 are noncytoplasmic proteins, which is inconsistent with the results obtained from TMHMM and Signal P (Fig. 16). It is possible that both b3634 and Mrub_1372 have regions of hydrophobic amino acids, but not be membrane-bound. The preponderance of evidence indicates that the predicated location for b3634 and Mrub_1372 is cytoplasm, and in this case, it is not wise to interpret the results of Phobius as an accurate prediction.

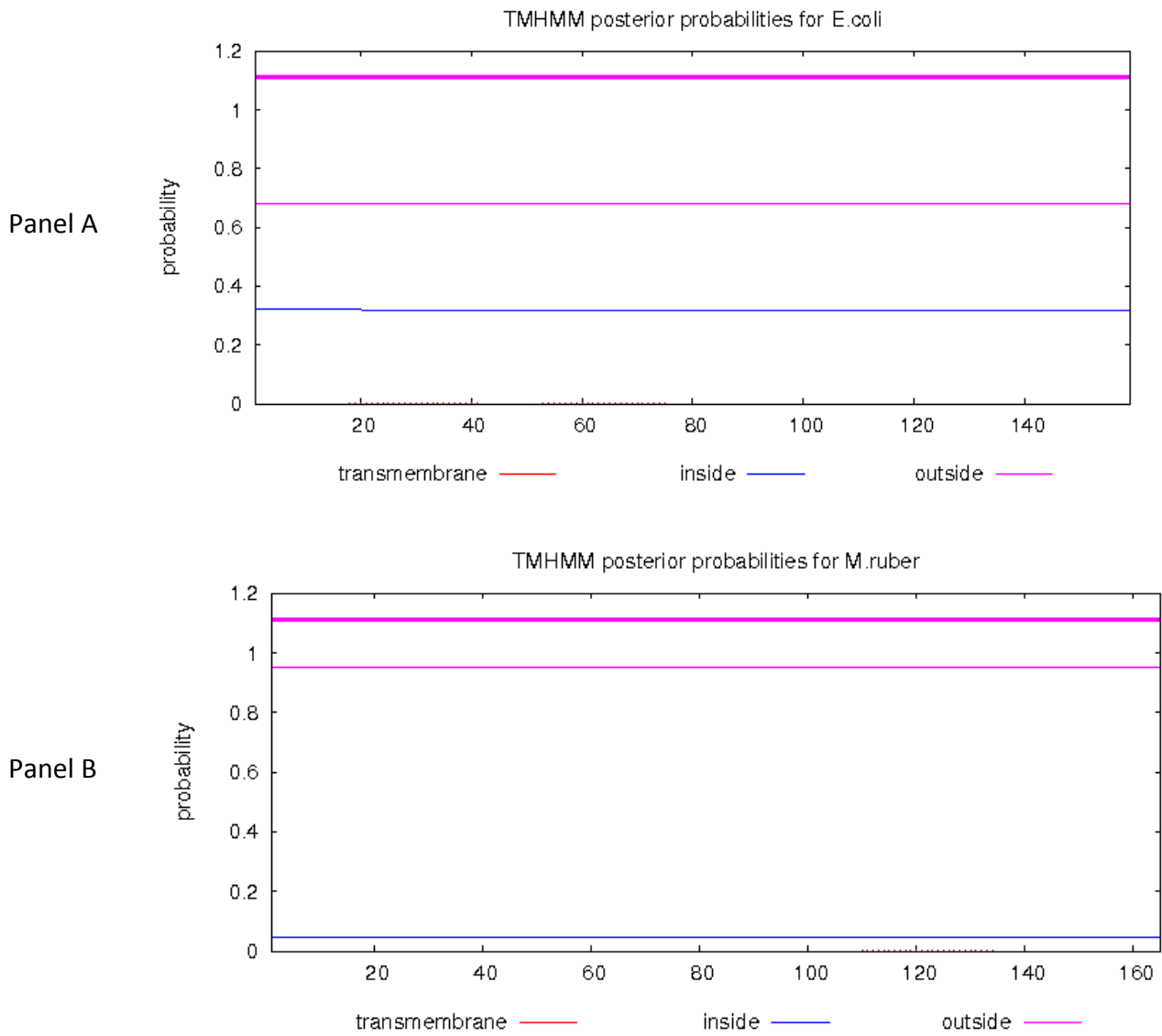


Figure 14. *E. coli coaD/b3634* and *Mrub_1372* do not contain TMH regions; a cytoplasmic location is predicted. Panel A=*E. coli coaD/b3634*; Panel B=*Mrub_1372*. TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM>) created this hydropathy plot.

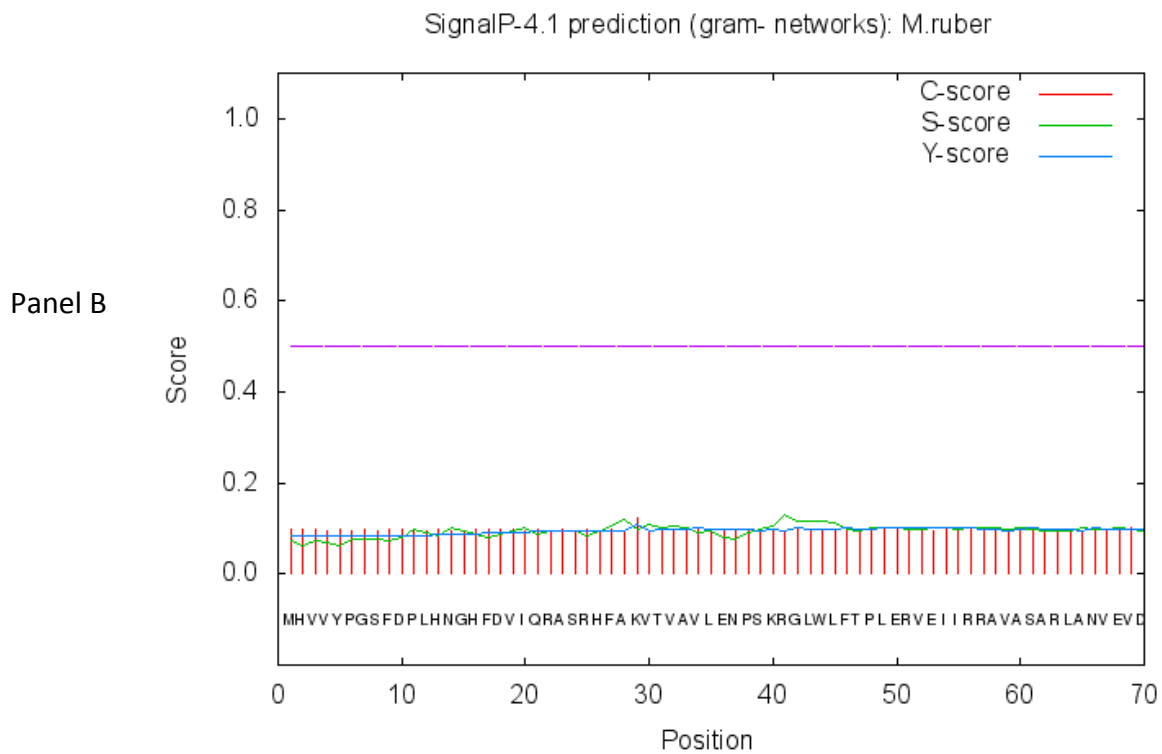
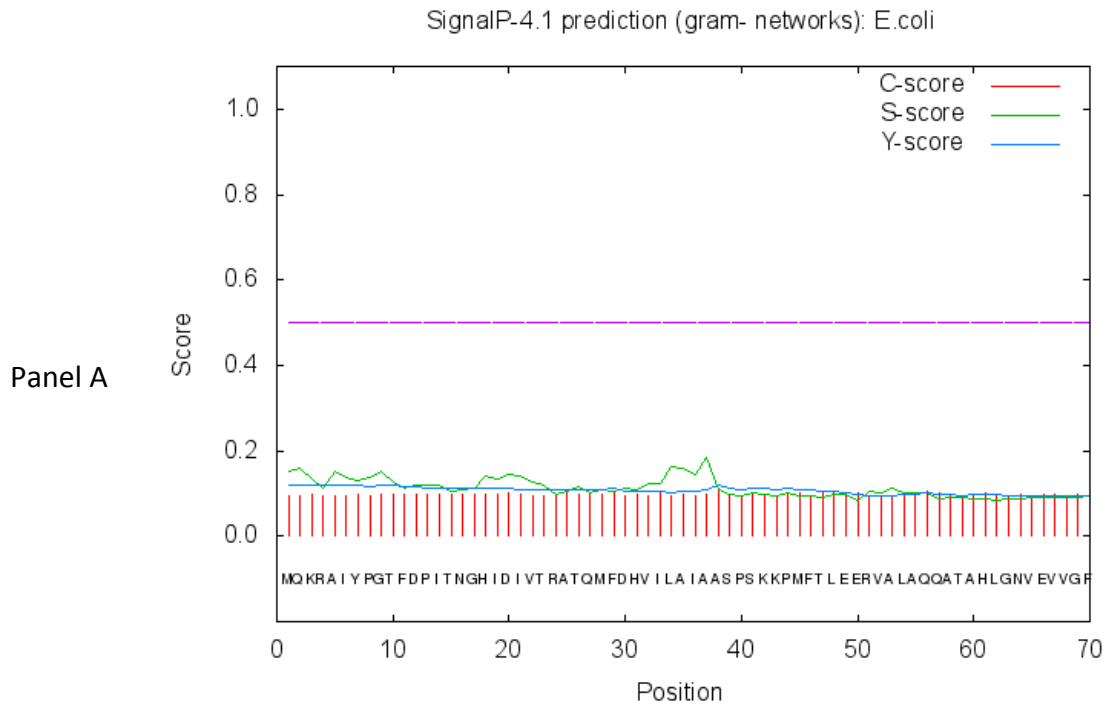
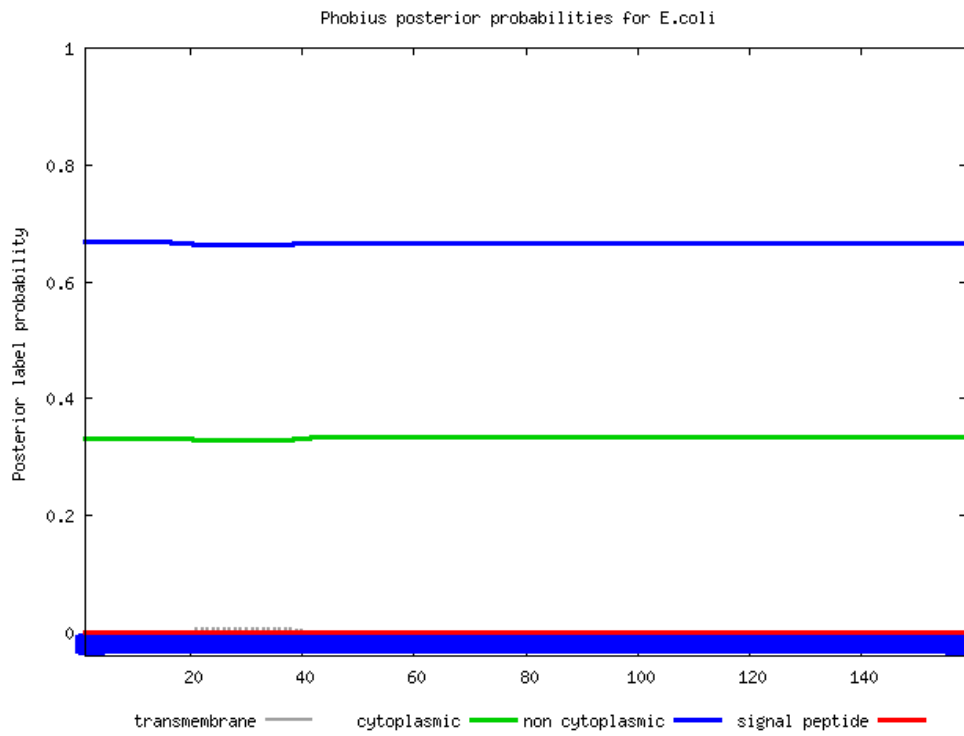


Figure 15. *E.coli* *coaD*/b3634 and Mrub_1372 do not contain signal peptides; a cytoplasmic location is predicted. Panel A=*E.coli* *coaD*/b3634; Mrub_1372. Signal IP (<http://www.cbs.dtu.dk/services/SignalP>) created this hydropathy plot.

Panel A



Panel B

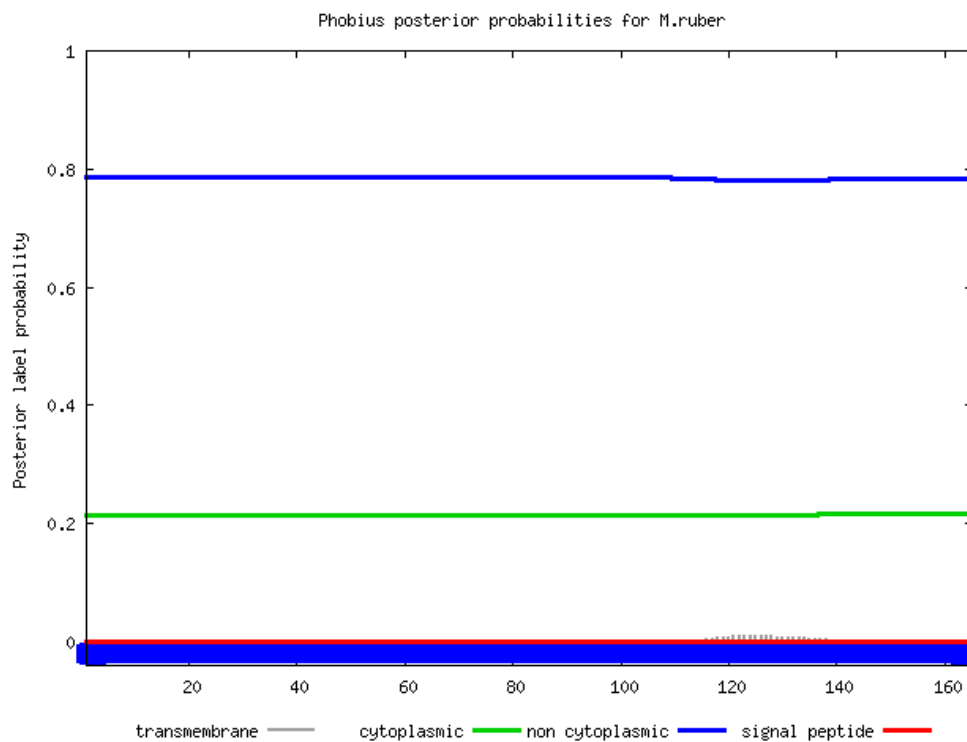


Figure 16. Both b3634 and Mrub_1372 is predicted to be noncytoplasmic. Panel A=*E.coli* coaD/b3634; Mrub_1372. Phobius (<http://phobius.sbc.su.se>) created this hydropathy plot.

Enzyme function

Figure 18 shows that both b3634 and Mrub_1372 involve in CoA biosynthesis. Both b3634 and Mrub_1372 are assigned with same enzyme commission number E.C. 2.7.7.3, which catalyzes the formation of intermediate Dephospho-CoA from 4'-phosphopantetheine. Each enzyme commission number is assigned to a specific biological reaction. Same enzyme commission number further confirms that b3634 and Mrub_1372 have similar function.

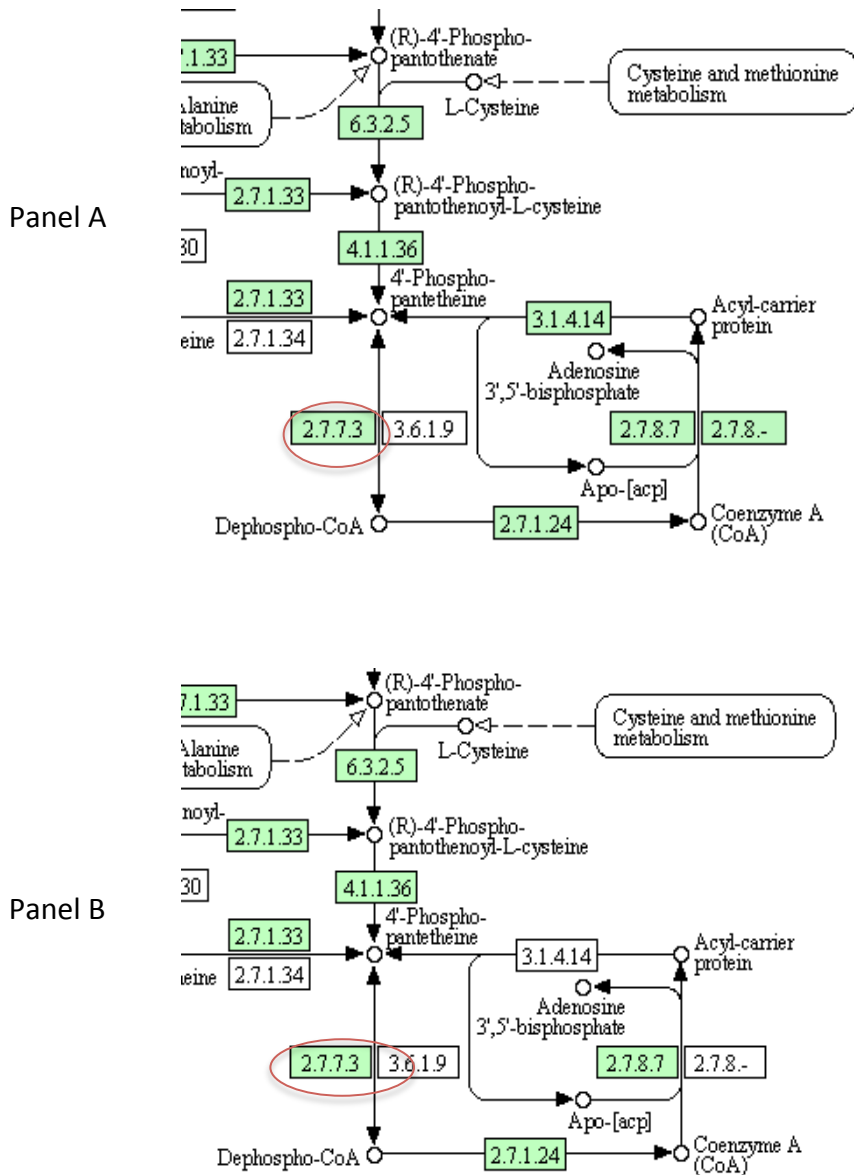


Figure 18. Both b3634 and Mrub_1372 involve in CoA biosynthesis. Both b3634 and Mrub_1372 are assigned with enzyme commission number E.C. 2.7.7.3. Panel A=*E.coli* coaD/b3634; Mrub_1372. KEGG (<http://www.genome.jp/kegg/pathway.html>) created this CoA metabolic pathway

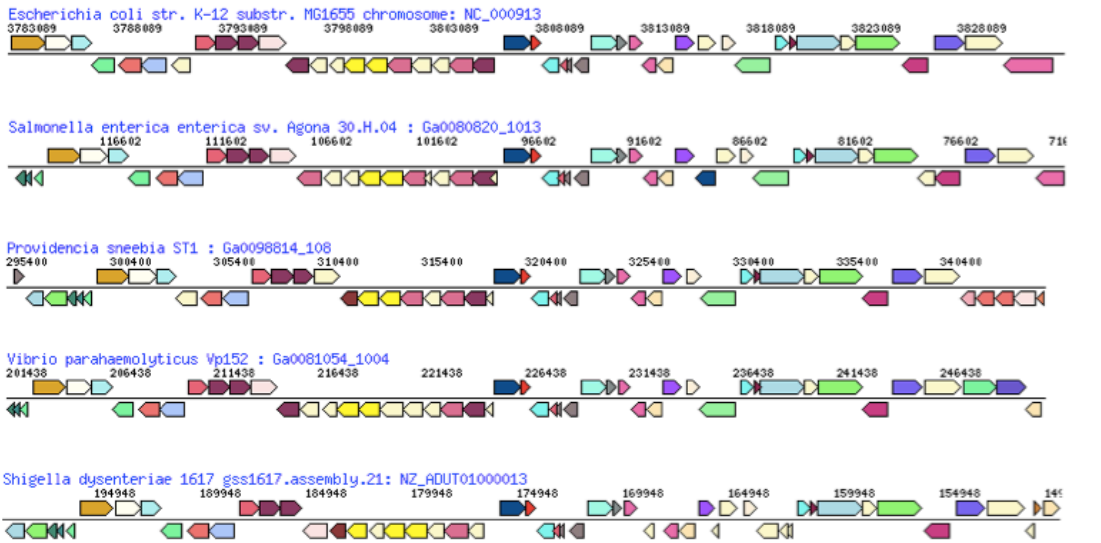
Evidence in gene context

E. coli b3634 (*coaD*) is denoted by red color, and it is next to gene *WaaA* (3-deoxy-D-manno-octulosonic-acid transferase), facing the same direction. Orthologs of *coaD* are seen in other organisms, and they are all positioned next to *WaaA* as well (Fig. 19. Panel A). According to KEGG, *WaaA* and *coaD* do not involve in the same metabolic pathway as they are denoted by different green colors (Fig. 20. Panel A). It is unlikely that *coaD* and *WaaA* form an operon. The Average GC content of b3634 in *E. coli* is 53%, which is about 2% higher than the average GC content of its genome, suggesting that there is possibility that b3634 might be acquired from horizontal gene transfer as well (Table 2). Phylogenetic tree shows that organisms containing most closely related b3634 (*coaD*) are closed relatives of *E. coli* and from the same phylum as *E. coli* which is Proteobacteria (Fig. 21 Panel A). Therefore, it is unlikely that b3634 has recently undergone horizontal gene transfer.

Mrub_1372 (*coaD*) is denoted by red color, and it is next to Mrub_1373 (delta-1-piperidine-6-carboxylate dehydrogenase) in brown color, facing the same direction (Fig. 19. Panel B). Orthologs of Mrub_1372 are seen in other organisms, and most of them are positioned next to Mrub_1373 (Fig. 19. Panel B). Although same gene order is conserved in most species, Mrub_1373 does not involve in the same metabolic pathway as Mrub_1372 as they are denoted by different KEGG colors (Fig. 20. Panel B). It is unlikely that Mrub_1372 and Mrub_1373 work together as an operon in those species. The Average GC content of Mrub_1372 is 59%, which is 4% lower than the average GC content of its genome (63%), strongly suggesting that Mrub_1372 might be acquired from horizontal gene transfer as well (Table 2.). Phylogenetic tree shows that some of species containing most closely related Mrub_1372 (*coaD*) are not closed relatives of *M. ruber* as they are not in the same phylum, but they are distant away from *M. ruber* (Fig. 20. Panel B). Although horizontal gene transfer might occur, it must occur long time ago.

In conclusion, b3634 and Mrub_1372 do not form operon with its neighbor genes. It is unlikely that b3634 and Mrub_1372 have recently undergone horizontal gene transfer.

Panel A



Panel B

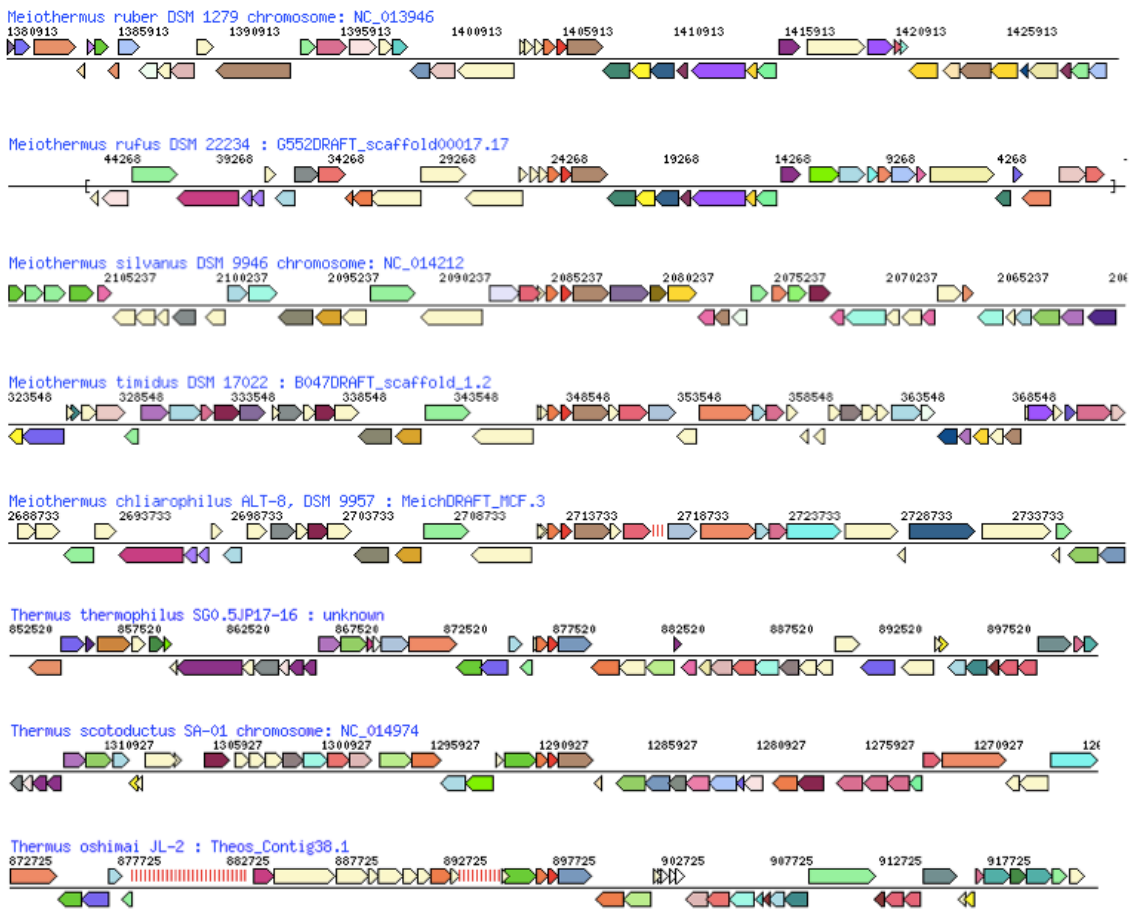


Figure 19. Gene contexts of b3634 and Mrub_1372 are conserved among several species. Panel A=*E. coli* *coaD*/b3634; Panel B=Mrub_1372. IMG/ER (<https://img.jgi.doe.gov/cgi-bin/er/main.cgi>) created the image of gene in context.

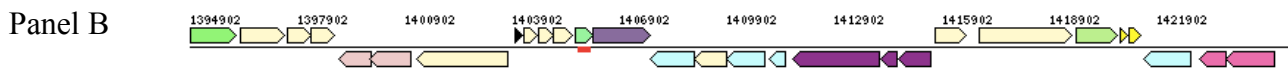
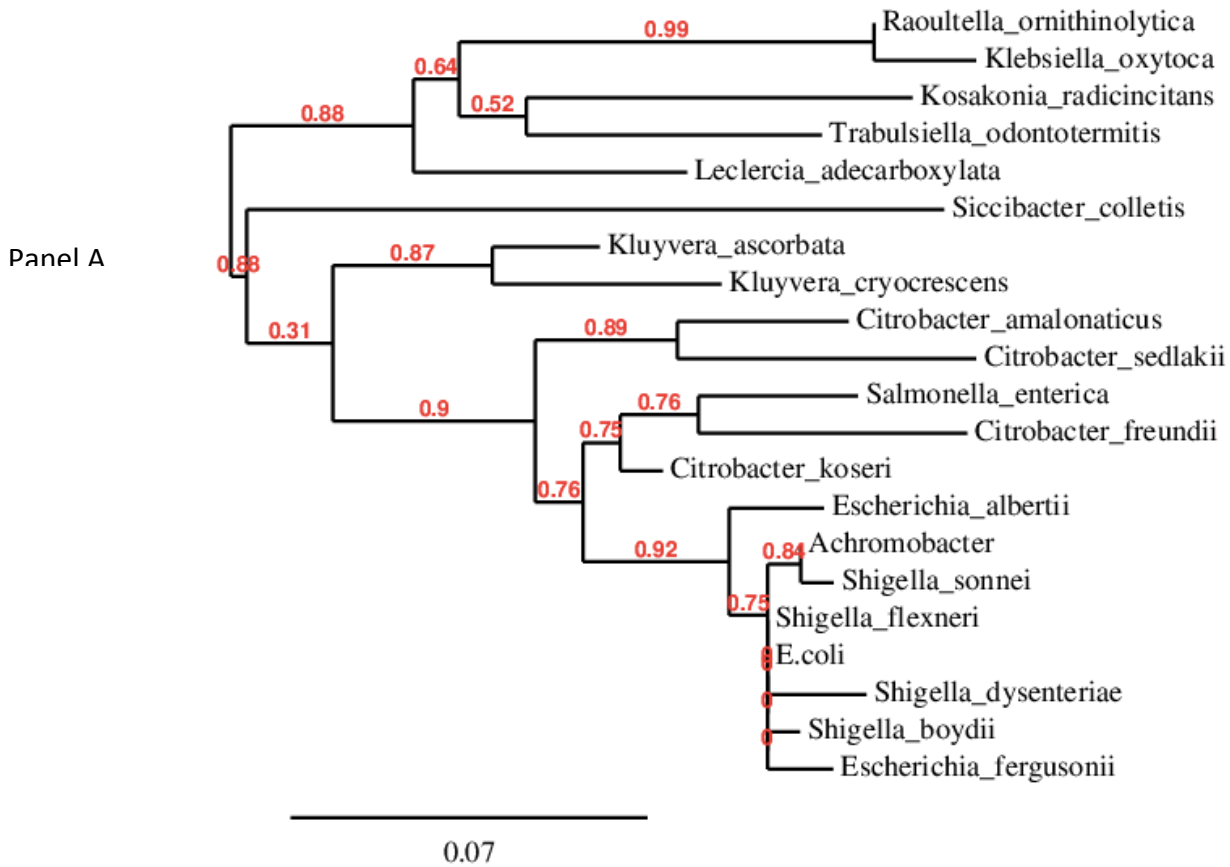


Figure 20. b3634 and WaaA do not form an operon, and Mrub_1372 does not belong to an operon. Panel A=*E. coli coaD/b3634*; Panel B=*Mrub_1372*. IMG/ER (<https://img.jgi.doe.gov/cgi-bin/er/main.cgi>) created this image of gene context.



Panel B

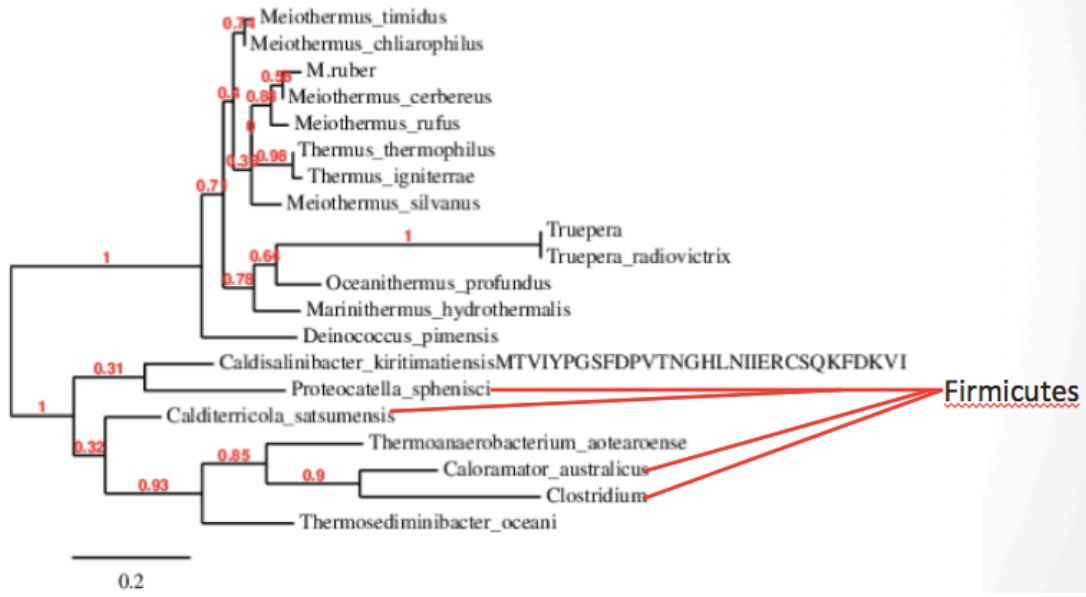


Figure 21. It is unlikely for both b3634 and Mrub_1372 have recently undergone horizontal gene transfer. Panel A=*E. coli coaD*/b3634; Panel B=Mrub_1372. Phylogeny. fr (<http://www.phylogeny.fr>) created this phylogenetic tree

Conclusion

The results of all bioinformatics analysis support the hypothesis that b3639 and Mrub_2047 are orthologs. BLAST result shows that amino acids sequence of b3639 and Mrub_2047 are similar, and the low E value suggests that the similarity in sequence is due to evolutionary relatedness not by chance. TMHMM, Sgnal P consistently predict that both b3639 and Mrub_2047 do not contain transmembrane regions and signal peptides. Although Phobius predicts that b3639 and Mrub_2047 are noncytoplasmic, it is not reliable in this case. Structural analysis shows that b3639 and Mrub_2047 belongs to same protein family as their predicted pfam, TIGRfam and CDD have same number and names. KEGG map suggest that b3639 and Mrub_2047 involve in pentothenate and coA synthesis. Horizontal gene transfer data shows it is unlikely that both b3639 and have recently undergone horizontal gene transfer (HGT). Gene context data suggests that Mrub_2047 belongs to an operon while b3639 is not a part of an operon. All in all, *E. coli* b3639 and Mrub_2047 are similar to each other both structurally and functionally; therefore, they are orthologs.

The results of all bioinformatics analysis also support the hypothesis that b3634 and Mrub_1327 are orthologs. BLAST result shows that the amino acids sequence of b3634 and Mrub_1372 are similar, and the low E value suggests that the similarities is due to evolutionary relatedness not by chance. TMHMM, Sgnal P consistently predicte that both b3633 and Mrub_1372 do not contain transmembrane regions and signal peptides. Although Phobius predicts that b3639 and Mrub_2047 are noncytoplasmic, it is not reliable in this case. Structural analysis shows that b3634 and Mrub_1372 belong to same protein family as their predicted pfam, TIGRfam and CDD have same number and names. KEGG map suggests that b3634 and Mrub_1372 involve in pentothenate and coA synthesis. Horizontal gene transfer data shows that it is unlikely that both b3634 and Mrub_21372 have recently undergone horizontal gene transfer (HGT). Gene context data suggests that neither b3634 nor Mrub_1372 are a part of an operon. All in all, *E. coli* b3634 and Mrub_1372 are similar to each other both structurally and functionally; therefore, they are orthologs.

Literature Cited

- Geerlof A, Lewendon A, Shaw WV. 1999. Purification and characterization of phosphopantetheine adenylyltransferase from *Escherichia coli*. *J Biol Chem* 274(38):27105-11.
- Kupke T, Uebele M, Schmid D, Jung G, Blaesche M, Steinbacher S. 2000. Molecular characterization of lantibiotic-synthesizing enzyme EpiD reveals a function for bacterial dfp proteins in coenzyme A biosynthesis. *J Biol Chem* 275(41):31838-46.
- Kupke T. 2001. Molecular characterization of the 4'-phosphopantothenoylcysteine decarboxylase domain of bacterial dfp flavoproteins. *J Biol Chem* 276(29):27597-604.
- Kupke T. 2002. Molecular characterization of the 4'-phosphopantothenoylcysteine synthetase domain of bacterial dfp flavoproteins. *J Biol Chem* 277(39):36137-45.
- Spitzer ED and Weiss B. 1985. Dfp gene of *Escherichia coli* K-12, a locus affecting DNA synthesis, codes for a flavoprotein. *J Bacteriol* 164(3):994-1003.
- Strauss E, Kinsland C, Ge Y, McLafferty FW, Begley TP. 2001. Phosphopantothenoylcysteine synthetase from *Escherichia coli* identification and characterization of the last unidentified coenzyme A biosynthetic enzyme in bacteria. *J Biol Chem* 276(17):13513-6.