

2-2016

Mrub_2765 is the version of *E. coli* FabZ in *Meiothermus ruber*, while Mrub_0266 is the version of *E. coli* FabI in *Meiothermus ruber*

Amanda M. Narkis
Augustana College, Rock Island Illinois

Dr. Lori Scott
Augustana College - Rock Island

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

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

Recommended Citation

Narkis, Amanda M. and Scott, Dr. Lori. "Mrub_2765 is the version of *E. coli* FabZ in *Meiothermus ruber*, while Mrub_0266 is the version of *E. coli* FabI in *Meiothermus ruber*" (2016). *Meiothermus ruber Genome Analysis Project*.
<http://digitalcommons.augustana.edu/biolmruber/16>

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.

Amanda Narkis

Department of Biology, Augustana College, Rock Island, IL, 61201

Project Title: Mrub_2765 is orthologous to *E. coli fabZ*, while Mrub_0266 is orthologous to *E. coli fabI* in *Meiothermus ruber*,

Introduction:

Meiothermus ruber (*M. ruber*) is a Gram-negative, nonspore-forming, red-pigmented, nonmotile, rod-shaped eubacteria that grows between 35-70°C under highly aerated conditions. *M. ruber* belongs to the bacterial phylum Deinococcus-Thermus which consists of organisms that typically live in high temperature environments otherwise known as thermophiles (Tindall et al. 2010). Not much is known about species in this phylum, but genome sequencing allows for better understanding of microorganisms, such as *M. ruber*, and the role they play in important processes. The GEBA project is an attempt to further our knowledge by sequencing genomes of diverse organisms to further fill in holes in the tree of life (Phylogenetic Diversity 2013). Overall, the GEBA project allows for the confirmation of identification through annotation and functional genomic analysis of all 3000+ genes predicted in the *M. ruber* genome as well as an authentic research experience for both students and faculty (Geni-science 2016).

Bioinformatics tools are used as sources to look at previously sequenced genomes of various organisms, allow for the study of diseases in humans, and are great assets for research. This project used GENI-ACT as a platform for collecting data. All collected data and observations were placed in a lab notebook on the GENI-ACT site. In using the different bioinformatics tools, *E. coli* was used as a positive control because of how well it is studied. This specific study looked at the genes for the enzymes involved in fatty acid biosynthesis which is catalyzed by the type II fatty acid synthase system where each step is carried out by an individual enzyme. The type II systems are found in the cytoplasm in most bacteria and plants where there is four chemical reactions needed to complete each round of fatty acid elongation (Heath RJ and Rock CO 1996).

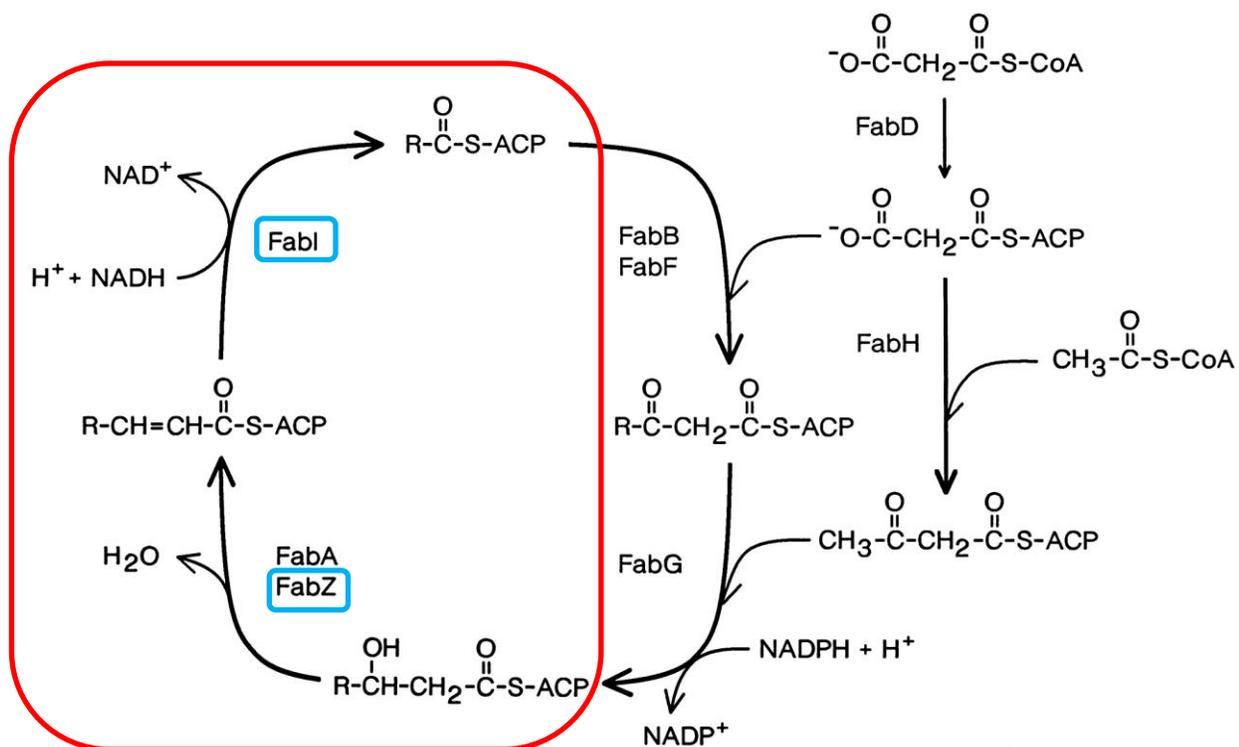


Figure 1. There are four reactions in each cycle of fatty acid elongation. The first cycle initiates with B-ketoacyl-carrier-protein (ACP) synthase II (FabH), which condenses malonyl-ACP with acetyl-coA. The following cycles begin with the condensation of malonyl-ACP with acyl-ACP catalyzed by B-ketoacyl-ACP synthases I and II (Fab B and FabF). B-ketoester is reduced by a single NADPH-dependent B-ketoacyl-ACP reductase (Fab G). This paper focuses on both the third and fourth step of the cycle. The third step is catalyzed by either the FabA or FabZ beta-hydroxyacyl-ACP dehydratase. In this study, the focus is on the enzyme FabZ. The final and fourth step in each cycle is catalyzed by a single NADH-dependent enoyl-ACP reductase (FabI) that converts *trans*-2-enoyl-ACP to acyl-ACP (Heath RJ and Rock CO 1996).

It is important to look at fatty acid biosynthesis because fatty acids are major sources of energy and are precursors for essential substances in our body both structurally and metabolically. For example, phospholipids are essential in all cell membranes and plasma lipoproteins. Also, defects in fatty acid metabolism can lead to different diseases such as defects in mitochondrial acyl CoA dehydrogenases which can prevent normal fatty acid oxidation (Bergler et al 1996).

Each step in the type II fatty acid synthase system is carried out by an individual enzyme. This paper focuses on the enzymes FabZ and FabI. FabZ is the enzyme responsible for the dehydration step of the dissociated (type II) fatty-acid biosynthesis system. Different forms of FabZ have preferences for substrates with different chain lengths. Gram-negative bacteria, such as *E. coli*, that produce unsaturated fatty acids have another form of the enzyme, Fab A, which prefers intermediate chain length. In the end, both forms can catalyze all steps leading to the

synthesis of palmitate. FabZ is unique because it can also accept unsaturated substrates which FabA cannot (Heath RJ and Rock CO 1996). FabI is the enzyme that catalyzes the reduction of the 2, 3 double bond in enoyl-acyl-[ACP] derivatives of the elongating fatty acid moiety (Figure 1). FabI will accept carbon chains with varying lengths depending on the organism it is in (Bergler et al 1996).

It appears that reduced activity of FabZ selectively increases the levels of intermediates in the saturated branch of the pathway. Every enzyme in the fatty acid biosynthesis pathway is very important in order to create fatty acids (Heath RJ and Rock CO 1996). It has also been previously found that the reduction of enoyl-ACP derivatives, such as FabI, is a key regulatory step in endogenous fatty acid biosynthesis. FabI and other derivatives are possibly involved in regulating the ration of saturated and unsaturated fatty acids and in the coordination of fatty acid and phospholipid syntheses (Bergler et al 1996).

Since *E. coli* is our positive control, the *fabZ* and *fabI* genes, which respectively encode for B-hydroxyacyl-acyl carrier protein (ACP) dehydratase and enoyl-[ACP] reductase, were compared to the possible *fabZ* and *fabI* *M. ruber* genes known as Mrub_2765 and Mrub_0266. It has been hypothesized that Mrub_2765 is orthologous to *E. coli* *fabZ* (b0180) and Mrub_0266 is orthologous to *E. coli* *fabI* (b1288).

Methods:

The data in this project were obtained through the bioinformatics programs within the GENI-ACT lab notebook and the results were also stored in the notebook. A listing and short description of all the programs used in GENI-ACT are provided here (<http://geni-act.org/>). There were some deviations in the listed protocols. For Tcoffee, the top ten to thirteen organisms were used excluding *E. coli* within the search. The BLAST results included in this paper were obtained by taking the *E. coli* sequence and blasting it against the *M. ruber* genome. The

pairwise alignment was saved in a separate document not included in the GENI-ACT lab notebook. Besides these listed deviations, all instructions were followed as listed on GENI-ACT.

Results:

Mrub_2765 compared against E. coli b0180 (FabZ)

Although there are some differences between genes, there are many similarities and results that lineup and lead to the assumption that *Mrub_2765* is orthologous to *E. coli fabZ*.

Table 1 displays many different characteristics that the two genes share. Part of the table displays E-values which are considered significant when they are 1e-3 or smaller. This significance means that the sequences that are being compared are highly similar rather than comparable by chance. All the E-values that were important and applicable are recorded in Table 1. A protein BLAST comparison between *E. coli* b0180 and *Mrub_2765* produced an E-value of 4e-49 and a bit score of 145 (Table 1). This BLAST result shows that b0180 and *Mrub_2765* are similar in sequence.

Table 1: *E. coli fabZ* and *Mrub_2765* are orthologs.

Description of evidence collected	<i>E. coli</i>	<i>M. ruber</i>
BLAST <i>E. coli</i> against <i>M. ruber</i>	Score: 145 bits ; E-value: 4e-49	
KEGG pathway	Biotin metabolism specifically involved in Fatty acid biosynthesis	
Cellular localization (lipop, P-sortB, Phobius)	cytoplasmic	
TIGRfam-protein family	TIGR01750 Beta-hydroxyl-ACP dehydratase	
	E=3.3e-88	E=2e-71
Pfam-protein family	PF07977 FabA Domain:HotDog	
	E=6.7e-40	E=3.1e-28
PDB	1U1Z -(3R)-hydroxyacyl-ACP dehydratase (FabZ)	
	E=2.61e-41	E=8.2887e-32
E.C. number	E.C.4.2.1.59	

	3-hydroxy-[acyl-carrier-protein] dehydratase	
Gene context	Part of an Operon	Not Part of an Operon
CDD (Cog category)	COG0764	
	3-hydroxymyristoyl/3-hydroxydecanoyl-(ACP) dehydratase	
	E=2.95e-68	E=7.02e-55
Gram Stain	Negative	

In general, both Mrub_2765 and b0180 are gram negative, part of the biotin pathway and specifically involved in fatty acid biosynthesis (Figure 2). The cropped figure shows the fatty acid biosynthesis pathway within biotin metabolism. The E.C. number for both genes is 4.2.1.59 which stands for 3-hydroxy-[ACP] dehydratase, the enzyme that catalyzes the dehydration of short chain B-hydroxyacyl-ACPs and long chain saturated and unsaturated B-hydroxyacyl-ACPs (Table 1). For both *E. coli* and *M. ruber*, the KEGG pathway shows that FabZ is in the same location in the fatty acid biosynthesis pathway.

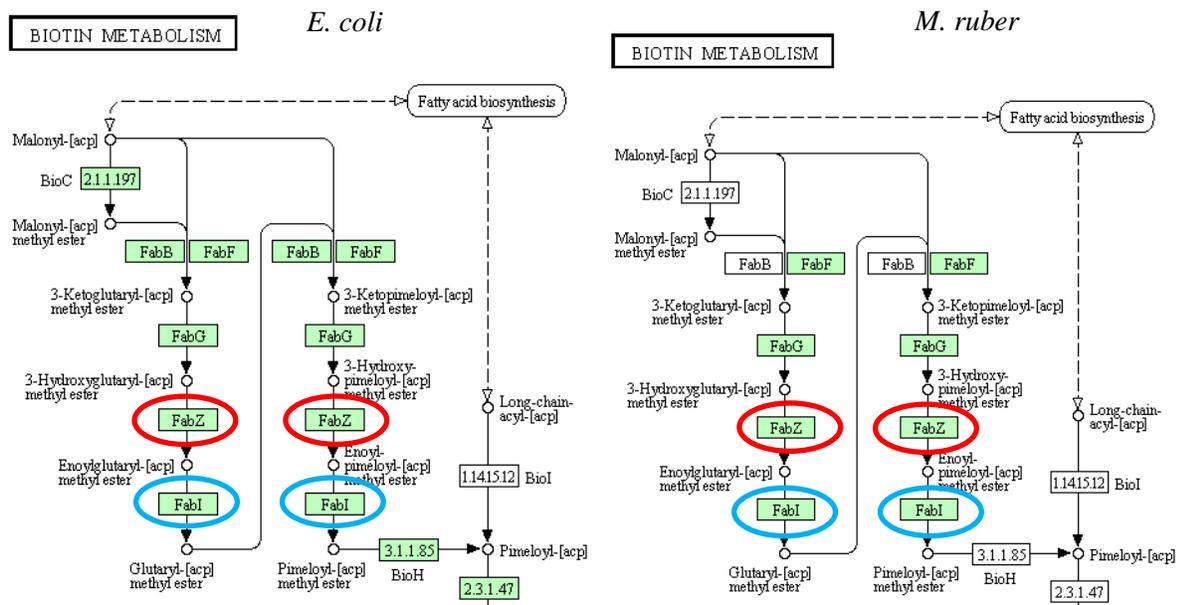


Figure 2. For both Mrub_2765 and *E. coli* FabZ, the location of the genes in the fatty acid biosynthesis pathway is the same as well as Mrub_0266 and *E. coli* FabI. The fabZ genes circled in red both code for 3-hydroxy-[ACP] dehydratase and the fabI genes circled in blue both code for enoyl- [ACP] reductase. This pathway was taken from the KEGG PATHWAY database (<http://www.genome.jp/kegg/pathway.html>).

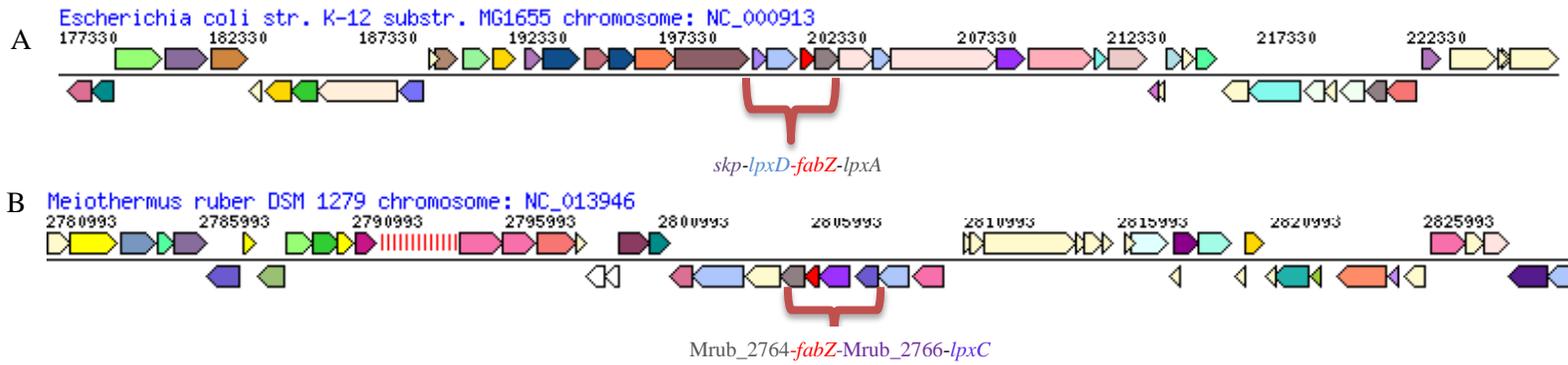


Figure 3. *E. coli fabZ* is part of an operon, while *Mrub_2765* is not. Panel A=*E. coli FabZ/b0180*; Panel B=*Mrub_2765*. The gene of interest (*FabZ*) is in red. *E. coli fabZ* has surrounding genes going to the right, while *Mrub_2765* has surrounding genes going to the left. The surrounding genes are marked by name. *Skp-lpxD-FabZ-lpxA* forms the genes to make up a transcription factor unit, while the genes in *M. ruber* do not make up an operon. IMG/ER (<https://img.jgi.doe.gov/cgi-bin/er/main.cgi>) created this image.

It appears that b0180 (*E. coli fabZ*) is part of an operon, while *Mrub_2765* is not (Figure 3). The gene of interest is in red in both maps. In panel A, *E. coli fabZ* has genes going in the same direction to the left and right of it, while panel B shows the same for *Mrub_2765* but all the genes are going to the left. With further research, it was discovered that *fabZ* is part of multiple operons including the transcription factor unit *hlpA-lpxD-FabZ-lapxA* which is marked in the figure below (Escherichia coli 2016). The genes match up with the name of the transcription unit besides *hlpA* being the promoter, but *skp* is the gene. *Skp* is the gene coding for the periplasmic chaperone for outer membrane proteins, *lpxD* is the gene coding for UDP-3-O[3-hydroxymyristoyl] glucosamine N-acyl transferase, *fabZ*, as mentioned previously, is the gene coding for 3-hydroxyacyl-[ACP] dehydratase, and *lpxA* is the gene for acyl-[ACP]—UDP-N-acetylglucosamine O-acyltransferase. *Mrub_2765* on the other hand has the set of genes *Mrub_2764-FabZ-Mrub_2766-lpxC*. *Mrub_2764* is the gene that codes for acyl-[ACP]—UDP-N-acetylglucosamine O-acyltransferase, *fabZ* codes for B-hydroxyacyl-[ACP] dehydratase, *Mrub_2766* is the gene that codes for an oxidoreductase domain-containing protein, and *lpxC* is the gene that codes for UDP-3-O-[3-hydroxymyristol] N –acetylglucosamine deacetylase. It appears that *Mrub_2765* in *M. ruber* does not form an operon.

Many of the bioinformatics programs used lead to both *E. coli* b0180 and Mrub_2765 being the sequences for Beta-hydroxyl-ACP dehydratase. The TIGRfam results for both genes were TIGR01750. Mrub_2765 produced an E-value of $2e-71$ and b0180 produced an E-value of $3.3e-88$. On the other hand, the Pfam results showed that the genes belonged to PF07977 (FabA) and both genes also belonged to the HotDog domain. Mrub_2765 produced an E-value of $3.1e-28$ and b0180 produced an E-value of $6.7e-40$. The protein database shows that both sequences had the PDB code of 1U1Z which stands for (3R)-hydroxyacyl-ACP dehydratase (FabZ). Mrub_2765 produced an E-value of $8.2887e-32$ and b0180 produced an E-value of $2.61e-41$. The Conserved Domain Database (CDD) search identified COG0764 (3-hydroxymyristoyl/3-hydroxydecanoyl-(ACP) dehydratase) to be the most significant COG number for both genes. Mrub_2765 produced an E-value of $7.02e-55$ and b0180 produced an E-value of $2.95e-68$.

The programs used for cellular localization such as lipoP, THMM, and P-sortB, show that both Mrub_2765 and b0180 are cytoplasmic. P-SortB had scores of 10.0 and 9.97 for the cytoplasmic scores for b0180 and Mrub_2765 indicating both are located in the cytoplasm (Table 1). Figure 4 shows no TMH regions in the hydropathy plot for b0180 (Panel A) and Mrub_2765 (Panel B) meaning there is no evidence for transmembrane helices.

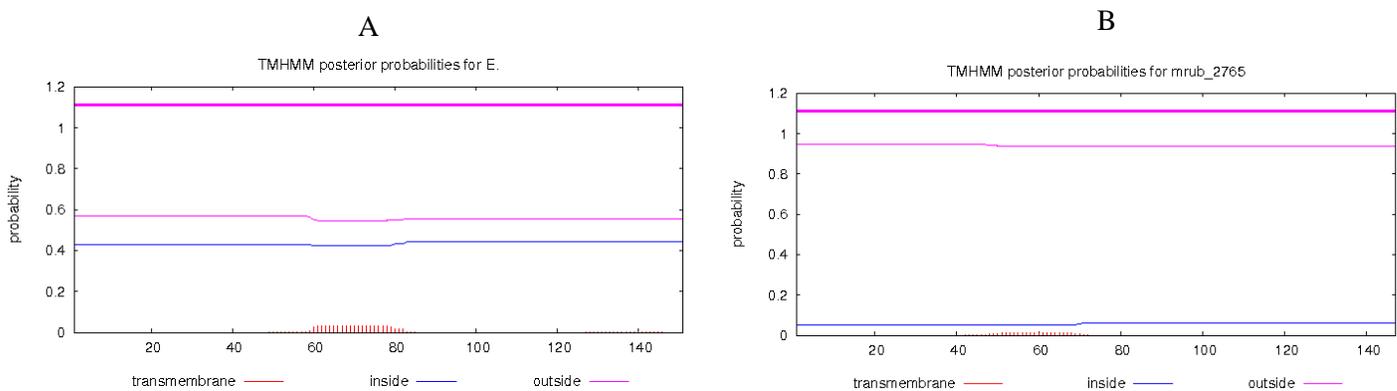


Figure 4. Mrub_2765 and *E. coli* FabZ do not contain TMH regions; a cytoplasmic location is predicted. Panel A=*E. coli* fabZ/b0180; Panel B=M_rub2765. TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM>) created this hydropathy plot.

SignalP results showed to have no cleavage sites or predicted signal peptides in both plots for Mrub_2765 and b0180 (Figure 5). Phobius results showed to have some regions of hydrophobicity indicated by vertical black lines, but these regions do not go out of the cytoplasm indicating the cellular localization to be in the cytoplasm (Figure 6). All of this data shows there is no evidence for transmembrane helices which further indicates a cytoplasmic location.

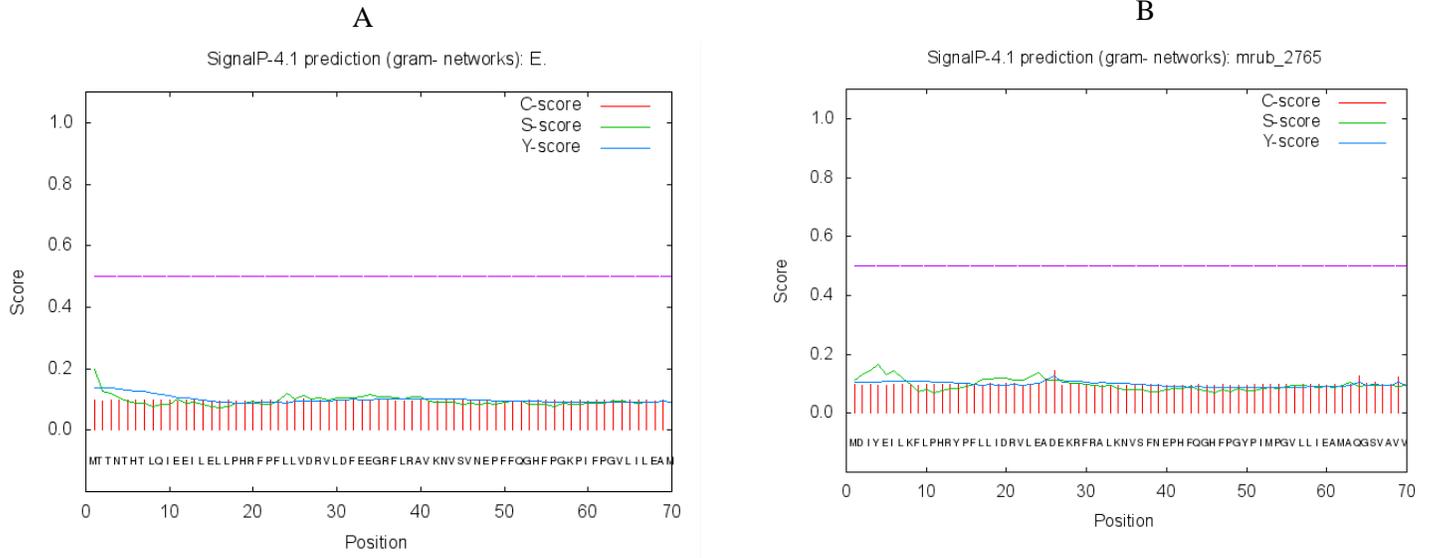


Figure 5. Mrub_2765 and *E. coli* FabZ do not have any cleavage sites and are predicted to have no signal peptides. This indicates there are no transmembrane helices present in either gene which indicates a cytoplasmic location. Panel A=*E. coli* FabZ/b0180; Panel B=M_rub2765. SignalP 4.1 server (<http://www.cbs.dtu.dk/services/SignalP>) created these plots.

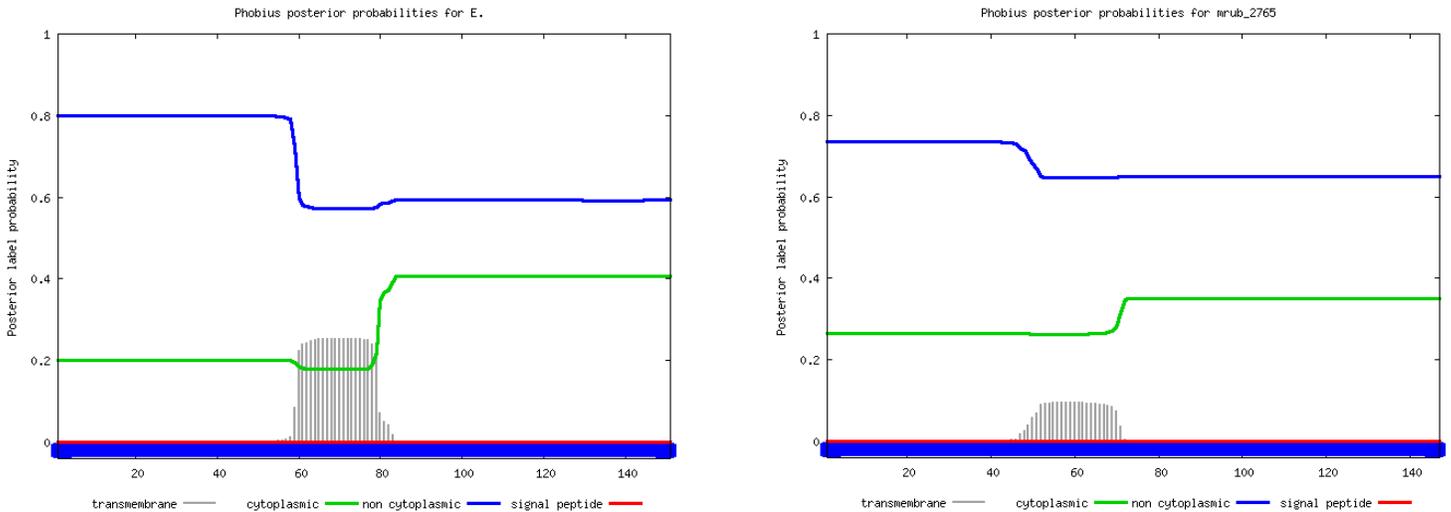


Figure 6. Mrub_2765 and *E. coli* FabZ do not have any transmembrane helices. There are regions of hydrophobicity indicated by the black vertical lines, but they do not go out of the cytoplasmic region. Since there are no transmembrane helices present in either gene, this indicates a cytoplasmic location. Panel A=*E. coli* FabZ/b0180; Panel B=M_rub2765. Phobius (<http://phobius.sbc.su.se/>) created these plots.

A

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value
				Start	End	Start	End	From	To			
FabA	FabA-like domain	Domain	CL0050	17	142	17	141	1	137	138	135.8	6.7e-40
#HMM	lpark.l111drvleieeegkfgkG1ivaekditpnewfFqgHFpgdPvmPGvLgleamaQ1lgfyllksggdegkgrank1vlfagvdevkfrgqvtPgdkk1r1eveikkiseekrvviiadaellvdgklyvea											
#MATCH	lp+r +11+drv1+e eg ++a k+++ne++FqgHFpg P+mPGvL+eamaQ++g+++ks g+ + g +l++fag+de+f+++v+Pg +++++v+++k +r+++ ++ +1vdgk+v+ea											
#PP	8*****99..666....79*****998888864...444.33689999*****99*****99*****9986665											
#SEQ	LPHRYpFLIDRVLEAD--EKR-----FRALKNVSFNEPHFQGHFPGYIIMPVLLIEAMAQGSVAVVTKQP--EFK-PGGLVFLVGVVEEARFKKPVVPGD-TLILEGELLA---YRRLGKVKVEARVEGELRAEA											

B

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites
				Start	End	Start	End	From	To				
FabA	FabA-like domain	Domain	CL0050	10	133	10	132	1	137	138	98.0	3.1e-28	n/a
#HMM	lpark.l111drvleieeegkfgkG1ivaekditpnewfFqgHFpgdPvmPGvLgleamaQ1lgfyllksggdegkgrank1vlfagvdevkfrgqvtPgdkk1r1eveikkiseekrvviiadaellvdgklyvea												
#MATCH	lp+r+ +11+drv1e + +++ +a k+++ne++FqgHFpg P+mPGvL+eamaQ + + ++k+ + + + lv+ +gv+e++f+++v+Pg +1+le e+ +r++++++e+ v+g+1 +ea												
#PP	8*****99..666....79*****998888864...444.33689999*****99*****99*****9986665												
#SEQ	LPHRYpFLIDRVLEAD--EKR-----FRALKNVSFNEPHFQGHFPGYIIMPVLLIEAMAQGSVAVVTKQP--EFK-PGGLVFLVGVVEEARFKKPVVPGD-TLILEGELLA---YRRLGKVKVEARVEGELRAEA												

Figure 7. The pairwise alignments for *E. coli* FabZ and Mrub_2765 have the same key functional groups except at slightly different positions. Panel A is the pairwise alignment for *E. coli* FabZ. Panel B is that for Mrub_2765. These alignments were obtained by using Pfam (<http://pfam.sanger.ac.uk/search>). The HMM line is the consensus sequence. The SEQ line is the protein sequence of the intended gene. The match line shows which functional groups match up between the two alignments, where the capital letters are highlighted as key functional groups between the two.

Lastly, the pairwise alignments for b0180 and Mrub_2765 show that both genes have the same key functional groups except at slightly different positions (Figure 7). The functional groups for *E. coli* FabZ are F51, H54, F55, P59, P62, G63, L65, Q72, P107, while those for Mrub_2765 are F43, H46, F47, P51, P54, G55, L57, Q73, and P98. Since both sequences have basically the same key functional groups, it is highly likely that Mrub_2765 is the gene that codes for FabZ in *M. ruber*.

Mrub_0266 compared against E. coli b1288 (FabI)

Although there are some differences between genes, there are many similarities and results that lineup and lead to the assumption that Mrub_0266 is orthologous to *E. coli fabI*. Table 2 displays many different characteristics that the two genes share. Part of the table displays E-values which are considered significant when they are $1e^{-3}$ or smaller. As mentioned previously, this significance means that the sequences that are being compared are highly similar rather than comparable by chance. All the E-values that were important and applicable are recorded in Table 2. A protein BLAST comparison between *E. coli* b1288 and Mrub_0266

produced an E-value of 5e-79 and a bit score of 231 (Table 2). This BLAST result shows that b1288 and Mrub_0266 are similar in sequence.

Table 2: *E. coli fabI* and Mrub_0266 are orthologs.

Description of evidence collected	<i>E. coli</i>	<i>M. ruber</i>
BLAST <i>E. coli</i> against <i>M. ruber</i>	Score: 231 bits ; E-value: 5e-79	
KEGG pathway	Biotin metabolism specifically involved in Fatty acid biosynthesis	
Cellular localization (lipoP, P-sortB, Phobius)	cytoplasmic	
TIGRfam-protein family	TIGR01830 3-oxoacyl-[ACP] reductase	
	E=2.3e-05	E=1.4e-09
Pfam-protein family	PF13561 enoyl-(ACP) reductase Clan:adh_short_C2	
	E=1.3e-79	E=2.4e-79
PDB	4CV2 Crystal structure of <i>E. coli</i> FabI in complex with NADH	2WYU High Resolution structure of Thermus Thermophilus Enoyl-ACP reductase
	E=7.02993e-153	E=1.51534e-90
E.C. number	E.C.1.3.1.9 Enoyl-[ACP] reductase (NADH)	
Gene context	Not Part of an Operon	
CDD (Cog category)	COG0623 Enoyl-[ACP] reductase (NADH)	
	E=1.99e-157	E=2.30e-131
Gram Stain	Negative	

In general, both Mrub_0266 and b1288 are gram negative and are both part of the biotin pathway and specifically involved in fatty acid biosynthesis (Figure 2). The cropped figure shows the fatty acid biosynthesis pathway within biotin metabolism. The E.C. number for both genes is 1.3.1.9 which stands for Enoyl-[ACP] reductase, the enzyme that catalyzes the reduction of the 2, 3 double bond in enoyl-acyl-[ACP] derivatives of the elongating

fatty acid moiety (Table 2). For both *E. coli* and *M. ruber*, the KEGG pathway shows that FabI is in the same location in the fatty acid biosynthesis pathway (Figure 2).

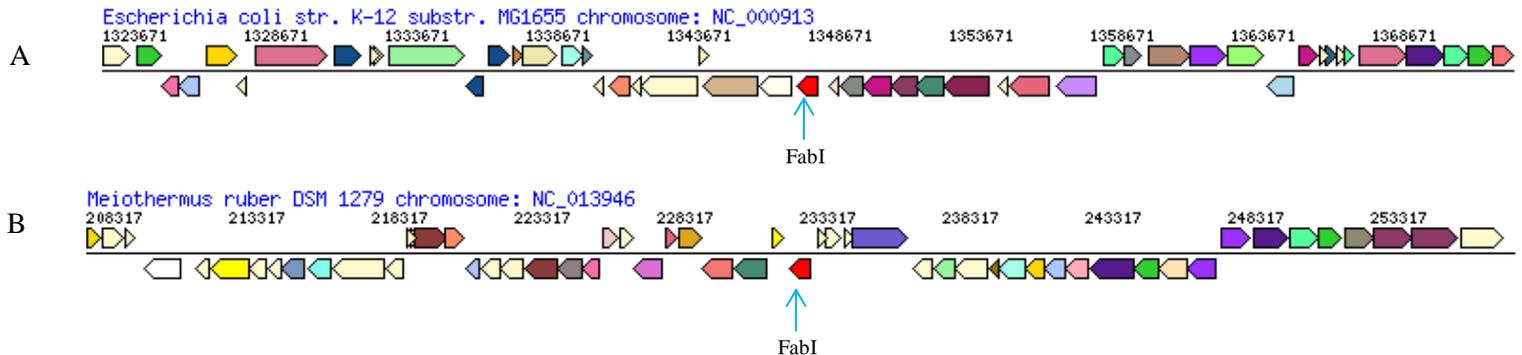


Figure 8. *E. coli fabI* and Mrub_0266 are both not part of operons while. Panel A=*E. coli fabI*/b1288; Panel B=M_rub0266. The gene of interest (*fabI*) is in red. In *M. ruber*, *fabI* does not have any genes next to it, while *fabI* in *E. coli* does have surrounding genes also going to the right. Further research discovered the neighboring genes next to *fabI* in *E. coli* do not form an operon including *fabI*. IMG/ER (<https://img.jgi.doe.gov/cgi-bin/er/main.cgi>) created this image.

It appears that b1288 (*E. coli fabI*) and Mrub_0266 are not part of operons. In *M. ruber fabI* does not have any surrounding genes going in the same direction next to it indicating no evidence for an operon. In *E. coli fabI* does have surrounding genes also pointed in the left direction. Further search lead to no results of *fabI* being part of an operon.

Many of the bioinformatics programs used lead to both *E. coli* b1288 and Mrub_0266 being the sequences for enoyl-[ACP] reductase. The TIGRfam number for both genes was TIGR01830. Mrub_0266 produced an E-value of 1.4e-09 and b1288 produced an E-value of 2.3e-05. The TIGRfam name of 3-oxoacyl-[ACP] reductase does not match entirely with the intended enzyme name of enoyl-[ACP] reductase, but they are both under the family of oxidoreductases. Both genes did have the same TIGRfam number which leads to evidence of Mrub_0266 and b1288 being related. On the other hand, the Pfam results showed that the genes belonged to PF13561 (enoyl-(ACP) reductase) and both genes also belonged to the adh_short_C2 clan. Mrub_0266 produced an E-value of 2.4e-79 and b1288 produced an E-value of 1.3e-79. The protein database shows that b1288 had the PDB code of 4CV2 which stands for

the crystal structure of *E. coli* FabI complex with NADH and had an E-value of 7.03e-153.

Mrub_0266 had the PDB code of 2WYU which stands for high resolution structure of *Thermus thermophilus* enoyl-ACP reductase and had an E-value of 1.51e-90. Although the genes did not have the same PDB code, both of the PDB results code for enoyl-ACP reductase otherwise known as FabI. The PDB results and significant E-values are evidence for Mrub_0266 and b1288 being related. The Conserved Domain Database (CDD) search identified COG0623 (enoyl-[ACP] reductase (NADH)) to be the most significant COG number for both genes. Mrub_0266 produced an E-value of 2.30 e-131 and b1288 produced an E-value of 1.99e-157 (Table 2).

The programs used for cellular localization such as lipoP and THMM show that both Mrub_0266 and b1288 are cytoplasmic. P-SortB had scores of 10 and 9.82 for the cytoplasmic membrane scores for b1288 and Mrub_0266 indicating both are located in the cytoplasmic membrane. Although P-SortB had a high score for the cytoplasmic membrane and Phobius had evidence for some hydrophobic regions, evidence on ecocyc and other evidence in the module lead to a more conclusive answer: FabI is indeed in the cytoplasm. Figure 9 shows no TMH regions in the hydropathy plot for b1288 (Panel A) and Mrub_0266 (Panel B) meaning there is no evidence for transmembrane helices. SignalP results showed to have no cleavage sites or any signal peptides reach the baseline region of significance. in both plots for Mrub_0266 and b1288 (Figure 10). There is some activity within the graph, but it is not over the cytoplasmic region. Phobius results showed to have some regions of hydrophobicity signaled by the red line going over the green cytoplasmic line, but the red line does not go out of the cytoplasm indicating the cellular localization to be in the cytoplasm (Figure 11). All of this data shows there is no evidence for transmembrane helices which further indicates a cytoplasmic location.

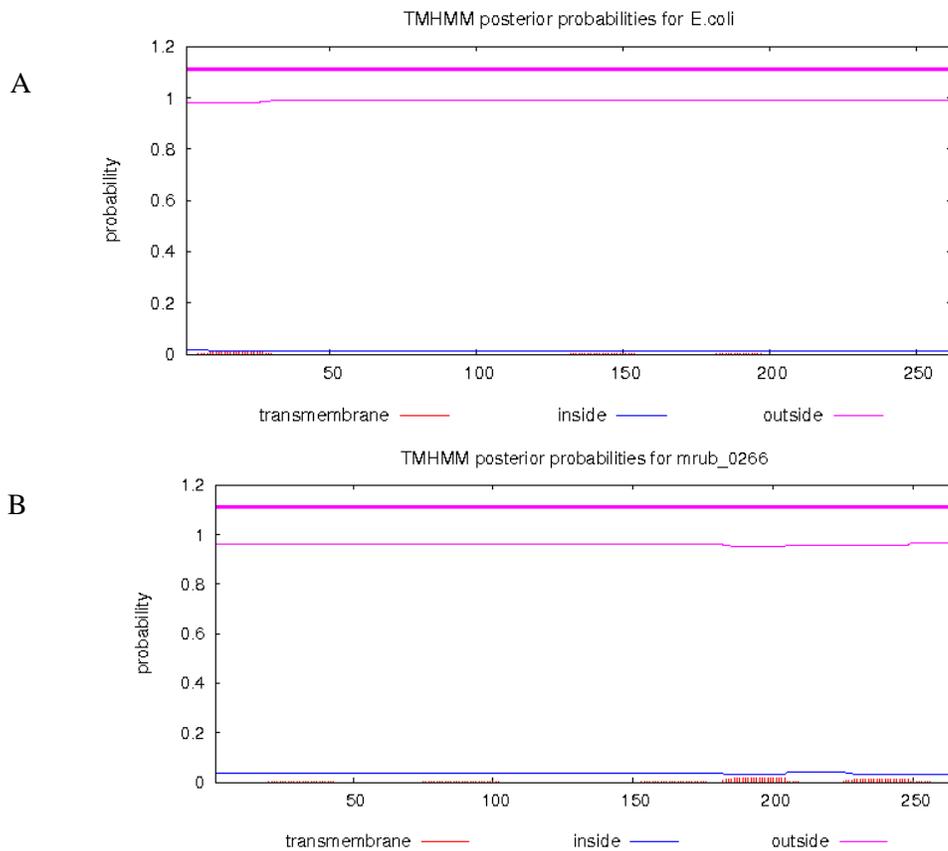


Figure 9. Mrub_0266 and *E. coli* FabI do not contain TMH regions; a cytoplasmic location is predicted. Panel A=*E. coli* fabI/b1288; Panel B=M_rub0266. TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM>) created this hydrophathy plot.

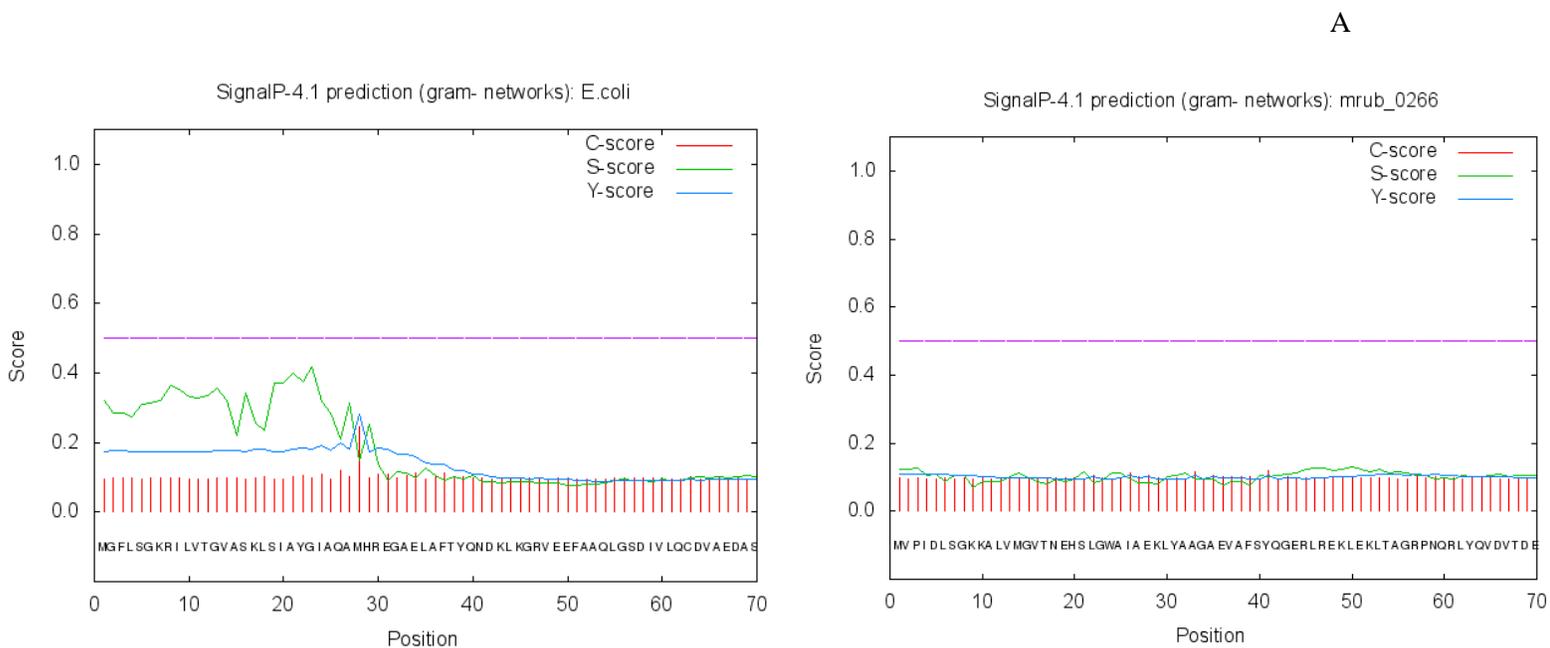
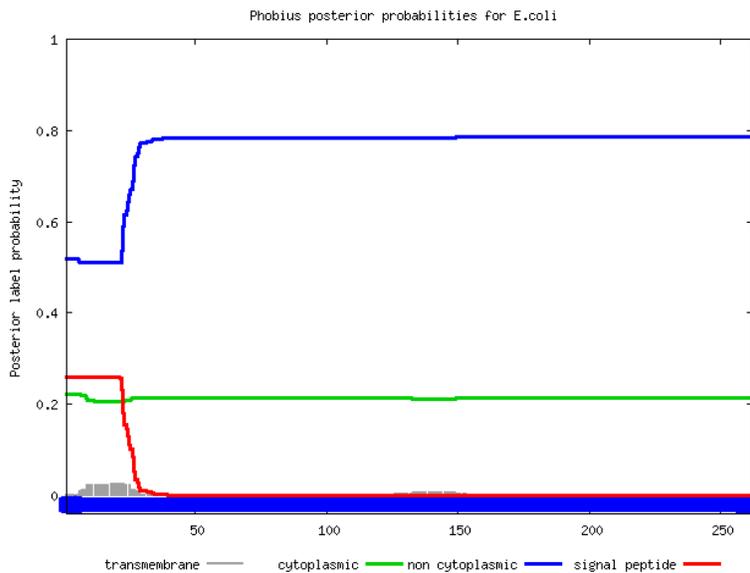


Figure 10. Mrub_0266 and *E. coli* FabI do not have any cleavage sites and predicted to have no signal peptides. This indicates there are no transmembrane helices present in either gene which indicates a cytoplasmic location. Panel A=*E. coli* FabI/b1288; Panel B=M_rub2765. SignalP 4.1 server (<http://www.cbs.dtu.dk/services/SignalP>) created these plots.

A



B

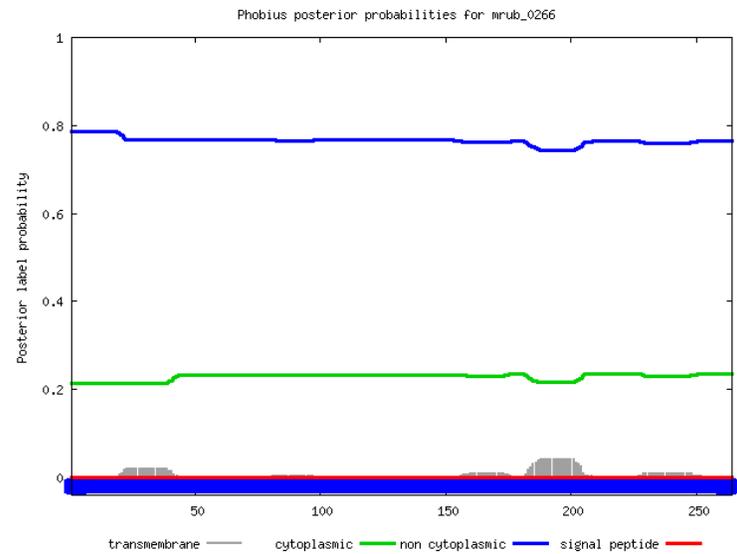


Figure 11. Mrub_0266 and *E. coli fabI* do not have any transmembrane helices. There are regions of hydrophobicity indicated by the red signal peptide line going over to cytoplasmic green line, but it does not go out of the cytoplasmic region. Since there are no transmembrane helices present in either gene, this indicates a cytoplasmic location. Panel A=*E. coli fabI*/b1288; Panel B=M_rub0266. Phobius (<http://phobius.sbc.su.se/>) created these plots.

A

Family	Description	Entry type	Clan	Envelope		Alignment	
				Start	End	Start	End
adh_short_C2	Enoyl-(Acyl carrier protein) reductase	Domain	CL0063	13	252	13	252
#HMM	GvaddnsigwaiaakalaeaGaeviltttwvpakragkvqelakelpad.vvplDatsdedvselaekvkedlggkiDilvhsiansp-ervekpllet.sregllaalsisayslvsllkaalpimneggSiva						
#MATCH	Gva + si+++ia+a+++Gaev+++t++++ +g+v+e+a++l+td v+++D++d ++++++++ + + k+D++vhsi+++p +++ +++++ +reg++ a++is+y+v++++ka+ +++n+g++++						
#PP	9999*****997,8*****56*****9,7*****965*****						
#SEQ	GVASKLSIAYGIAQAMHREGAELAFYQNDKLLKGRVEEFAAQLGSDIVLQCDVAEDASIDTIFAELEGKLNPKFDGFVHSIGFAPGDQDLDGVDVIAVTRREGFKIAHDISSYFVAMAKACRSMLNPGSALL						
HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment	
From	To						
1	241	241	267.4	1.3e-79	n/a	<input type="button" value="Hide"/>	
valsylaservipgygmsaKaALEsItrVlAvelgrkkgiRvntisaGplktraaakaggiekmieyseenapLgkelsaeveaaafllsdlasaiTgetlyvDgGln ++syl++er+ip+y++m+ aKa LE ++R++A+ +g+ +g+Rvn+isaGp++t aa+i++++km+++++ +p++++++e+v++ aafL+sd1+ +i ge+++vDgG++ *****96 LTLsYLGAERAIpHYNVMGILAKASLEANVRVYMANAMGP-EGVRVNAISAGPIRTLAASGDKFRKMLAHCEAVTPIRRVTIETDVGNsAAFLCSDLsAGISGEVHVHVDGGS							

B

Family	Description	Entry type	Clan	Envelope		Alignment	
				Start	End	Start	End
adh_short_C2	Enoyl-(Acyl carrier protein) reductase	Domain	CL0063	15	253	15	253
#HMM	GvaddnsigwaiaakalaeaGaeviltttwvpakragkvqelakelpad.vvplDatsdedvselaekvkedlggkiDilvhsiansp-ervekpllet.sregllaalsisayslvsllkaalpimneggSiva						
#MATCH	Gv+++++gwaiaa+l+++Gaev++++ ++ ++k++l++ p++ +++D+t++ aa +++ + + +g +D+lvhsia++p ++e ++++++ + +al++sayslv+++ a+p++ eggS+v+						
#PP	9*****987,689999998885559*****7*****99*****						
#SEQ	GVtNEHSLGWATAEKLYAAGAEEVAFSYQGERLREKLEKLTAGRPNQMLYQVDVTDAAALKAMFADLQQA-NGLDYLVHSIAFAPRAAMEGRFIDTIAADWNITALQVSAYSLVAVAREAEPLREGGSLVTI						
HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment	
From	To						
1	241	241	266.6	2.4e-79	n/a	<input type="button" value="Hide"/>	
valsylaservipgygmsaKaALEsItrVlAvelgrkkgiRvntisaGplktraaakaggiekmieyseenapLgkelsaeveaaafllsdlasaiTgetlyvDgGln v+l+y ase+v+p y++m+ aKaALE ++R+lA+elg+ k+iRvn+isaG ++t aa i+g++k++ + + + apLg+ +++eev++ ++lLs+l+s++Tg+t+yvD G++ *****9g *****99*****96 VTLTYAYASEKVVPRYNVMGIAKAALASVRYLAYELGKLNKIRVNAISAGAIRTVAAAMSIPGFRKMWAKYNATAPLGRMITHEEVGNLGLYLLSPLSSGTGTQTVYVDAGYS							

Figure 12. The pairwise alignments for *E. coli* FabI and Mrub_0266 have the same key functional groups except at slightly different positions. Panel A is the pairwise alignment for *E. coli* FabI. Panel B is that for Mrub_0266. These alignments were obtained by using Pfam (<http://pfam.sanger.ac.uk/search>). The HMM line is the consensus sequence. The SEQ line is the protein sequence of the intended gene. The match line shows which functional groups matchup between the two alignments, where the capital letters are highlighted as key functional groups between the two.

Lastly, the pairwise alignments for b1288 and Mrub_0266 show that both genes have mostly the same key functional groups except at slightly different positions (Figure 12). The functional groups for *E. coli* FabI are G13, G32, D64, D86, K163, A164, L166, E167, R171, A174, R183, G190, D247, and G250, while those for Mrub_0266 are G15, G34, D66, D88, S141, K164, A165, L167, E168, R172, A175, R184, G191, L217, L233, T242, D249, and G251. Since both sequences have mostly the same key functional groups, it is highly likely that Mrub_0266 is the gene that codes for FabI in *M. ruber*.

Conclusion:

Through the evidence gathered and analyzed, it appears that Mrub_2765 is orthologous to *E. coli* fabZ and Mrub_0266 is orthologous to *E. coli* FabI. The bioinformatics programs used show that both Mrub_2765 and *E. coli* FabZ are cytoplasmic, involved in the biotin metabolism and specifically the fatty acid biosynthesis, have the same E.C. number, have the same TIGRfam, Pfam, PDB and COG numbers and they both have the same key functional groups, but just at different positions. All the recorded E-values for both Mrub_2765 and b0180 are significant since they are all less than $1e-3$. This indicates that the sequences are related rather than lined up by chance. The bioinformatics programs used also show that both Mrub_0266 and *E. coli* FabI are cytoplasmic, involved in the biotin metabolism and specifically the fatty acid biosynthesis, have the same E.C. number, both are not part of an operon, have the same TIGRfam, Pfam, and COG numbers and they both have mostly the same key functional groups, but just at different positions. All the recorded E-values for both Mrub_0266 and b1288 are significant since they are all less than $1e-3$. This indicates that the sequences are related rather than lined up by chance. Although there were many similarities, the differences between the two pairs of genes were very minute and did not give any reason for the genes not to be orthologs.

All of the included evidence shows that Mrub_2765 is indeed *M. ruber fabZ* and codes for the enzyme B-hydroxyacyl-acyl carrier protein (ACP) dehydratase in the type II fatty acid synthase system and Mrub_0266 is indeed *M. ruber fabI* and codes for the enzyme enoyl-[ACP] reductase in the type II fatty acid synthase system..

References

- (1) Tindall BJ, Sikorski J, Lucas S, Goltsman E, Copeland A, Glavina DR, Nolan M, Tice H, Cheng J, Han C, et al. Complete genome sequence of *meiothermus ruber* type strain (21(T)). *Standards in Genomic Sciences* 2010 07/29;3(1):26-36.
- (2) Phylogenetic Diversity. [Internet]. 2013. Joint Genome Institute; [2016 Feb 1]. Available from:<http://jgi.doe.gov/our-science/science-programs/microbial-genomics/phylogenetic-diversity/>
- (3) Geni-science. [Internet]. Unknown. Geni-science; [2016 Feb 1]. Available from:<http://geni-science.org/secure/projects/view/>.
- (4) Geni-Act. [Internet]. Unknown. GENI-ACT; [2016 Feb 1]. Available from: <http://geni-act.org/>.
- (5) Heath RJ, Rock, CO. Roles of the FabA and FabZ B-Hydroxyacyl-Acyl Carrier Protein Dehydratases in *Escherichia coli* Fatty acid biosynthesis. *JBC* 1996 11/1;271(44):27795-27801.
- (6) Bergler H, Fuchsbichler S, Högenauer G, Turnowsky F. The enoyl-[acyl-carrier-protein] reductase (FabI) of *escherichia coli*, which catalyzes a key regulatory step in fatty acid biosynthesis, accepts NADH and NADPH as cofactors and is inhibited by palmitoyl-CoA. *European Journal of Biochemistry* 1996;242(3):689-94.
- (7) *Escherichia coli* K-12 substr. MG1655 Transcription Unit: hlpA-lpxD-fabZ-lpxA. [Internet]. Unknown. Biocyc; [2016 Feb 1]. Available from: <http://biocyc.org/ECOLI/NEW-IMAGE?type=OPERON&object=TU0-4146>.