

2-2016

A Bioinformatics Study on Whether or Not Mrub_2763 gene in *M. ruber* is Similar to the LpxB Gene in *E. coli* and if Mrub_2768 is Similar to the LpxD gene in *E. coli*.

Frank J. Habura

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

Habura, Frank J. and Scott, Dr. Lori. "A Bioinformatics Study on Whether or Not Mrub_2763 gene in *M. ruber* is Similar to the LpxB Gene in *E. coli* and if Mrub_2768 is Similar to the LpxD gene in *E. coli*." (2016). *Meiothermus ruber Genome Analysis Project*. <http://digitalcommons.augustana.edu/biolmruber/14>

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.

A Bioinformatics Study on Whether or Not Mrub_2763 gene in *M. ruber* is Similar to the LpxB Gene in *E. coli* and if Mrub_2768 is Similar to the LpxD gene in *E. coli*.

Frank Habura

Introduction

Meithermus ruber is a Gram-negative bacteria that lives in relatively hot environments (Tindall et. al., 2010). Not much is known about *M. ruber* because it has not yet been studied in depth. One way to study an unknown bacteria like *M. ruber* is to compare it to a well-known bacteria like *Escherichia coli*. Scientists have sequenced the entire *E. coli* genome, so it is a perfect candidate to compare DNA sequences to (Blattner et. al., 1997). This also make *E. coli* a great control because it is easy to find all the information on it. A way to compare these two organisms is using bioinformatics tools. Bioinformatics is a division of science that allows scientists to study, evaluate, and explain biological evidence (Pujari, n.d.). It is a great way to condense huge amounts of data like entire genomes and protein sequences and have them available at any time (Pujari, n.d.).

The pathway being examined in this study is the Lipid A biosynthesis pathway (partly depicted in Figure 1). Gram-negative bacteria have two membranes, a Cytoplasmic Membrane/Inner Membrane (IM) and an Outer Membrane (OM) (Slonczewski and Foster, 2014). Lipopolysaccharides (LPS) (Figure 2) are embedded into the OM and are very important to Gram-negative bacteria because they help stabilize the OM structure and prevent the bacteria from being engulfed by other bacteria/cells (Slonczewski and Foster, 2014). Lipid A is a component of a LPS that resides in the OM of Gram-negative bacteria (Raetz et. al. 2009). Lipid A has many hydrophobic fatty acids attached to it that help anchor the LPS into the OM (Slonczewski and Foster, 2014).

Lipid A biosynthesis is part of the LPS biosynthesis pathway as depicted in the KEGG map (Figure 3). The focus of this study is the most conserved portion of the Lipid A biosynthesis pathway, which is the first 6 steps shown in Figure 1 (Raetz et. al., 2009). The two genes being examined are LpxD and LpxB. LpxD is the third enzyme in the pathway and cleaves a(3R)-3-hydroxymyristoyl-[acp] into [acp] and (3R)-3-hydroxymyristoyl (Anderson et. al., 1988). LpxD then attaches (3R)-3-hydroxymyristoyl, one of the 4 fatty acid chains in Lipid A, onto UDP-3-O-(3-hydroxymyristoyl)- α -D-glucosamine to form UDP-2-N, 3-O-bis[(3R)-3-hydroxytetradecanoyl]- α -D-glucosamine (Anderson et. al., 1988). LpxB, in short, takes 2 products of LpxD (UDP-2-N, 3-O-bis[(3R)-3-hydroxytetradecanoyl]- α -D-glucosamine) and joins them together through a hydrogen bond (Ray et. al., 1984). The result of this reaction is lipid A disaccharide (Figure 1).

This study is important for multiple reasons. Researching a poorly studied organism like *M. ruber* may result in unknown discoveries that may be beneficial to society. It is also a great opportunity to just gain knowledge on these unknown organisms to better understand them and

the world around us. However, something that is known about Lipid A is that it is an endotoxin that excites the immune system in humans to such a degree that it can result in death (Slonczewski and Foster, 2014). If scientists can somehow prevent Lipid A from even forming by inhibiting one of these constitutive enzymes, then maybe it can save lives of those infected with Gram-negative bacteria.

The purpose of this study is to compare Mrub_2763 gene from *M. ruber* to the LpxB gene of *E. coli* as well as the Mrub_2768 gene to LpxD using bioinformatics tools. It is hypothesized that Mrub_2763 and Mrub_2768 encodes the LpxB and LpxD enzymes respectively.

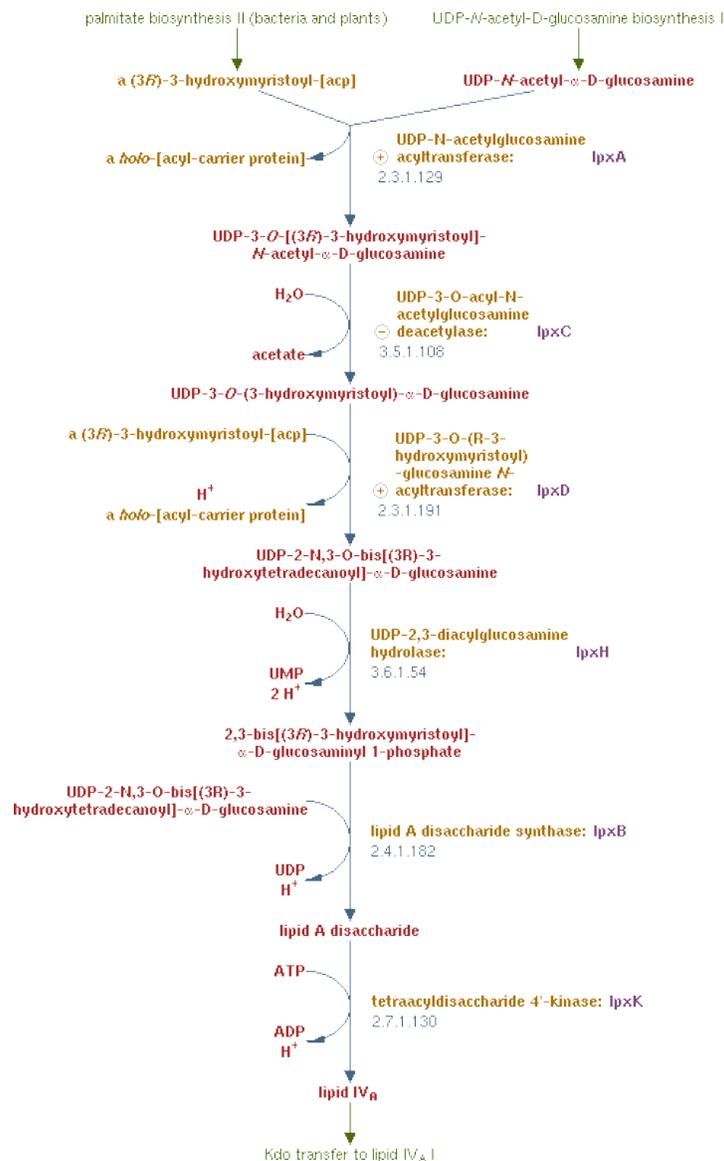


Figure 1. Lipid IV_A biosynthesis pathway. The reactants, intermediates, and product are in red and the enzymes are gold with alternate names in purple. Image from MetaCyc.

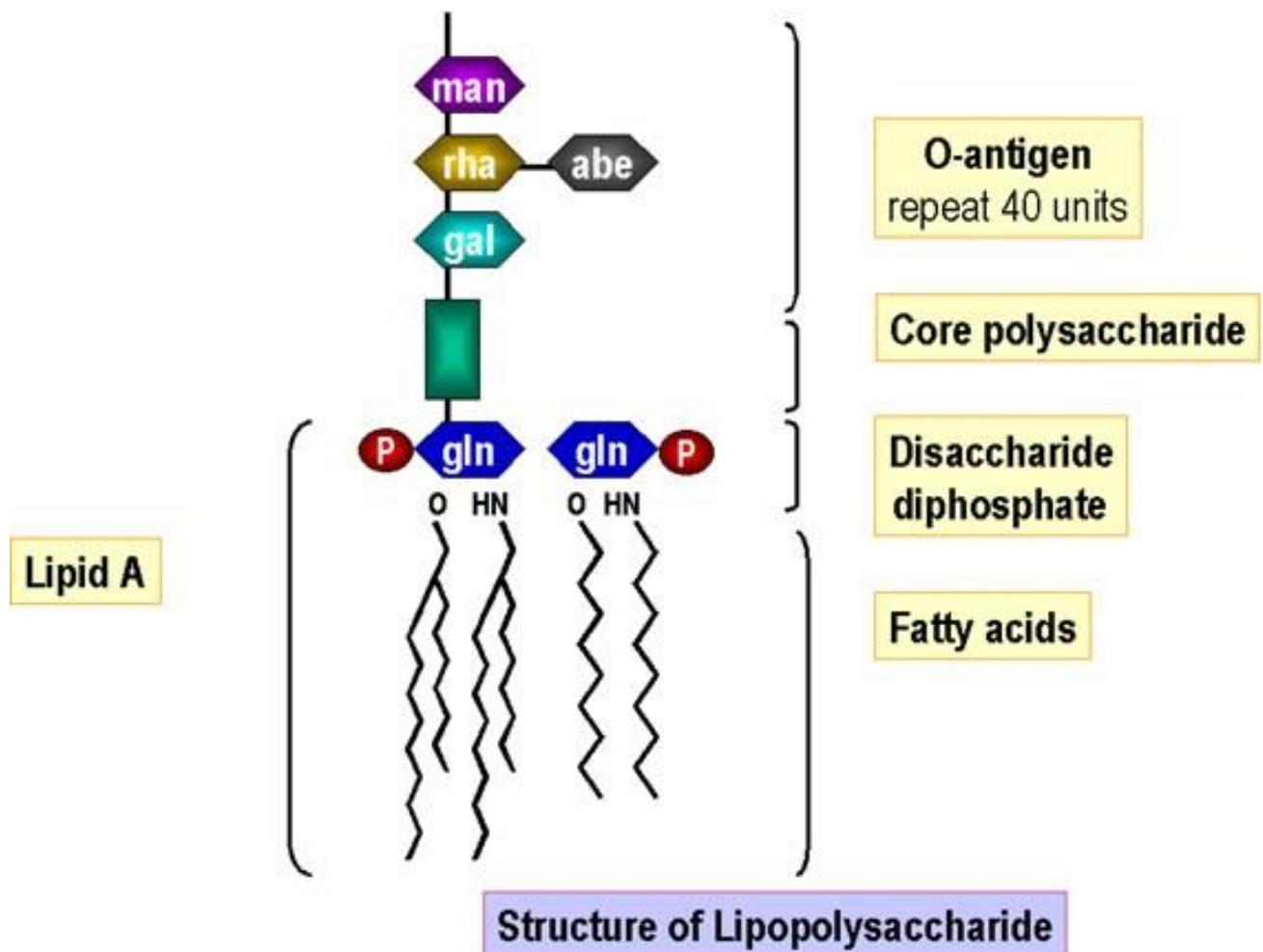


Figure 2. Structure of a Lipopolysaccharide. Lipid A is embedded into the OM of the bacteria because it has multiple hydrophobic fatty acids. Image from: <http://i0.wp.com/microbeonline.com/wp-content/uploads/2013/04/lipopolysaccharides.jpg>

Methods

The methods and various bioinformatics tools that were used are summarized in this URL < <http://www.geni-act.org/education/main/> > within the GENI-ACT system. There were, however, some deviations from this protocol. For the T-Coffee module under *E. coli*, instead of using only 10 sequences I used 15. I also excluded *E. coli* from the sequences as to get a wider variety of organisms. The paralog module was excluded from this study. I also added a BLAST sequence of *E. coli* versus *M. ruber* first in order to determine if *M. ruber* did in fact have a gene similar to *E. coli*'s LpxB and LpxD. The KEGG pathway map was also altered slightly to include colored E.C. numbers for the enzymes involved in *E. coli* and *M. ruber* Lipopolysaccharide biosynthesis. The phylogeny module was also altered. Instead of using their database, I used the top 15 hits from BLAST to create a phylogenetic tree for both Mrub_2768 and Mrub_2763. Both *E. coli* genes do not have a phylogenetic tree in the Horizontal Gene Transfer module.

Results

LpxD

The results from the various bioinformatics tests that were run were testing the hypothesis that Mrub_2768 and LpxD(b0179) in *E. coli* are orthologs. A protein BLAST comparison between *E. coli* LpxD and Mrub_2768 produced an E-value of 1e-32, bit score of 130, and a 27% identity (Table 1). Both *E. coli* LpxD and *M. ruber* Mrub_2768 are part of the lipopolysaccharide biosynthesis pathway according to KEGG (Table 1). A protein family test was run using TIGRfam that showed that both *E. coli* LpxD and Mrub_2768 belong in the same family (TIGR01853) and had low E-values of 3.9e-203 and 2.5e-64 with bit scores of 686.1 and 225 respectively (Table 1). Another protein domain test was run using Pfam that determined whether or not there are similar domains in the protein. *E. coli* LpxD had a different hit than Mrub_2768, however it is important to note that *E. coli* LpxD did share a hit with Mrub_2768. That second hit was Hexapep (PF00132) with both genes having relatively low bit scores and high E-values (Table 1). The protein domains were also compared using the Conserved Domain Database (CDD) to determine if the two proteins belonged to the same Cluster of Orthologs (COG) group. The two had the same COG groups with significant E-values as shown in Table 1.

Table 1: *E. coli* LpxD and Mrub_2768 are orthologs

Description of evidence collected	<i>E.coli</i>	<i>M.ruber</i>
Cellular localization(Module 3)	Cytoplasmic	
BLAST <i>E. coli</i> against <i>M. ruber</i>	Score: 130 bits; E-value: 1e-32; 27% identity	
KEGG pathway	Lipopolysaccharide Biosynthesis	
Pfam – protein family	PF04613 LpxD (E=7.9e-22; score: 76.7) AND PF00132 Hexapep (E=3.2e-10; score 39.2)	PF00132 Hexapep (E=3.8e-6; score: 26.2)
CDD (COG category)	COG1044 (E=4.81e-156) LpxD	COG1044 (E=3.82e-79) LpxD
TIGRfam – protein family	TIGR01853 (E=3.9e-203; score: 686.1) Lipid A lpxD: UDP-3-O-[3-hydroxymyristoyl]	TIGR01853 (E=2.5e-64; score: 225.0) Lipid A lpxD: UDP-3-O-[3-hydroxymyristoyl]
E.C. number	E.C. 2.3.1.191 UDP-3-O-(3-hydroxymyristoyl)glucosamine N-acyltransferase	

PDB	3EH0 (E=0e00) Crystal Structure of LpxD from <i>E. coli</i>	3PMO (E=4.26e-39) The structure of LpxD from <i>Pseudomonas aeruginosa</i> at 1.3 Å resolution
------------	---	--

Table 1 also summarizes all of the content in module 3, which concluded that both LpxD and Mrub_2768 reside in the cytoplasm and have no transmembrane helices (Figure 4). Figure 5 shows that both LpxD and Mrub_2768 do not have any signal peptides, so they are not integrated into the membrane. The signal peptide probability for Mrub_2768 was $D=.102$ and for *E. coli* LpxD it was $D=.109$. PSORT-B predicted the subcellular locations of both LpxD and Mrub_2768 to be in the cytoplasm. Their cytoplasmic scores were 9.26/10 and 9.97/10 respectively. Mrub_2768 and LpxD *E. coli* both have the same E.C. number of 2.3.1.191 suggesting they have the same function (Table 1). The E.C. number is also depicted in Figure 3 using a KEGG pathway map of the Lipid A disaccharide biosynthesis pathway for both Mrub_2768 and *E. coli* LpxD. The same map came up for both *E. coli* and *M. ruber* with the lipid A disaccharide biosynthesis pathway being highlighted in green.

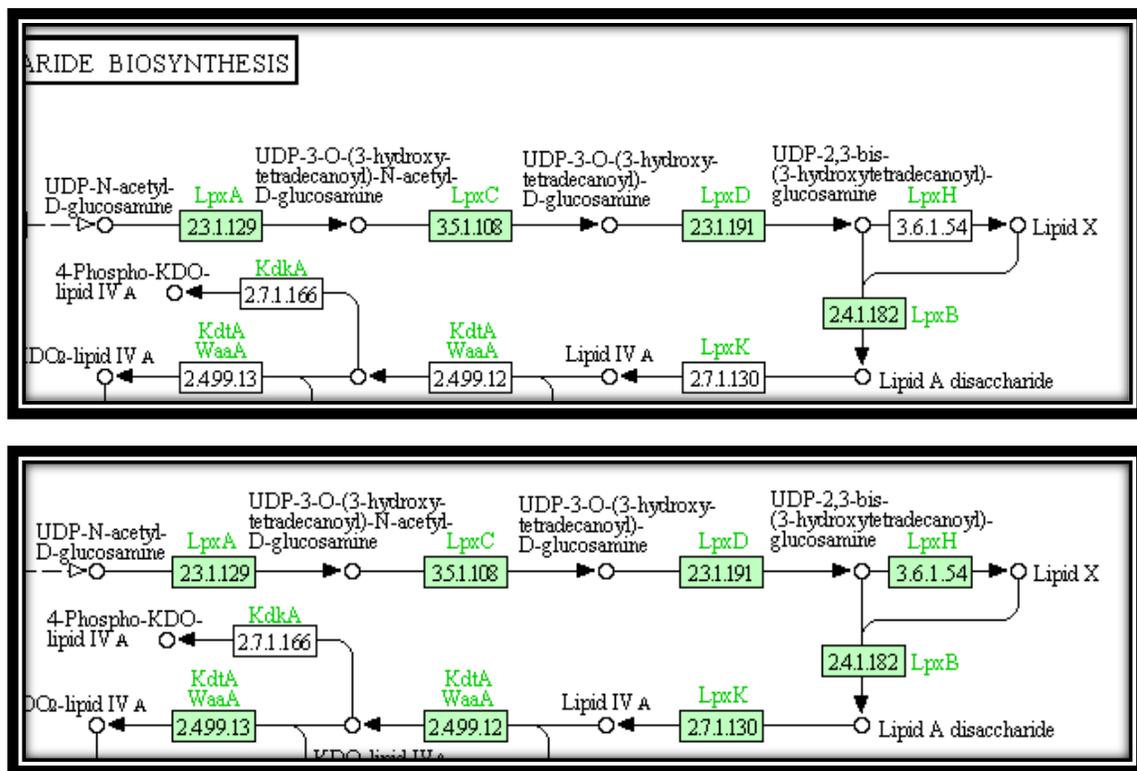


Figure 3. *M. ruber* (Top) and *E. coli* Lipopolysaccharide biosynthesis pathway (Bottom). Enzymes are shown in green text with their corresponding E.C. numbers listed below. These images were taken from KEGG and cropped in order to focus on Lipid A disaccharide biosynthesis.

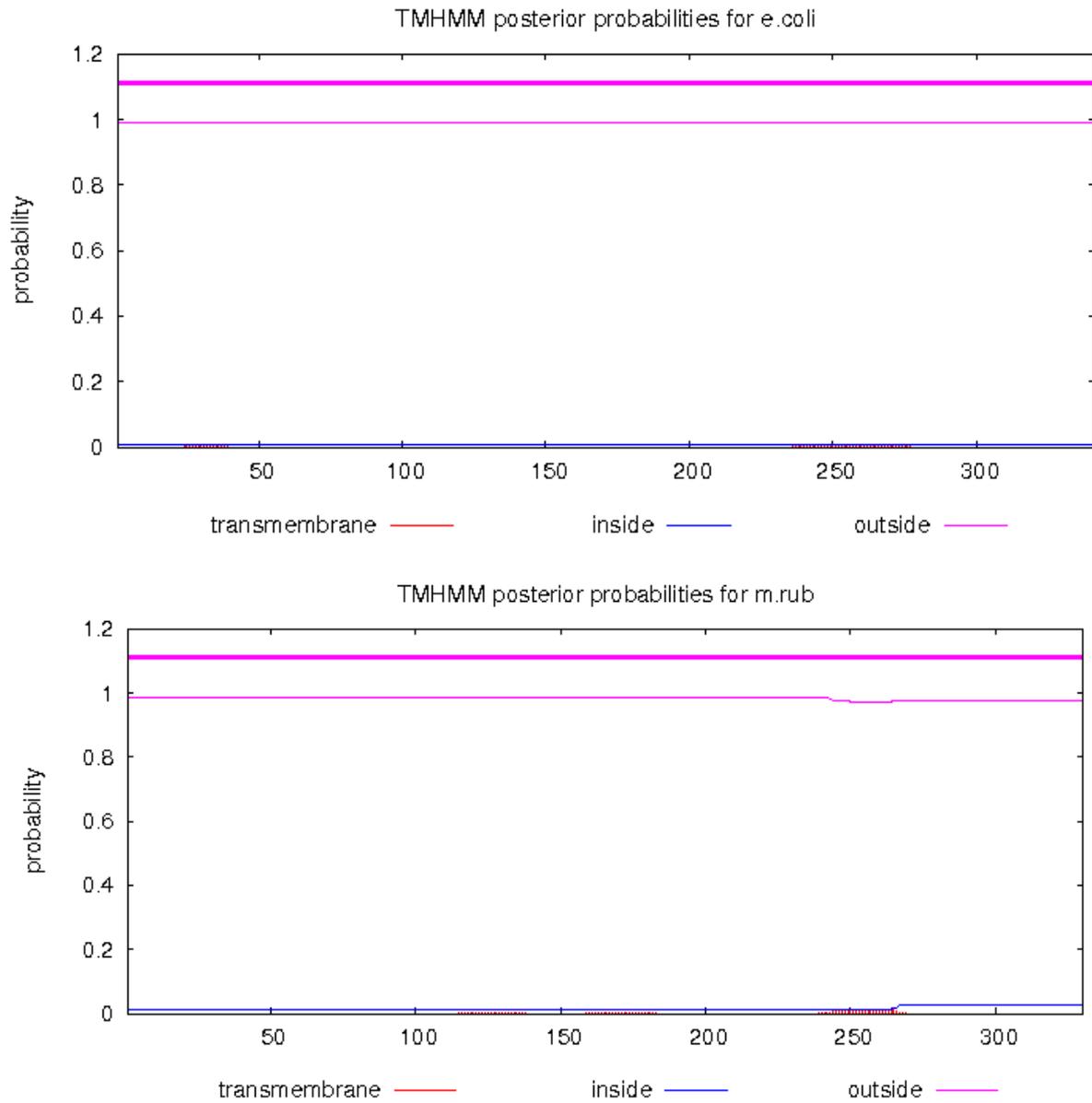


Figure 4. *E. coli* LpxD (top) and Mrub_2768 (bottom) do not contain TMH regions; a cytoplasmic location is predicted. THHMM server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) created this hydropathy plot.

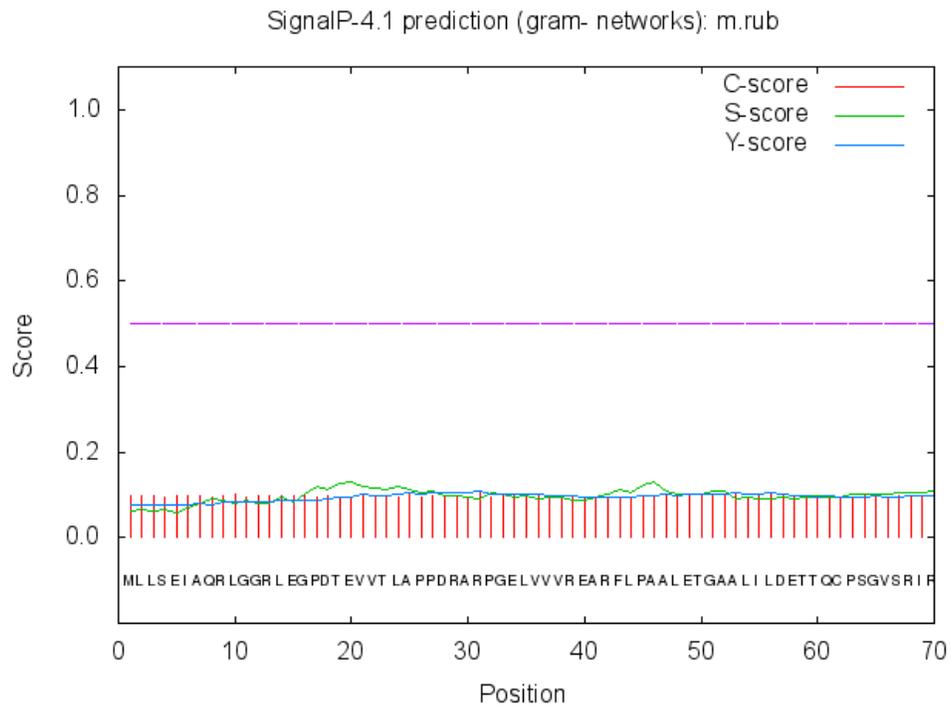
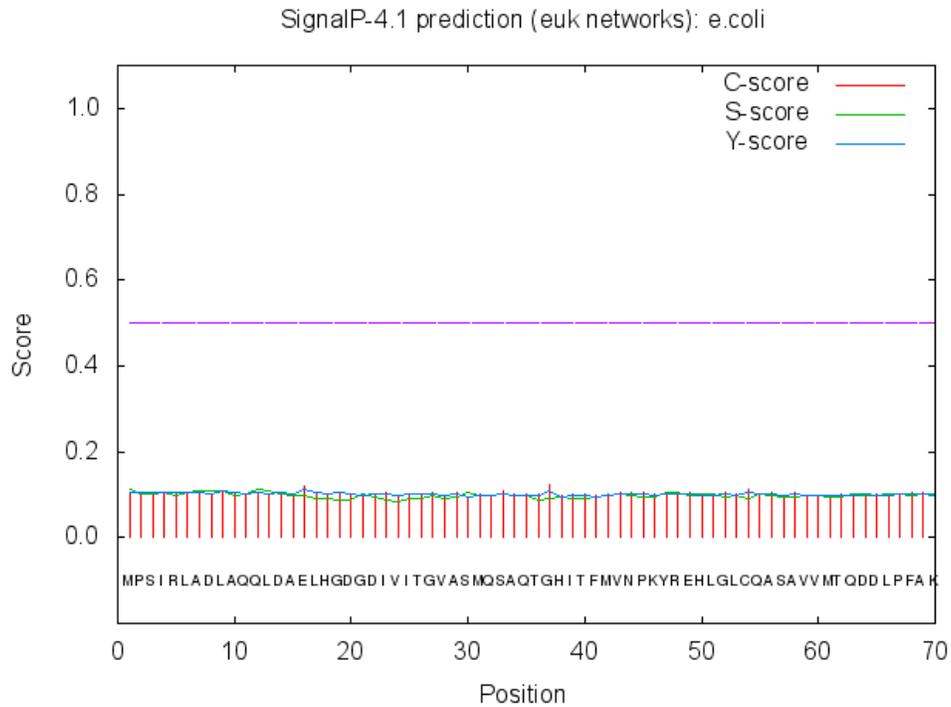


Figure 5. *E. coli* LpxD(top) and Mrub_2768(bottom) do not contain signal peptides; no membrane integration is predicted. Signal P Server 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>) created these graphs.

An HMM logo was created from Pfam to determine if the conserved parts of the sequences for LpxD and Mrub_2768 families. The first hit for LpxD did not match that of Mrub_2768, however the second hit for LpxD did match it (Table 1). The similar family's first 10 conserved amino acid sequence is shown in Figure 6 with E-values shown in Table 1. A pairwise alignment was examined to determine if *E. coli* and *M. ruber* had similar conservative sequences for the Hexapep family that the two had as shown in Figure 7. The pairwise alignment shows that within the 35 amino acids, both sequences match the consensus sequence at positions G4, G22, and I27.

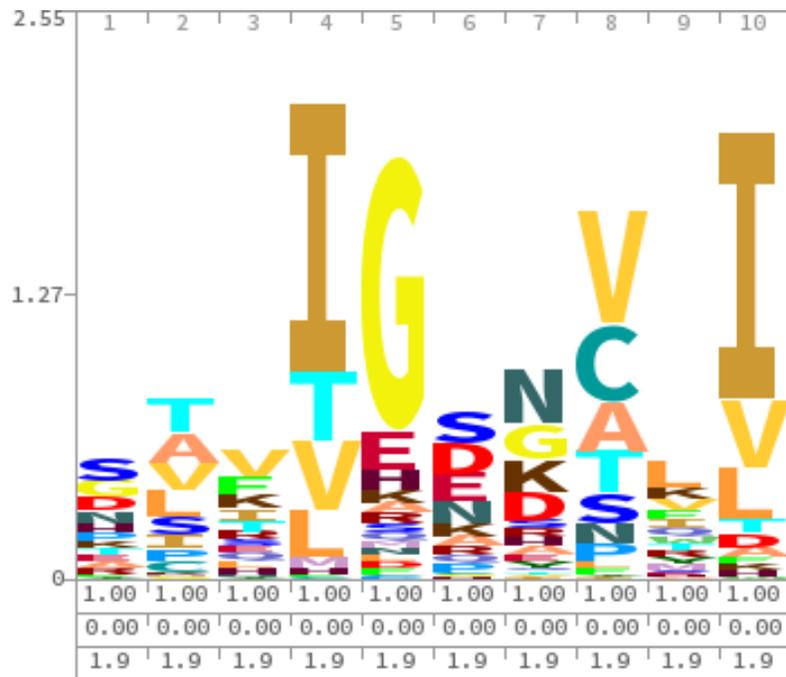


Figure 6. The second hit for *E. coli* LpxD matched the first hit for Mrub_2768 showing that there may be some similar conserved sequences according to HMM. PFAM (<http://pfam.xfam.org/search/sequence>) was used to create these logos.

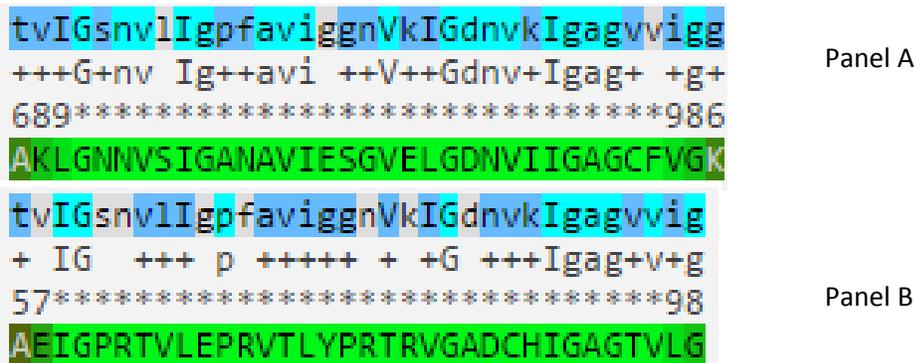


Figure 9. Mrub_2768 and *E. coli* LpxD have very similar conservative sequences according to their alignments with the Hexapep family. Panel A=*E. coli* pairwise alignment to consensus sequence. Panel B=*M. ruber* pairwise alignment to consensus sequence. PFAM (<http://pfam.xfam.org/search/sequence>) was used to create these alignments.

LpxB

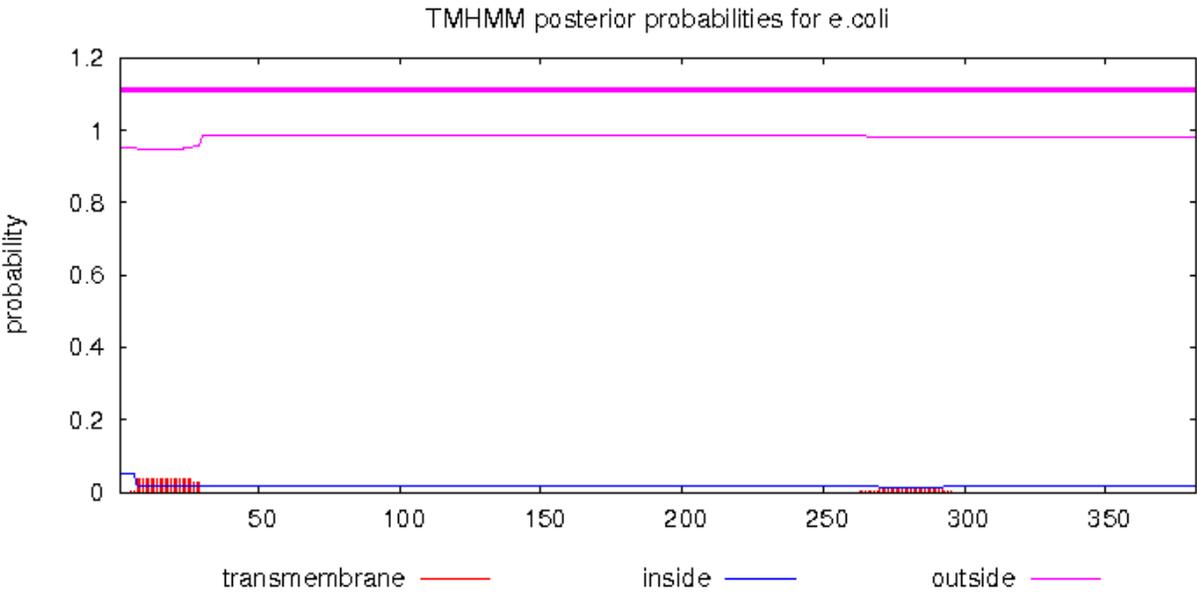
Another part of this study was determining whether or not M_rub2763 is the *M. ruber* gene for LpxB(b0182) in *E. coli* using bioinformatics tools. A protein BLAST comparison between *E. coli* LpxB and Mrub_2763 produced a very high E-value of .29 (Table 2). Both *E. coli* LpxB and *M. ruber* Mrub_2763 are part of the lipopolysaccharide biosynthesis pathway according to KEGG (Table 2). A protein family test was run using TIGRfam and both genes yielded different results. *E. coli* LpxB was shown to be a part of TIGR00215 (E=6.9e-190; score: 642.1) LpxB Lipid A disaccharide synthase and Mrub_2763 first hit was TIGR03492 (E=.9; score: -293.7) conserved hypothetical protein (Table 2). Another protein domain test was run using Pfam that, once again, determined whether or not two had similar domains in the protein. *E. coli* LpxB's first hit was PF02684 LpxB with an E=8.6e-159 and a bit score of 76.7 as shown in Table 2. Mrub_2763 did not come up with any hits. The protein domains were also compared using the CDD to determine if the two proteins belonged to the same COG group. The two had different COG groups as shown in Table 2, but the two shared the same name of Lipid A disaccharide synthetase with LpxB having an E-value of 0 and Mrub_2763 having an E-value of 2.07e-13 (Table 2).

Table 2: Comparing *E. coli* LpxB and Mrub_2763

Description of evidence collected	<i>E.coli</i>	<i>M.ruber</i>
Cellular localization(Module 3)	Cytoplasmic	
BLAST <i>E. coli</i> against <i>M. ruber</i>	37% identity, E-value=.26	
KEGG pathway	Lipopolysaccharide Biosynthesis	
Pfam – protein family	PF02684 LpxB (E=8.6e-159; score: 76.7)	No hits in database
CDD (COG category)	COG0763 (E=0.00) LpxB Lipid A disaccharide synthetase	COG1044 (E=2.07e-13) Lipid A disaccharide synthetase
TIGRfam – protein family	TIGR00215 (E=6.9e-190; score: 642.1) LpxB Lipid A disaccharide synthase	TIGR03492 (E=.9; score: -293.7) conserved hypothetical protein
E.C. number	E.C. 2.4.1.182 Lipid A disaccharide synthase	
PDB	3EHB (E=.652) A D-Pathway Mutation Decouples the Paracoccus Denitrificans Cytochrome c	No hits in database

	Oxidase by Altering the side chain orientation of a distant, conserved Glutamate	
--	--	--

Table 2 also summarizes all of the content of module 3, which concluded that both LpxB and Mrub_2763 reside in the cytoplasm and have no transmembrane helices (Figure 8). Figure 9 shows that both LpxB and Mrub_2763 do not have any signal peptides, so they are not integrated into the membrane. The signal peptide probability for Mrub_2763 was D=.116 and for *E. coli* LpxB it was D=.109. PSORT-B predicted the subcellular locations of both LpxD and Mrub_2763 to be in the cytoplasm. Their cytoplasmic scores were 10/10 and 8.96/10 respectively. Mrub_2763 and LpxB *E. coli* both have the same E.C. number of 2.4.1.182 suggesting they have the same function (Table 2).



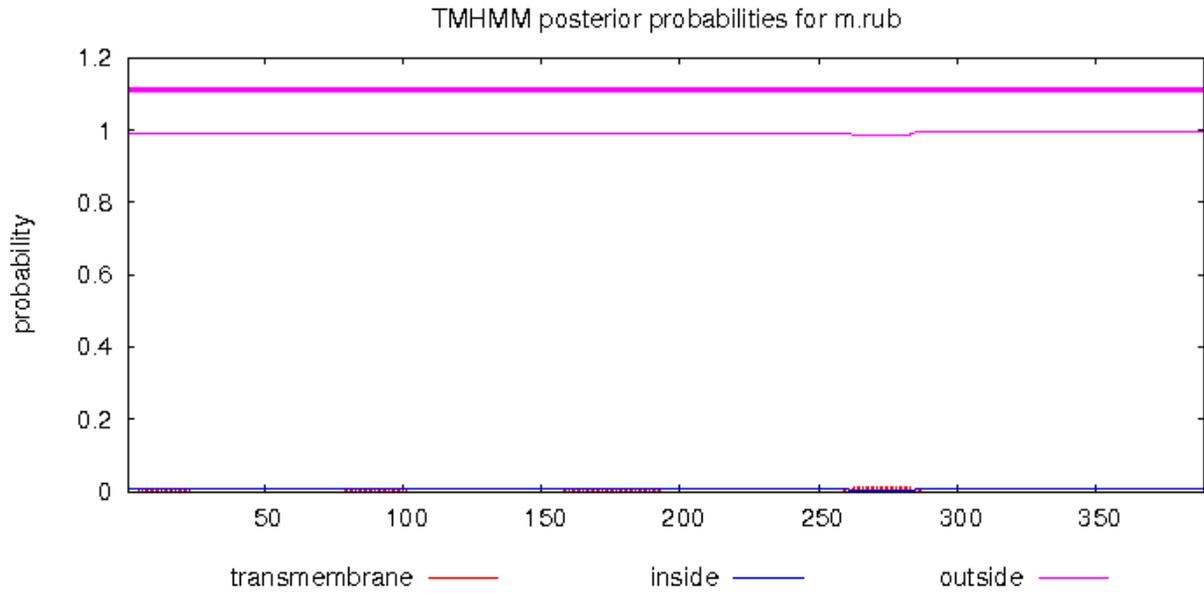
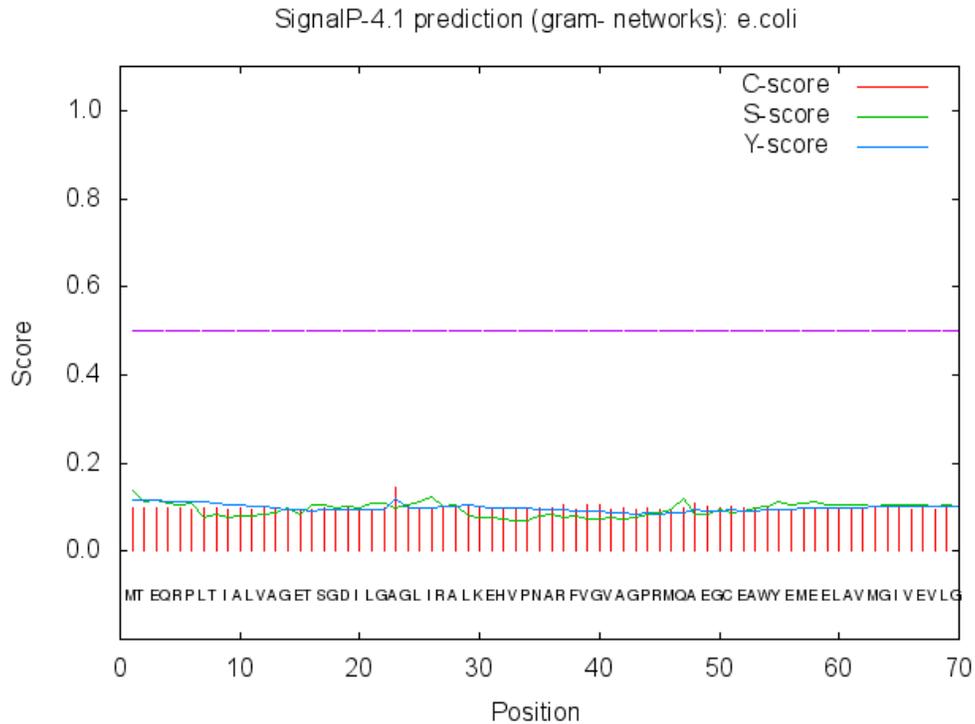


Figure 8. *E. coli* LpxB (top) and Mrub_2763 (bottom) do not contain TMH regions; a cytoplasmic location is predicted. THMM server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) created this hydropathy plot.



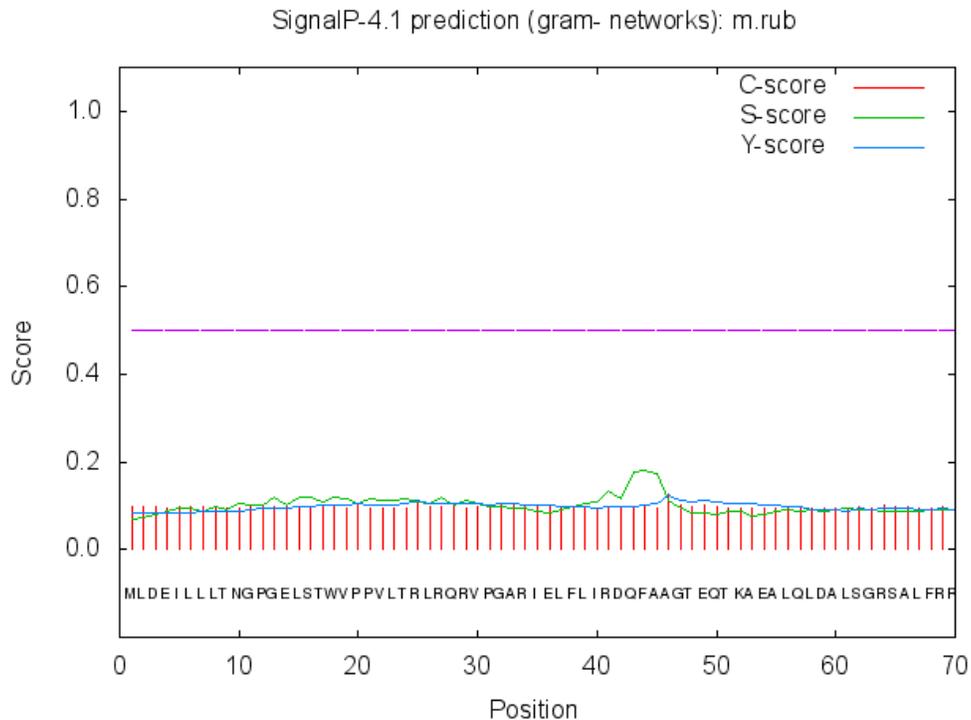


Figure 9. *E. coli* LpxB (top) and Mrub_2763 (bottom) do not contain signal peptides; no membrane integration is predicted. Signal P Server 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>) created these graphs.

The E.C. number is also depicted in Figure 3 using a KEGG pathway map of the Lipid A disaccharide biosynthesis pathway for both Mrub_2763 and *E. coli* LpxB. The same map came up for both *E. coli* and *M. ruber* with the lipid A disaccharide biosynthesis pathway being highlighted in green. Because there were no hits under Pfam for Mrub_2763, there is not an HMM logo for LpxB. It is important to note that the first two hits when BLASTing Mrub_2763 came up to be an enzyme that did not match LpxB. However, the enzyme is a synthetase just like LpxB and the gene product name for LpxB is seen further down the list as shown in Figure 10.

	Des	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	sugar synthetase [Meiothermus cerbereus]	627	627	99%	0.0	84%	WP_027877254.1
<input type="checkbox"/>	sugar synthetase [Meiothermus rufus]	504	504	100%	2e-174	71%	WP_027883039.1
<input type="checkbox"/>	sugar synthetase [Marinithermus hydrothermalis]	467	467	98%	8e-160	66%	WP_013703154.1
<input type="checkbox"/>	sugar synthetase [Meiothermus chliarophilus]	442	442	99%	1e-149	62%	WP_051304056.1
<input type="checkbox"/>	hypothetical protein [Meiothermus timidus]	441	441	99%	1e-149	62%	WP_018466809.1
<input type="checkbox"/>	sugar synthetase [Meiothermus silvanus]	432	432	98%	4e-146	62%	WP_013157365.1
<input type="checkbox"/>	glycosyl transferase family 19 [Truepera radiovictrix DSM 17093]	194	194	91%	9e-54	42%	ADI13450.1
<input type="checkbox"/>	hypothetical protein [Truepera radiovictrix]	167	167	86%	5e-44	41%	WP_049772908.1
<input type="checkbox"/>	glycosyl transferase family protein, lipid-A-disaccharide synthase [Armatimc	144	144	98%	5e-35	32%	KRT75645.1
<input type="checkbox"/>	hypothetical protein AMS17_07960 [Spirochaetes bacterium DG_61]	142	142	96%	2e-34	30%	KPJ87644.1
<input type="checkbox"/>	lipid-A-disaccharide synthase [Cyanotheca sp. PCC 7424]	142	142	96%	3e-34	32%	WP_015954781.1
<input type="checkbox"/>	lipid-A-disaccharide synthase [Synechocystis sp. PCC 7509]	133	133	96%	3e-31	30%	WP_028954382.1
<input type="checkbox"/>	lipid-A-disaccharide synthase [Stanieria cyanosphaera]	133	133	97%	3e-31	29%	WP_015194044.1
<input type="checkbox"/>	lipid-A-disaccharide synthase [Phormidium sp. OSCR]	133	133	97%	4e-31	32%	KPQ38085.1
<input type="checkbox"/>	hypothetical protein [Leptolyngbya boryana]	132	132	99%	4e-31	32%	WP_017290947.1
<input type="checkbox"/>	lipid-A-disaccharide synthase [Planktothrix agardhii]	132	132	99%	9e-31	28%	WP_042156703.1
<input type="checkbox"/>	lipid-A-disaccharide synthase [Raphidiopsis brookii]	131	131	96%	1e-30	30%	WP_009341484.1
<input type="checkbox"/>	hypothetical protein [Dactylococcopsis salina]	130	130	98%	3e-30	31%	WP_015229937.1

Figure 10. BLAST results of Mrub_2763. The first few hits are not the same as LpxB, but lower down the list is Lipid-A-disaccharide (LpxB) with significant E-values. NCBI blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) created these results.

Conclusion

LpxD

The hypothesis that Mrub_2768 is the *M. ruber* version of the LpxD gene in *E. coli* is supported through the various bioinformatics tests run. Module three of the experiment determined that both Mrub_2768 and LpxB were cytoplasmic and were not attached to the membrane. A BLAST of LpxB versus Mrub_2768 showed that the two had a relatively similar amino acid sequence. Although the first hit for *E. coli* LpxB in Pfam was not the same for Mrub_2763, the second hit did match showing that there is a connection. The two also had the same names and numbers for their COG groups, which is a very good indication that the two are related. TIGRfam also showed that they belonged to the same protein family and had significant E-values. Their E.C. numbers were exactly the same. The two genes did not have the same PDB name or code, however

LpxD was in the name of each. All of this evidence supports the hypothesis that Mrub_2768 is similar to LpxD and it can be safely concluded that they are orthologs.

LpxB

The hypothesis that Mrub_2763 is the *M. ruber* version of the LpxB gene in *E. coli* is inconclusive according to the various bioinformatics tests run. There was some evidence that did support the hypothesis. Module three of the experiment determined that both Mrub_2763 and LpxB were cytoplasmic and not attached to the membrane. A BLAST of Mrub_2763 resulted in hits that matched LpxB (Lipid A disaccharide synthase) just not LpxB for *E. coli* specifically as shown in Figure 10. The LpxB gene was highlighted in the KEGG pathway as shown in Figure 3 indicating that the enzyme exists in *M. ruber*. Their E.C. numbers were also exactly the same. The two had different COG numbers, but they had the same enzyme name (Lipid A disaccharide synthase) in both. LpxB for *E. coli* had “LpxB” in the COG name indicating that it could be *E. coli* specific and that is why the two COG numbers were different. There were no hits in the Pfam or PDB databases for Mrub_2763. TIGRFam had a hit, but it had no real name and was not significant. These results do not go against the hypothesis, but they do not support it either. The lack of hits from the databases does not refute the hypothesis, there is just no information in the databases on the LpxB enzyme yet. Therefore, there is not enough evidence to support or refute the hypothesis that Mrub_2763 is similar enough to LpxB to be considered orthologs; the results are inconclusive. Further research needs to be done after databases are more up to date because there is evidence that this enzyme exists in *M. ruber*.

References

- Anderson MS, Robertson AD, Macher I, Raetz CR. 1988. Biosynthesis of lipid A in *Escherichia coli*: identification of UDP-3-O-[(R)-3-hydroxymyristoyl]-alpha-D-glucosamine as a precursor of UDP-N2,O3-bis[(R)-3-hydroxymyristoyl]-alpha-D-glucosamine. *Biochemistry*. 22;27(6):1908-17.
- Blattner FR, Plunkett G, Bloch CA, Perna NT, Valerie B, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The Complete Genome Sequence of *Escherichia coli* K-12. *Science*. 277.
- Fichman Y, Gerdes SY, Kovács H, Szabados L, Zilberstein A and Csonka LN. 2014. Evolution of proline biosynthesis: Enzymology, bioinformatics, genetics, and transcriptional regulation. *Biol Rev*.
- Pujari, S. “Bioinformatics: A useful essay on bioinformatics & biotechnology.” n.d. (Internet address: <http://www.yourarticlelibrary.com/essay/bioinformatics-an-useful-essay-on-bioinformaticsbiotechnology/29374/> ; accessed 12-15-2015).
- Raetz CRH, Guan Z, Ingram BO, Six DA, Song F, Want X, Zhao J. 2009. Discovery of new biosynthetic pathways: the lipid A story. *J Lipid Res*. 50. From: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2674688/>

Ray BL, Painter G, Raetz CRH. 1984. The biosynthesis of gram-negative endotoxin. Formation of lipid A disaccharides from monosaccharide precursors in extracts of *Escherichia coli*. *J Biol Chem.* 259(8): 4852-9. From: <http://www.ncbi.nlm.nih.gov/pubmed/6370995>

Slonczewski JL, Foster JW. 2014. *Microbiology: An Evolving Science*. 3rd Edition. W. W. Norton Company, Inc.

Tindall B, Sikorski J, Lucas S, Goltsman E, Copeland A, Del Rio TG, Nolan M, Tice H, Cheng J, Han C, Pitluck S, Liolios K, Ivanova N, Mavromatis K, Ovchinnikova G, Pati A, Fährnich R, Goodwin L, Chen A, Palaniappan K, Land M, Hauser L, Chang Y, Jeffries CD, Rohde M, Göker M, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk H, and Lapidus A. 2010. Complete genome sequence of *meiothermus ruber* type strain. *Standards in Genomic Sciences.* 3(26).