2018

Meiothermus ruber mrub_0320 gene is an ortholog of the b3452 gene, mrub_0321 gene is an ortholog of the b3451 gene, mrub_0322 gene is an ortholog of the b3453 gene, mrub_2366 gene is an ortholog of the b3450 gene found in *Escherichia coli*, which encode for components of an ABC transporter involved in sn-glycerol - 3-phosphate

Jenna Hall
Augustana College, Rock Island Illinois

Dr. Lori Scott
Augustana College, Rock Island Illinois

Meiothermus ruber mrub_0320 gene is an ortholog of the b3452 gene, mrub_0321 gene is an ortholog of the b3451 gene,mrub_0322 gene is an ortholog of the b3453 gene, mrub_2366 gene is an ortholog of the b3450 gene found in *Escherichia coli*,which encode for components of an ABC transporter involved in *sn*-glycerol - 3-phosphate

Augustana College

Jenna Hall and Joel Nelson

Introduction

Membrane Transport is crucial for all organisms. Transport systems bring in external nutrients and eliminates harmful toxins. These transporters are both importers and exporters. Transporters are diverse since they must achieve many different tasks. ATP-binding cassettes ,ABC transporters, have ATP binding domains with the presence of a a phosphate-binding loop and a short consensus sequence “LSGGQ” (1). These transporters are characterized by the ATP binding subunit. An ABC transporters are integral membrane proteins that have two nucleotide-binding domains (NBD) and two transmembrane domains (TMDs). ABC transporters are located in the plasma membrane but often go through conformational changes thus allowing them to change directions of transport. ABC transports are increasingly significant as they make up the largest protein family. In humans there are 48 ABC transporters which carry importance since they have been linked to many disease such as cystic fibrosis and cancer.

Glycerol-3-phosphate (Gro3P) is participates in glycolysis, gluconeogenesis, lipid synthesis and Gro3P electron transfer to mitochondria (2). Gro3P transport plays a key role in metabolic regulation. New research has identified that there is a Glycerol- 3-phosphate phosphatase (Gro3PP) that converts Gro3P to glycerol. This research can now be used to identify treatments for diabetic individuals.

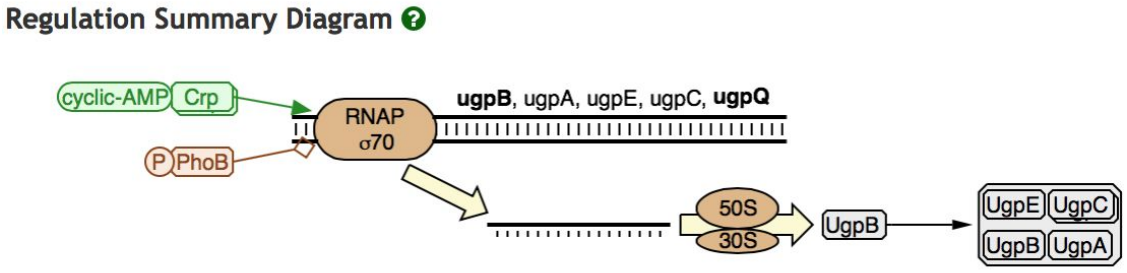


Figure 1: Pathway of the glycerol-3-phosphate transporters in *E.coli*. (3)
This supports *ugpB, ugpA, ugpE,* and *ugpC* are a part of an operon. To promote transcription of the operon cyclic-AMP Crp, P-PhoB. The genes ugpB, ugpA, ugpE, ugpC are a part of an operon as supported by the IGM/M database results. The 50s and 30s represent ribosomal subunits which all for translation to occur. The end results is the formation of proteins UgpA, UgpB, UgpC and UgpE.

The *E.coli* K12 MG1655 locus tags involved with glycerol-3-phosphate transport are b3450 ,b3451,b3452 and b3453 (3). Throughout this paper locus tags beginning with b indicate an *Escherichia coli* gene. The genes involved in this system are *ugpA, ugpB, ugpC,* and *ugpE* which can be identified in Figure 1. These genes compose an operon as seen in Figure 1. The proteins involved are UgpA, UgpB, UgpC and UgpE. These genes are found in *E. coli* and in *M.ruber (26).* The *ugpA* is b3452 in *E.coli* and mrub_0320 in *M.ruber* which functions as the an ABC transporter permease. The *ugpB* is b3453 in *E.coli* and mrub_0322 in *M.ruber* which functions as the ABC transporter periplasmic binding protein. The *ugpC* is b3450 which functions as the ABC transporter ATPase. The *ugpE* is b3451 in *E.coli* and mrub_321 in *M.ruber* which functions as an ABC transporter permease.

Meiothermus ruber is the study organism used in this research to compare to *Escherichia coli. M.ruber* is studied because the organism has phylogenetic diversity with respect to *Escherichia coli* . Using *M. ruber* as a study model organism has many benefits, which include improved identification of orthologs, improved phylogenetic anchoring, gene discovery, and classification of microbes (4). *M. ruber* physical features allow for effective

research (5). One feature is its ability to grow at high temperatures allowing for flexibility in research conditions. Since the organism is a thermophile its able to resist denaturing thus allowing for better results (7). It is important to study these ABC transporters because they are used in metabolism of drugs. This research can be related to how humans become resistant to certain drugs(10).

Purpose

The purpose of this study is to identify orthologs between *E.coli* and *M.ruber* within the glycerol-3-phosphate transport system. It is proposed that mrub_0320 gene is an ortholog of the b3452 gene, mrub_0321 gene is an ortholog of the b3451 gene, mrub_0322 gene is an ortholog of the b3453 gene, mrub_2366 gene is an ortholog of the b3450 gene found in Escherichia coli coding for Glycerol-3-phosphate. Once orthologous genes are confirmed, we can then begin to look into amino acid substitutions within these proteins (specifically alanine) to see how they affect the functionality of the protein (6).

Material and Methods

To begin the comparison between the *E. coli* and *M. ruber* genes of interest, a BLAST was performed, comparing the amino acid sequence of both in order to determine whether they were orthologous to one another (11, 12). This produce some an E-value, as well as a Bit Score, which helped determine exactly how similar the two genes were to each other.

We then began the process of confirming the location of the *M. ruber* GOI's start codon. The *E. coli* GOI did not have to go through this process, because their start codon has been previously confirmed. IMG/M (27), BLAST (11, 12), T-Coffee (29), and Weblogo (30) were used to look at the possible open-reading frames, obtain a variety of similar sequences from other known species, compare with a multiple sequence alignment, and to create a visual

representation of this visual comparison to understand the conservation of the amino acids, respectively. With the combination of these programs, and the final visual of results, we are able to have a better understanding of how conserved these genes are relative to other species.

To have a better understanding of the possible orthologs between the *E. coli* and *M. ruber* GOI, we used multiple programs to determine more about the protein structure and function. To begin, TMHMM was used to determine how many, if at all, transmembrane domains there were for each protein (15). SignalP (19) was used to estimate if a signal peptide was a part of the protein, and LipoP (20) was used to see if there was a signal peptide or lipoprotein as a part of the structure of the protein. P-SORT B (13) and Phobius (14) we used to identify the probability of where exactly these proteins were located within the cell membrane. We then looked at the protein families of our GOI's proteins were a part of. CDD (31) produced a possible member of the COG family, Pfam (22, 23) produced a possible Pfam family connection, TIGRfam (21) produced a possible TIGRfam family connection, and finally PDB (24, 25) was able to give us a possible protein domains that were connected to our proteins of interest.

Lastly, the possibility of these genes being in an operon was researched. IMG/M (27) database was used, and with the help of KEGG (26), images were produced in order to have a better idea of the function and location of the genes. Top COG hits were also used and compared in order to have better confirmation of a presence of an operon (27). In all these comparisons, each of these steps were taken to get a complete viewpoint of the compared genes.

Results

Comparing *E.coli* b3450 and *M. ruber* Mrub_ 2366

When forming a comparison between the pairs of genes in question, tables were formed, like Table 1, which summarizes the results obtained from numerous bioinformatic tools to compare

E.coli b3450 and mrub_ 2366. When looking at the comparison of E-values, the closer to 0.0 indicates that the genes are unlikely to be similar by chance. The original BLAST performed was against the two specific genes of interest, b3450 and mrub_2366 (11). The COG number identified through the CDD database was the same for both genes, with slightly varied E-values (31). Using the cellular localization databases (TMHMM (15), SignalP (19), LipoP (20), PSORT-B (13), and Phobius (14)) both genes were confirmed to have the same location. Both TIGRfam (21) and Pfam (22, 23) database show the same protein family between the two genes, with similar E-values. Protein Database showed different proteins of similarities (24, 25). For both *E. coli* and *M. rub* systems, the genes were found within the ABC transporter page (KEGG map 02010) (26).

Bioinformatics Tool Used	**b3450**	**mrub_2366**
Blast against opposite genome	Score: 357 E-value:2e-123	
CDD Data	Cog Number: COG3839 ABC-type sugar transport system, ATPase component [Carbohydrate transport and metabolism]	
	E-value: 1.89e-180	E-value: 0
Cellular Localization	Embedded in the inner plasma membrane	
TIGRfam - Protein Family	TIGR00968 sulfate ABC transporter, ATP-bin	
	E-value: 8.9e-94	E-value: 5.5e-90
Pfam - Protein Family	PF00005 ABC_trans	
	E-value: 4.4e-31	E-value: 1.3e-28
Protein Database	Entity 3 containing Chain S, T E-value: 3.72838E-99	Sugar binding protein E-value: 3.7532E-111
KEGG Pathway Map	ABC transporters (02010)	

Table 1. Comparison of b3450 and mrub_2366

To begin the analysis of these sets of gene in the sn-glycerol-3-phosphate ABC transporter pathway, multiple databases were used. Through the use of KEGG, the amino acid sequences of both b3450 and mrub_2366 were found and used for analysis (26). A slight variation in this specific analysis comes from the fact that mrub_2366 does not seem to be a part of the *sn*-glycerol-3-phosphate specifically, as seen in Figure 2. Through further research into the KEGG data, as well as a BLAST was performed to identify what possible gene could be a part of this system,, mrub_2366 was found to be the ATP-binding domain for many different monosaccharides, including *sn*-glycerol-3-phosphate (11).

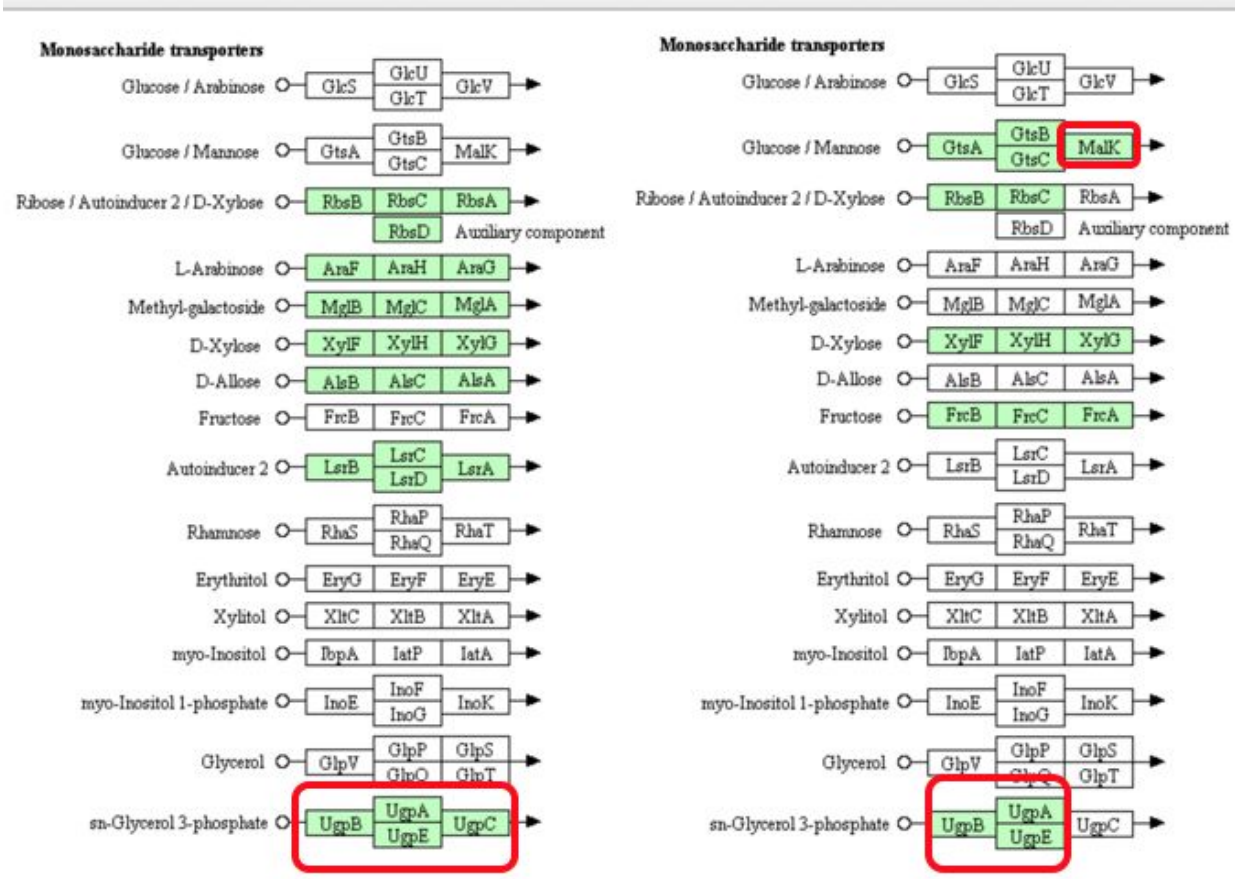
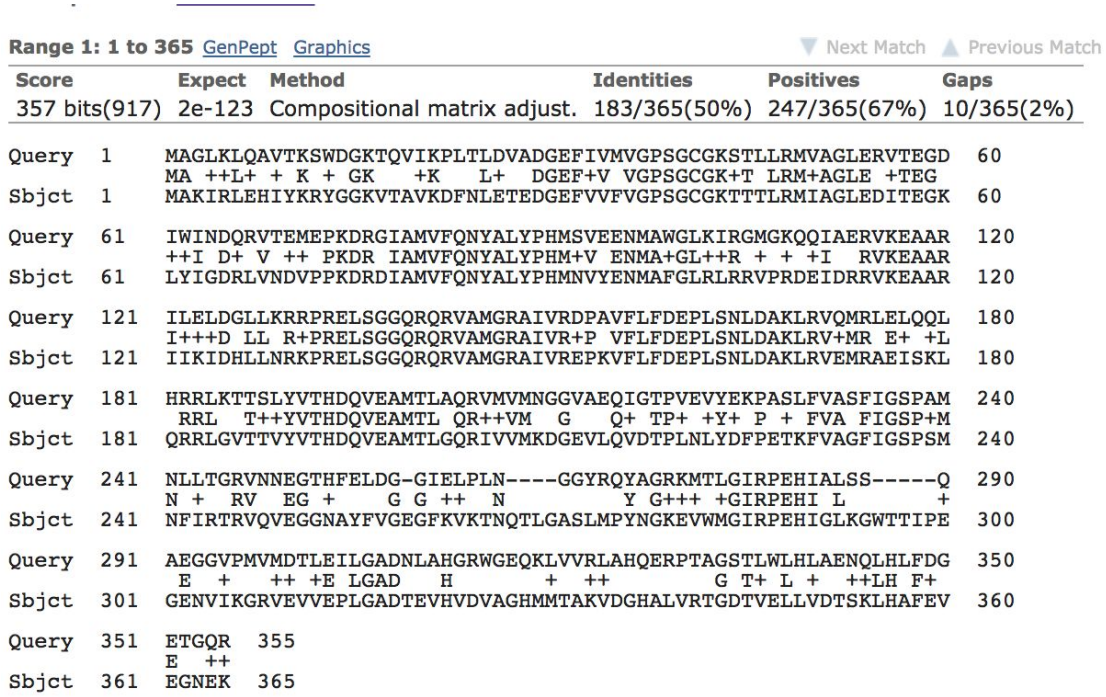


Figure 2. KEGG map of *E. coli* and *M. ruber* monosaccharide ABC transporters. Shown above are the structures for the ABC transporters of various monosaccharides in both the *E. coli* (left) and *M. ruber* (right) genomes. Our GOIs are found within UgpB (b3453), UgpA (b3452), UgpE (b3451), and UgpC (b3450) in the *E. coli* genome, and UgpB (mrub_0322), UgpA (mrub_0321), UgpE (mrub_0320), and MalK (mrub_2366) in the *M. ruber* genome, all circled in red(26).

Through the use of the BLAST database the images in Figure 3 are produced(11). The sequence alignment between b3450 and mrub_2366 shows the similarities in their amino acid sequences. The query lines represent the amino acid sequence of b3450 and subject line represents mrub_2366. The line between the two shows the highly conserved amino acids with that amino acid in the space. The addition sign indicates that the two amino acids are similar in functional groups. The E-value is 2e-123 which is very low indicating that the similarity in results are not due to chance. This similarity in amino acid sequences suggest that b3450 and mrub_2366 are orthologs.



Range 1: 1 to 365 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
357 bits(917)	2e-123	Compositional matrix adjust.	183/365(50%)	247/365(67%)	10/365(2%)

```
Query  1    MAGLKLQAVTKSWDGKTQVIKPLTLDVADGEFIVMVGPSGCGKSTLLRMVAGLERVTEGD  60
            MA ++L+ + K + GK   +K   L+  DGEF+V VGPSGCGK+T LRM+AGLE +TEG
Sbjct  1    MAKIRLEHIYKRYGGKVTAVKDFNLETEDGEFVVFVGPSGCGKTTTLRMIAGLEDITEGK  60

Query  61   IWINDQRVTEMEPKDRGIAMVFQNYALYPHMSVEENMAWGLKIRGMGKQQIAERVKEAAR  120
            ++I D+ V ++ PKDR IAMVFQNYALYPHM+V ENMA+GL++R + + +I  RVKEAAR
Sbjct  61   LYIGDRLVNDVPPKDRDIAMVFQNYALYPHMNVYENMAFGLRLRRVPRDEIDRRVKEAAR  120

Query  121  ILELDGLLKRRPRELSGGQRQRVAMGRAIVRDPAVFLFDEPLSNLDAKLRVQMRLELQQL  180
            I+++D LL R+PRELSGGQRQRVAMGRAIVR+P VFLFDEPLSNLDAKLRV+MR E+ +L
Sbjct  121  IIKIDHLLNRKPRELSGGQRQRVAMGRAIVREPKVFLFDEPLSNLDAKLRVEMRAEISKL  180

Query  181  HRRLKTTSLYVTHDQVEAMTLAQRVMVMNGGVAEQIGTPVEVYEKPASLFVASFIGSPAM  240
             RRL  T++YVTHDQVEAMTL QR++VM  G   Q+ TP+ +Y+ P + FVA FIGSP+M
Sbjct  181  QRRLGVTTVYVTHDQVEAMTLGQRIVVMKDGEVLQVDTPLNLYDFPETKFVAGFIGSPSM  240

Query  241  NLLTGRVNNEGTHFELDG-GIELPLN----GGYRQYAGRKMTLGIRPEHIALSS-----Q  290
            N +  RV  EG +    G G ++  N         Y G+++ +GIRPEHI L       +
Sbjct  241  NFIRTRVQVEGGNAYFVGEGFKVKTNQTLGASLMPYNGKEVWMGIRPEHIGLKGWTTIPE  300

Query  291  AEGGVPMVMDTLEILGADNLAHGRWGEQKLVVRLAHQERPTAGSTLWLHLAENQLHLFDG  350
             E  +   ++ +E LGAD   H       +  ++        G T+ L +  ++LH F+
Sbjct  301  GENVIKGRVEVVEPLGADTEVHVDVAGHMMTAKVDGHALVRTGDTVELLVDTSKLHAFEV  360

Query  351  ETGQR  355
            E  ++
Sbjct  361  EGNEK  365
```

Figure 3. The BLAST database showed that when the protein sequence of b3450 is BLAST'ed against *Meiothermus ruber* the results show mrub_ 2366 has a low E-value indicating similarities within the two protein amino acid sequences, showing a high chance for orthologous genes between the two species (11).

Having confirmed that b3450 and mrub_2366 have significant sequence similarity, the focus

then turned to learning more about these proteins in regards to cell localization. In order to identify whether or not these proteins have transmembrane domains, the TMHMM database was used (15-18). Figure 4 shows the results of the TMHMM analysis. Using the known amino acid sequences, the database calculated which sections would be found throughout the cellular membrane. In the case of b4350 and mrub_2366, there were no transmembrane domains found. This result is expected since the function of both of these genes involve binding to ATP. In order to bind, they need to be located outside of the membrane.

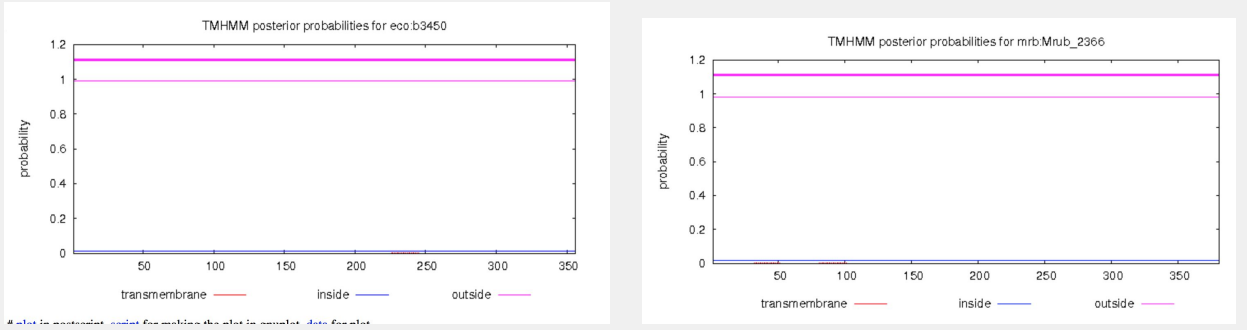


Figure 4. The bioinformatics database TMHMM was used to develop two graphs indicating cellular localization of b3450 and mrub_ 2366. The left image indicates the *E. coli* b3450 and the right image indicates mrub_ 2366. The similarities between the two graphs indicate a similarity in cellular localization between genes. Although these two graphs show no deviation, if there were any deviations present through the graph, it would indicate the presence of transmembrane helices. (15-18)

Also important for cell localization, is understanding whether our GOI create proteins that contain signal peptides, or some derivative of them. For this question, the database SignalP was used (19,20). Again, the amino acid sequences of b3450 and mrub_2366 were used within the database. Figure 5 shows the results of both of these analyses. In both cases, there is minimal strength of signal peptide probability. The SignalP results indicate that there is too low of a probability for both genes to have signal peptides in their protein sequences.

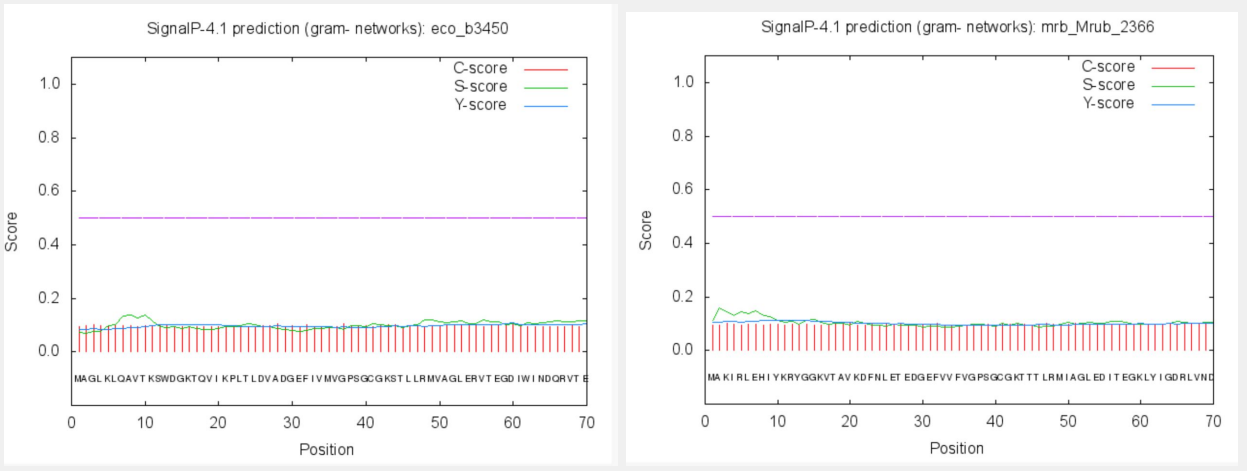


Figure 5. The bioinformatics database SignalP was used to develop two graphs indicating cellular localization of b3450 and mrub_ 2366. The left image indicates the *E. coli* b3450 and the right image indicates mrub_ 2366. The similarities between the two graphs indicate a similarity in cellular localization between genes. Strong signals in these graphs would indicate the possible presence of a signal peptide on the proteins. (19-20)

To help in the conclusion of cellular location of these proteins, PSORT-B and Phobius were two other databases used (13,14). Again, amino acid sequences of b3450 and mrub_2366 were used. The databases would use theoretical amino acid sequences to determine possible intra-protein interactions that would determine the location relative to the cell membrane. With both of these genes, the PSORT-B indicated that both proteins were located on or around the cytoplasmic membrane. The Phobius graphs, shown in Figure 6, also indicate that these proteins are located in the non-cytoplasmic region. The combination of these two results help confirm that b3450 and mrub_2366 are truly ATP-binding domains, that are located directly on the cytoplasmic membrane, with a very small amount actually located in the cytoplasm of the cell.

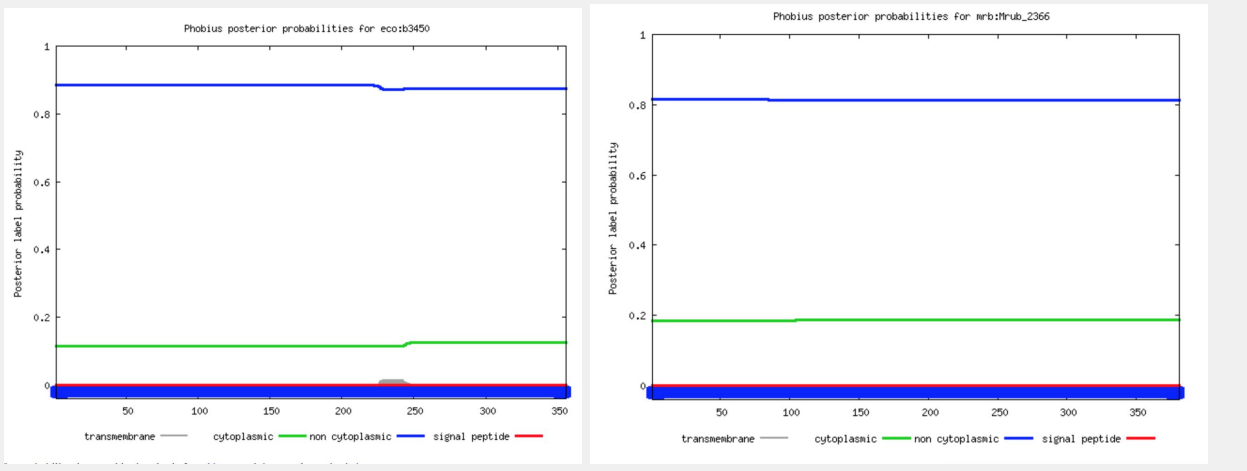


Figure 6. The bioinformatics database Phobius was used to develop two graphs indicating cellular localization of b3450 and mrub_ 2366. The left image indicates the *E. coli* b3450 and the right image indicates mrub_ 2366. The similarities between the two graphs indicate a similarity in cellular localization between genes(14).

With a better idea of where b3450 and mrub_2366 are localized as proteins, the Pfam database was used in order to further clarify how similar the two genes are (22). As seen in Figure 7, when comparing b3450 and mrub_2366 to a consensus sequence, there is a large amount of similarity, which can be seen by the large amount of shades of green found in the 4th row of each sequence. This helps confirm that the two GOI are orthologous to one another.

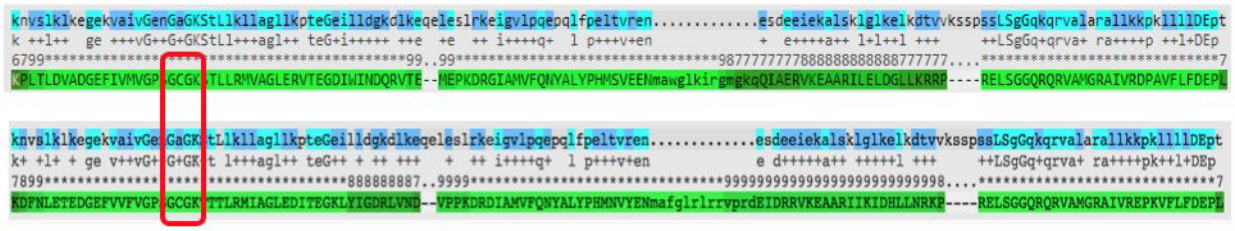


Figure 7. The database Pfam was used to create aligned sequences of b3450 and mrub_ 2366 for comparison. The top sequence is b3450 and the bottom is mrub_2366. An example of conserved amino acid sequences can be seen in the red box. As seen by the green line of both images the two sequences which supports b3450 and mrub_ 2366 as orthologs (22).

Through the IMG/M database the genes b3450 and mrub_2366 were analyzed using color by KEGG as shown in Figure 8 (26,27). Color by KEGG coordinates function of genes with

colors. The two GOI are indicated by the red line above the gene. The top panel is the *E.coli* gene b3450 and the bottom is mrub_2366 which are colored green and purple respectively. Both of these colors indicate they are involved in membrane transport. This confirms these genes are ABC transporters and that they are orthologs since they have the same function.

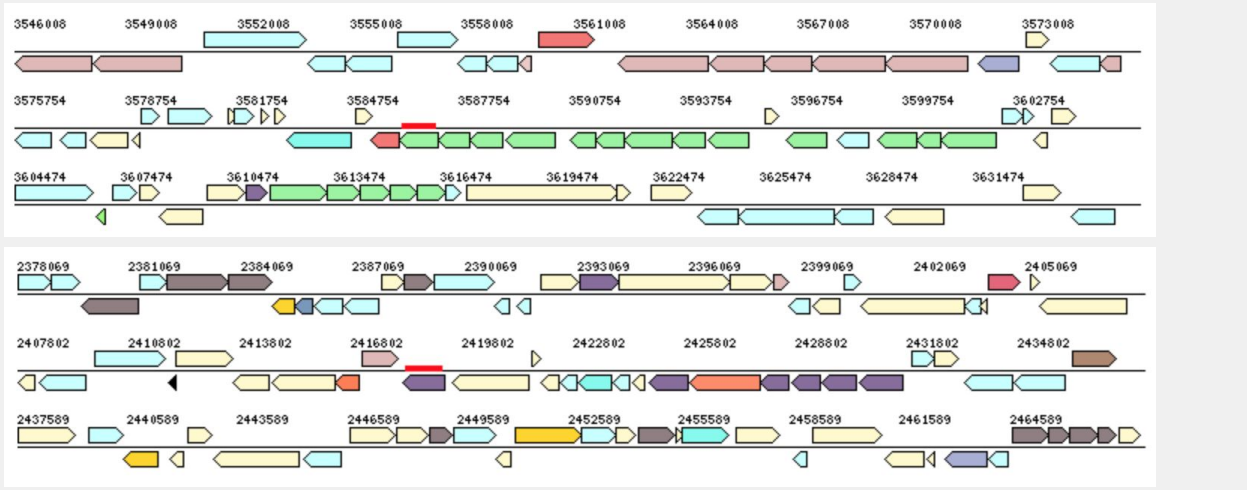


Figure 8. The IMG/M database and the color by KEGG function was used to analyze b3450 and mrub_ 2366. The top image represents b3450 and the bottom image represents mrub_2366. The red line indicates the gene of interest. The color of the genes represent their function. The *E.coli* green gene and the *M. ruber* purple gene represent the fact that these gene's function involves membrane transport (26,27).

Comparing *E.coli* b3451 and *M. ruber* mrub_0321

We continue to do the same process defined above with a new set of genes, b3451 and mrub_0321, both of which are classified as UgpE in their respective genomes. Again, for ease of understanding, Table 2 summarizes the results obtained from numerous bioinformatic tools to compare *E.coli* b3451 and mrub_0321. When looking at the comparison of E-values, the closer to 0.0 indicates that the genes are unlikely to be similar by chance. The original BLAST performed was against the two specific genes of interest, b3451 and mrub_0321 (11). The COG number identified through the CDD database was the same for both genes, with slightly varied

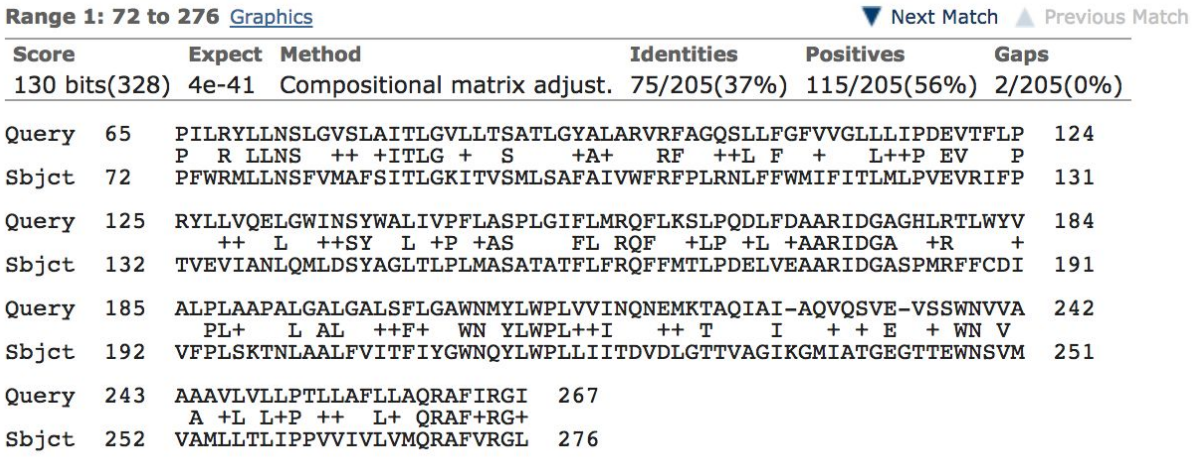
E-values (31). Using the cellular localization databases (TMHMM (15), SignalP (19), LipoP (20), PSORT-B (13), and Phobius (14)), both genes were confirmed to have the same location. TIGRfam for b3451 did not identify a protein family, but it did for mrub_0321 (21). Pfam database show the same protein family between the two genes, with similar E-values (21, 22). Protein Database showed identical proteins for both genes (24, 25). For both *E. coli* and *M. rub* systems, the genes were found within the ABC transporter page (KEGG map 02010) (26).

Bioinformatics Tool Used	**E.coli Gene b3451**	**mrub_0321**
BLAST genome	Score: 130 bits E-value: 4e-41	
CDD Data	COG Number: COG0345 ABC-type glycerol-3-phosphate transport system, permease component	
	E-value: 5.52e-77	E-value: 1.29e-57
Cellular Localization	Transmembrane in the inner plasma membrane	
TIGRfam- Protein Family	No TIGRfam found	TIGR01581 NifC-like ABC-type porter
	E-value: N/a	E-value: 6.1e-07
Pfam- Protein Family	Pf00528 BPD_trans_1	
	E-value: 2.1e-22	E-value: 2.1e-17
Protein Database	4TQU - Entity 2 containing Chain N	
	E-value:7.90769E-7	E-value: 6.3694E-8
KEGG Pathway Map	ABC transporters (02010)	

Table 2. Comparison of b3451 and mrub_0321

Through the use of KEGG, the amino acid sequences of both b3451 and mrub_0321 were found and used for analysis (26). This time, they are clearly identified in Figure 2 in the *sn*-glycerol-3-phosphate section of the pathway, marked as UgpE in their respective genomes.

Through the use of the BLAST database the images in Figure 9 are produced (11). The sequence alignment between b3451 and mrub_0321 shows the similarities in their amino acid sequences. The query lines represent the amino acid sequence of b3451 and subject line represents mrub_0321. The line between the two shows the highly conserved amino acids with that amino acid in the space. The addition sign indicates that the two amino acids are similar in functional groups. The E-value, which is close to zero, indicates that the similarity in results are not due to chance. This similarity in amino acid sequences suggest that b3451 and mrub_0321 are orthologs.



Range 1: 72 to 276 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
130 bits(328)	4e-41	Compositional matrix adjust.	75/205(37%)	115/205(56%)	2/205(0%)

```
Query  65   PILRYLLNSLGVSLAITLGVLLTSATLGYALARVRFAGQSLLFGFVVGLLLIPDEVTFLP  124
            P  R LLNS  ++ +ITLG +  S    +A+   RF  ++L F  +   L++P EV   P
Sbjct  72   PFWRMLLNSFVMAFSITLGKITVSMLSAFAIVWFRFPLRNLFFWMIFITLMLPVEVRIFP  131

Query  125  RYLLVQELGWINSYWALIVPFLASPLGIFLMRQFLKSLPQDLFDAARIDGAGHLRTLWYV  184
             ++  L  ++SY  L +P +AS    FL RQF  +LP +L +AARIDGA  +R    +
Sbjct  132  TVEVIANLQMLDSYAGLTLPLMASATATFLFRQFFMTLPDELVEAARIDGASPMRFFCDI  191

Query  185  ALPLAAPALGALGALSFLGAWNMYLWPLVVINQNEMKTAQIAI-AQVQSVE-VSSWNVVA  242
              PL+   L AL  ++F+  WN YLWPL++I   ++ T    I   + + E  + WN V
Sbjct  192  VFPLSKTNLAALFVITFIYGWNQYLWPLLIITDVDLGTTVAGIKGMIATGEGTTEWNSVM  251

Query  243  AAAVLVLLPTLLAFLLAQRAFIRGI  267
             A +L L+P ++  L+ QRAF+RG+
Sbjct  252  VAMLLTLIPPVVIVLVMQRAFVRGL  276
```

Figure 9. The BLAST database showed that when the protein sequence of b3451 is BLAST'ed against *Meiothermus ruber* the results show mrub_ 0321 has a low E-value indicating similarities within the two protein amino acid sequences, showing a high chance for orthologous genes between the two species (11).

Having confirmed that b3451 and mrub_0321 have significant sequence similarity, the focus then turned to learning more about these proteins in regards to cell localization. In order to identify whether or not these proteins have transmembrane domains, the TMHMM database was used (15-18). Figure 10 shows the results of the TMHMM analysis for this gene comparison. Using the known amino acid sequences, the database calculated which sections would be found throughout the cellular membrane. In the case of b4351 and mrub_0321, there

were 6 transmembrane domains found. This result is expected for these two genes, because they have been previously identified as the permease component of the ABC transporter, meaning that there needs to be transmembrane domains in order the protein to transport products (26).

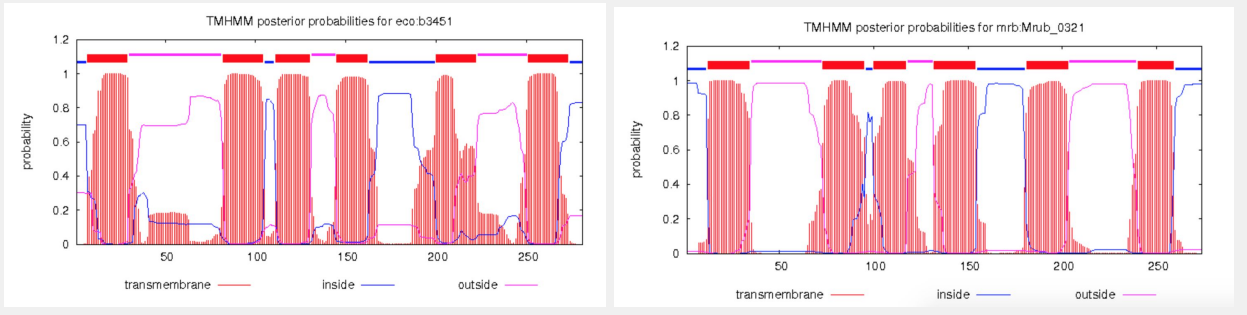


Figure 10. The bioinformatics database TMHMM was used to develop two graphs indicating cellular localization of b3451 and mrub_ 0321. The left image indicates the *E. coli* b3451 and the right image indicates mrub_ 0321. The similarities between the two graphs indicate a similarity in cellular localization between genes. Both graphs show clear deviations in the graph that define a transmembrane domain within the protein species. A total number of 6 can be found on each graph (15-18).

Also important for cell localization, is understanding whether our GOI create proteins that contain signal peptides, or some derivative of them. For this question, the database SignalP was used (19,20). Again, the amino acid sequences of b3451 and mrub_0321 were used within the database. Figure 11 shows the results of both of these analyses. In both cases, there is minimal strength of signal peptide probability. The SignalP results indicate that there is too low of a probability for both genes to have signal peptides in their protein sequences.

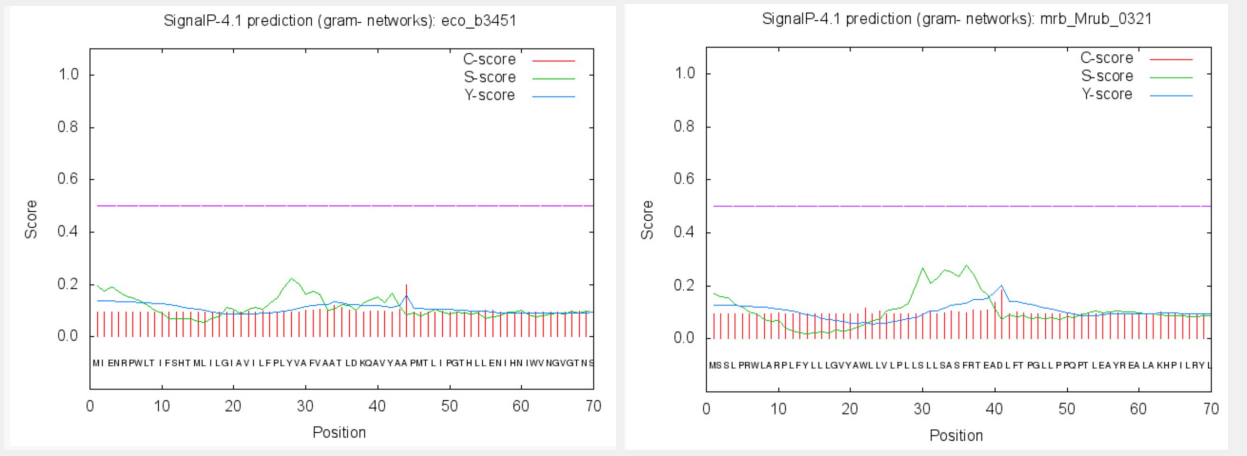
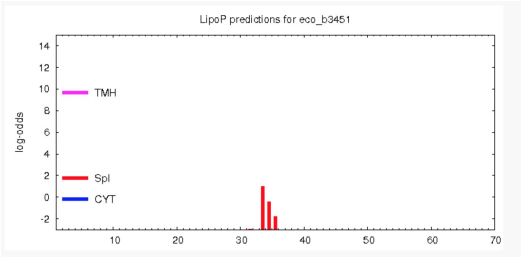


Figure 11. The bioinformatics database SignalP was used to develop two graphs indicating cellular localization of b3451 and mrub_ 0321. The left image indicates the *E. coli* b3451 and the right image indicates mrub_ 0321. The similarities between the two graphs indicate a similarity in cellular localization between genes. Strong signals in these graphs would indicate the possible presence of a signal peptide on the proteins, which is clearly not seen in both of these cases (19-20).

To confirm the previous conclusion of no signal peptide, another database was used, known as LipoP (20). This database is similar to SignalP, but focuses more clearly on lipoproteins. Figure 12 shows the LipoP results from both b3451 and mrub_0321. Although the results from b3451 do show some higher probability levels for signal peptides in the form of Signal Peptide I (SpI), it is still not high enough to be very probable. The probability for mrub_0321 is much, much lower.




```
# mrb_Mrub_0321 TMH score=7.46796 margin=7.668873
# Cut-off=-3
mrb_Mrub_0321   LipoP1.0:Best   TMH     1       1       7.46796
mrb_Mrub_0321   LipoP1.0:Margin TMH     1       1       7.668873
mrb_Mrub_0321   LipoP1.0:Class  CYT     1       1       -0.200913



# NO PLOT made - less than 4 putative cleavage sites predicted
```

Figure 12. The bioinformatics database LipoP was used to develop two graphs indicating probability of lipoproteins for b3451 and mrub_ 0321. The left image indicates the *E. coli* b3451 and the right image indicates mrub_ 0321. The b3451 graph shows a very slight probability for a signal protein to be present, but the data above the graph shows the probability is still relatively low. In the case of mrub_0321, there was no graph present, which indicates an extremely low chance of any signal peptide or lipoprotein to be present. Strong signals in these graphs would indicate the possible presence of a lipoprotein as a part of the proteins, which is clearly not seen in these cases (20).

To help in the conclusion of cellular location of these proteins, PSORT-B and Phobius were two other databases used (13,14). Again, amino acid sequences of b3451 and mrub_0321 were used. The databases would use theoretical amino acid sequences to determine possible intra-protein interactions that would determine the location relative to the cell membrane. With both of these genes, the PSORT-B indicated that both proteins were located on or around the cytoplasmic membrane. The Phobius graphs, shown in Figure 13, also indicates that these proteins are located both in the non-cytoplasmic and cytoplasmic region, further confirming the presence of transmembrane domains. The combination of these two results help confirm that b3451 and mrub_0321 are truly permease domains, that are located throughout the inner membrane of the respective bacteria.

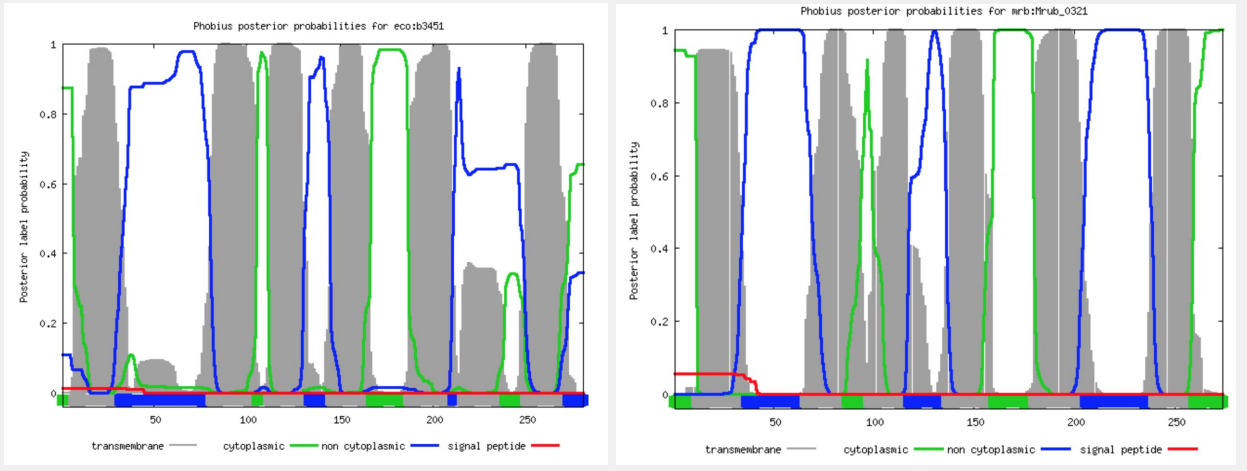


Figure 13. The bioinformatics database Phobius was used to develop two graphs indicating cellular localization of b3451 and mrub_ 0321. The left image indicates the *E. coli* b3451 and the right image indicates mrub_ 0321. The similarities between the two graphs indicate a similarity in cellular localization between genes. The constant change between cytoplasmic (green) and non cytoplasmic (blue) helps confirm the presence of transmembrane domains located on the inner bacterial membrane (14).

With a better idea of where b3451 and mrub_0321 are localized as proteins, the Pfam database was used in order to further clarify how similar the two genes are (22). As seen in Figure 14, when comparing b3451 and mrub_0321 to the same consensus sequence, there is a large

amount of similarity, which can be seen by the large amount of shades of green found in the 4th row of each sequence. This helps confirm that the two GOI are orthologous to one another.

Figure 14. The database Pfam was used to create aligned sequences of b3451 and mrub_ 0321 for comparison. The top sequence is b3451 and the bottom is mrub_0321. An example of conserved amino acid sequences can be seen in the red box. As seen by the green line of both images the two sequences which supports b3451 and mrub_ 0321 as orthologs (22).

Through the IMG/M database the genes b3451 and mrub_0321 were analyzed using color by KEGG as shown in Figure 15 (26,27). Color by KEGG coordinates function of genes with colors. The two GOI are indicated by the red line above the gene. The top panel is the *E.coli* gene b3451 and the bottom is mrub_0321 which are colored green and purple respectively. Both of these colors indicate they are involved in membrane transport. This confirms these genes are ABC transporters and that they are orthologs since they have the same function.

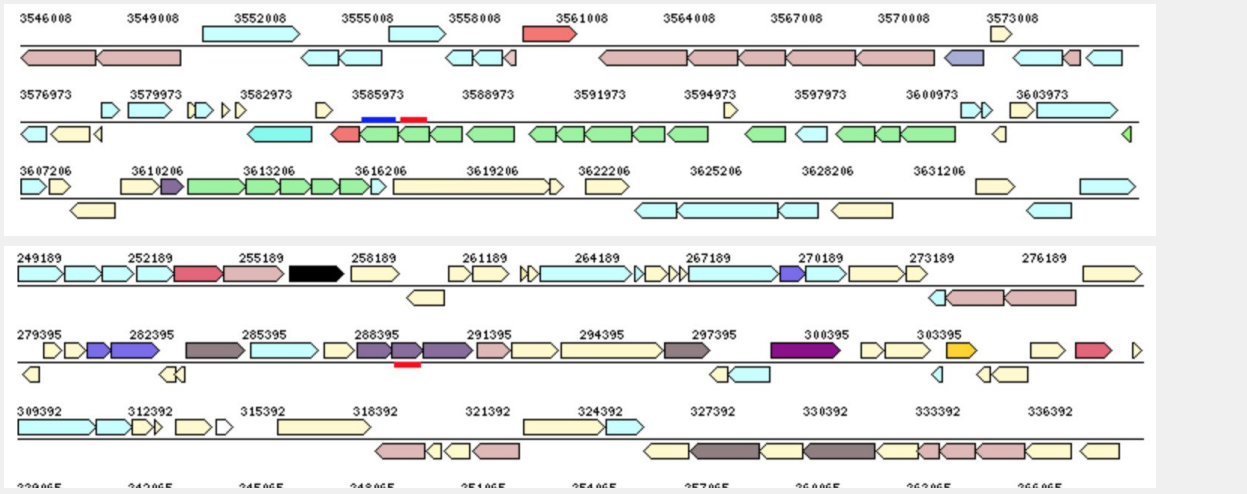


Figure 15. The IMG/M database and the color by KEGG function was used to analyze b3451 and mrub_ 0321. The top image represents b3451 and the bottom image represents mrub_0321. The red line indicates the gene of interest. The color of the genes represent their function. The *E.coli* green gene and the *M. ruber* purple gene represent the fact that both of these gene's function involves membrane transport (26,27).

Comparing *E. coli* b3452 and *M. ruber* mrub_0320

We continue to do the same process defined above with a new set of genes, b3451 and mrub_0321, both of which are classified as UgpE in their respective genomes. Again, for ease of understanding, Table 3 summarizes the results obtained from numerous bioinformatic tools to compare *E.coli* b3452 and mrub_ 0320. When looking at the comparison of E-values, the closer to 0.0 indicates that the genes are unlikely to be similar by chance. The original BLAST performed was against the two specific genes of interest, b3452 and mrub_0320 (11). The COG number identified through the CDD database was the same for both genes, with slightly varied E-values (31). Using the cellular localization databases (TMHMM (15), SignalP (19), LipoP (20), PSORT-B (13), and Phobius (14)), both genes were confirmed to have the same location. TIGRfam has varied results between the two genes, with two different TIGRfam numbers showing up (21). Pfam database shows the same protein family between the two genes, with similar E-values (22, 23). Protein Database showed different proteins of similarities (24, 25). For

both *E. coli* and *M. rub* systems, the genes were found within the ABC transporter page (KEGG map 02010) (26).

Bioinformatics Tool Used	**E.coli Gene b3452**	**mrub_0320**
BLAST genome	Score: 134 bits E-value: 1e-41	
CDD Data	COG Number: COG1175 ABC-type sugar transport system, permease component	
	E-value: 2.10e-82	E-value: 1.62e-74
Cellular Localization	Transmembrane in the inner plasma membrane	
TIGRfam- Protein Family	TIGR01581	TIGR00969
	E-value: 0.00015	E-value: 3.9e-09
Pfam- Protein Family	Pf00528 Binding-protein-dependent transport system inner membrane	
	E-value: 5.2e-20	E-value: 1.5e-19
Protein Database	3FH6 Entity 1 containing Chain F, H	
	E-value:4.22747E-15	E-value: 5.33363E-11
KEGG Pathway Map	ABC transporters (02010)	

Table 3. Comparison of b3452 and mrub_0320

Through the use of KEGG, the amino acid sequences of both b3452 and mrub_0320 were found and used for analysis (26). This time, they are clearly identified in Figure 2 in the *sn*-glycerol-3-phosphate section of the pathway, marked as UgpA in their respective genomes. Through the use of the BLAST database the image in Figure 16 is produced (11). The sequence alignment between b3452 and mrub_0320 shows the similarities in their amino acid sequences. The query lines represent the amino acid sequence of b3452 and subject line represents mrub_0320. The line between the two shows the highly conserved amino acids with that amino

acid in the space. The addition sign indicates that the two amino acids are similar in functional groups. The E-value, which is close to zero, indicates that the similarity in results are not due to chance. This similarity in amino acid sequences suggest that b3452 and mrub_0320 are orthologs.

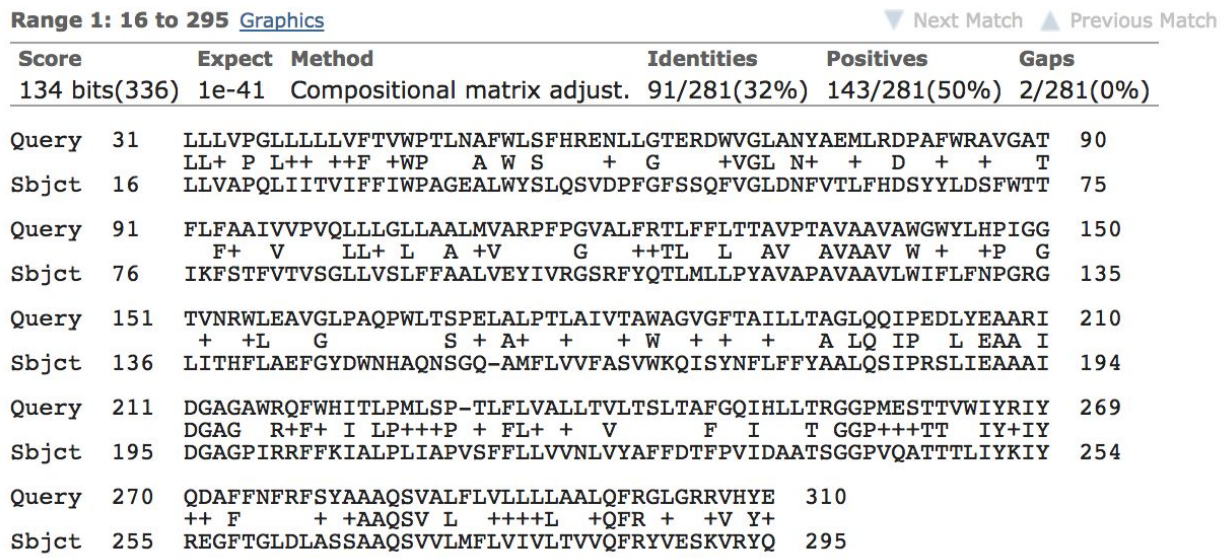


Figure 16. The BLAST database showed that when the protein sequence of b3452 is BLAST'ed against *Meiothermus ruber* the results show mrub_ 0320 has a low E-value indicating similarities within the two protein amino acid sequences, showing a high chance for orthologous genes between the two species (11).

Having confirmed that b3452 and mrub_0320 have significant sequence similarity, the focus then turned to learning more about these proteins in regards to cell localization. In order to identify whether or not these proteins have transmembrane domains, the TMHMM database was used (15-18). Figure 17 shows the results of the TMHMM analysis for this gene comparison. Using the known amino acid sequences, the database calculated which sections would be found throughout the cellular membrane. In the case of b4352 and mrub_0320, there were 6 transmembrane domains found. This result is expected for these two genes, because they have been previously identified as the permease component of the ABC transporter,

meaning that there needs to be transmembrane domains in order the protein to transport products (26).

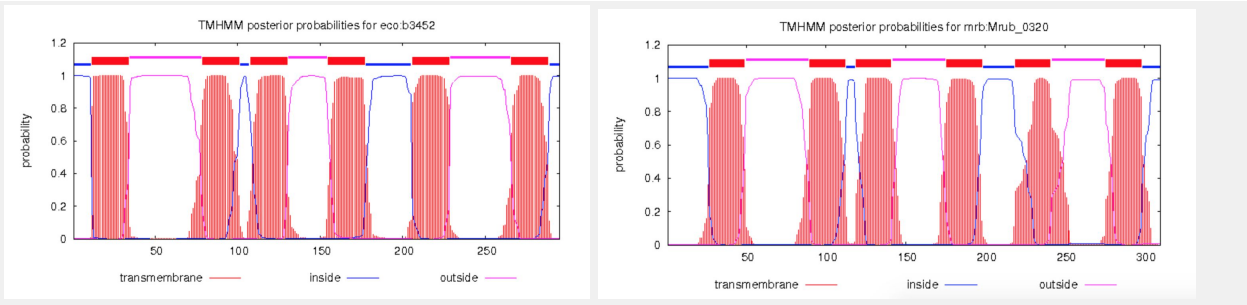


Figure 17. The bioinformatics database TMHMM was used to develop two graphs indicating cellular localization of b3452 and mrub_ 0320. The left image indicates the *E. coli* b3452 and the right image indicates mrub_ 0320. The similarities between the two graphs indicate a similarity in cellular localization between genes. Both graphs show clear deviations in the graph that define a transmembrane domain within the protein species. A total number of 6 can be found on each graph (15-18).

Also important for cell localization, is understanding whether our GOI create proteins that contain signal peptides, or some derivative of them. For this question, the database SignalP was used (19,20). Again, the amino acid sequences of b3452 and mrub_0320 were used within the database. Figure 18 shows the results of both of these analyses. In both cases, there is minimal strength of signal peptide probability. The SignalP results indicate that there is too low of a probability for both genes to have signal peptides in their protein sequences.

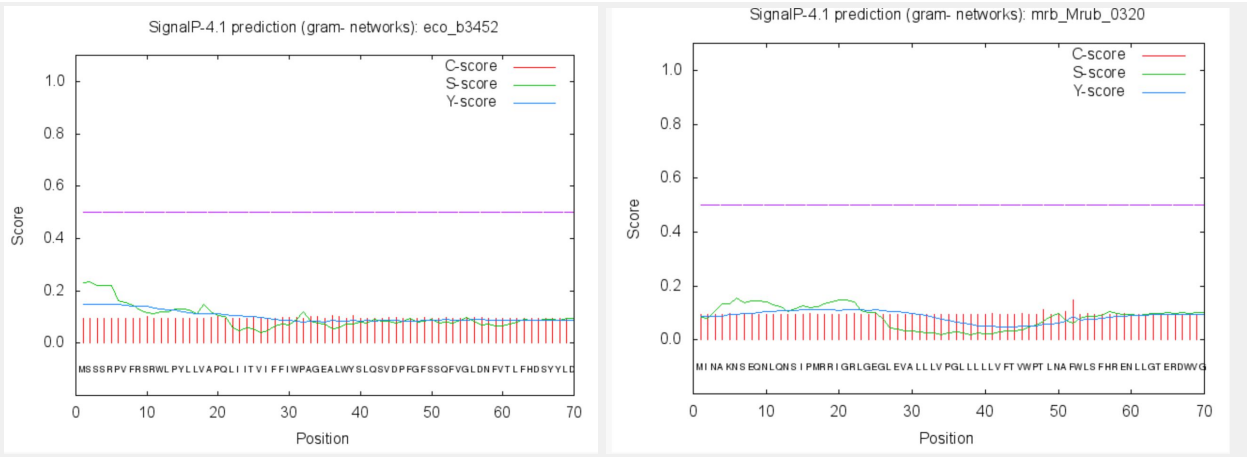


Figure 18. The bioinformatics database SignalP was used to develop two graphs indicating cellular localization of b3452 and mrub_ 0320. The left image indicates the *E. coli* b3452 and

the right image indicates mrub_ 0320. The similarities between the two graphs indicate a similarity in cellular localization between genes. Strong signals in these graphs would indicate the possible presence of a signal peptide on the proteins, which is clearly not seen in both of these cases (19-20).

To help in the conclusion of cellular location of these proteins, PSORT-B and Phobius were two other databases used (13,14). Again, amino acid sequences of b3452 and mrub_0320 were used. The databases would use theoretical amino acid sequences to determine possible intra-protein interactions that would determine the location relative to the cell membrane. With both of these genes, the PSORT-B indicated that both proteins were located on or around the cytoplasmic membrane. The Phobius graphs, shown in Figure 19, also indicates that these proteins are located both in the non-cytoplasmic and cytoplasmic region, further confirming the presence of transmembrane domains. The combination of these two results help confirm that b3452 and mrub_0320 are truly permease domains, that are located throughout the inner membrane of the respective bacteria.

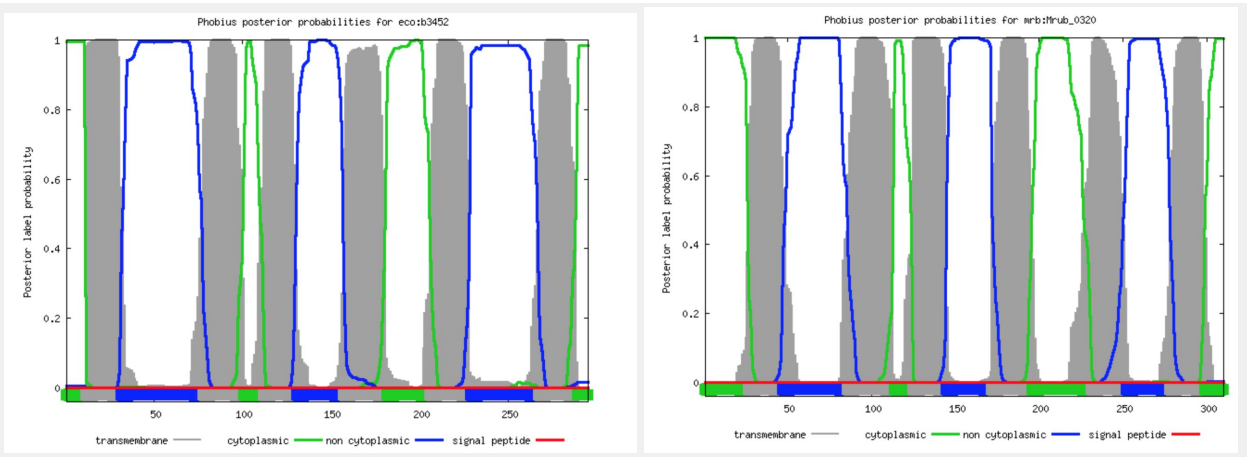


Figure 19. The bioinformatics database Phobius was used to develop two graphs indicating cellular localization of b3452 and mrub_ 0320. The left image indicates the *E. coli* b3452 and the right image indicates mrub_ 0320. The similarities between the two graphs indicate a similarity in cellular localization between genes. The constant change between cytoplasmic (green) and noncytoplasmic (blue) helps confirm the presence of transmembrane domains located on the inner bacterial membrane (14).

With a better idea of where b3452 and mrub_0320 are localized as proteins, the Pfam database was used in order to further clarify how similar the two genes are (22). As seen in Figure 20, when comparing b3452 and mrub_0320 to the same consensus sequence, there is a large amount of similarity, which can be seen by the large amount of shades of green found in the 4th row of each sequence. This helps confirm that the two GOI are orthologous to one another/.

Figure 20. The database Pfam was used to create aligned sequences of b3452 and mrub_ 0320 for comparison. The top sequence is b3452 and the bottom is mrub_0320. An example of conserved amino acid sequences can be seen in the red box. As seen by the green line of both images the two sequences which supports b3452 and mrub_ 0320 as orthologs (22).

Through the IMG/M database the genes b3452 and mrub_0320 were analyzed using color by KEGG as shown in Figure 21 (26,27). Color by KEGG coordinates function of genes with colors. The two GOI are indicated by the red line above the gene. The top panel is the *E.coli* gene b3452 and the bottom is mrub_0320 which are colored green and purple respectively. Both of these colors indicate they are involved in membrane transport. This confirms these genes are ABC transporters and that they are orthologs since they have the same function.

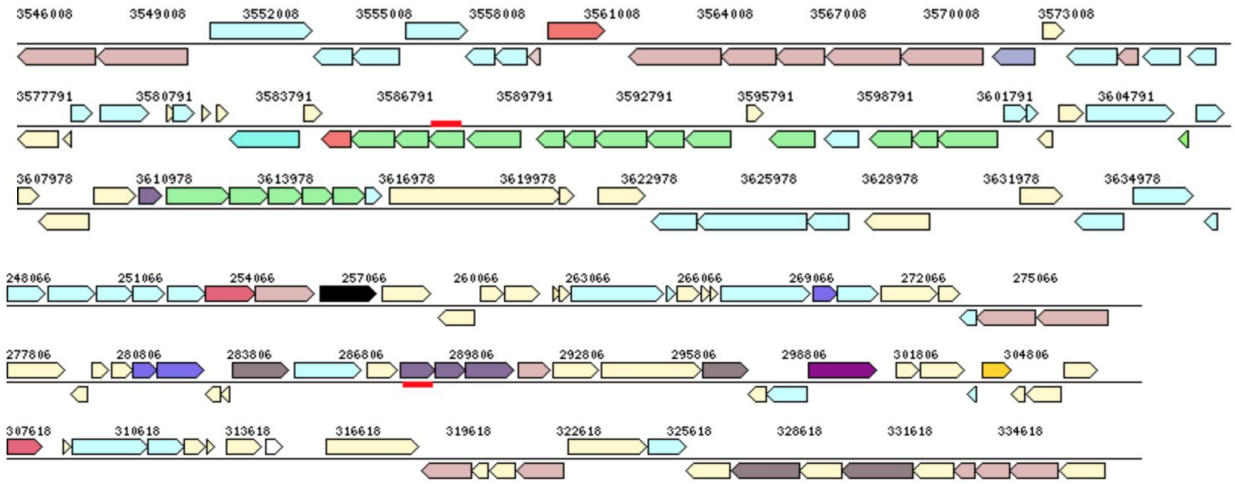


Figure 21. The IMG/M database and the color by KEGG function was used to analyze b3452 and mrub_ 0320. The top image represents b3452 and the bottom image represents mrub_0320. The red line indicates the gene of interest. The color of the genes represent their function. The *E.coli* green gene and the *M. ruber* purple gene represent the fact that both of these gene's function involves membrane transport (26,27).

Comparing *E. coli* b3453 and *M. ruber* mrub_0322

We continue to do the same process defined above with a new set of genes, b3453 and mrub_0322, both of which are classified as UgpB in their respective genomes. Again, for ease of understanding, Table 4 summarizes the results obtained from numerous bioinformatic tools to compare *E.coli* b3453 and mrub_ 0322. When looking at the comparison of E-values, the closer to 0.0 indicates that the genes are unlikely to be similar by chance. The original BLAST performed was against the two specific genes of interest, b3453 and mrub_0322 (11). The COG number identified through the CDD database was the same for both genes, with slightly varied E-values (31). Using the cellular localization databases (TMHMM (15), SignalP (19), LipoP (20), PSORT-B (13), and Phobius (14)), both genes were confirmed to have the same location. Pfam (22, 23) database shows the same protein family between the two genes, with similar E-values, while neither gene had a hit with TIGRfam (21). Protein Database showed different proteins of

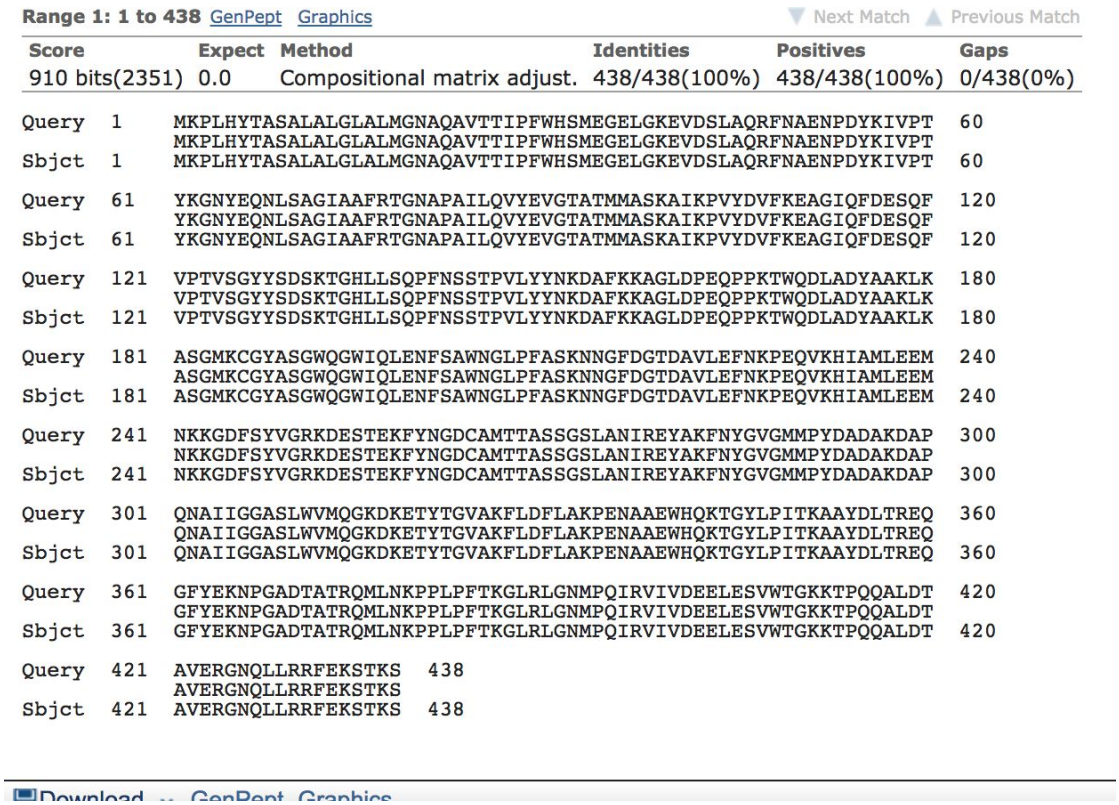
similarities. For both *E. coli* and *M. rub* systems, the genes were found within the ABC transporter page (KEGG map 02010) (26).

Bioinformatics Tool Used	**E.coli Gene b_3453**	**mrub_0322**
BLAST genome	Score: 187 bits E-value: 5e-59	
CDD Data	COG Number: COG1653 ABC-type glycerol-3-phosphate transport system, periplasmic component	
	E-value: 4.91e-72	E-value: 3.76e-60
Cellular Localization	Anchored to the periplasmic side of the inner membrane	
TIGRfam- Protein Family	No TIGRfam Number Found	
	E-value: n/a	E-value: n/a
Pfam- Protein Family	PF13416 SPB_bac_8	
	E-value: 9.2e-46	E-value: 1.3e-45
Protein Database	Entity 1 containing Chain A	
	E-value: 0.0	E-value: 5.22839e-46
KEGG Pathway Map	ABC transporters (02010)	

Table 4. Comparison of b_3453 and mrub_0322

Through the use of KEGG, the amino acid sequences of both b3453 and mrub_0322 were found and used for analysis (26). This time, they are clearly identified in Figure 2 in the *sn*-glycerol-3-phosphate section of the pathway, marked as UgpB in their respective genomes. Through the use of the BLAST database the image in Figure 22 is produced (11). The sequence alignment between b3453 and mrub_0322 shows the similarities in their amino acid sequences. The query lines represent the amino acid sequence of b3453 and subject line represents mrub_0322. The line between the two shows the highly conserved amino acids with that amino acid in the space. The addition sign indicates that the two amino acids are similar in functional

groups. The E-value, which is close to zero, indicates that the similarity in results are not due to chance. This similarity in amino acid sequences suggest that b3453 and mrub_0322 are orthologs.



Range 1: 1 to 438 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
910 bits(2351)	0.0	Compositional matrix adjust.	438/438(100%)	438/438(100%)	0/438(0%)

```
Query  1    MKPLHYTASALALGLALMGNAQAVTTIPFWHSMEGELGKEVDSLAQRFNAENPDYKIVPT  60
            MKPLHYTASALALGLALMGNAQAVTTIPFWHSMEGELGKEVDSLAQRFNAENPDYKIVPT
Sbjct  1    MKPLHYTASALALGLALMGNAQAVTTIPFWHSMEGELGKEVDSLAQRFNAENPDYKIVPT  60

Query  61   YKGNYEQNLSAGIAAFRTGNAPAILQVYEVGTATMMASKAIKPVYDVFKEAGIQFDESQF  120
            YKGNYEQNLSAGIAAFRTGNAPAILQVYEVGTATMMASKAIKPVYDVFKEAGIQFDESQF
Sbjct  61   YKGNYEQNLSAGIAAFRTGNAPAILQVYEVGTATMMASKAIKPVYDVFKEAGIQFDESQF  120

Query  121  VPTVSGYYSDSKTGHLLSQPFNSSTPVLYYNKDAFKKAGLDPEQPPKTWQDLADYAAKLK  180
            VPTVSGYYSDSKTGHLLSQPFNSSTPVLYYNKDAFKKAGLDPEQPPKTWQDLADYAAKLK
Sbjct  121  VPTVSGYYSDSKTGHLLSQPFNSSTPVLYYNKDAFKKAGLDPEQPPKTWQDLADYAAKLK  180

Query  181  ASGMKCGYASGWQGWIQLENFSAWNGLPFASKNNGFDGTDAVLEFNKPEQVKHIAMLEEM  240
            ASGMKCGYASGWQGWIQLENFSAWNGLPFASKNNGFDGTDAVLEFNKPEQVKHIAMLEEM
Sbjct  181  ASGMKCGYASGWQGWIQLENFSAWNGLPFASKNNGFDGTDAVLEFNKPEQVKHIAMLEEM  240

Query  241  NKKGDFSYVGRKDESTEKFYNGDCAMTTASSGSLANIREYAKFNYGVGMMPYDADAKDAP  300
            NKKGDFSYVGRKDESTEKFYNGDCAMTTASSGSLANIREYAKFNYGVGMMPYDADAKDAP
Sbjct  241  NKKGDFSYVGRKDESTEKFYNGDCAMTTASSGSLANIREYAKFNYGVGMMPYDADAKDAP  300

Query  301  QNAIIGGASLWVMQGKDKETYTGVAKFLDFLAKPENAAEWHQKTGYLPITKAAYDLTREQ  360
            QNAIIGGASLWVMQGKDKETYTGVAKFLDFLAKPENAAEWHQKTGYLPITKAAYDLTREQ
Sbjct  301  QNAIIGGASLWVMQGKDKETYTGVAKFLDFLAKPENAAEWHQKTGYLPITKAAYDLTREQ  360

Query  361  GFYEKNPGADTATRQMLNKPPLPFTKGLRLGNMPQIRVIVDEELESVWTGKKTPQQALDT  420
            GFYEKNPGADTATRQMLNKPPLPFTKGLRLGNMPQIRVIVDEELESVWTGKKTPQQALDT
Sbjct  361  GFYEKNPGADTATRQMLNKPPLPFTKGLRLGNMPQIRVIVDEELESVWTGKKTPQQALDT  420

Query  421  AVERGNQLLRRFEKSTKS  438
            AVERGNQLLRRFEKSTKS
Sbjct  421  AVERGNQLLRRFEKSTKS  438
```

Download ... GenPept Graphics

Figure 22. The BLAST database showed that when the protein sequence of b3453 is BLAST'ed against *Meiothermus ruber* the results show mrub_ 0322 has a low E-value indicating similarities within the two protein amino acid sequences, showing a high chance for orthologous genes between the two species (11).

Having confirmed that b3453 and mrub_0322 have significant sequence similarity, the focus then turned to learning more about these proteins in regards to cell localization. In order to identify whether or not these proteins have transmembrane domains, the TMHMM database was used (15-18). Figure 23 shows the results of the TMHMM analysis for this gene comparison. Using the known amino acid sequences, the database calculated which sections would be found throughout the cellular membrane. In the case of b4353 and mrub_0322, there were no transmembrane domains found. This result is expected for these two genes, because

they have been previously identified as the substrate binding component of the ABC transporter, meaning that the protein needs to be on the outside of the cell membrane in order to interact with the substrate (26).

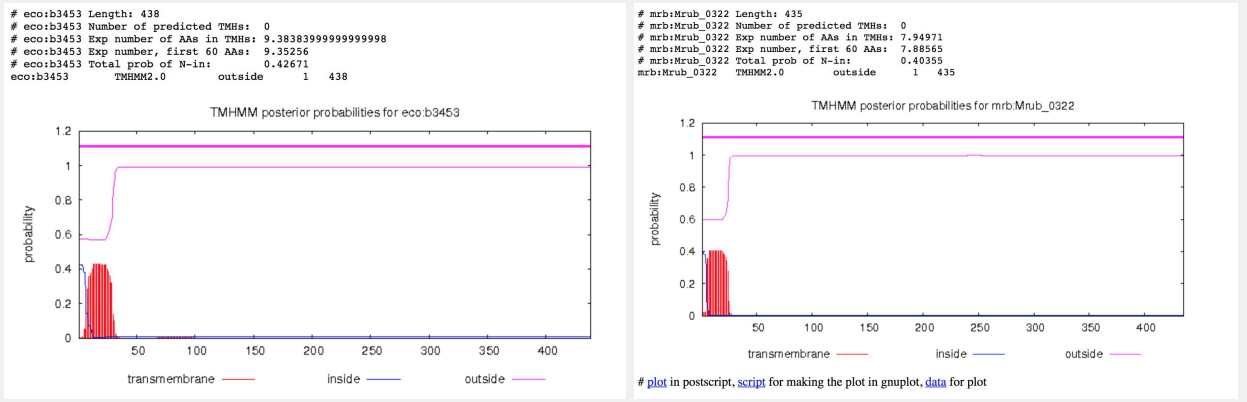


Figure 23. The bioinformatics database TMHMM was used to develop two graphs indicating cellular localization of b3453 and mrub_0322. The left image indicates the *E. coli* b3453 and the right image indicates mrub_ 0322. The similarities between the two graphs indicate a similarity in cellular localization between genes. Although these two graphs show no deviation, if there were any deviations present through the graph, it would indicate the presence of transmembrane helices. (15-18)

Also important for cell localization, is understanding whether our GOI create proteins that contain signal peptides, or some derivative of them. For this question, the database SignalP was used (19,20). Again, the amino acid sequences of b3453 and mrub_0322 were used within the database. Figure 24 shows the results of both of these analyses. In the case of b3453, there is a significant probability that the protein does contain a signal peptide. The SignalP results indicate that there is too low of a probability for the mrub_0322 gene to have signal peptides in their protein sequences.

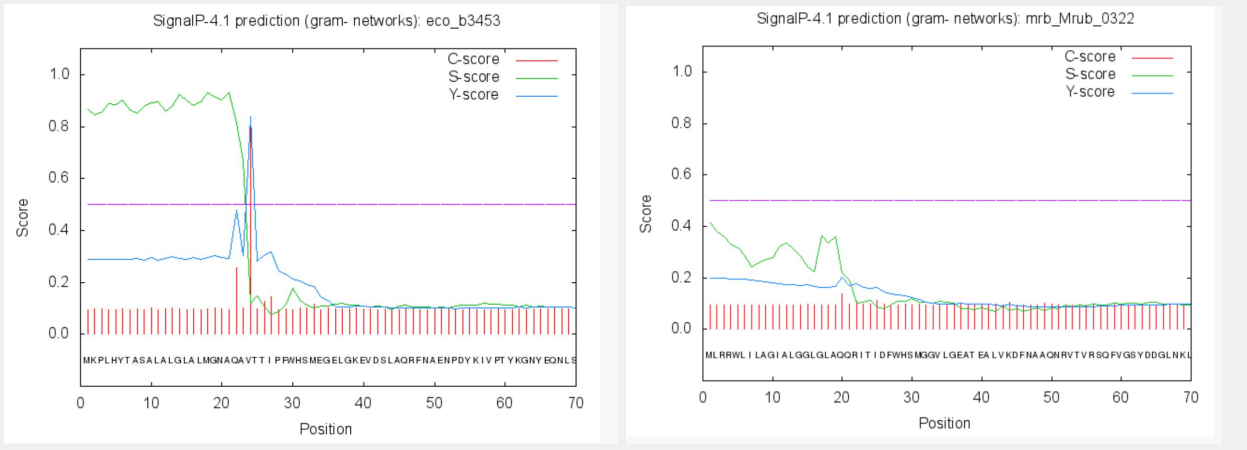


Figure 25. The bioinformatics database SignalP was used to develop two graphs indicating cellular localization of b3453 and mrub_ 0322. The left image indicates the *E. coli* b3453 and the right image indicates mrub_ 0322. The similarities between the two graphs indicate a similarity in cellular localization between genes. Strong signals in these graphs would indicate the possible presence of a signal peptide on the proteins, which is clearly seen in b3453, but not in mrub_0322 (19-20).

To confirm the previous conclusion of a presence of signal peptide, another database was used, known as LipoP (20). This database is similar to SignalP, but focuses more clearly on lipoproteins. Figure 26 shows the LipoP results from both b3453 and mrub_0322. The results from b3453 do show some higher probability levels for signal peptides in the form of Signal Peptide I (SpI). The probability for mrub_0322 is much lower.

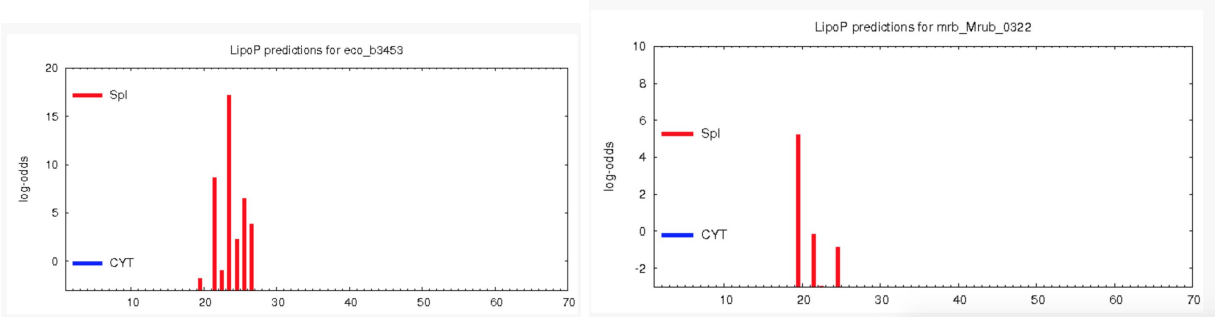


Figure 26. The bioinformatics database LipoP was used to develop two graphs indicating probability of lipoproteins for b3453 and mrub_ 0322. The left image indicates the *E. coli* b3453 and the right image indicates mrub_ 0322. The b3453 graph shows significant probability for a signal protein to be present, but the data for mrub_0322 shows the probability is still relatively low. Strong signals in these graphs would indicate the possible presence of a lipoprotein as a part of the proteins, which is clearly not seen in these cases (20).

To help in the conclusion of cellular location of these proteins, PSORT-B and Phobius were two other databases used (13,14). Again, amino acid sequences of b3453 and mrub_0322 were used. The databases would use theoretical amino acid sequences to determine possible intra-protein interactions that would determine the location relative to the cell membrane. With both of these genes, the PSORT-B indicated that both proteins were located on or around the periplasm. The Phobius graphs, shown in Figure 27, also indicates that these proteins are located in the non-cytoplasmic region, further confirming the idea of them being in the periplasm. The combination of these two results help confirm that b3453 and mrub_0322 are truly substrate domains, that are located throughout the periplasm of the respective bacteria to help bring in the desired substrate, in this case *sn*-glycerol-3-phosphate.

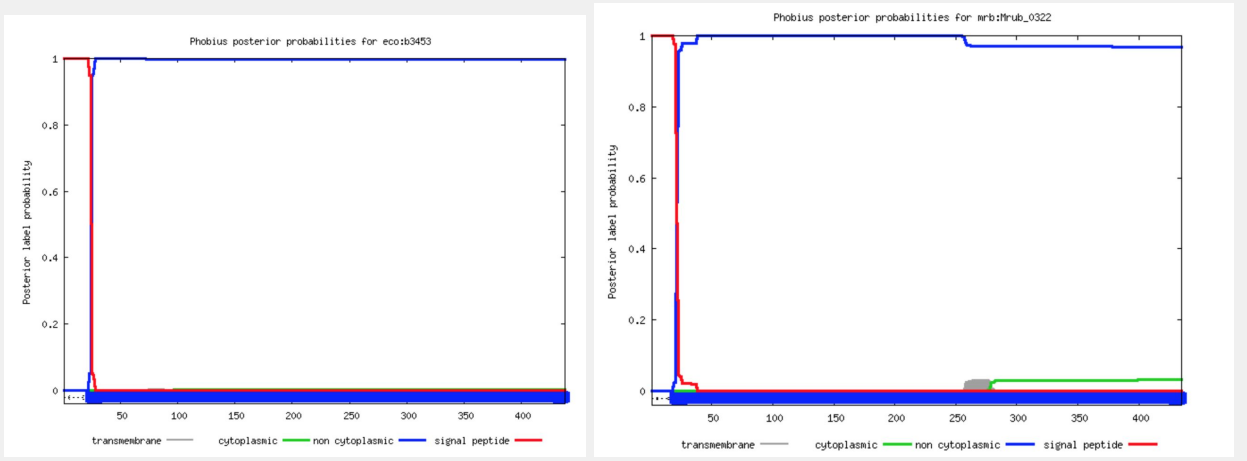


Figure 27. The bioinformatics database Phobius was used to develop two graphs indicating cellular localization of b3453 and mrub_ 0322. The left image indicates the *E. coli* b3453 and the right image indicates mrub_0322. The similarities between the two graphs indicate a similarity in cellular localization between genes(14).

With a better idea of where b3453 and mrub_0322 are localized as proteins, the Pfam database was used in order to further clarify how similar the two genes are (22). As seen in Figure 28, when comparing b3453 and mrub_0322 to a consensus sequence, there is a large amount of similarity, which can be seen by the large amount of shades of green found in the 4th row of each sequence. This helps confirm that the two GOI are orthologous to one another.

Figure 28. The database Pfam was used to create aligned sequences of b3453 and mrub_ 0322 for comparison. The top sequence is b3453 and the bottom is mrub_0322. As seen by the green line of both images the two sequences which supports b3453 and mrub_ 0322 as orthologs (22).

Through the IMG/M database the genes b3453 and mrub_0322 were analyzed using color by KEGG as shown in Figure 29 (26,27). Color by KEGG coordinates function of genes with colors. The two GOI are indicated by the red line above the gene. The top panel is the *E.coli* gene b3453 and the bottom is mrub_0322 which are colored green and purple respectively. Both of these colors indicate they are involved in membrane transport. This confirms these genes are ABC transporters and that they are orthologs since they have the same function.

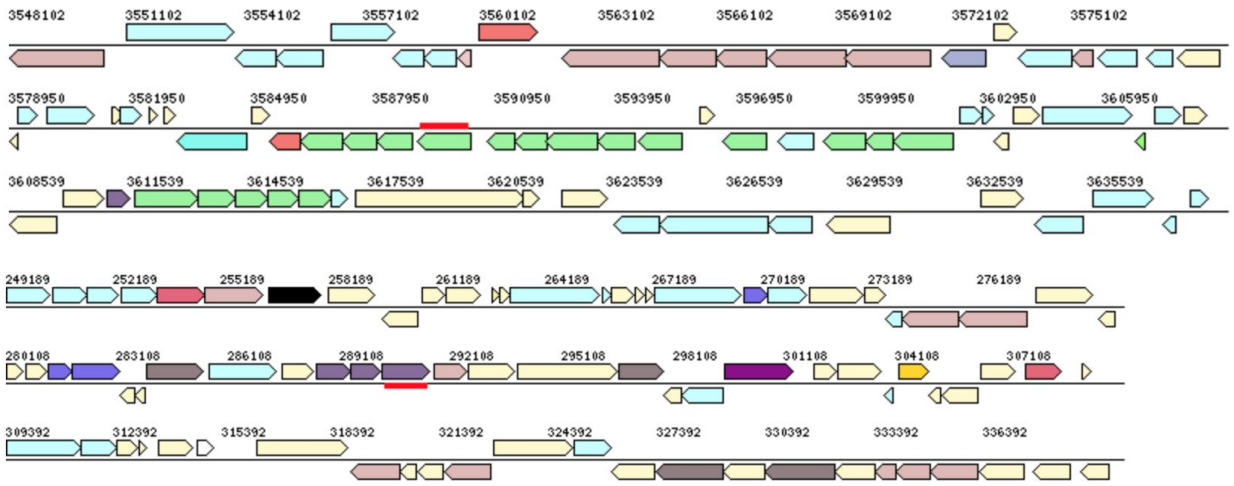


Figure 29. The IMG/M database and the color by KEGG function was used to analyze b3453 and mrub_ 0322. The top image represents b3453 and the bottom image represents mrub_0322. The red line indicates the gene of interest. The color of the genes represent their function. The *E.coli* green gene and the *M. ruber* purple gene represent the fact that both of these gene's function involves membrane transport (26,27).

Conclusion

Throughout the use of many bioinformatics tools it can be concluded that mrub_0320 gene is an ortholog of the b3452 gene, mrub_0321 gene is an ortholog of the b3451 gene,mrub_0322 gene is an ortholog of the b3453 gene, mrub_2366 gene is an ortholog of the b3450 gene. Although there are some instances of differences between results of ortholog pairs the majority of evidence indicates their distinct similarities. Since the two organisms are distantly related it may have contributed to the slight differences in analysis. All of the following genes are part of operons that involve membrane transport as shown in figure 8,15,21 and 29 by the color by KEGG database(26,27).

The first gene pair analyzed was mrub_2366 and b3450 to confirm they are orthologs. Numerous bioinformatics tools were used to support that these two genes are orthologs. The first task done to begin analysis was to blast the nucleotide sequences of mrub_2366 against b3450. The BLAST results gave a low E-value indicating these genes are highly related and not

due to chance (11). These genes have the same COG (31), TIGRfam (21) and Pfam (22, 23) numbers which is another piece of evidence proving they are orthologs. The last key piece of evidence is that all of the cellular localization database data matches. The overall cellular localization is the inner plasma membrane for both genes (TMHMM (15, 16, 17, 18), SignalP (19), LipoP (20), PSORT-B (15), Phobius (14)). For both *E. coli* and *M. rub* systems, the genes were found within the ABC transporter page (KEGG map 02010) (26).The only difference in data found was through the protein database where the two genes represent different proteins which could be due to an incomplete database (25).This piece can be disregarded because of the significant number of evidence gathered that confirms mrub_2366 and b3450 being orthologs.

The same process defined above is completed with the second set of genes, b3451 and mrub_0321, both of which are classified as UgpE in their respective genomes. As seen in Table 2 the BLAST results show a low E-value indicating theses genes are indeed orthologs and not likely similar by chance(11). The COG and Pfam numbers are the same for both genes (22,23). The cellular localization databases (TMHMM (15), SignalP (19), LipoP (20), PSORT-B (13), and Phobius (14)), both genes were confirmed to have the same location on the inner plasma membrane. Protein Database showed identical proteins for both genes (24, 25). For both *E. coli* and *M. rub* systems, the genes were found within the ABC transporter page (KEGG map 02010) (26). The only piece of data that differed between mrub_0321 and b3451 was the TIGRfam database did not give the same value since b3451 did not yield any hits (21). This result does not conclude against their similarity but just yields inconclusive. The overall conclusion remains that b3451 and mrub_0321 are orthologs.

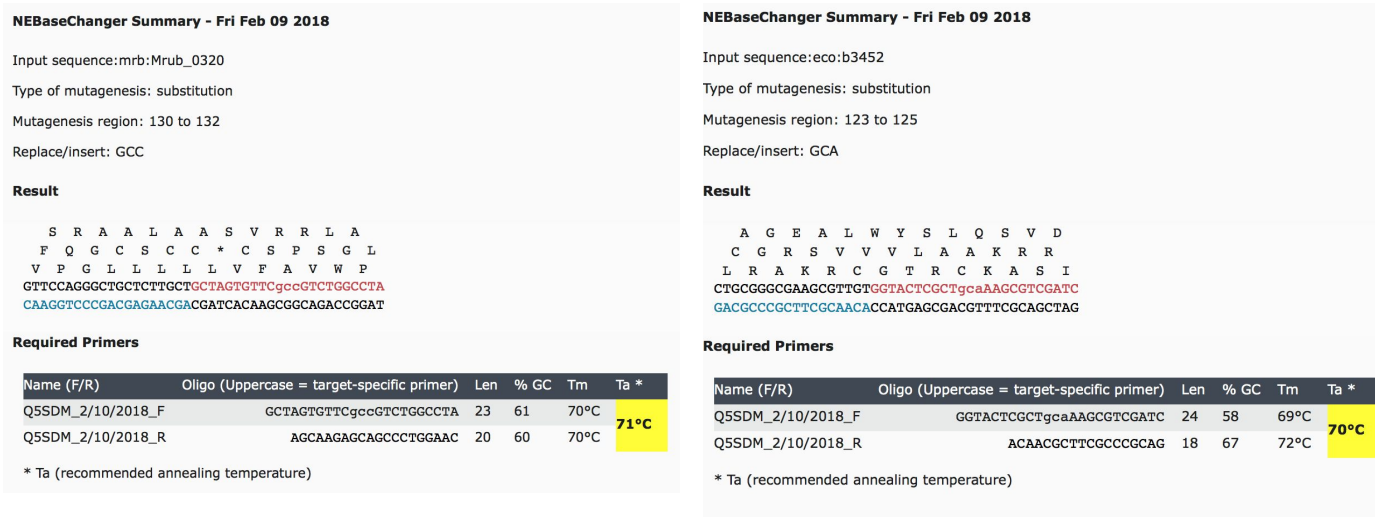
The third gene pair analyzed was mrub_0320 and b3452 both of which are classified as UgpA in their respective genomes (26). Again, for ease of understanding, Table 3 summarizes

the results obtained from numerous bioinformatic tools to compare *E.coli* b3452 and mrub_ 0320. The BLAST performed comparing these genes produced a low E-value indicating the genes to be orthologs (11). The COG number identified through the CDD database was the same for both genes, with slightly varied E-values (31). Using the cellular localization databases (TMHMM (15), SignalP (19), LipoP (20), PSORT-B (13), and Phobius (14)), both genes were confirmed to have the same location. TIGRfam has varied results between the two genes, with two different TIGRfam numbers showing up (21). Pfam database shows the same protein family between the two genes, with similar E-values (22, 23). Protein Database showed different proteins of similarities (24, 25). For both *E. coli* and *M. rub* systems, the genes were found within the ABC transporter page (KEGG map 02010) (26). Nearly all forms of data helps conclude that b3452 and mrub_0320 are orthologous; Regardless of TIGRfam results, this connection is strong.

The last gene pair analyzed was mrub_0322 and b3453, both of which are classified as UgpB in their respective genomes. Again, for ease of understanding, Table 4 summarizes the results obtained from numerous bioinformatic tools to compare *E.coli* b3453 and mrub_ 0322. The BLAST database gave a low E-value indicating the genes to be orthologs not by chance (11). The COG number identified through the CDD database was the same for both genes, with slightly varied E-values (31). Using the cellular localization databases (TMHMM (15), SignalP (19), LipoP (20), PSORT-B (13), and Phobius (14)), both genes were confirmed to have the same location. Pfam (22, 23) database shows the same protein family between the two genes, with similar E-values, while neither gene had a hit with TIGRfam (21). No hit in TIGRfam leads to inconclusive results but because the several other pieces of evidence are strong it can be concluded they are orthologs. Protein Database showed the same protein with different

E-values (24,25). For both *E. coli* and *M. rub* systems, the genes were found within the ABC transporter page (KEGG map 02010) (26).

Finally, site directed mutagenesis was studied by using NEB changer bioinformatics to change a highly conserved amino acid into alanine (28). The substitution was completed on *ugpA* (mrub_0320) at amino acid F44 to an alanine. The result of this change is shown in the left side of Figure 30 below. The Threonine in the amino acid sequence from KEGG is at position 44 since that is a highly conserved amino acid found by the weblogo database. The nucleotide sequence was substituted at position 130 to 132. Then the same procedure was done with b3452 however the substitution was from position 123 to 125 from a Threonine to an alanine shown in Figure 30 on the right side.



NEBaseChanger Summary - Fri Feb 09 2018

Input sequence:mrb:Mrub_0320

Type of mutagenesis: substitution

Mutagenesis region: 130 to 132

Replace/insert: GCC

Result

```
   S  R  A  A  L  A  A  S  V  R  R  L  A
  F  Q  G  C  S  C  C  *  C  S  P  S  G  L
V  P  G  L  L  L  L  L  V  F  A  V  W  P
GTTCCAGGGCTGCTCTTGCTGCTAGTGTTCgccGTCTGGCCTA
CAAGGTCCCGACGAGAACGACGATCACAAGCGGCAGACCGGAT
```

Required Primers

Name (F/R)	Oligo (Uppercase = target-specific primer)	Len	% GC	Tm	Ta *
Q5SDM_2/10/2018_F	GCTAGTGTTCgccGTCTGGCCTA	23	61	70°C	**71°C**
Q5SDM_2/10/2018_R	AGCAAGAGCAGCCCTGGAAC	20	60	70°C	

* Ta (recommended annealing temperature)

NEBaseChanger Summary - Fri Feb 09 2018

Input sequence:eco:b3452

Type of mutagenesis: substitution

Mutagenesis region: 123 to 125

Replace/insert: GCA

Result

```
   A  G  E  A  L  W  Y  S  L  Q  S  V  D
  C  G  R  S  V  V  V  L  A  A  K  R  R
 L  R  A  K  R  C  G  T  R  C  K  A  S  I
CTGCGGGCGAAGCGTTGTGGTACTCGCTgcaAAGCGTCGATC
GACGCCCGCTTCGCAACACCATGAGCGACGTTTCGCAGCTAG
```

Required Primers

Name (F/R)	Oligo (Uppercase = target-specific primer)	Len	% GC	Tm	Ta *
Q5SDM_2/10/2018_F	GGTACTCGCTgcaAAGCGTCGATC	24	58	69°C	**70°C**
Q5SDM_2/10/2018_R	ACAACGCTTCGCCCGCAG	18	67	72°C	

* Ta (recommended annealing temperature)

Figure 30. The site-directed mutagenesis of mrub_0320 and b3452 to substitute a Threonine for Alanine. The left side of the image represents the *M.ruber* gene mrub_0320 and the right side is the *E.coli* ortholog b3452.

The sequences above must then undergo functional genomics to allow analysis of the data. Functional genomics, such as the complementation test, will show the substitution in the

highly conserved amino acid, threonine, to alanine will cause a change in protein function. This change in function detected indicates the substitution caused a change in protein structure ultimately affecting its function. This study can be used to further research on orthologous genes in *Meiothermus ruber and Escherichia coli* ABC transporters involved in *sn*-glycerol - 3-phosphate. This research has been shown to be linked to ABC transport systems involving *sn*-glycerol - 3-phosphate. These transport systems are linked to diabetes, cancer, and many other metabolic diseases research. These disease have high prevalence around the world indicated further research is necessary to continue to understand their metabolic pathways to develop more efficient treatment.

Works Cited:

(1) Wilkens S. Structure and mechanism of ABC transporters. F1000Prime Reports. 2015 [accessed 2018 Feb 6];7.

(2) Mugabo Y, Zhao S, Seifried A, Gezzar S, Al-Mass A, Zhang D, Lamontagne J, Attane C, Poursharifi P, Iglesias J, et al. Identification of a mammalian glycerol-3-phosphate phosphatase: Role in metabolism and signaling in pancreatic β-cells and hepatocytes. Proceedings of the National Academy of Sciences. 2016 [accessed 2018 Feb 6];113(4):E430–E439.

(3) Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muniz-Rascado L, Ong Q, Paley S, Schroder I, Shearer A, Subhraveti P, Travers M, Weerasinghe D, Weiss V, Collado-Vides J, Gunsalus RP, Paulsen I, Karp PD. 2013. EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res. 41:D605-612.

(4) Phylogenetic Diversity. Joint Genome Institute. [accessed 2018 Feb 6]. https://jgi.doe.gov/our-science/science-programs/microbial-genomics/phylogenetic-diversity/

(5) Tindall BJ, Sikorski J, Lucas S, Goltsman E, Copeland A, Del Rio TG, Lapidus A. 2010. Complete genome sequence of Meiothermus ruber type strain (21T). Standards in Genomic Sciences, 3(1):26–36.

(6) Betts MJ and Russell RB. 2003. Amino-Acid Properties and Consequences of Substitutions. Bioinformatics for Geneticists. 311–342.

(7) Kataoka M, Yamaoka A, Kawasaki K, Shigeri Y, Watanabe K. 2014. Extraordinary denaturant tolerance of keratinolytic protease complex assemblies produced by meiothermus ruber H328. Appl Microbiol Biotechnol 98(7):2973-80.

(8) Quay SC, Dick TE and Oxender, DL. 1977. Role of transport systems in amino acid metabolism: leucine toxicity and the branched-chain amino acid transport systems.

(9) Journal of Bacteriology, 129(3):1257-1265.

(10) Karin Ejendal and Christine Hrycyna. 2002. Multidrug resistance and cancer: The role of the human ABC transporter ABCG2. Current Protein & Peptide Science 3(5):503-11.

(11) Madden T. 2002 [Updated 2003 Aug 13]. The BLAST Sequence Analysis Tool [Internet] In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information. Available from http://www.ncbi.nlm.nih.gov/books/NBK21097/ BLAST tool: BLASTp tool from https://blast.ncbi.nlm.nih.gov/Blast.cgi

(12) Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403-410.

(13) Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FSL. 2010. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, Bioinformatics. 26(13):1608-1615.

(14) Kall L, Krogh A, Erik LL. Sonnhammer E. 2007. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res. 35:W429-32.

(15) Krogh A, Rapacki K. 2016. TMHMM Server, v. 2.0. Cbs.dtu.dk. [Internet]. Denmark: Technical University of Denmark. [cited 2018 Feb 6]. Available from http://www.cbs.dtu.dk/services/TMHMM/

(16) Sonnhammer E, von Heijne G, Krogh A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. In J. Glasgow, T. Littlejohn, F. Major, R. Lathrop, D. Sankoff, and C. Sensen, editors, Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology. Menlo Park, CA: AAAI Press. p. 175-182.

(17) Krogh A, Larsson B, von Heijne G, Sonnhammer E. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol, 305(3):567-580.

(18) Kall L, Krough A, Sonnhammer E. 2004. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 338(5):1027-1036.

(19) Petersen T, Brunak S, von Heijne G, Nielsen H. 2011. Discriminating signal peptides from transmembrane regions. Nat Methods, 8:785-786. Available from: http://www.cbs.dtu.dk/services/SignalP.

(20) Juncker A, Willenbrock H, von Heijne G, Nielsen H, Brunak S and Krogh A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. 12(8):1652-62. Available at: http://www.cbs.dtu.dk/services/LipoP/.

(21) Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, White O. 2001. TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res 29(1):41-3.

(22) Finn RD, Bateman A, Clements J, et al. 2014. Pfam: the protein families database. Nucleic Acids Research [Internet]. [cited 2018 Feb 6]. 42 (Database issue):D222-D230. Available from http://pfam.xfam.org/

(23) Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future: Nucleic Acids Res. [Internet]. [cited 2018 Feb 6] 44:D279-D285. Available from: http://pfam.xfam.org/

(24) Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank Nucleic Acids Research, 28: 235-242.

(25) Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. [Internet]. The Protein Data Bank; [cited 2018 Feb 6]. Available from: http://www.rcsb.org/.

(26) Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44:D457–D462. Available from: http://www.genome.jp/kegg/

(27) Markowitz VM, Chen IA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, et al. 2012. IMG: The integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 40(D1):D115-22. Available from: http://nar.oxfordjournals.org/content/40/D1/D115.full.

(28) Biolabs, New England Home - NEB | New England Biolabs. Home - NEB | New England Biolabs. Available from: https://www.neb.com/

(29) Notredame C, Higgins DG, Herringa, J. **2000**. T-Coffee: A novel method for multiple sequence alignments, *J. Mol. Biol.* 302 p.205-217

(30) Crooks, GE, Hon G, Chandonia JM, Brenner SE. **2004**. WebLogo: A sequence logo generator. *Genome Research*, 14:1188-1190.

(31) Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43 (Database issue):D222-2. Doi: 10.1093/nar/gku1221. [PubMed PMID: 25414356]