2018

Mrub_1199 & Mrub_2272 of *Meiothermus ruber* are orthologous genes to the b0262 gene in *Escherichia coli* while Mrub_1200, Mrub_1201, Mrub_2015 & Mrub_2271 are not orthologous to the b0262 gene coding for the Iron (Fe3+) ABC Transport System

Kumail Hussain
Augustana College, Rock Island Illinois

Dr. Lori Scott
Augustana College, Rock Island Illinois

Mrub_1199 & Mrub_2272 of *Meiothermus Ruber* are orthologous genes to the b0262 gene in *Escherichia coli* while Mrub_1200, Mrub_1201, Mrub_2015 & Mrub_2271 are not orthologous to the b0262 gene coding for the Iron (Fe3+) ABC Transport System

Kumail Hussain
Dr. Lori R. Scott Laboratory
Biology Department, Augustana College
639 38th Street, Rock Island, IL 61201

Introduction

Importance of *M. Ruber*

Typically, bacteria thrive in humid and moist environments. The bacterium, *Meiothermus Ruber*, thrives on the extreme version of these environments. *Meiothermus Rube*r is part of the Genus *Meiothermus* and produces a unique red pigment. It is thermophilic and a gram-negative bacteria (Tindall et. al 2010). The species was first grouped under genus *Thermus* until later transferred to the genus *Meiothermus*. The "ruber" part of the name describes the red appearance of the species. Species of the genus can be found in hot springs or artificial thermal environments (Tindall et. al 2010). *Meiothermus ruber* is overshadowed by the numerous publications for other bacterium such as *E. coli* and *Salmonella*. Luckily, projects that highlight the importance of research of lesser known bacteria exist to explore possibilities of variable genes and processes. The Genomic Encyclopedia of Bacteria and Archaea (GEBA) are one of such projects (DOE JGI.2018). The *E. coli* bacteria are an incredibly useful organism to compare the *M. ruber* genome, since they are similar yet different at the same time. The thermophilic ability of *M. ruber* makes it a favorite organism to explore for example, one study that utilized *M.ruber* as the organism to compare its enzymatic activity has won much praise. The study analyzed the activity of an enzyme in *M. rube*r that catalyzes the biological reaction for production of butanol (Reiße 2015).

E. Coli containing potential orthologs

E. coli is a common organism used in the laboratory for comparing similar bacteria's function since, it is very easy to handle and is thoroughly researched (Blount 2015). The KEGG pathway for ABC transporters show the *E. coli* genome has a single gene coding for ATP Binding Cassette (ABC) transport of Iron (Fe3+) which is labeled as b0262, while the *Meiothermus ruber* genome contains 6 genes that encode the proteins for the Iron ABC Transport system. A detailed

explanation is found in Figure 2. Orthologs are genes that have similar protein function, but the gene will be from another species. It is a result of evolutionary change and likely means conserved function among organisms (Studer 2009). There are many genes in *M. rube*r that can be similar to *E. coli*, but not necessarily be characterized as orthologs. The purpose of this project is determine which genes are orthologs and which are not.

Transport of Iron (Fe3+) through ABC Transporters

Figure 1 below outlines the general framework for ABC transporters in *Escherichia coli* and *Meiothermus ruber*. A typical iron-uptake ABC transporter consists of a ferric ion-binding protein, located in the periplasm, two transmembrane proteins that form a pathway for ferric ions and two peripheral ATP binding proteins located on the cytoplasmic side of the inner membrane (Wang 2014). Since the bacteria that are being researched is gram negative, there is a periplasm followed by the outer membrane of the cell wall. Ions such as Fe3+ will travel across the outer membrane and will require the ABC transport system to facilitate its passage through the inner cell membrane. A specific periplasmic binding protein will guide the Fe3+ to the transmembrane domains to allow transport of the ion (Moussatova et. al 2008).

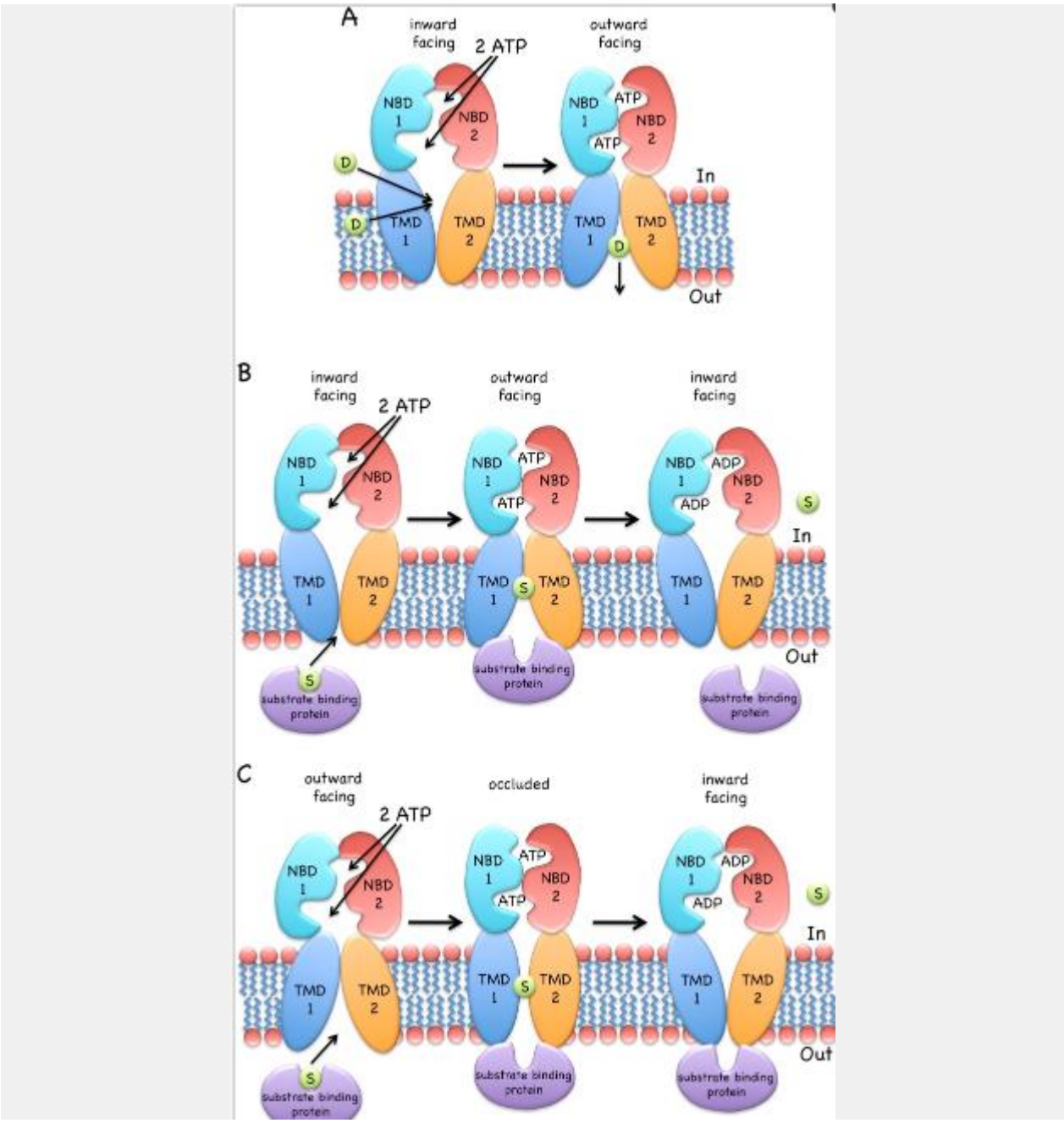


Figure 1. General diagram of ABC Transporter System in gram-negative bacteria. The protein system has 4 domains and a substrate binding Protein: 2 Nucleotide Binding Domains (NBD) and 2 Transmembrane Domain (TMD) with 1substrate binding protein. The 2 nucleotide binding domains lie in the cytoplasm in close proximity with the cytoplasmic membrane. The 2 transmembrane domains are embedded in the membrane and coordinate with the nucleotide binding domains to transport the Iron (Fe3+) in the cell. A substrate binding protein, located in the periplasm of the bacteria, is utilized to bring the Fe3+ to the transport system. Sections B &

C of Figure 1 depict the mechanism of the transport system. Iron is brought to the TMD and then brought halfway through the membrane. By that time, the NBD has ATP bound to the nucleotide binding sites. The hydrolysis of the nucleotides releases energy and drives the complete transport of Iron in the cell. Image taken from: (Wilken 2015).

The role of Iron in gram-negative bacteria has recently become more understood. The versatility of Iron includes electron carriers and catalytic centers. Thus, it is vital to biological processes such as photosynthesis, the Krebs Cycle, gene regulation, etc. (Krewulak 2008). Iron will usually be assimilated from host proteins or other sources of Iron in the bacteria. Iron will be extracted from transferrin, siderophores, and hemoglobin, which are host carrier proteins (Krewulak 2008).

Bioinformatics

Bioinformatics allow biologists to conduct experiments without an expensive laboratory, basically allowing experiments to be conducted on a computer (Bioinformatics). A multitude of bioinformatics techniques were implicated in the gene annotation of the selected genes. There will be instances when bioinformatics tools do not produce the desired results. The learning experience of the project stems from the understanding of why this may have occurred. Despite the numerous bioinformatics information produced during the project, there will only be discussion of relevant findings. Such relevant findings will describe structure, localization, and function.

Purpose/ Hypothesis

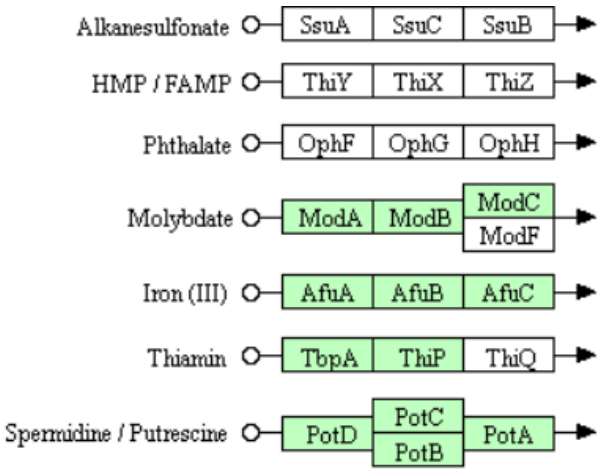
This project will determine if there are any orthologs to the b062 (afu*C*) gene of *E. coli* in the *M. rube*r genome. The a*fuC* gene of *E. coli* is the sole gene for the ABC transport system of Iron (Fe3+) while the *M. rube*r gene has 6 known genes: Mrub_1199, Mrub_1200, Mrub_1201, Mrub_2015, Mrub_2271, and Mrub_2272. The reason for the project is to shed light on important transport systems in these organisms so that we can better understand the mechanism behind it. Similarities between these organisms can lead to the discovery of essential processes and eventually be used outside of prokaryotes. The *afuC* genes between the organisms are Mrub_1199 & Mrub_2272 & b0262. The remaining genes are grouped into two gene names *afuA* (Mrub_1201, Mrub_2015) and *afuB* (Mrub_1200, Mrub2271).

Methods

The gene annotation of the project was performed based on GENI-ACT gene annotation website instructions (http://www.geni-act.org/education/main/). Since the project is also disproving orthologs, the instructions and included results are modified slightly. After comparing the KEGG pathway, a BLAST was then performed on all *M. ruber* genes and was compared against the *E. coli* genome to determine any closely related genes. Cell localization websites were then used to determine the predicted location of the protein domains. Then, bioinformatics websites for analyzing protein function in other organisms were used to determine similarities or dissimilarities between b0262 and respective *M. ruber* genes. Lastly, a search for possible operons of the genes was done through the site IMG/M to determine if genes were part of operons. These steps were done to prove orthology and to disprove orthology.

Results

Panel A



Panel B

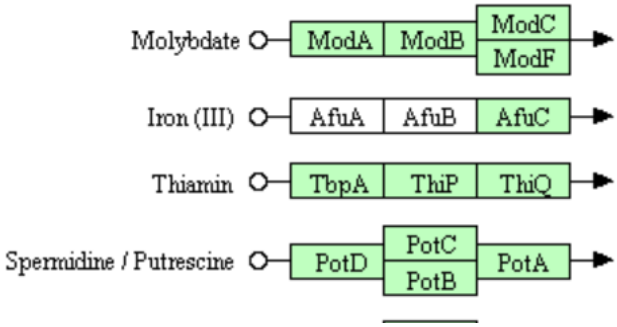


Figure 2. ABC Transport pathways between *M. ruber* and *E. coli.* share similarities. The images above are taken from the KEGG pathway for the predicted genes of the ABC Transport System. Panel A includes the genes used for the transport of Iron (Fe3+) in *M. ruber*. Panel B includes the genes for the transport of Iron in *E. coli*. Pathways were generated through The Kyoto Encyclopedia of Genes and Genomes (KEGG) database at http://www.genome.jp/kegg/.

Figure 2 is the first indicator of potential orthologs for ABC transporters between *M. ruber* and *E. coli*. The three green highlighted genes in Panel A are *afuA, afuB, afuC*. Each these genes contain their own pair of genes as follows: *afuA* (Mrub_1201, Mrub_2015), *afuB* (Mrub_1200, Mrub2271) and *afuC* (Mrub_1199, Mrub_2272). The highlighted gene in Panel B is *afuC* and contains a single gene involved with the transport of Iron: labeled as b0262. The *afuC* is present in both bacteria, which suggests the genes might be related to a common ancestor. The genes *afuA* and *afuB* are expected to be absent in *E. coli*. Further tests are required to determine the function and relation of these genes.

Table 1: *E.coli* b0262 and *M. ruber* Mrub_2272 & Mrub_1199 are orthologs

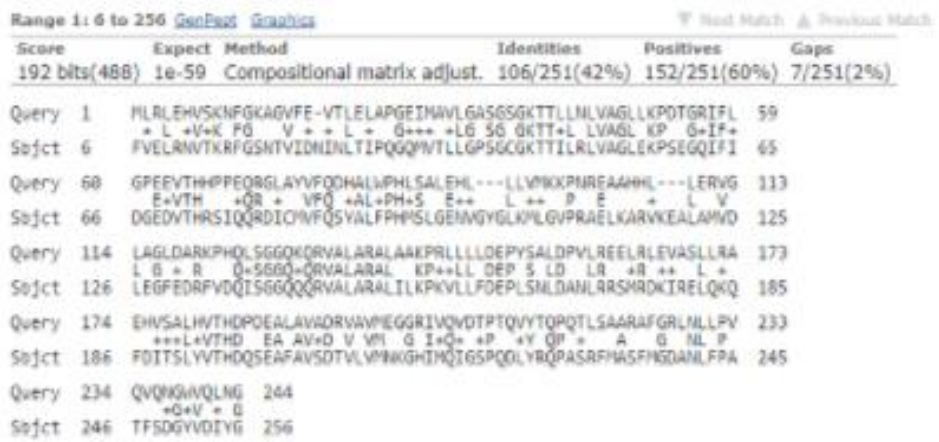
Bioinformatics tool used	***E. coli* b0262 gene**	**Mrub_2272 gene**	**Mrub_1199**
BLAST against opposite genome	Mrub_2272 Score: 240 E-value: 8e-78	b0262 Score: 240 E-value: 1e-77	b0262 Score: 192 E-value: 1e-59
CDD Data	COG3842 ABC-type Fe3+/spermidine/putrescine transport systems, ATPase components		COG3839 ABC-type sugar transport system, ATPase component
	E-value: 3.88e-165	E-value: 1.08e-137	2.75e-88
Cell localization	Localized partly in cytoplasmic membrane		
TIGRfam - Protein family	sulfate ABC transporter ATP binding TIGR00968	Polyamine ABC transporter, ATP binding TIGR01187	sulfate ABC transporter ATP binding TIGR00968
	E= 1.3e-113	E= 9.6e-101	E= 1.5e-69
Pfam - Protein family	PF0005 CL0023 P-loop_NTPase	PF0005 CL0023 P-loop containing nucleoside triphosphate hydrolase superfamily	PF0005 CL0023 P-loop containing nucleoside triphosphate hydrolase superfamily

Pfam - Protein family			
	E= 1.6e-37	E= 7.5e-36	E= 4.99e-33
Protein Database (PDB)	1V43 Crystal Structure of ATPase subunit of ABC Sugar Transporter E= 6.68272E-74	1VCI Crystal structure of the ATP-binding cassette of multisugar transporter from Pyrococcus horikoshii OT3 complexed with ATP E= 6.34144E-61	1VCI Crystal structure of the ATP-binding cassette of multisugar transporter from Pyrococcus horikoshii OT3 complexed with ATP E= 7.28893E-43
KEGG Pathway Map	ABC Transporters KEGG Number: 02010		

Table 1 summarizes the results between Mrub_1199 and Mrub_2272 and the supposed ortholog b0262. First row shows the BLAST alignment. The *M. ruber* genes both produced hits with the *E. coli* b0262 gene with low e-values. The b0262 did produce a hit for Mrub_2272 with a low e-value. This is the first type of evidence that shows an evolutionary relationship between the genes. The CDD yielded the same COG hit for b0262 and Mrub_2272. COG2842 refers to an ATPase component of the ABC transporter of Iron/ Spermidine/ Putrescine. Mrub_1199 produced a hit for COG3839, which is also an ATPase component, but for an ABC transporter of sugar. A similarity in structure and function is first observed. Cell localization predicts all the genes above to be located in in the cytoplasm. There is no cleavage site for the protein domains and there are no transmembrane helices either. Further interpretation of cell localization results are illustrated in figures 4 & 5 through TMHM and SignalP. TIGRfam pulled TIGR00968 for both b0262 and Mrub_1199, but pulled TIGR01187 for Mrub_2272. All TIGRfam hits are ATP binding domains and are a consistent function between the genes. Pfam further proved an orthologous relationship through the same family and clan names. The family represents ABC transporter and the clan represents nucleoside triphosphate hydrolase. PDB results had hits that were consistent with ATP binding domains. Even though the PDB displayed hits for an ABC sugar transporter, this does not disregard orthology since there are studies that show sugar transporters also transporting Iron (Wilkens 2015). The table concludes that b0262 is orthologous to Mrub_1199 and Mrub_2272. The overwhelming evidence confirms similarity in structure, function, and localization of the protein domain. Interestingly enough the *M. ruber*

genes are code for the 2 nucleotide binding domains described in Figure 1. while the single *E. coli* gene in the ABC transport system codes only for the nucleotide binding domain.

Panel A

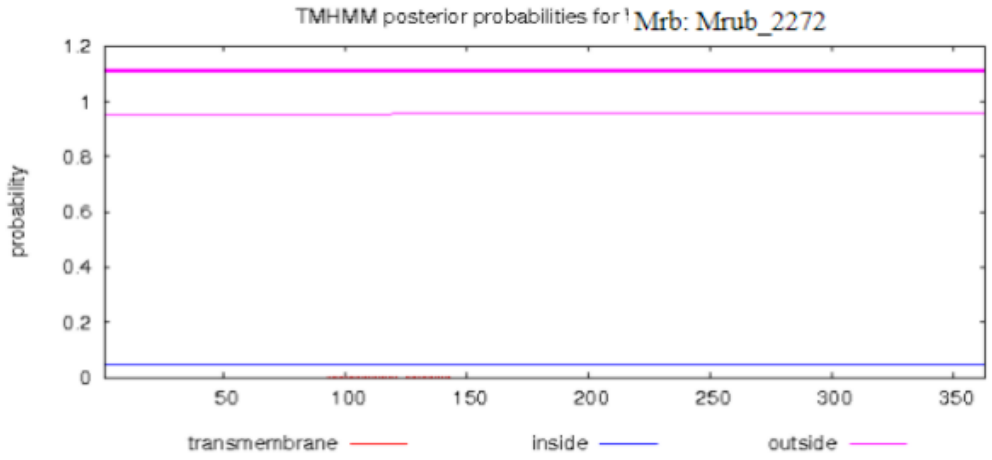


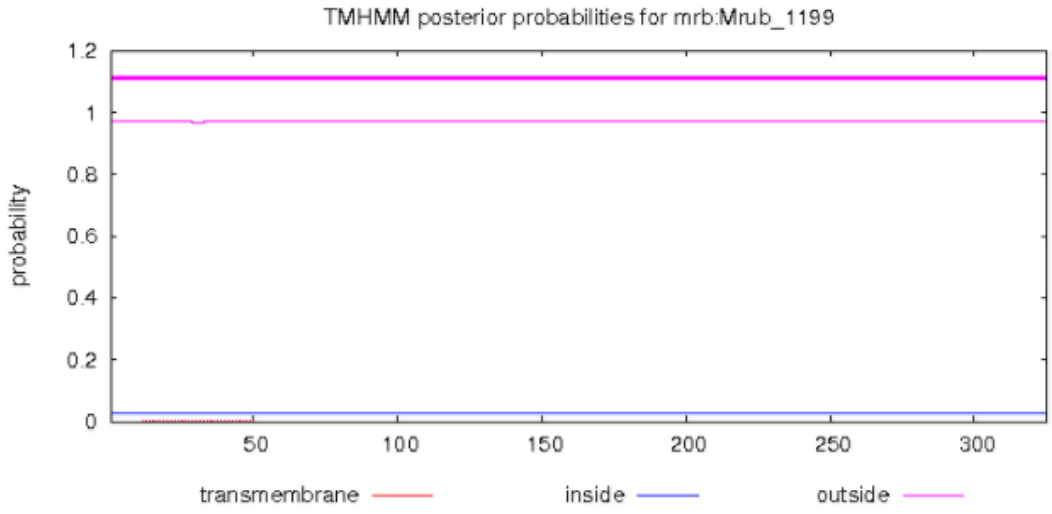
Panel B

Figure 3. Successful BLAST search of Mrub_1199 & Mrub_2272 against *E. coli* genome yielding hits for b0262. Panel A shows the BLAST search of Mrub_1199 (query sequence) against the *E. coli* genome (search set) yielding a hit for the sought out gene b0262. This was not the first hit, but the low e-value and high bit score indicate a relationship not based on chance. Panel B shows the BLAST search of Mrub_2272 (query sequence) against the *E. coli* genome (search set) yielding a hit for b0262. Again, this was not the first hit, but still has an acceptable e-

value and bit score to demonstrate a significant relationship between the genes. Analysis was performed using the NCBI BLAST bioinformatics tool at http://www.ncbi.nlm.nih.gov

Figure 3 shows the next piece of evidence to prove the orthology of the *M. ruber* genes. Panel A has an e-value of 1e-59 & Panel B has an e-value of 1e-77. These are low e-values and both yield a hit matching the gene for the transport of Iron (*afuC*) in *E. coli*. There is a low chance this is due to random alignment. These results prove structural similarities between the proteins the genes code for. Evidence for the function of the genes can help further prove if these genes are orthologs.





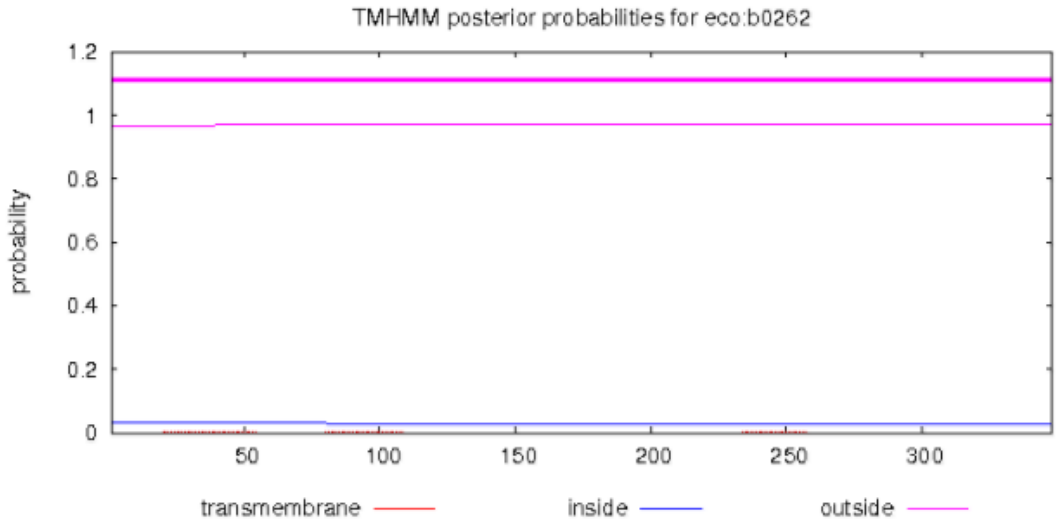
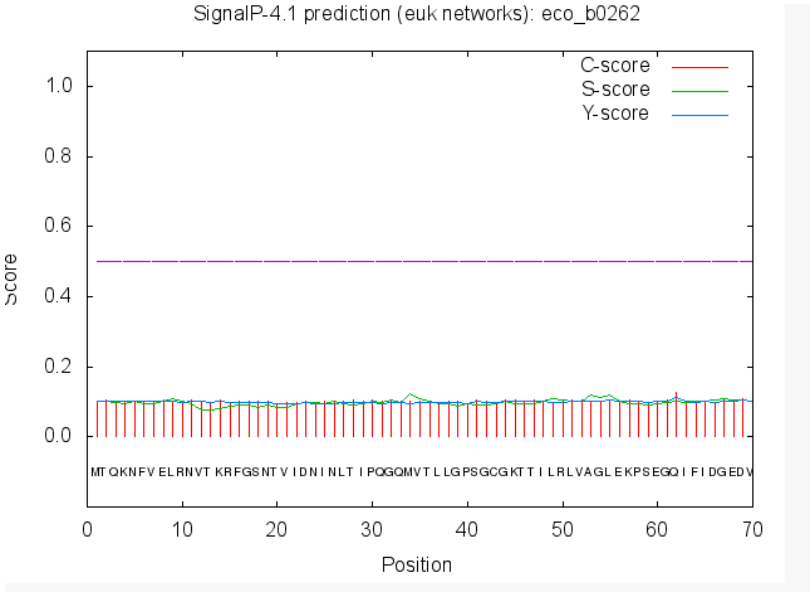
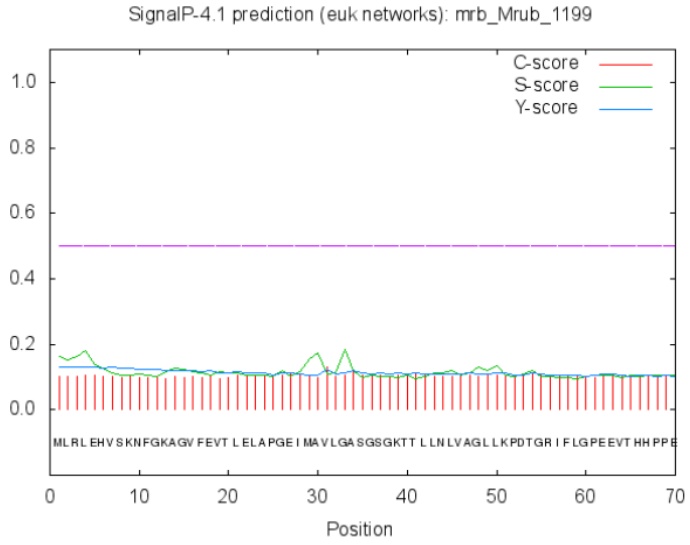


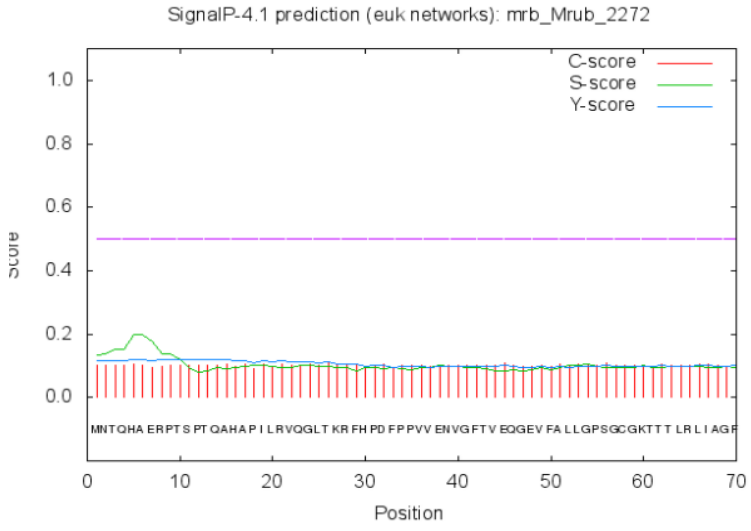
Figure 4. Mrub_2272 (top), Mrub_1199 (middle) b0262 and (bottom) do not contain any transmembrane domains. Titles of the hydropathy plots indicate the gene it refers to. The light pink line near the top of each graph demonstrates the absence of transmembrane domains. The data shows similarity in localization of protein domains and hints at the possibility of similar function. TMHMM Server v 2.0 http://www.cbs.dtu.dk/services/TMHMM was used to create hydropathy plots.

This bioinformatics tool is useful for determining location and any membrane components. None of the genes of interest show any transmembrane domains and suggest localization within the cytoplasm of the cell. The relationship between these genes is furthered proved by understanding the location of the protein domains are the alike throughout all the genes. It is becoming more apparent that an orthologous relationship might exist between the *E. coli* genes and *M. rub* genes.




```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C     62       0.125
  max. Y     62       0.112
  max. S     34       0.124
  mean S      1-61    0.097
       D      1-61    0.104   0.450   NO
Name=eco_b0262  SP='NO' D=0.104 D-cutoff=0.450 Networks=
```






```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C     26       0.109
  max. Y     11       0.121
  max. S      5       0.199
  mean S      1-10    0.154
       D      1-10    0.139   0.450   NO
Name=mrb_Mrub_2272      SP='NO' D=0.139 D-cutoff=0.450 I
```

```
easure  Position  Value    Cutoff   signal peptide?
ax. C     31       0.129
ax. Y     12       0.123
ax. S     33       0.184
ean S      1-11    0.133
     D     1-11    0.128   0.450   NO
e=mrb_Mrub_1199      SP='NO' D=0.128 D-cutoff=0.450 Net
```

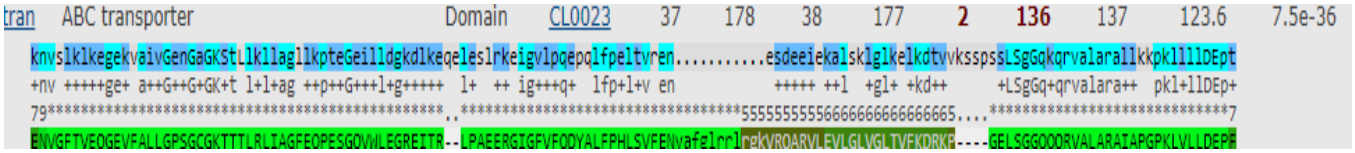
Figure 5. b0262, Mrub_1199 & Mrub_2272 indicate no cleavage sites in any of the genes based on SignalP output. Titles of the genes are labeled at the top of each output. D values were below

cutoffs for all plots. Plots created by Signal P server v. 4.1
http://www.cbs.dtu.dk/services/SignalP

All D-values were below the cutoff of 0.450. This shows the absence of any cleavage sites for all genes. The function of all genes is consistent. These plots are useful for comparing the localization of the genes and determining similarity in function.

Panel A

Panel B



Panel C

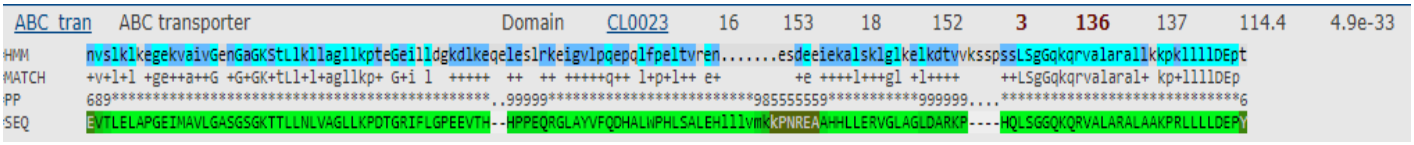


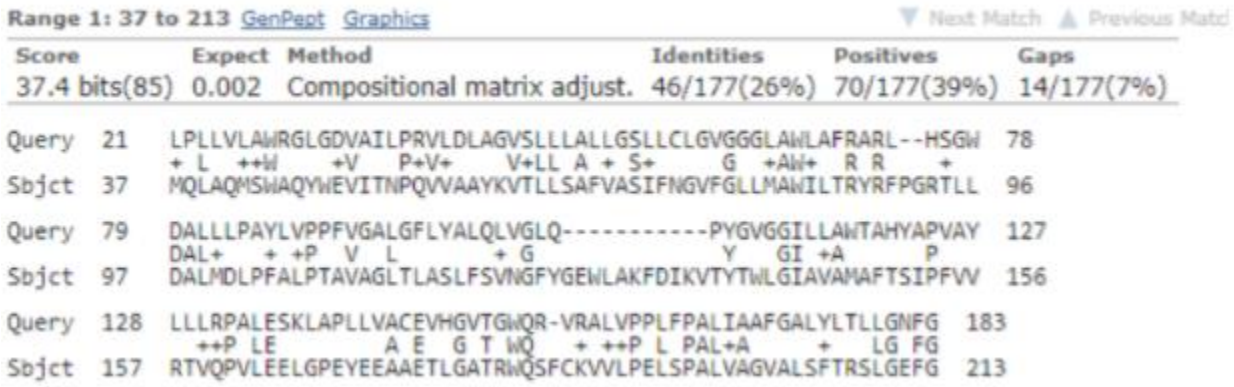
Figure 6. Mrub_1199 (Panel C), Mrub_2272 (Panel B) & b0262 (Panel A) have similarities in highly conserved amino acids and code for the same domain**.** The Pfam results show all genes in the same clan CL0023 which is part of the nucleotide triphosphate hydrolase family. There are several amino acids that are conserved in all three alignments. Panel A has a glycine at the 9th position that is also present in Panel B at position 10 and in Panel C at position 9. The pairwise alignment was created using the Pfam website http://pfam.sanger.ac.uk/search.

Figure 6 is able to tell us the clan number about the genes, which corresponds to a nucleotide binding domain. These genes seem to be consistent with the function as a nucleotide binding domain. The Pfam identifies the genes as part of the nucleotide triphosphate hydrolase family. The presence of the conserved amino acids between the genes further proves similarity in function and structure. The results thus far have provided adequate evidence to provide an orthologous relationship between the genes.

Figure 7. The genes b0262 (top), Mrub_1199 (middle), Mrub_2272 (bottom) are not part of operons according to color by COG. The gene is identified by a red line. All genes are red since they have similar functions. Non of the surrounding genes are the same color in any image. Images were created by IMG/M website: http://img.jgi.doe.gov/ and color coordinated by COG hits.

Figure 7 is the last piece of evidence to confirm orthology between the *M. ruber* genes and the *E. coli genes.* The images show that none of the genes are part of operons when coloring by COG hits. This suggests that the function of the proteins of these genes is closely related.

Panel A



Range 1: 37 to 213 GenPept Graphics

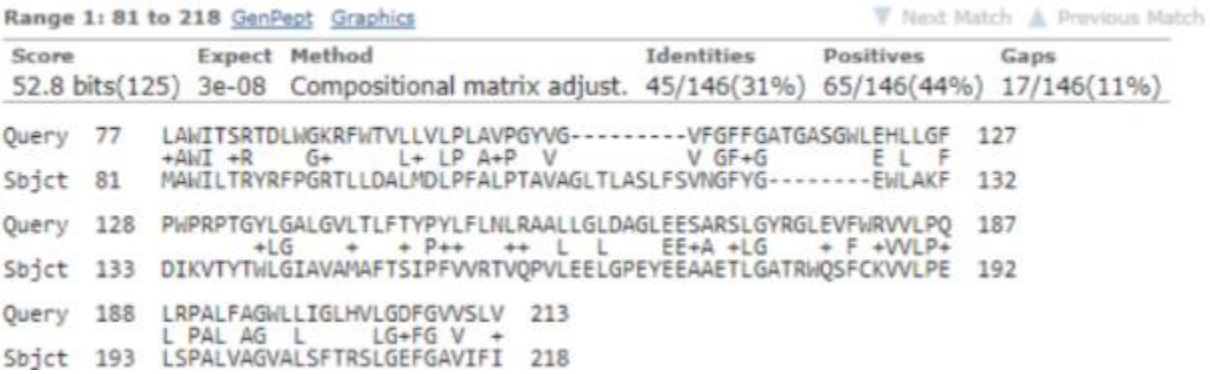
Score	Expect	Method	Identities	Positives	Gaps
37.4 bits(85)	0.002	Compositional matrix adjust.	46/177(26%)	70/177(39%)	14/177(7%)

```
Query  21   LPLLVLAWRGLGDVAILPRVLDLAGVSLLLALLGSLLCLGVGGGLAWLAFRARL--HSGW  78
            + L  ++W    +V   P+V+    V+LL A + S+     G  +AW+  R R    +
Sbjct  37   MQLAQMSWAQYWEVITNPQVVAAYKVTLLSAFVASIFNGVFGLLMAWILTRYRFPGRTLL  96

Query  79   DALLLPAYLVPPFVGALGFLYALQLVGLQ-----------PYGVGGILLAWTAHYAPVAY  127
            DAL+   + +P  V  L        + G                Y   GI +A      P
Sbjct  97   DALMDLPFALPTAVAGLTLASLFSVNGFYGEWLAKFDIKVTYTWLGIAVAMAFTSIPFVV  156

Query  128  LLLRPALESKLAPLLVACEVHGVTGWQR-VRALVPPLFPALIAAFGALYLTLLGNFG  183
             ++P LE         A E  G T WQ   + ++P L PAL+A      +   LG FG
Sbjct  157  RTVQPVLEELGPEYEEAAETLGATRWQSFCKVVLPELSPALVAGVALSFTRSLGEFG  213
```

Panel B



Range 1: 81 to 218 GenPept Graphics

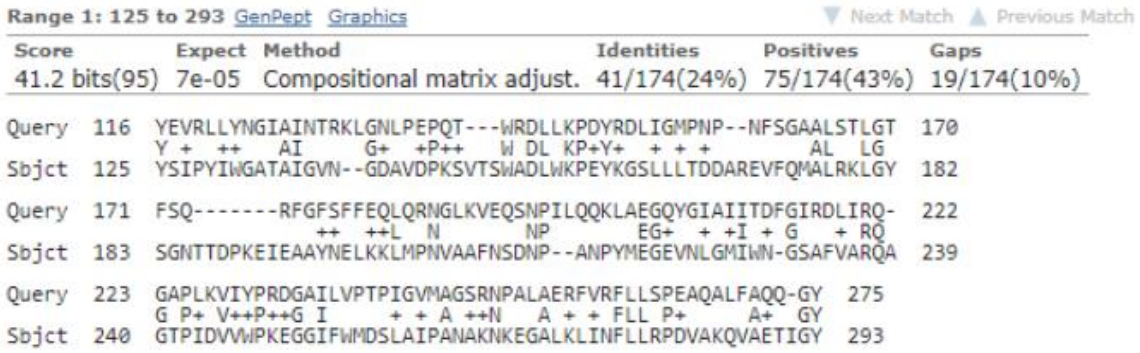
Score	Expect	Method	Identities	Positives	Gaps
52.8 bits(125)	3e-08	Compositional matrix adjust.	45/146(31%)	65/146(44%)	17/146(11%)

```
Query  77   LAWITSRTDLWGKRFWTVLLVLPLAVPGYVG---------VFGFFGATGASGWLEHLLGF  127
            +AWI +R    G+     L+ LP A+P  V          V GF+G        E L  F
Sbjct  81   MAWILTRYRFPGRTLLDALMDLPFALPTAVAGLTLASLFSVNGFYG--------EWLAKF  132

Query  128  PWPRPTGYLGALGVLTLFTYPYLFLNLRAALLGLDAGLEESARSLGYRGLEVFWRVVLPQ  187
                    +LG    +   + P++    ++  L  L    EE+A +LG    + F +VVLP+
Sbjct  133  DIKVTYTWLGIAVAMAFTSIPFVVRTVQPVLEELGPEYEEAAETLGATRWQSFCKVVLPE  192

Query  188  LRPALFAGWLLIGLHVLGDFGVVSLV  213
            L PAL AG  L     LG+FG V  +
Sbjct  193  LSPALVAGVALSFTRSLGEFGAVIFI  218
```

Panel C



Range 1: 125 to 293 GenPept Graphics

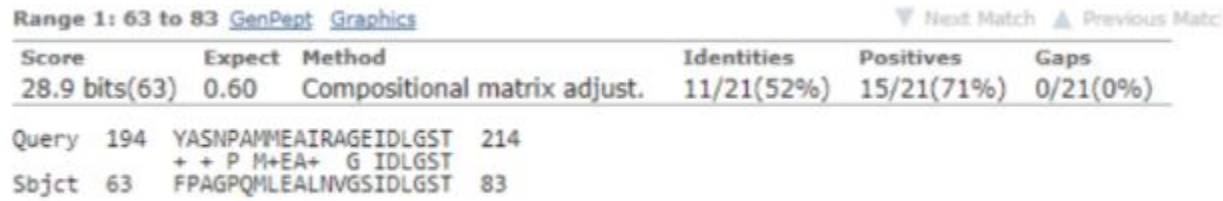
Score	Expect	Method	Identities	Positives	Gaps
41.2 bits(95)	7e-05	Compositional matrix adjust.	41/174(24%)	75/174(43%)	19/174(10%)

```
Query  116  YEVRLLYNGIAINTRKLGNLPEPQT---WRDLLKPDYRDLIGMPNP--NFSGAALSTLGT  170
            Y +  ++   AI     G+  +P++   W DL KP+Y+  + + +         AL  LG
Sbjct  125  YSIPYIWGATAIGVN--GDAVDPKSVTSWADLWKPEYKGSLLLTDDAREVFQMALRKLGY  182

Query  171  FSQ-------RFGFSFFEQLQRNGLKVEQSNPILQQKLAEGQYGIAIITDFGIRDLIRQ-  222
                      ++  ++L  N       NP       EG+  + +I + G   + RQ
Sbjct  183  SGNTTDPKEIEAAYNELKKLMPNVAAFNSDNP--ANPYMEGEVNLGMIWN-GSAFVARQA  239

Query  223  GAPLKVIYPRDGAILVPTPIGVMAGSRNPALAERFVRFLLSPEAQALFAQQ-GY  275
            G P+ V++P++G I     + + A ++N   A + + FLL P+     A+  GY
Sbjct  240  GTPIDVVWPKEGGIFWMDSLAIPANAKNKEGALKLINFLLRPDVAKQVAETIGY  293
```

Panel D



Range 1: 63 to 83 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
28.9 bits(63)	0.60	Compositional matrix adjust.	11/21(52%)	15/21(71%)	0/21(0%)

```
Query  194  YASNPAMMEAIRAGEIDLGST  214
            + + P M+EA+  G IDLGST
Sbjct  63   FPAGPQMLEALNVGSIDLGST  83
```

Figure 8. BLAST of Mrub_1200, Mrub_1201, Mrub_2271, Mrub_2015 against the *E. coli* genome yielding no hits resembling b0262. Panel A is the BLAST for Mrub_1200 that showed a first hit for b_2424 with a very high e-value of 0.002. This suggests no evolutionary relationship with gene. Panel B is the BLAST for Mrub_2271 that also yielded b_2424 with a high e-value of 3e-08. This might also show the absence of an evolutionary relationship, but requires further investigation. Panel C shows the BLAST for Mrub_1201 with the first hit being b_1123 with an e-value of 7e-06. This value does not provide concrete evidence of an evolutionary relationship and also require further study. Panel D is the BLAST for Mrub_2015 which shows its first hit as b0936 with a very high e-value of 0.60. This is gene definitely does not share a common ancestor with *E. coli.* Analysis was performed using the NCBI BLAST bioinformatics tool at http://www.ncbi.nlm.nih.gov

Figure 8 demonstrates that the genes Mrub_1201 and Mrub_2015 are not related to any genes in the *E. coli* genome, thus excluding the genes as orthologs. Mrub_2271 and Mrub_1201 have lower e-values which might suggest similarities in structure however; the gene that is pulled by BLAST does not match the description or function of the *M. ruber* genes. These genes may most likely not be labeled as orthologs, but require bioinformatics tools to further explain the function and structure of these genes.

Table 2*:* **The functions of *M. Ruber* Mrub_2271, Mrub_2015, Mrub_1200 & Mrub_1201**

Bioinformatics tools used	**Mrub_2271 gene**	**Mrub_2015 gene**	**Mrub_1200 gene**	**Mrub_1201 gene**
BLAST against opposite genome	b2424 Score: 52.8 E-value: 3e-08	b0936 Score: 28.9 E-value: 0.60	b2424 Score: 37.4 E-value:0.002	b1123 Score: 41.2 E-value: 7e-05
CDD Data	COG1178 ABC-type Fe3+ transport system, permease component	COG1840 ABC-type Fe3+ transport system, periplasmic component	COG1178 ABC-type Fe3+ transport system, permease component	COG1840 ABC-type Fe3+ transport system, periplasmic component
	E-value: 4.26e-91	E-value: 2.03e-40	E-value: 2.29e-22	E-value: 1.88e-48
Cell localization	Cytoplasmic membrane	Periplasm	cytoplasmic membrane	Equal chance of being located near cytoplasmic membrane or periplasmic membrane
TIGRfam - Protein family	ABC transporter, permease pro PF00528	ABC transporter periplasmic binding protein TIGR01254	2-aminoethylphosphonate ABC transport TIGR03255	putative 2-aminoethylphosphonate ABC TIGR03261
	E-value: 6.8e-12	E-value: 4e-06	E-value: 4.1e-05	E-value: 2.9e-09
Pfam - Protein family	BPD transporter like PF00528	Bacterial extracellular solute-binding protein F13531	Binding-protein-dependent transport system inner membrane component PF00528	Bacterial extracellular solute-binding protein F13531
	E= 8.7e-10	E-value= 5.44e-22	E- value= 3e-07	E-value= 2.9e-09

Protein Database (PDB)	**3D31** ModBC from Methanosarcina acetivorans E= 0.123396	**3WAE** X-ray structure of Fe(III)-bicarbonates-ttfbpa, a ferric ion-binding protein from thermus thermophilus HB8 E-value:8.8996E-121	**No Hits**	**4R72** Structure of the periplasmic binding protein AfuA from Actinobacillus pleuropneumoniae (apo form) E= 1.255E-16
KEGG Pathway Map	ABC Transporters KEGG Number: 02010			

Table 2 shows the function between Mrub_2271, Mrub_2015, Mrub_1200 & Mrub_1201 in order to prove these genes are not any orthologous to *E. coli.* The BLAST search did not provide compelling evidence for potential orthologs in *E. coli*, see Figure 8 provides a further explanation. The CDD pulled COG1178 for Mrub_2271 & Mrub_1200, which signifies a permease component to the ABC Iron transport system. COG1840 was hit for Mrub_2015 and Mrub_1201 and indicates a periplasmic component to the ABC Iron transport system. Cell localization, which includes PSORT-B and TMHM, predict Mrub_2271 and Mrub_1200 to be in the cytoplasmic membrane. Cell localization predicts Mrub_2015 and Mrub_1201 to be in the periplasmic space. TIGRfam produces hits for Mrub_2271 and Mrub_2015 as a permease and periplasmic protein, respectively. Mrub_1200 and Mrub_1201 produce hits for ABC transporter domains, but don't match the system for Iron transport. Pfam further proves the function of the *M. ruber* genes. The family PF00528 for a binding protein dependent domain, located in the membrane, corresponds to genes Mrub_2271 & Mrub_1200. The family F13531 for solute-binding proteins corresponds to genes Mrub_2015 & Mrub_1201. Lastly, the PDB does produce a periplasmic binding protein for Mrub_1201 and an Iron binding protein for Mrub_2015, which is consistent with a solute binding protein. The PDB hits for the rest of the genes does not further prove the function of the genes as permease domains, but this piece of evidence can be overlooked. The table concludes that not only is there any compelling evidence for orthologs to these genes in *E. coli*, it also predicts that Mrub_2271 and Mrub_1200 are the permease domains in the membrane & Mrub_2015 and Mrub_1201 function as solute binding proteins within the periplasm.

Panel A

Panel B

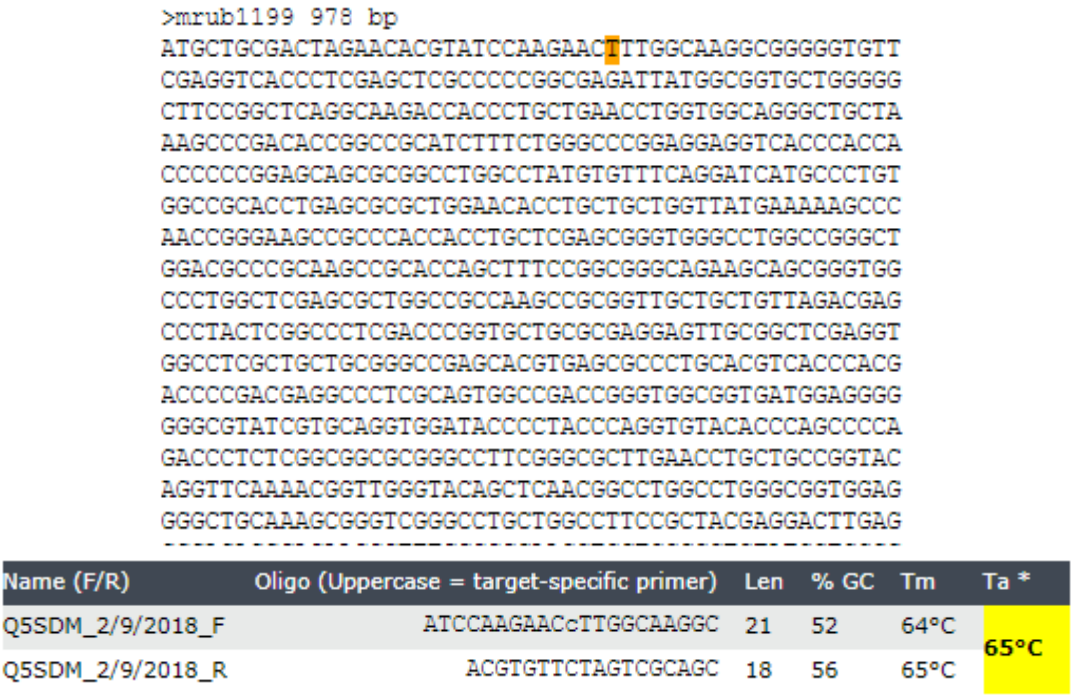
Panel C

Figure 9. Mrub_2015 is not part of an operon (Panel A). Mrub_2271 and Mrub_2272 (left to right) are part of the same operon (Panel B). Mrub_1199, Mrub_1200, Mrub_1201 (left to right) are part of an operon (Panel C).

Panel A shows Mrub_2015 not part of an operon, which might be due to its location in the periplasm so it does not have to be directly in contact with another domain. In Panel B, Mrub_2271 is the permease and Mrub_2272 is the nucleotide binding domain and are transcribed in the same orientation. In Panel C, Mrub_1199 is the nucleotide binding domain, Mrub_1200 is the permease, and Mrub_1201 is the substrate binding protein. These three genes are transcribed in the same orientation. This proves the function of all the genes and orientation of the coded genes

Site Directed Mutagenesis

Further research will include a point mutation at a highly conserved amino acid. A highly conserved amino acid in Mrub_1199 was located at position F28 by the HMM logo. In the HMM logo the alignment begins at position 18. The Phenylalanine at the 11th position has the DNA sequence TTT which corresponds to the mRNA sequence UUU. In order to mutate the gene through a single point nucleotide substitution, the 31st position of the gene is mutated from T -->C to substitute from a Phenylalanine to a Leucine. The figure below shows the substitution and the primer required to flank the region and substitute the amino acid. The mutation should create a loss of function or a drastic change in function (Betts and Russel 2003).



>mrub1199 978 bp
ATGCTGCGACTAGAACACGTATCCAAGAACTTTGGCAAGGCGGGGGTGTT
CGAGGTCACCCTCGAGCTCGCCCCCGGCGAGATTATGGCGGTGCTGGGGG
CTTCCGGCTCAGGCAAGACCACCCTGCTGAACCTGGTGGCAGGGCTGCTA
AAGCCCGACACCGGCCGCATCTTTCTGGGCCCGGAGGAGGTCACCCACCA
CCCCCCGGAGCAGCGCGGCCTGGCCTATGTGTTTCAGGATCATGCCCTGT
GGCCGCACCTGAGCGCGCTGGAACACCTGCTGCTGGTTATGAAAAAGCCC
AACCGGGAAGCCGCCCACCACCTGCTCGAGCGGGTGGGCCTGGCCGGGCT
GGACGCCCGCAAGCCGCACCAGCTTTCCGGCGGGCAGAAGCAGCGGGTGG
CCCTGGCTCGAGCGCTGGCCGCCAAGCCGCGGTTGCTGCTGTTAGACGAG
CCCTACTCGGCCCTCGACCCGGTGCTGCGCGAGGAGTTGCGGCTCGAGGT
GGCCTCGCTGCTGCGGGCCGAGCACGTGAGCGCCCTGCACGTCACCCACG
ACCCCGACGAGGCCCTCGCAGTGGCCGACCGGGTGGCGGTGATGGAGGGG
GGGCGTATCGTGCAGGTGGATACCCCTACCCAGGTGTACACCCAGCCCCA
GACCCTCTCGGCGGCGCGGGCCTTCGGGCGCTTGAACCTGCTGCCGGTAC
AGGTTCAAAACGGTTGGGTACAGCTCAACGGCCTGGCCTGGGCGGTGGAG
GGGCTGCAAAGCGGGTCGGGCCTGCTGGCCTTCCGCTACGAGGACTTGAG

Name (F/R)	Oligo (Uppercase = target-specific primer)	Len	% GC	Tm	Ta *
Q5SDM_2/9/2018_F	ATCCAAGAACcTTGGCAAGGC	21	52	64°C	65°C
Q5SDM_2/9/2018_R	ACGTGTTCTAGTCGCAGC	18	56	65°C	

Figure 6. Single point mutation (T->C) at position 31 of the Mrub_1199 gene. The primers needed are labeled above and must be annealed at 65℃. The amino acid substitute is well conserved among species and will most likely result in a drastic change in function. Primers produced from NEB Base Changer.

Conclusion

Proving Orthologs

The results of the project proved Mrub_1199 & Mrub_2272 are orthologs to b0262 and that Mrub_1200, Mrub_1201, Mrub_2271, and Mrub_2015 do not have orthologs to *E. coli*. First is proving the relationship for the first set of genes. The evidence for common ancestry between b0262 and Mrub_1199 & Mrub_2272 first started with the KEGG pathway comparison showing that the *afuC* in *E. coli* (b0262) and *M. ruber* (Mrub_1199 &Mrub_2272) were predicted to be orthologous genes coding for Iron (Fe3+) transport. A BLAST was then performed for the predicted orthologous *M. ruber* genes, which both yielded hits for the b0262 gene. Despite the results not being the first hit for BLAST, it is well within acceptable e-values. The COG hits are consistent with ATPase domains, but not strictly for Iron, hits included Spermidine/ Putrescine, and sugar. As mentioned before a study showed these kinds of transporters also transporting Iron (Wilkens 2015). Cell localization was then performed using TMHM and SignalP and predicted the genes in the cytoplasm. TIGRfam and PFAM both yielded the same predicted function of a nucleotide binding domain for all genes with low e-values. Pfam also displayed conserved amino acids between the genes. The operon comparison by IMG/M by COG coloring showed similarity in operon structure and function. These results, which prove the similarity in structure, localization, and function, provide adequate evidence that these genes are orthologs. When referring back to Figure 1, which describes the general, structure of an ABC transporter, these genes match the 2 nucleotide binding domains described in a general ABC transporter.

Proving genes that are not orthologs

The KEGG pathway map also predicted that there are no orthologs to *M. ruber's afuA* and *afuB* in *E. coli*. That being said, the genes Mrub_1200, Mrub_1201, Mrub_2015, Mrub_ 2271 do not have orthologs to *E. coli* and serve a separate function from b0262. A BLAST search of these genes against the *E. coli* genome yielded different genes in *E. coli*, but with high e-values. This does not suggest an evolutionary relationship between the genes and *E. coli*. COG hits for Mrub_2271 & Mrub_1200 showed the permease domain for Iron transport. The rest of the bioinformatics tools including TMHM, Pfam, and TIGRfam all predict the genes Mrub_2271 & Mrub_1200 to be the transmembrane domain (permease portion) of the ABC transporter diagrammed in Figure 1. The function and location are not consistent with any genes used for Iron transport in *E. coli.*

Lastly, the genes Mrub_2015 and Mrub_1201 are proven to be the substrate binding protein based on the bioinformatics tools (CDD, TIGRfam, Pfam). Specifically, the COG hits for these genes are both for Iron transport and represent solute binding proteins, which is confirming

evidence for similar function. Cell localization (TMHM and SignalP) states the genes lay within the periplasm and all produced hits for protein families are consistent with this function. Interestingly enough, Figure 1 states there will typically be one substrate binding protein in the ABC transport system of Iron, but the *M.ruber* genome contains 2 genes for two proteins used for substrate binding. The operon structure for the genes based on KEGG shows that the periplasmic substrate binding Mrub_2015 is not part of an operon. The permease Mrub_2271 and nucleotide binding domain are part of the same operon. The permease Mrub_1199, nucleotide binding domain Mrub_1200, and substrate binding Mrub_1201 are part of the same operon. The purpose of the operon structure is to ensure the predicted genes that are not orthologs have a confirmed function in *M.rube*r. In this case there were similar functions between the genes. Overall, the results confirm that *afuA* (Mrub_2015 and Mrub_1201) and *afuB* (Mrub_1200, Mrub2271) from *M. ruber* do not have orthologs in *E. coli.*

Literature Cited

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

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/.

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.

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

Biolabs, N. E. Home - NEB | New England Biolabs. Home - NEB | New England Biolabs. Available from: https://www.neb.com/.

Bioinformatics [Internet]. London: Springer London; c2015.

Blount ZD. 2015. The unexhausted potential of *E. coli*. eLife 4.

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/

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/

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.

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

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/

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.

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/

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

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. [Internet]. [cited 2018 Feb 6] 43(Database issue):D222-2. Available from https://www.ncbi.nlm.nih.gov/pubmed/25414356

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.

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.

Phylogenetic Diversity. [Internet] U.S. Department of Energy Joint Genome Institute; [2015 Dec 16]. Available from http://jgi.doe.gov/our-science/science-programs/microbialgenomics/phylogenetic-diversity/
Reiße S, Garbe D, Brück T. 2015. Identification and characterization of a highly thermostable crotonase from meiothermus ruber. Journal of Molecular Catalysis B: Enzymatic 112:40-4.

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.

Studer RA and Robinson-Rechavi M. 2009. How confident can we be that orthologs are similar, but paralogs differ? Trends in Genetics 25(5):210-6.

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.

Tindall et al. 2010. Complete genome sequence of Meiothermus ruber type strain. Stand Genomic Sci 3(1): 26-36

Wang S, Ogata M, Horita S, Ohtsuka J, Nagata K, Tanokura M. 2014. A novel mode of ferric ion coordination by the periplasmic ferric ion- binding subunit FbpA of an ABC- type iron transporter from thermus thermophilus HB8. Acta Crystallographica Section D 70(1):196-202.

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

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.