Mrub_2642, Mrub_1054, and Mrub_1059 genes are orthologs of the *Escherichia coli* genes b2942, b0159, and b2687 genes, respectively, which code for methionine adenosyltransferase, adenosylhomocysteine nucleosidase, and S-ribosylhomocysteine lyase

Nicholas M. Orslini  
*Augustana College, Rock Island Illinois*

Dr. Lori R. Scott  
*Augustana College, Rock Island Illinois*

Follow this and additional works at: [http://digitalcommons.augustana.edu/biolmruber](http://digitalcommons.augustana.edu/biolmruber)

Part of the Biochemistry Commons, Bioinformatics Commons, Biology Commons, Evolution Commons, Genomics Commons, Molecular Biology Commons, and the Molecular Genetics Commons

Augustana Digital Commons Citation

Orslini, Nicholas M. and Scott, Dr. Lori R.. "Mrub_2642, Mrub_1054, and Mrub_1059 genes are orthologs of the *Escherichia coli* genes b2942, b0159, and b2687 genes, respectively, which code for methionine adenosyltransferase, adenosylhomocysteine nucleosidase, and S-ribosylhomocysteine lyase" (2017). *Meiothermus ruber Genome Analysis Project.*  
[http://digitalcommons.augustana.edu/biolmruber/20](http://digitalcommons.augustana.edu/biolmruber/20)

This Student Paper is brought to you for free and open access by the Biology at Augustana Digital Commons. It has been accepted for inclusion in *Meiothermus ruber Genome Analysis Project* by an authorized administrator of Augustana Digital Commons. For more information, please contact digitalcommons@augustana.edu.
Mrub_2642, Mrub_1054, and Mrub_1059 genes are orthologs of the *Escherichia coli* genes b2942, b0159, and b2687 genes, respectively, which code for methionine adenosyltransferase, adenosylhomocysteine nucleosidase, and S-ribosylhomocysteine lyase.

Nicholas M. Orslni

*Dr. Lori R. Scott Laboratory*

Biology Department, Augustana College

639 38th Street, Rock Island, IL 61201

**Introduction**

**Why should we study *M. ruber***?

*Meiothermus ruber* is a gram-negative, rod-shaped, aerobic, red-pigmented, and thermophilic bacterium that thrives in temperatures ranging from 35-70°C (Tindall *et al.*, 2010). Interestingly, this microorganism was first isolated in hot springs of Kamchatka, Russia, as the hot springs provide a preferential environment for optimal growth (Loginova *et al.*, 1975). This organism is important to study because there are over 30,000 publications of *E. coli* on PubMed, where there are only 28 publications of *M. ruber* (Scott, personal communication). Also, there isn’t as much information available regarding the genes inside its genome and the role that these genes play in the various functions of this organism. Some may question why even study these less well-known bacteria, but there is a need to understand the similarities and differences between model organisms, such as *E. coli*, and bacteria like *M. ruber*. These organisms may provide insight to new genes or variants of traditionally understood processes that could lead to a better understanding of the intricacies of bacterial genomes and their remarkable capabilities. In fact, the Joint Genome Institute (JGI), has started a project called the Encyclopedia of Bacteria and
Archaea (GEBA) with the goal of trying to find new important research on bacteria who haven’t been well-studied (Wu et al., 2009). With this said, it is important that we do not put aside these bacterial species that could potentially provide us with some new insight. By using *E. coli* as a control, we attempt to unravel new information that doesn’t currently exist with regards to the *M. ruber* genome.

**Methionine degradation**

Methionine adenosyltransferase is an enzyme that catalyzes the formation of a sulfonium compound called S-adenosyl-L-methionine (SAM or AdoMet) (Markham *et al.*, 1980). This unusual reaction cleaves the triplyphosphate chain of the ATP to pyrophosphate and phosphate prior to sulfonium production. The sulfonium compound created is a high energy reagent that can transfer its methyl group to a large variety of acceptor substrates, such as nucleic acids, proteins, phospholipids, amines, etc., it becomes a very important biological compound for the organism. For example, quorum sensing as described in the next section is a biological reaction that this compound takes place in. The enzyme takes the shape of a homotetramer, essentially a dimer of dimers with active sites located at each interface, which is required for its catalytic role. Some important cofactors other than ATP are K⁺ and Mg²⁺, aiding in the catalysis of this reaction by making the homotetramer more stable in its active site. As seen in Figure 1, the reaction includes the substrates, ATP, L-methionine, and water, a reaction catalyzed by methionine adenosyltransferase (metK) to yield the products SAM, inorganic phosphate, and diphosphate.

![Figure 1](https://ecocyc.org/ECOLI/NEW-IMAGE?type=REACTION&object=S-ADENMETSYN-RXN)

Figure 1. Methionine degradation reaction involving the metK enzyme in *E. coli* K12 to produce the intracellular alkylating agent S-adenosyl-L-methionine (SAM). Substrates, products, genes/enzymes, and the direction of catalysis most favored in this reaction are shown in this figure. Image was taken from [https://ecocyc.org/ECOLI/NEW-IMAGE?type=REACTION&object=S-ADENMETSYN-RXN](https://ecocyc.org/ECOLI/NEW-IMAGE?type=REACTION&object=S-ADENMETSYN-RXN)
The \textit{metK} gene is extremely important to the growth of \textit{E. coli} K12 and cell division. It breaks down methionine into a product that can be used in other biochemical reactions, making it critical to cell survival. A deletion of the \textit{metK} gene is lethal unless a rescue plasmid carrying a functional \textit{metK} gene is present. With limited gene expression, genomic DNA methylation decreased and cell division was disrupted (Wei \textit{et al.}, 2002). When methionine adenosyltransferase activity falls below a certain threshold, cells end up producing long filaments and regularly dispersed nucleoids. Furthermore, expression of a plasmid-carried \textit{metK} gene will restore normal growth and stop filamentation (Newman \textit{et al.}, 1998). Due to the importance of this enzyme, it has been the target for development of antimicrobial and anticancer drugs (Markham \textit{et al.}, 1980). As this enzyme plays a great role in many organism’s enzymatic systems, it is essential that we study and understand how it is produced. Some of these mechanisms include methylation, as well as other epigenetic modifications, affecting transcription, gene stability, and parental imprinting. It directly impacts chromatin structure and can modulate gene transcription, or even completely silence or activate genes, without mutation to the gene itself. Though the mechanisms of this genetic control are complex, hypo- and hypermethylation of DNA is implicated in many diseases. Also, methylation of proteins has a regulatory role in protein-protein interactions, protein-DNA interactions, and protein activation (Wei \textit{et al.}, 2002). We specifically will focus on SAM production in the \textit{Meiothermus ruber} bacteria and the role that this sulfonium product plays in the production of autoinducer II: an important molecule in the activation of quorum sensing.

\textbf{Autoinducer (AI-2) biosynthesis I (\textit{mtn/luxS} genes)}

The exchange of extracellular signalling molecules, called autoinducers, is named quorum sensing or cell-to-cell communication. This process allows cells to coordinate gene expression once a certain density of cells is obtained. Factors, such as bioluminescence, virulence factor expression, antibiotic production, and biofilm development benefit from this community cooperation. The \textit{mtn} gene encoding 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase enzyme yields a reaction product, S-ribosyl-L-homocysteine, that in turn produces the autoinducer 2 (AI-2) from the next enzyme being examined in this research paper, \textit{luxS} or S-ribosylhomocysteine lyase. AI-2 production depends on several growth conditions: nutrients, pH, osmolarity, oxygen, growth rate, and stress factors leading to control of both intraspecies and interspecies signalling (Cornell \textit{et al.}, 1998). With this said, \textit{luxS} is thus involved in the biosynthesis of AI-2: the hormone-like signal that mediates cell-cell communication in quorum sensing. It should also be noted that \textit{luxS} also participates in the recycling of S-adenosylhomocysteine through the S-adenosyl-L-methionine cycle I (Surette \textit{et al.}, 1999).

As shown in Figure 2, the initial substrate, S-adenosyl-L-homocysteine, reacts with water and is catalyzed adenosylhomocysteine nucleosidase (\textit{mtn}) to produce S-ribosyl-L-homocysteine and
adenine. Then, luxS catalyzes the formation of AI-2 by cleavage of S-ribosyl-L-homocysteine, leaving L-homocysteine as a product that can ultimately enter the SAM cycle, while AI-2 will activate transcription of the lsr operon encoding an Lsr ABC transporter-the LsrB protein- that serves as a AI-2 receptor in E. coli K12 (Rezzonico et al., 2008).

(Methionine Degradation) SAM

![Image of the SAM pathway](https://biocyc.org/ECOLI/NEW-IMAGE?type=PATHWAY&object=PWY-6153)

Figure 2. Autoinducer-2 biosynthesis pathway I showing the reactants, products, and genes/enzymes involved in the catalysis of the mtn and luxS catalyzed reactions being studied. Image was taken from https://biocyc.org/ECOLI/NEW-IMAGE?type=PATHWAY&object=PWY-6153

**Bioinformatics**

Having knowledge about bioinformatic programs is particularly important in biological careers and biological research because they provide easy access for genomic information about many different types of organisms (Persidis, 1999). Not only are these programs available online for free, they save a lot of time for those who understand how to navigate them and interpret the data outputs presented. Since advances in technology are only expanding, data will be increasingly stored in these databases. Therefore, bioinformatics programs will be continue to be critical to having success in biology-related fields and played a huge role in this project.

**Purpose/Hypothesis**

In this project, we utilize various bioinformatic programs to discover if the Mrub_2642 gene is an ortholog of the E. coli b_2942 (metK) gene; if Mrub_1054 gene is an ortholog of the E. coli b_0159 (mtn) gene; and if Mrub_1059 gene is an ortholog of the E. coli b_2687 (luxS) gene.
These programs help determine whether similarities and differences exist between the different genes. Doing so, requires that a thorough knowledge of E-values, a value that is generated by the bioinformatics program to help formulate a conclusion based on the significance of the results. A high E-value indicate that the sequences entered in the program were only aligned due to chance alone; whereas, a low E-value is significant because it indicates that the sequences were not aligned due to chance (Madden, 2002). This project collects additional bioinformatics evidence to support our hypothesis for the *M. ruber* genes. We performed an initial BLAST search of *E. coli metK*, *E. coli mtn*, and *E. coli luxS* against the *Mrub_2642*, *Mrub_1054*, and *Mrub_1059* genes, resulting in three extremely low E-values: 3e-168, 3e-22, and 6e-48 respectively. Based on these data, we hypothesize that the three genes are orthologous to the corresponding *E. coli* genes.

### Methods

In order to address the gaps in our understanding of the *M. ruber* bacterial genome, we use the model organisms, such as *E. coli*, as our positive control. Since *E. coli* is relatively easy to grow in the laboratory, it has been very well-studied and we can use it to help analyze similar genes in our microorganism. To make a coherent argument about these three genes being orthologous and gather data, the GENI-ACT gene annotation website instructions were followed ([http://www.geni-act.org/education/main/](http://www.geni-act.org/education/main/)) with some deviations. An initial BLAST search was used for each of the three genes of *E. coli* versus *M. ruber* (*Mrub_2942*, *Mrub_1054*, and *Mrub_1059*) to determine if there were similarities between the sequences (Madden, 2002). Once we had established that these sequences were similar to each other, we filled out the modules on the GENI-ACT site, utilizing the various bioinformatic programs associated with the site. These include: CDD (Marchler-Bauer et al., 2016), T-Coffee (Notredame et al., 2000), KEGG (Kanehisa et al., 2016), Web Logo (Crooks et al., 2004), PDB (Berman et al., 2000), LipoP (Juncker et al., 2003), IMG (Markowitz et al., 2012), TMHMM (Krogh et al., 2016), TIGR fam (Haft et al., 2001), PSORTB (Yu et al., 2010), Pfam (Finn et al., 2016), EcoCyc (Keseler et al., 2013), Phobius (Kall et al., 2004), and SignalP (Petersen et al., 2011). A deviation from the original instructions was that instead of using the top 10 BLAST hits for the T-coffee analysis, we used between 12 and 20 hits. These hits were obtained by first excluding all proteobacteria from the search results and then excluding *E. coli* for the second search. Also, instead of using the MetaCyc website to find the biosynthesis and degradation pathways for *M. ruber* we used the EcoCyc website. Another deviation is that we used the map colored by KEGG for the Gene Context portion of the Horizontal Gene Transfer module. Finally, the Gene Duplication and Degradation module had no paralogs or pseudogenes so it became irrelevant. Therefore, the reason that we use *E. coli* as our control is because it not only more well-studied and is easy to grow in lab, it also contains genes that may be orthologous to the *Mrub_2942*, *Mrub_1054*, and *Mrub_1059* genes we are interested in (Cooper, 2000).
## Results for *E. coli metK* and *Mrub_2642*

### Table 1. *E. coli metK* and *Mrub_2642* are orthologs

<table>
<thead>
<tr>
<th>Bioinformatics tools used</th>
<th><em>E. coli b2942</em> gene (<em>metK</em>)</th>
<th><em>M. ruber Mrub_2642</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>BLAST <em>E. coli</em> against <em>M. ruber</em></td>
<td>Score: 464 E-value: 3e-168</td>
<td></td>
</tr>
<tr>
<td>CDD Data (COG category)</td>
<td>COG number: COG0192 S-adenosylmethionine synthetase E-value: 0.0 E-value: 0.0</td>
<td></td>
</tr>
<tr>
<td>Cellular Localization</td>
<td>Cytoplasm of the cell</td>
<td></td>
</tr>
<tr>
<td>TIGRfam-protein family</td>
<td>TIGR01034 MetK: methionine adenosyltransferase E-value: 0.0 E-value: 2.5e-249</td>
<td></td>
</tr>
<tr>
<td>Pfam-protein family</td>
<td>PF00438 S-adenosylmethionine synthetase E-value: 5.1e-43 E-value: 3.3e-42</td>
<td></td>
</tr>
<tr>
<td>Protein Database (PDB)</td>
<td>1FUG S-adenosylmethionine synthetase E-value: 0.0 E-value: 1.36e-130</td>
<td></td>
</tr>
<tr>
<td>Enzyme Commission Number</td>
<td></td>
<td>EC 2.5.1.6 Methionine adenosyltransferase</td>
</tr>
<tr>
<td>KEGG Pathway Map</td>
<td></td>
<td>Cysteine and Methionine Metabolism Pathway</td>
</tr>
</tbody>
</table>
Table 1 summarizes the results from a variety of bioinformatics tools that were used to compare the *E. coli* **metK** gene to **Mrub_2642**. The first row of data shows the initial BLAST comparison as stated in the introduction. Although a high bit score was seen when annotating these genes, the bit score is not particularly relevant in this situation because the proteins vary in length. More important is the E-value of 3e-168, which is very close zero so we can be certain that these two sequences didn’t align due to random chance alone and that they share many of the same amino acids. These genes share a common ancestor and are likely orthologs. Also, the CDD came up with the same COG number (COG0192) and name (S-adenosylmethionine synthetase) from the database, both with very small E-values (0.0) indicating significant similarity. This shows that the genes both encode the same enzyme in the methionine degradation pathway. All of the bioinformatic tools used in the analysis of cellular location (TMH, Signal P, LipoP, and PSORT-B) suggested that both proteins are located in the cytoplasm of the cell and that neither contain a cleavage site. Therefore, the cellular location of the enzyme encoded by **metK** and **Mrub_2642** are the same, which further confirms that these genes may be orthologs. Moreover, the TIGRfam number (TIGR01034) pulled from the database for both protein sequences was the same and was called metK: methionine adenosyltransferase. Pfam then found that both proteins share the same domain and number with very low E-value numbers, (PF00438) and S-adenosylmethionine synthetase. The protein database (PDB) also provided the same number and name for both sequences and the enzyme commission number of EC 2.5.1.6 was the exact same for both genes. Finally, both genes were predicted to be involved in the same step of the methionine degradation pathway as determined by KEGG.
Figure 3. Mrub_2642 and E. coli metK have similar protein sequence. Query sequence is metK; Subject sequence is Mrub_2642. Analysis was performed by using NCBI BLAST bioinformatics tool at https://blast.ncbi.nlm.nih.gov/Blast.cgi

Figure 3 shows the results of a BLAST alignment of E. coli metK versus Mrub_2642. As we can see from the data, 59% of the amino acids were exactly the same between both sequences, and 286 of the amino acids were similar. The calculated E-value for the BLAST was 3e-168, which shows that the two sequences would not be aligned due to random chance alone and are very similar because the value is so close to zero. Therefore, E. coli metK and Mrub_2642 appear to share some primary structural similarities. This is our first piece of evidence to support the hypothesis that the two genes might be orthologs.
Figure 4. *E. coli* metK and *Mrub_2642* do not contain TMH regions; a cytoplasmic location is predicted for both proteins. Panel A shows the TMHMM for *E. coli metK/b_2942*; Panel B shows the TMHMM data for *Mrub_2642*. TMHMM Server v 2.0 [http://www.cbs.dtu.dk/services/TMHMM](http://www.cbs.dtu.dk/services/TMHMM) was used to create these two hydropathy charts.

The images in figure 4 above displays TMH hydropathy plots for *E. coli metK* and *Mrub_2642*. Red peaks that appear on the plot indicate the presence of transmembrane helices. In this case there were no red peaks so there were no significant results. Furthermore, the THM plots of both
organisms are consistent with each other, each predicting that the protein encoded by these genes is present in the cytoplasm of the cell as opposed to the membrane.

Panel A
Panel B

Figure 5. *E. coli* metK and *Mrub_2642* do not contain cleavage sites; the D-value (D=0.098) for *E. coli* metK and the D-value (D=0.112) *Mrub_2642* were below the cutoff value. Panel A shows the plot for *E. coli* metK; Panel B shows the plot for *Mrub_2642*. Signal P server v 4.1 http://www.cbs.dtu.dk/services/SignalP created these plots.

The above plots in Figure 5 are SignalP graphs were generated by both *E. coli* metK and *Mrub_2642*. This bioinformatics tool is used to predict protein cleavage sites by assigning each protein a D-value, which is calculated using the Y-score and S-score, and a cutoff value, which is indicated by the purple line. For *E. coli* metK (Panel A), the D-value (0.098) is lower than the cutoff value (0.450), which means the protein does not have any cleavage sites. The same can be determined for *Mrub_2642* (Panel B), as the D-value of 0.112 is also below the cutoff value. Therefore, these data are consist for both genes, suggesting that neither have cleavage sites and neither is attached to or crosses through the cell membrane.
Also, it should be noted that no alternative open reading frames were proposed as determined by JGI/IMG for these two genes and that the original start codon was in a good distance away from the proposed Shine Dalgarno sequence in the correct reading frame (Markowitz et al., 2012).

Figure 6. *E. coli* metK and *Mrub_2642* are present in the same biochemical pathway. Panel A shows the KEGG pathway for *E. coli* metK. Panel B shows the KEGG pathway after selecting for *Meiothermus ruber*. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database at [http://www.genome.jp/kegg/pathway.html](http://www.genome.jp/kegg/pathway.html) was used to located the cysteine and methionine biosynthesis/degradation pathway maps.

Figure 6 shows the methionine degradation biochemical pathway that both of these genes are involved in. Enzymes that are colored in green are thought to be present in the organism. From the pathways, we can see that both genes play a role in the step that converts L-methionine to S-adenosyl-L-methionine (SAM), both encoding the methionine adenosyltransferase (metK) enzyme. This provides even more evidence that these two genes are orthologous.
Panel A

Figure 7. Mrub_2642 and E. coli metK have the same highly conserved amino acids and code for the same domain, S-adenosylmethionine synthetase. Panel A shows the pairwise alignment for E. coli metK. Panel B shows the pairwise alignment for Mrub_2642. The pairwise alignment was created using the Pfam website http://pfam.sanger.ac.uk/search.

As we can see from the pairwise alignments in figure 7, both organisms contain the highly conserved valine, glycine, proline, aspartic acid, and histidine residues near the beginning of the protein sequence. Rather than the two sequences being compared to each other as in the BLAST search, this pairwise alignment takes a query sequence and compares it to a consensus sequence created from hundreds of proteins. Since both E. coli metK and Mrub_2642 pulled out very similar consensus sequences gives us even more information to support an argument that the genes are orthologous to each other.

Panel A
Panel B

Figure 8. *E. coli* metK and *Mrub_2642* genes are not part of an operon. Note that the gene being examined is highlighted in red. Chromosome viewer maps were colored by KEGG. Panel A: *E. coli* metK Chromosome Viewer; Panel B *Mrub_2642* Chromosome Viewer. Images were taken from [http://img.jgi.doe.gov/](http://img.jgi.doe.gov/).

Although the above images in Figure 8 show that both genes are not part of an operon, they provide us with further data that these two genes are orthologs of each other based on color. One is a purple color and the other is a lighter purple color, indicating that these genes are both part of coenzyme transport and metabolism (Markowitz et al., 2012). Therefore, the Chromosome Viewer map adds evidence that these genes are orthologous.
Figure 9. *E. coli* and *M. ruber* genes show no evidence of HGT. Panel A shows the phylogenetic tree for *E. coli*; Panel B shows the phylogenetic tree for *M. ruber*. Images were created by http://www.phylogeny.fr.

The images above in Figure 9 support the traditional phylogenetic relationships for *E. coli* and *M. ruber*. All the organisms in the phylogenetic tree in panel A are a part of the proteobacteria phylum and most of the organisms in panel B are a part of the either the firmicutes or the Deinococcus-Thermus phylums. Therefore, no evidence of HGT is indicated because all species near *M. ruber* are within the Deinococcus-Thermus phylum.

**Results for *E. coli mtn* and *Mrub_1054***

**Table 2. *E. coli mtn* and *Mrub_1054* are orthologs**

<table>
<thead>
<tr>
<th>Bioinformatics tools used</th>
<th><em>E. coli b0159 gene (mtn)</em></th>
<th><em>M. ruber Mrub_1054</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>BLAST <em>E. coli</em> against <em>M. ruber</em></td>
<td>Score: 78.2</td>
<td>E-value: 3e-22</td>
</tr>
<tr>
<td>CDD Data (COG category)</td>
<td>COG number: COG0775 Nucleoside phosphorylase</td>
<td>E-value: 1.39e-84</td>
</tr>
<tr>
<td>Cellular Localization</td>
<td>Cytoplasm of the cell</td>
<td></td>
</tr>
<tr>
<td>TIGRfam-protein family</td>
<td>TIGR01704 MTA/SAH-Nsdase: MTA/SAH nucleosidase</td>
<td>E-value: 3.1e-200</td>
</tr>
<tr>
<td>Pfam-protein family</td>
<td>PF01048 Phosphorylase superfamily</td>
<td>E-value: 6.2e-49</td>
</tr>
<tr>
<td>Protein Database (PDB)</td>
<td>304V: MTA/SAH nucleosidase in complex with</td>
<td></td>
</tr>
</tbody>
</table>
Table 2 summarizes the results from a variety of bioinformatics tools that were used to compare the *E. coli* *mtn* gene to *Mrub_1054*. The first row of data shows the initial BLAST alignment as stated in the introduction. Although a high bit score was seen when annotating these genes, the bit score is not particularly relevant in this situation because the proteins vary in length. More important is the E-value of 3e-22, which is very close zero, we can conclude that these two sequences don’t align due to random chance alone and are orthologs. The CDD database search came up with the same COG number (COG0775) and name (Nucleoside phosphorylase), both with very small E-values indicating significance. This shows that the genes both likely encode the same enzyme in the autoinducer-2 pathway. All of the bioinformatic tools used in the analysis of cellular location (TMH, Signal P, LipoP, and PSORT-B) suggested that both proteins are located in the cytoplasm of the cell and that neither contain a cleavage site. Therefore, the cellular location of the enzyme encoded by *mtn* and *Mrub_1054* are the same, which further confirms that these genes may be orthologs. Moreover, the TIGRfam number (TIGR01704) pulled from the database for both protein sequences was the same and was called MTA/SAH nucleosidase. Pfam then found that both proteins share the same domain and number with very low E-value numbers, (PF01048) and phosphorylase superfamily. The protein database (PDB) provided different numbers, but with similar names for both sequences and the enzyme commission number of EC 3.2.2.9 was the exact same for both genes. Finally, both genes were predicted to be involved in the same step of the autoinducer-2 pathway as predicted by EcoCyc.

<table>
<thead>
<tr>
<th>Enzyme Commission Number</th>
<th>EC 3.2.2.9 Adenosylhomocysteine nucleosidase</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG Pathway Map</td>
<td>Cysteine and Methionine Metabolism Pathway</td>
</tr>
</tbody>
</table>
Figure 10. Mrub_1054 and E. coli mtn have similar protein sequence. Query sequence is mtn; Subject sequence is Mrub_1054. Analysis was performed by using NCBI BLAST bioinformatics tool at https://blast.ncbi.nlm.nih.gov/Blast.cgi

Figure 10 shows the results of a BLAST search of E. coli mtn versus Mrub_1054. As we can see from the data, 29% of the amino acids were exactly the same between both sequences, and 103 of the amino acids were similar in character. The calculated E-value for the BLAST was 3e-22, which shows that the two sequences would not be aligned due to random chance alone. Therefore, we can conclude that E. coli mtn and Mrub_1054 appear to share some primary structural similarities. This is our first piece of evidence to support the hypothesis that the two genes might be orthologs.

Panel A
Panel B

Figure 11. *E. coli mtn* and *Mrub_1054* do not contain TMH regions; a cytoplasmic location is predicted for both proteins. Panel A shows the TMHMM for *E. coli mtn/b_1059*; Panel B shows the TMHMM data for *Mrub_1054*. TMHMM Server v 2.0 [http://www.cbs.dtu.dk/services/TMHMM](http://www.cbs.dtu.dk/services/TMHMM) was used to create these two hydropathy charts.

The images in Figure 11 above display TMH hydropathy plots for *E. coli mtn* and *Mrub_1054*. Red peaks that appear on the plot indicate the presence of transmembrane helices. In this case there were no red peaks so there were no significant results. Furthermore, the THM plots of both organisms are consistent with each other, each predicting that the protein encoded by these genes is present in the cytoplasm of the cell as opposed to the membrane.

Panel A
Figure 12. *E. coli mtn* and *Mrub_1054* do not contain cleavage sites; the D-value (D=0.110) for *E. coli mtn* and the D-value (D=0.120) for *Mrub_1054* were below the cutoff value. Panel A shows the plot for *E. coli mtn*; Panel B shows the plot for *Mrub_1054*. Signal P server v 4.1 [http://www.cbs.dtu.dk/services/SignalP](http://www.cbs.dtu.dk/services/SignalP) created these plots.

The above plots in Figure 12 are SignalP graphs were generated by both *E. coli mtn* and *Mrub_1054*. This bioinformatics tool is used to predict protein cleavage sites by assigning each protein a D-value, which is calculated using the Y-score and S-score, and a cutoff value, which is indicated by the purple line. For *E. coli mtn* (Panel A), the D-value (0.110) is lower than the cutoff value (0.450), which means the protein does not have any cleavage sites. The same can be determined for *Mrub_1054* (Panel B), as the D-value of 0.120 is also below the cutoff value. Therefore, these data are consist for both genes, suggesting that neither have cleavage sites and neither is attached to or crosses through the cell membrane.
Also, it should be noted that no alternative open reading frames were proposed as determined by JGI/IMG for these two genes and that the original start codon was in a good distance away from the proposed Shine Dalgarno sequence in the correct reading frame (Markowitz et al., 2012).

Panel A

Panel B

Figure 13. *E. coli* mtn and Mrub_1054 are present in the same biochemical pathway. Panel A shows the KEGG pathway for *E. coli* mtn. Panel B shows the KEGG pathway after selecting for *Meiothermus ruber*. The Kyoto Ency wholepedia of Genes and Genomes (KEGG) database at http://www.genome.jp/kegg/pathway.html was used to located the cysteine and methionine biosynthesis/degradation pathway maps.

Figure 13 shows the autoinducer-2 biochemical pathway predicted to include these enzymes. Enzymes that are colored in green are thought to be present in the organism. From the pathways, we can see that both genes are predicted to play a role in the step that converts S-adenosyl-L-homocysteine to S-D-ribosyl-L-homocysteine, both encoding the
adenosylhomocysteine nucleosidase (mtn) enzyme. This provides even more evidence that these organisms have orthologous genes.

Panel A

Panel B

Figure 14. Mrub_1054 and E. coli mtn have the same highly conserved amino acids and code for the same domain, phosphorylase superfamily. Panel A shows the pairwise alignment for E. coli mtn. Panel B shows the pairwise alignment for Mrub_1054. The pairwise alignment was created using the Pfam website http://pfam.sanger.ac.uk/search.

As we can see from the pairwise alignments in figure 14, both organisms contain the highly conserved glycine residues seen throughout the entire protein sequences, indicated by a capital G. Rather than the two sequences being compared to each other as in the BLAST search, this pairwise alignment takes a provided sequence and compares it to a consensus sequence created from hundreds of proteins. Since both E. coli mtn and Mrub_1054 pulled out very similar consensus sequences gives us even more information to support an argument that the genes are orthologous to each other.
Figure 15. *E. coli mtn* and *Mrub_1054* genes are not part of an operon. Note that the gene being examined is highlighted in red. Chromosome viewer maps were colored by KEGG. Panel A: *E. coli mtn* Chromosome Viewer; Panel B *Mrub_1054* Chromosome Viewer. Images were taken from http://img.jgi.doe.gov/.

Although the above images in Figure 15 show that both genes are not part of an operon, they provide us with further data that these two genes are orthologs of each other based on color. One is a purple color and the other is a brownish color, indicating that these genes are both part of nucleotide transport and metabolism (Markowitz et al., 2012). Therefore, the Chromosome Viewer map adds evidence that these genes are orthologous.

**Results for *E. coli luxS* and *Mrub_1059***

**Table 3. *E. coli luxS* and *Mrub_1059* are orthologs**

<table>
<thead>
<tr>
<th>Bioinformatics tools used</th>
<th><em>E. coli b2687 gene (luxS)</em></th>
<th><em>M. ruber Mrub_1059</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>BLAST <em>E. coli</em> against <em>M. ruber</em></td>
<td>Score: 140 E-value: 6e-48</td>
<td></td>
</tr>
<tr>
<td>CDD Data (COG category)</td>
<td>COG number: COG1854 S-ribosylhomocysteine lyase LuxS autoinducer biosynthesis E-value: 2.89e-74</td>
<td>E-value: 1.92e-106</td>
</tr>
</tbody>
</table>
Table 3 summarizes the results from a variety of bioinformatics tools that were used to compare the *E. coli luxS* gene to *Mrub_1059*. The first row of data shows the initial BLAST search as stated in the introduction. Although a high bit score was seen when annotating these genes, the bit score is not particularly relevant in this situation because the proteins vary in length. More important is the E-value of 6e-48, which is very close zero so we can be certain that these two sequences didn’t align due to random chance alone and that they share many of the same amino acids. With this evidence, we can argue that these genes are orthologs. Also, the CDD database search came up with the same COG number (COG1854) and name (*S*-ribosylhomocysteine lyase) from the database, both with very small E-values indicating significance. This shows that the genes both encode the same enzyme in the autoinducer-2 pathway, which aids in quorum sensing and signal transduction mechanisms. All of the bioinformatic tools used in the analysis of cellular location (TMH, Signal P, LipoP, and PSORT-B) suggested that both proteins are located in the cytoplasm of the cell and that both also do not contain a cleavage site. Therefore, the cellular location of the enzyme encoded by *luxS* and *Mrub_1059* are the same, which further confirms that these genes may be orthologs. There
was no TIGRfam number for both organisms, but Pfam found that both proteins share the same domain and number with very low E-value numbers, (PF02664) and S-ribohomocysteine lyase (LuxS). The protein database (PDB) provided different numbers, but with similar names for both sequences and the enzyme commission number of EC 4.4.1.21 was the exact same for both genes. Finally, both genes were predicted to be involved in the same step of the autoinducer-2 biosynthesis pathway as determined by KEGG.

**Mrub_1059**
Sequence ID: Query_234421  Length: 161  Number of Matches: 1

<table>
<thead>
<tr>
<th>Range</th>
<th>Score</th>
<th>Expect</th>
<th>Method</th>
<th>Identities</th>
<th>Positives</th>
<th>Gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 to 158</td>
<td>140 bits(353)</td>
<td>6e-48</td>
<td>Compositional matrix adjust.</td>
<td>66/156(42%)</td>
<td>99/156(63%)</td>
<td>6/156(3%)</td>
</tr>
</tbody>
</table>

| Query 5 | DSFTVDR1RMEAPA0VRAKTMTPHGD9AITYVDFLIRCVFNKEVMPKFG1HTEELPAGTFM | 64 |
| Sbjct 9 | ESFRLDTHKVRAPYVRLGKTEPQKGLIEFYDRLFAQ5PMQEA5LGA1HTEELLATYI | 68 |
| Query 65 | RNHLNNGVIEIIIISP1MGCRCTGFYMFHCDEQRTAVDAKAMDAVLDKQVQDQTPNL | 124 |
| Sbjct 69 | RHSLEG---VVDISP1MGCRCTGFYVLQVPQ1PQVLEAFRATQLDV---VNHTEVPGVS | 122 |
| Query 125 | VYOCGTYQHSLQEAOQD1ARSFVDRVINSNEELA | 160 |
| Sbjct 123 | ELECGNLYDHPQSAKAWAEHELSNLHNLQETTEIA | 158 |

Figure 16. **Mrub_1059** and *E. coli luxS* have similar protein sequence. Query sequence is *luxS*; Subject sequence is *Mrub_1059*. Analysis was performed by using NCBI BLAST bioinformatics tool at [https://blast.ncbi.nlm.nih.gov/Blast.cgi](https://blast.ncbi.nlm.nih.gov/Blast.cgi)

Figure 16 shows the results of a BLAST search of *E. coli luxS* versus *Mrub_1059*. As we can see from the data, 42% of the amino acids were exactly the same between both sequences, and 66 of the amino acids were similar. The calculated E-value for the BLAST was 6e-48, which shows that the two sequences would not be aligned due to random chance alone. Therefore, we can see that *E. coli luxS* and *Mrub_1059* appear to share some primary structural similarities. This is our first piece of evidence to support the hypothesis that the two genes might be orthologous.

Panel A

<table>
<thead>
<tr>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>transmembrane</td>
</tr>
</tbody>
</table>

TMHMM posterior probabilities for *E. coli* 2687
Figure 17. *E. coli luxS* and *Mrub* _1059_ do not contain TMH regions; a cytoplasmic location is predicted for both proteins. Panel A shows the TMHMM for *E. coli luxS/b* _2687_; Panel B shows the TMHMM data for *Mrub* _1059_. TMHMM Server v 2.0 http://www.cbs.dtu.dk/services/TMHMM_ was used to create these two hydropathy charts.

The images in Figure 17 above display TMH hydropathy plots for *E. coli mtn* and *Mrub* _1054_. Red peaks that appear on the plot indicate the presence of transmembrane helices. In this case there were no red peaks so there were no significant results. Furthermore, the THM plots of both organisms are consistent with each other, each predicting that the protein encoded by these genes is present in the cytoplasm of the cell as opposed to the membrane.
Panel B

Figure 18. *E. coli luxS* and *Mrub_1059* do not contain cleavage sites; the D-value (D=0.101) for *E. coli luxS* and the D-value (D=0.107) for *Mrub_1059* were below the cutoff value. Panel A shows the plot for *E. coli luxS*; Panel B shows the plot for *Mrub_1059*. Signal P server v 4.1 http://www.cbs.dtu.dk/services/SignalP created these plots.

The above plots in Figure 18 are SignalP graphs were generated by both *E. coli luxS* and *Mrub_1059*. This bioinformatics tool is used to predict protein cleavage sites by assigning each protein a D-value, which is calculated using the Y-score and S-score, and a cutoff value, which is indicated by the purple line. For *E. coli luxS* (Panel A), the D-value (0.101) is lower than the cutoff value (0.450), which means the protein does not have any cleavage sites. The same can be determined for *Mrub_1059* (Panel B), as the D-value of 0.107 is also below the cutoff value. Therefore, these data are consist for both genes, suggesting that neither have cleavage sites and neither is attached to or crosses through the cell membrane.
Also, it should be noted that no alternative open reading frames were proposed as determined by JGI/IMG for these two genes and that the original start codon was in a good distance away from the proposed Shine Dalgarno sequence in the correct reading frame (Markowitz et al., 2012).

Panel A

![Panel A](image1)

Panel B

![Panel B](image2)

Figure 19. *E. coli* luxS and *Mrub_1059* are present in the same biochemical pathway. Panel A shows the KEGG pathway for *E. coli* LuxS. Panel B shows the KEGG pathway after selecting for *Meiothermus ruber*. The Kyoto Encyclopaedia of Genes and Genomes (KEGG) database at [http://www.genome.jp/kegg/pathway.html](http://www.genome.jp/kegg/pathway.html) was used to located the cysteine and methionine biosynthesis/degradation pathway maps.
Figure 19 shows the autoinducer-2 biochemical pathway predicted to include these enzymes. Enzymes that are colored in green are thought to be present in the organism. From the pathways, we can see that both genes are predicted to play a role in the step that converts S-D-ribosyl-L-homocysteine to L-homocysteine, both encoding the S-ribosylhomocysteine lyase (LuxS) enzyme. This provides even more evidence that these genes are orthologous.

Panel A

![Panel A](image)

Panel B

![Panel B](image)

Figure 20. *Mrub_1059* and *E. coli luxS* have the same highly conserved amino acids and code for the same domain, phosphorylase superfamily. Panel A shows the pairwise alignment for *E. coli luxS*. Panel B shows the pairwise alignment for *Mrub_1059*. The pairwise alignment was created using the Pfam website [http://pfam.sanger.ac.uk/search](http://pfam.sanger.ac.uk/search).

As we can see from the pairwise alignments in Figure 20, both organisms contain the highly conserved glycine, proline, cysteine, and threonine highlighted in light blue near the middle of the protein sequence. Rather than the two sequences being compared to each other as in the BLAST search, this pairwise alignment takes a provided sequence and compares it to a consensus sequence created from hundreds of proteins. Since both *E. coli luxS* and *Mrub_1059* pulled out very similar consensus sequences gives us even more information to support an argument that the genes are orthologous to each other.

Panel A

![Panel A](image)
Panel B

Figure 21. *E. coli luxS* and *Mrub_1059* genes are not part of an operon. Chromosome viewer maps were colored by KEGG. Panel A: *E. coli luxS* Chromosome Viewer; Panel B *Mrub_1059* Chromosome Viewer. Images were taken from [http://img.jgi.doe.gov/](http://img.jgi.doe.gov/).

The above images in Figure 21 provide us with further data that these two genes are orthologs of each other. As the genes around them are represented by different colors, neither of these genes is part of an operon. Both genes are light blue colored, indicating that these genes are both part of signal transduction mechanisms and AI-2 biosynthesis (Markowitz et al., 2012). Therefore, the Chromosome Viewer map adds evidence that these genes are orthologous.

**Conclusion**

In conclusion, the results obtained from this study revealed that *E. coli metK* and *Mrub_2642* are orthologous genes; *E. coli mtn* and *Mrub_1054* are orthologous genes; and *E. coli luxS* and *Mrub_1059* are orthologous genes, which means that these organisms are related through a common ancestry. Evidence for this genetic linkage was first determined by a BLAST analysis comparing the protein sequences of each *E. coli* gene and *M. ruber* gene. Further confirming this result were cellular location bioinformatics tools such as TMH, SignalP, LipoP, Phobius and PSORT-B, all indicated a cytoplasmic cellular location for all genes being annotated in this study. Additionally, TIGRfam and Pfam consistently matched the protein sequences of both organism’s three genes to that of methionine adenosyltransferase, adenosylhomocysteine nucleosidase, and S-ribosylhomocysteine lyase, including the domains of which each protein is composed. Also, the phylogenetic tree generated for *M. ruber* confirmed that it was a part of the deinococcus-thermus phylum and that the data supported the traditional phylogenetic tree for this organism because all the other organisms were either a part of the same phylum or the firmicutes phylum, which is closely related. There were also additional bioinformatics programs utilized for this project that also presented the same results for both of the genes being studied. In fact, none of the bioinformatics programs used for this project showed any deviations between all the genes. We can conclude, based on the frequency of the bioinformatic tools matching up all three genes for both organisms, that each gene is orthologous to each other: *E. coli metK* and *Mrub_2642*, *E. coli mtn* and *Mrub_1054*, *E. coli luxS* and *Mrub_1059*. 
Site-directed mutagenesis (SDM) is a method to create specific, targeted changes in double stranded plasmid DNA. There are many reasons to make specific DNA alterations (insertions, deletions and substitutions), including: to study changes in protein activity that occur as a result of the DNA manipulation; to select or screen for mutations (at the DNA, RNA or protein level) that have a desired property; to introduce or remove restriction endonuclease sites or tags (http://nebasechanger.neb.com/).

Panel A

Panel B

<table>
<thead>
<tr>
<th>#HNN</th>
<th>YlEtsezVseCHFkzAcq1sdadldalKkdpkzvacet1vtt1vsvzcvzkkvzdieklvzrevklgYkzeka7gfdakzctezlvaleeqzp</th>
</tr>
</thead>
<tbody>
<tr>
<td>#NA/TCE</td>
<td>1 tsezV+eGHkz+zad+isda+ldall+dparkzvacet1vtt1vsvzcvzkkvzdieklvzrevklgYkzeka7gfdakzctezlvaleeqzp</td>
</tr>
<tr>
<td>#PP</td>
<td>S85*************************************************************************************************7</td>
</tr>
<tr>
<td>#SEQ</td>
<td>DAV7SESVIEGHPKLAISRSDAVIDAILAQCDFKARVACETLVTTGIVMEGITTVSVPRLVRQTVLEVGYTRAQGFDONCAVLTAIDFS</td>
</tr>
</tbody>
</table>
Panel C

Mrub_2642 1176 bp

<table>
<thead>
<tr>
<th>Substitution</th>
<th>Insertion</th>
<th>Deletion</th>
</tr>
</thead>
</table>

Find:

Start and end positions included in substitution.

Start (5') 40   End (3') 42

Desired Sequence

GCC

<table>
<thead>
<tr>
<th>Common Peptide Tags</th>
</tr>
</thead>
</table>

Result

<table>
<thead>
<tr>
<th>Name (F/R)</th>
<th>Oligo (Uppercase = target-specific primer)</th>
<th>Len</th>
<th>% GC</th>
<th>Tm</th>
<th>Ta *</th>
</tr>
</thead>
<tbody>
<tr>
<td>QSSDM_1/25/2017_F</td>
<td>0ACCGAGG0G0G0CCCG0CAAAC</td>
<td>23</td>
<td>74</td>
<td>65°C</td>
<td></td>
</tr>
<tr>
<td>QSSDM_1/25/2017_R</td>
<td>ACCGATTCCGAGTTACC</td>
<td>18</td>
<td>50</td>
<td>61°C</td>
<td>63°C</td>
</tr>
</tbody>
</table>

Required Primers

Figure 22. Creating a missense mutation Mrub_2642 by substituting GCC for CAC at positions 40 and 42 in the DNA sequence, which changes histidine to alanine. Panel A shows the HMM logo for Mrub_2642 gene, which includes all the amino acids in its protein sequence—the most conserved being the tallest letters and the least conserved being the shortest letters. Histidine is in position 14 on the HMM logo and is the tallest letter, indicating that it is a highly conserved amino acid. Panel B confirms that histidine at position 14 in the protein sequence is highly conserved in M. ruber, E. coli, and all other related organisms that the database found had this amino acid in common for this protein. Panel C shows the missense mutation along with the primers that would be needed to make this DNA change in lab. The website http://nebasechanger.neb.com/ was used to create this SDM mutation.

The image above in figure 22 shows an example of alanine mutagenesis applied to the Mrub_2642 gene. This mutation was a missense mutation substituting alanine for by changing CAC to GCC at positions 40 and 42. Since alanine lacks unusual backbone dihedral angle preferences and is a very dull amino acid, substituting this amino acid in for one like histidine that is a polar, charged amino acid involved in proton transfers and catalytic binding sites, will affect the protein function of this enzyme. Histidine has more flexibility than alanine and a pKa
close to physiological pH, which explains why it aids in proton transfers. As Mrub_2642 or methionine adenosyltransferase is involved in proton transfers, substituting a hydrophobic, nonpolar amino acid like alanine would inhibit function inside the active site and this loss of function could have potentially detrimental effects on the methionine degradation pathway that this enzyme catalyzes to make SAM (Betts and Russell, 2003).
Works Cited


Scott LR. Meiothermus ruber Genome Analysis Project. [Internet]. GENI-ACT; [2015 Dec 16]. Available from:http://geni-science.org/secure/projects/vi


