Winter 2-2016

Possible orthologs of trpA and trpB genes between *E. coli* (b1260 and b1261) and *M. ruber* (Mrub_1512 and Mrub_1511)

John J. Stenger
Augustana College, Rock Island Illinois

Dr. Lori Scott
Augustana College, Rock Island Illinois

John Stenger

Bio 375: Molecular Genetics

Dr. Lori Scott

2-6-2016

Possible orthologs of *trp*A and *trp*B genes between *E. Coli* (b1260 and b1261) and *M. Ruber* (Mrub_1512 and Mrub_1511)

Introduction:

This study was in an effort to determine if the tryptophan A and B genes in *Meiothermus ruber* (Mrub_1512 [*trp*A] and Mrub_1511[*trp*B]) are orthologous to the respective genes in *Escherichia coli* (b1260 [*trp*A] and b1261[*trp*B]). This study uses *E. coli* as the model organism for comparison and *M. ruber* for the genes in question. The bacteria *Meiothermus ruber* is a species that is non-motile and has a distinctive red pigmentation. Additionally, *M. ruber* is a thermophile, has a gram-negative stain, and prefers a growth medium of potato peptone at an incubation temperature around 55°C or 60°C. The bacterium was discovered in a hot spring in Russia in 1984 (Tindall et al. 2010).

The genes in question, *trp*A and *trp*B, are part of the biosynthesis pathway that creates the amino acid tryptophan. The biosynthetic pathway can be viewed in Figure1. below where that *trp*A and *trp*B are circled. The process starts with chromismate and goes through a series of reactions until it reaches the last two step involving *trp*A and *trp*B. Then it is eventually transformed into L-tryptophan as shown below in Figure 1.

For some background on tryptophan or, L-tryptophan, it is one of the twenty essential amino acids found in all living things. Each amino acid is usually annotated with a single letter and L-tryptophan is annotated with the letter W and was discovered in 1901(Palego et. al 2016).

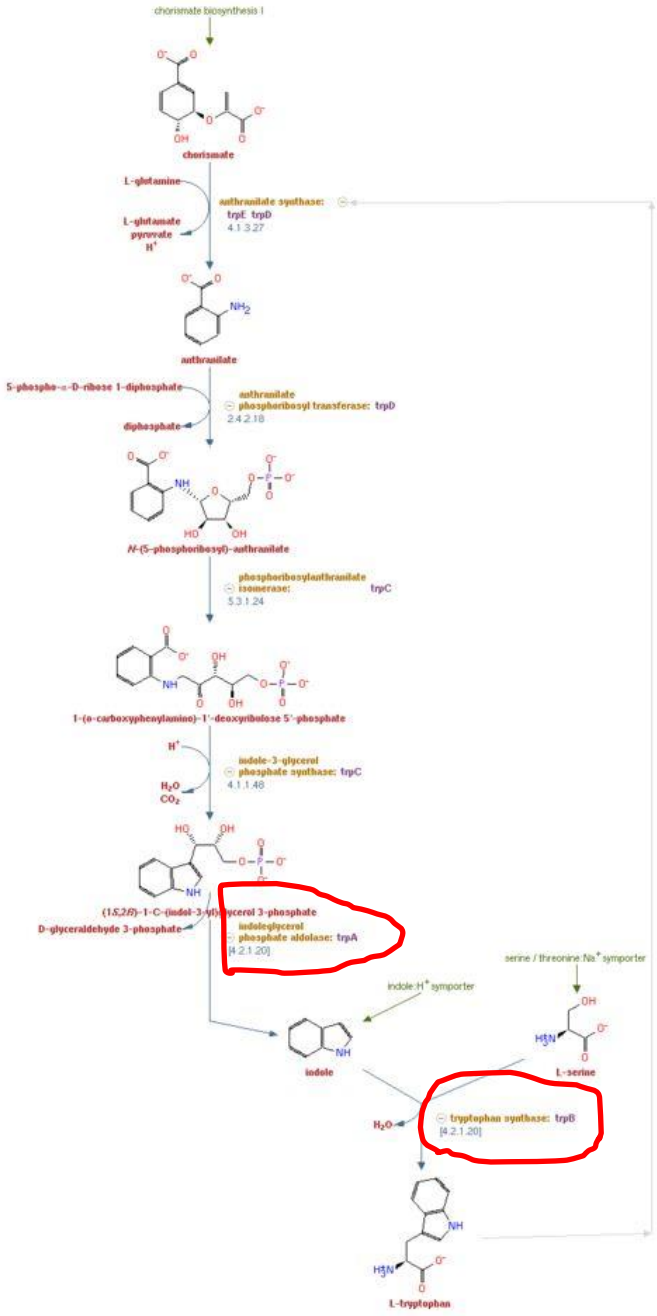


Figure 1. This is a diagram of the biosynthetic pathway of tryptophan. *Trp*A and *Trp*B are circled. Available from: http://ecocyc.org/

However it was not until 1974 that *trp*A and *trp*B were successfully purified and crystalized from strains of E. coli bacteria (Adachi et. al 1974).

As was said earlier tryptophan is an essential amino acid in the human body making it a vital building block. One of the things it is an essential part of is the neurotransmitter serotonin (5-hydroxy-tryptamine), which affects human mood and behavior among other things (Palego et. al 2016). So not only is it an essential amino acid found in every living thing but it is also significant piece of an extremely important building block of serotonin and countless other things in our bodies.

All this these facts aside, this study is exploring the possibility of orthologous genes between *E.coli* and *M. ruber.* This study is important because it furthers our knowledge of bacteria's genome and how they are related. The expansion of our knowledge is always valuable and this research may be useful to future scientists one day. After comparison through bioinformatics sties, it is predicted that there will be evidence supporting that [b1260 and Mrub_1512] are orthologs and [b1261 and Mrub_1511] are also orthologs.

Methods:

Utilizing the guidance of geni-act.org information was collected from a number of different bioinformatics websites. Geni-Act provides clear instruction for all of the bioinformatics sites used, which provided the results. All of the bioinformatics sites used can be found at the following link: http://www.geni-act.org/education/main/. The same instructions we followed for all 4 genes and were not deviated from. After all the genes were annotated in geni-act they were then analyzed and compared to each other.

Results:

Below is a collection of some of the important results from these different gene annotations. First will be the comparison of the results for *trp*A genes b1260 (*E.coli*) and Mrub_1512 (*M. Ruber*).

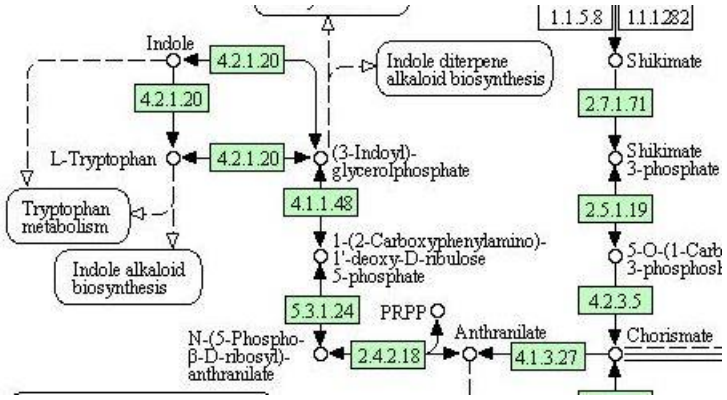
Table 1: *Trp*A (*E. coli* b1260 and Mrub_1512) are Orthologs

Description of Evidence Collected	*E. coli*	*M. ruber*
Cellular Location	Cytoplasmic	Cytoplasmic
KEGG pathway	Phenylalanine, tyrosine and tryptophan biosynthesis	Phenylalanine, tyrosine and tryptophan biosynthesis
Pfam – Protein Family	PF00290 Tryptophan synthase alpha chain (E=2.9e-102)	PF00290 Tryptophan synthase alpha chain (E=7.2e-83)
CDD (COG category)	COG0159 Tryptophan synthase alpha chain [Amino acid transport and metabolism] (E= 3.97e-124)	COG0159 Tryptophan synthase alpha chain [Amino acid transport and metabolism] (E= 5.01e-77)
TIGRfam – protein family	TIGR00262 *trp*A: tryptophan synthase, alpha subunit (E=2.3e-119)	TIGR00262 *trp*A: tryptophan synthase, alpha subunit (E=5.7e-69)
E.C. number	4.2.1.20 Tryptophan synthase	4.2.1.20 Tryptophan synthase
PDB	1V7Y tryptophan synthase alpha-subunit from *Escherichia coli* (E=4.84176E-140)	1UJP Tryptophan Synthase A-Subunit From *Thermus thermophiles* (E=2.41674E-70)

Table 1 shows that in many of the modules tested the results yielded very similar or exact matches. For example, in Pfam, COG, and TIGRfam they all resulted in matching numbers with slightly different E-values. The lowest E-value in Table 1 is the TIGRfam result for Mrub_1512 which is 5.7e-69. That being said, it is still a fairly significant number. The only category that

produced a different result was the PDB coming up with IV7Y for b1260 and 1UJP for Mrub_1512. However, the names attached to these numbers reveal that they are simply specific numbers for *trp*A and code for the respective bacteria. IUJP shows it as the alpha subunit of *Thermus thermophiles* which is a relative to *Meiothermus ruber*.

Panel A



Panel B

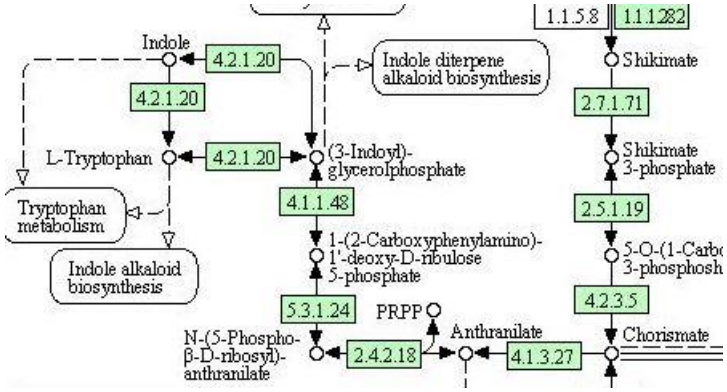
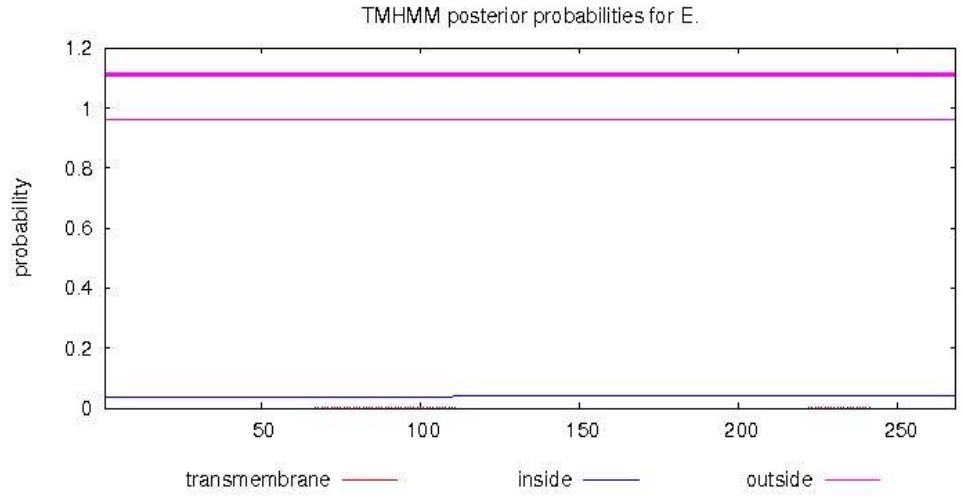


Figure 2. These KEGG pathways show that tryptophan synthesis follows the same biosynthetic pathway for both *E.coli* (panel A) and *M. ruber*(panel B). *trp*A and *trp*B are in the part of the pathway where the E.C. number 4.2.1.20. KEGG pathways created at: http://www.genome.jp/kegg-bin/show_pathway?org.

As mentioned above there were many similarities, so it was important to see how closely related their chemical pathways are. Above, in Figure 2. is a comparison of the tryptophan synthesis pathways in *E. coli* and *M. ruber*. The results show that both bacteria code for tryptophan in the same pathway (KEGG 2016). This shows a good amount of comparison between the two bacteria but there are many other aspects to look at.

Panel A

```
# E. Length: 268
# E. Number of predicted TMHs:  0
# E. Exp number of AAs in TMHs: 0.07173
# E. Exp number, first 60 AAs:  0.00018
# E. Total prob of N-in:        0.03829
E.      TMHMM2.0        outside      1   268
```



Panel B

```
# M. Length: 263
# M. Number of predicted TMHs:  0
# M. Exp number of AAs in TMHs: 14.2042
# M. Exp number, first 60 AAs:  0.00709
# M. Total prob of N-in:        0.68093
M.      TMHMM2.0        outside      1   263
```

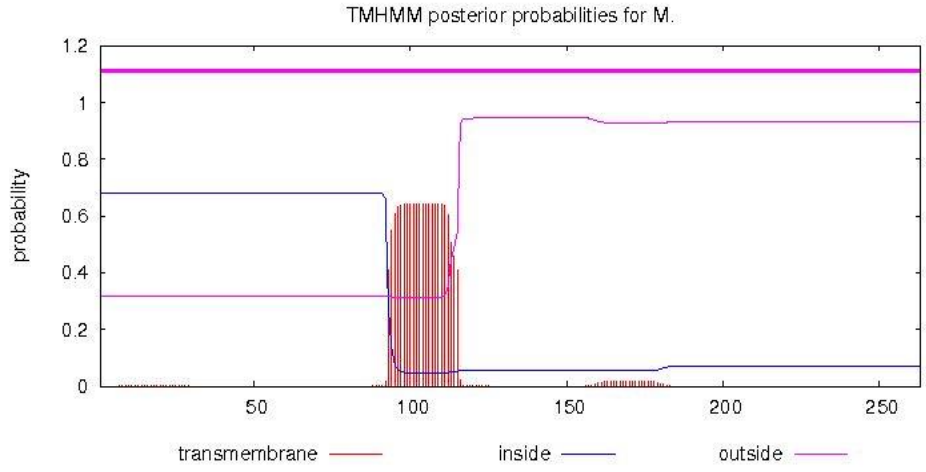
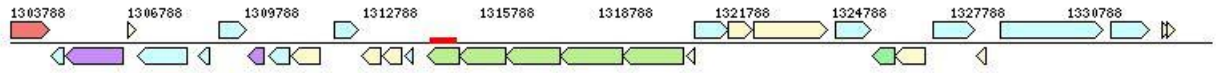


Figure 3. The TMHMM plots shows that there are no trans-membrane helices predicted for either b1260 (panel A) or Mrub_1512 (panel B). There is a small spike in the trans-membrane region of Mrub_1512 but it is apparently not enough to warrant predicted trans-membrane helices. TMHMM plots created at: http://www.cbs.dtu.dk/services/TMHMM.

Another important aspect is determining where tryptophan is located in the cell for both bacteria. One way of looking for location is looking for trans-membrane helices. As shown in Figure 3. It appears tryptophan shows no trans-membrane helices for both *E. coli* and *M. ruber* which shows another parallel of similarity between the two genes (TMHMM 2016).

Panel A



Panel B

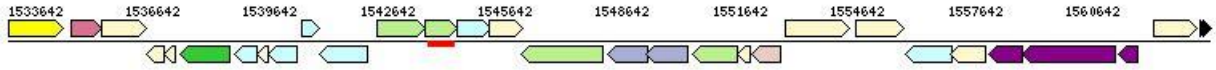


Figure 4. These show the *trp*A genes b1260 (panel A) and Mrub_1512 (panel B) and that they are contained in operons. They both have operons containing *trp*A and B, but *E. coli*'s operon additionally contains *trp*C, D, and E. Gene context images created at: https://img.jgi.doe.gov/cgi-bin/er/main.cgi.

Looking at the gene of interest in context lends more useful information when comparing the two genes. Here in Figure 4. a couple differences arise between the two bacteria. The *E. coli's* context shows *trp*A as part of an operon also containing *trp*B, C, D, and E. However, *M. Ruber's* gene context only shows that its operon only contains *trp*A and B. These show a difference between the two bacteria, but the difference is not huge. They have different numbers of genes contained in their operons but both are operons that code for tryptophan. Additionally, the gene codes in opposite directions. This does not make any difference in the end product but it simply codes in the opposite direction (IMG 2016). Now that there is some established similarity between these *trp*A genes (b1260 and Mrub_1512), next a closer look at *trp*B genes(b1261 and Mrub_1512) should be taken.

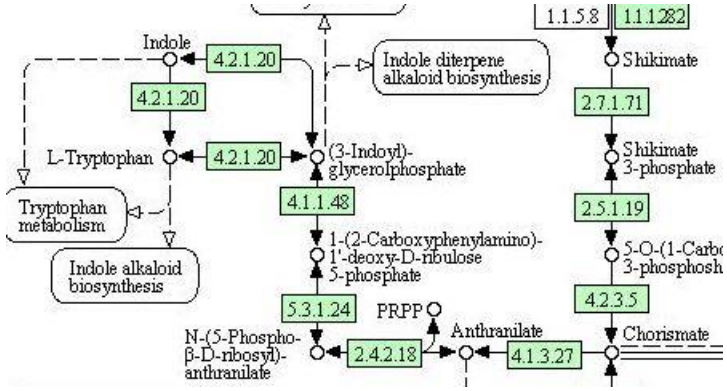
Table 2: *Trp*B (*E. coli* b1261 and Mrub_1511) are Orthologs

Description of Evidence Collected	*E. coli*	*M. ruber*
Cellular Location	Cytoplasmic	Cytoplasmic
KEGG pathway	Phenylalanine, tyrosine and tryptophan biosynthesis 00400	Phenylalanine, tyrosine and tryptophan biosynthesis 00400
Pfam – Protein Family	PF00291 Pyridoxal-phosphate dependent enzyme (E= 1.3e-50)	PF00291 Pyridoxal-phosphate dependent enzyme (E= 6.9e-53)
CDD (COG category)	COG0133 Tryptophan synthase beta chain [Amino acid transport and metabolism] (E=0.0)	COG0133 Tryptophan synthase beta chain [Amino acid transport and metabolism] (E=0.0)
TIGRfam – protein family	TIGR00263 *trp*B: tryptophan synthase, beta subunit (E=9e-282)	TIGR00263 *trp*B: tryptophan synthase, beta subunit (E=7.4e-274)
E.C. number	4.2.1.20 Tryptophan synthase	4.2.1.20 Tryptophan synthase
PDB	1A50 Wild-Type Tryptophan Synthase (E=0.00)	1X1Q tryptophan synthase beta chain from *Thermus thermophilus* (E= 7.97287E-178)

Table 2 shows some of the modules tested the results yielded very similar matches. For example, in COG and TIGRfam resulted in matching numbers with the correct title of *trp*B with significant E-values. The Pfam results yielded the same number but it was not with the title *trp*B, but instead was Pyridoxal-phosphate dependent enzyme. It turns out this molecule can be folded into different things, one of which being tryptophan synthase (Elliot and Kirsch 2016). This explains the potential difference in the results as the E values are in the e-50 range so it is kind of significant but not overwhelming. Again PDB showed that different results. b1261 resulted in a

Wild-type tryptophan synthase and Mrub_1511 resulted in a match with *trp*B but for Thermus thermophilus as it did for the *trp*A gene.

Panel A



Panel B

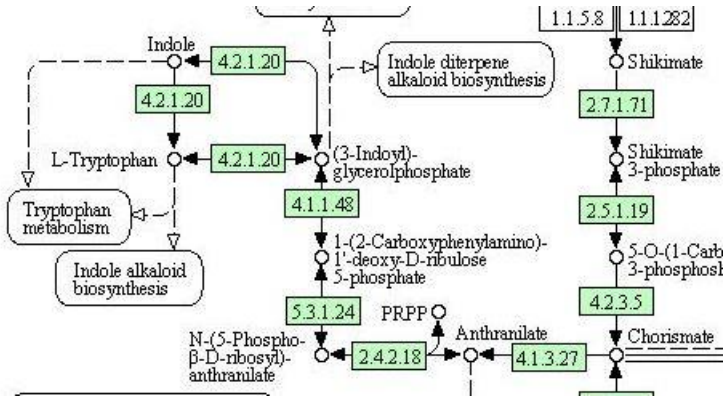
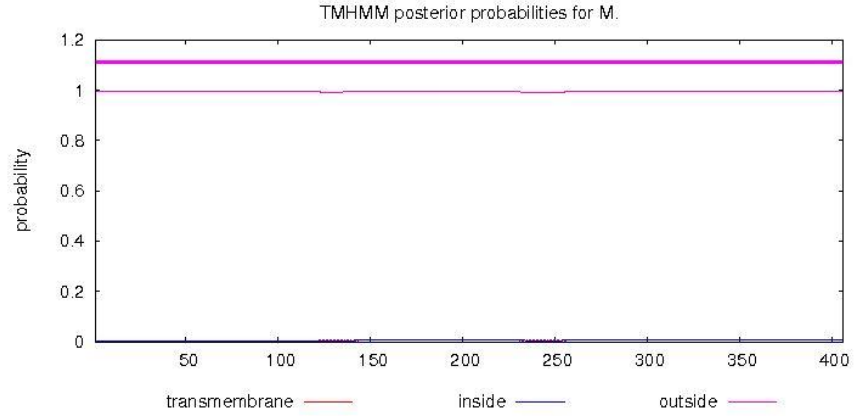


Figure 5. These KEGG pathways show that tryptophan synthesis follows the same biosynthetic pathway for both *E.coli* (panel A)and *M. ruber*(panel B). *trp*A and *trp*B are in the part of the pathway where the E.C. number 4.2.1.20. KEGG pathways created at: http://www.genome.jp/kegg-bin/show_pathway?org.

Again it is important to see how closely related the chemical pathways are. Above, Figure 5. shows the comparison of the tryptophan synthesis pathways in *E. coli* and *M. ruber* (KEGG 2016). These are the same as above for the *trp*A genes as they are for the *trp*B genes here.

Panel A

```
# M. Length: 406
# M. Number of predicted TMHs:  0
# M. Exp number of AAs in TMHs: 0.24162
# M. Exp number, first 60 AAs:  0
# M. Total prob of N-in:        0.00222
M.      TMHMM2.0        outside      1   406
```



Panel B

```
# E. Length: 397
# E. Number of predicted TMHs:  0
# E. Exp number of AAs in TMHs: 0.12201
# E. Exp number, first 60 AAs:  0.05425
# E. Total prob of N-in:        0.00560
E.      TMHMM2.0        outside      1   397
```

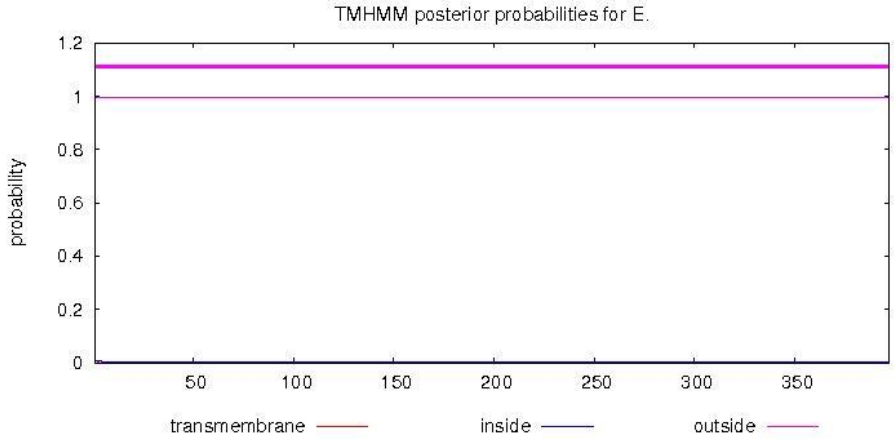
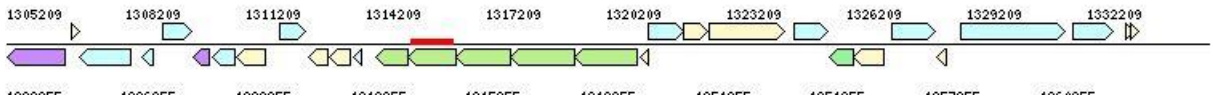


Figure 6. The TMHMM plots shows that there are no trans-membrane helices predicted for either b1261 (panel A) or Mrub_1511 (panel B). There are no abnormal spikes as noted on the previous *M. ruber* gene. TMHMM plots created at: http://www.cbs.dtu.dk/services/TMHMM.

Tryptophan location is searched for again here by looking for trans-membrane helices. As shown in Figure 6. It appears tryptophan shows no trans-membrane helices for both *E. coli* and *M. ruber*. The lack of trans-membrane helices shows another parallel of similarity between the two genes for cellular location (TMHMM 2016). This is the same results found for the *trp*A genes above.

Panel A



Panel B

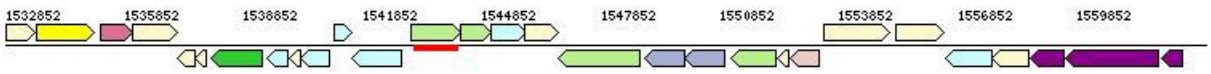


Figure 7. These show the *trp*A genes b1261 (panel A) and Mrub_1511 (panel B) and that they are contained in operons. They both have operons containing *trp*A and B, but *E. coli*'s operon additionally contains *trp*C, D, and E. Gene context images created at: https://img.jgi.doe.gov/cgi-bin/er/main.cgi

Looking at the gene of interest in context with other genes lends more useful information when comparing the two genes. Here in Figure 7. The same difference arises as above between the two

bacteria. The *E. coli's* context shows *trp*A as part of an operon also containing *trp*B, C, D, and E. Additionally, the gene codes in opposite directions. However, *M. Ruber's* gene context only shows that its operon only contains *trp*A and B (IMG 2016).

Conclusion:

It appears that the hypothesis that [b1260 and Mrub_1512] and [b1261 and Mrub_1511] are orthologous to each other seems to be correct. There were small differences between the different sets of genes but overwhelmingly they evidence points to the genes being orthologs. One of these differences is seen in the gene's context. *E. coli's trp*A and B are contained in a larger operon that also contains *trp*C, D, and E and codes in reverse. *M.ruber's* operon only contains *trp*A and B and codes forward (IMG 2016). Another difference was the different results from PDB: 1A50 for b1261 and 1X1Q for Mrub_1511. While different the name clarified they were both *trp*B but simply different codes for the slight difference between bacteria. The same was true for b1260 (1V7Y) and Mub_1512 (1UJP) as they yielded the same sort of results for *trp*A (RSCB PDB 2016). However, there were many shared similarities between the genes. Some of these include the same tryptophan biosynthetic pathway for all four of the genes of interest (KEGG 2016). Also all four genes had no predicted trans-membrane helices (TMHMM 2016). Other small thing such as matching TIGRfam number for *trp*A (TIGR00262) and *trp*B (TIGR00263) help solidify the commonality (TIGRfam 2016). There are also many other similarities that can be found in the results section above. The majority of the data points in the same direction and makes it fairly easy to conclude that [b1260 and Mrub_1512] and [b1261 and Mrub_1511] are orthologs to each other.

Literature Cited:

Adachi, O, Kohn, LD, Miles, EW. 1974. Crystalline alpha2 beta2 complexes of tryptophan synthetase of Escherichia coli. A comparison between the native complex and the reconstituted complex. (serial online). (cited 2/5/2016). 249(24):7756-63. Available from: Pubmed

Elliot, Andrew C, and Kirsch, Jack F. 2004. Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations. (serial online). (cited 2/5/2016). 73:383-415. Available from: Ebscohost

Palego, Lionella, Betti, Laura, Rossi, Alessandra, and Giannaccini Gino. 2016. Tryptophan Biochemistry: Structural, Nutritional, Metabolic, and Medical Aspects in Humans. Journal of Amino Acids. (serial online). (cited 2/4/2016). 1-13. Available from Ebscohost

Tindall, Brian J, Johannes Sikorski, Susan Lucas, Eugene Goltsman, Alex Copeland, Tijana Glavina Del Rio, Matt Nolan, et al. 2010. Complete genome sequence of *Meiothermus ruber* type strain. Standards in Genomic science. (serial online). (cited 2/4/16). 3:26-36. Available from: Ebscohost

Bioinformatics Cited:

Ecocyc. *E. coli* Database. 2016. (cited 2/4/16). Available from: http://ecocyc.org/

ExPASY. Bioinformatics Resource Portal. 2016.(cited 2/4/16). Available from: http://enzyme.expasy.org/enzyme-search-ec.html

Geni-act. 2016 (cited 12/4/16). Available from: http://www.geniact.org/gene/locus/CP001743/Mrub_1080/

IMG. 2016. (cited 2/4/16). Available from: https://img.jgi.doe.gov/cgi-bin/er/main.cgi

KEGG. KEGG Pathway Database. 2016. (cited 2/4/16). Available from: http://www.genome.jp/kegg/pathway.html

NCBI. 2016 (cited 2/4/16). Available from: http://www.ncbi.nlm.nih.gov/pubmed

NCI Blast. 2016 (cited 2/4/16). Available from: http://blast.ncbi.nlm.nih.gov/Blast.cgi

Pfam. 2016. (cited 2/4/16). Available from: http://pfam.xfam.org/search

RSCB PDB. Protein Data Bank. 2016. (cited 2/4/16). Available from: http://www.rcsb.org/pdb/home/home.do

TIGRfam. J. Craig Venter Institute. 2016. (cited 2/4/16). Available from: http://blast.jcvi.org/web-hmm/

TMHMM. CBS Prediction Servers. 2016. (cited 2/4/16). Available from: http://www.cbs.dtu.dk/services/TMHMM

WebLogo. 2016. (cited 2/4/16). Available from: http://weblogo.berkeley.edu/