2-2016

# Pyruvate metabolism in *m. ruber* when compared to *e. coli*

Amanda M. Johnson
*Augustana College, Rock Island Illinois*

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

**Introduction:**

Glucose is broken down in the body via a process called glycolysis. Glycolysis is a series of 10 enzyme-catalyzed reactions that produce a high energy molecule, pyruvate (AAMC 2008). Those pyruvic acid molecules then migrate to the cytoplasm. The cytoplasm contains protein complex called the pyruvate dehydrogenase complex (PDHC). The PDHC has three enzyme functions caused by three separate proteins of the complex referred to as E1, E2, and E3 (Berg, Tymoczko, Stryer 2002) shown in figure 1 below.

Henning et. al. cloned the E1, E2, and E3 genes and performed a series of reactions to determine their role in *E. coli* (1997). The first reaction is catalyzed by the thiamin pyrophosphate (TPP) dependent decarboxylating component of the complex, EC 1.2.4.1 (E1) (Hennig 1997). E1 removes a carbon and 2 oxygens from the pyruvate molecule, releasing one molecule of $CO_2$ (Berg, Tymoczko, Stryer 2002). In the second reaction, the hydroxyethyl group of pyruvate is transferred to E2, producing E2-dihydrolipoamide. The remaining acetyl group of pyruvate is then free to react with CoA to produce acetyl-coenzyme A (Henning 1997). The acetyl CoA molecule is oxidized by E3, which allows NADH to be reduced to NADH+. This NADH+ is then used in the electron transport chain (ETC) important for cellular respiration. For each molecule of glucose metabolized, two molecules of acetyl-CoA are produced and used in the Citric Acid Cycle (CAC) (Lodish et al 2008). The CAC is responsible for oxidizing fuel molecules to be used in the ETC (Berg, Tymoczko, Stryer 2002).

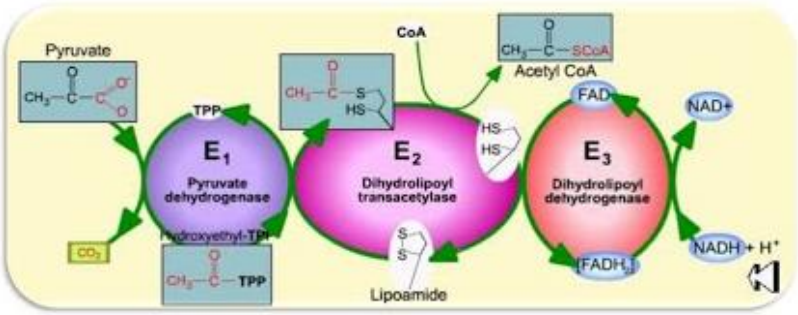



**Figure 1.** Pyruvate is converted to Acetyl CoA by a series of three reactions catalyzed by E1, E2, and E3 in the pyruvate dehydrogenase complex. Image from:https://www.youtube.com/watch?v=-d1bOSSzB_0.

A key reaction to central metabolism is the conversion of pyruvate to acetyl-CoA because it links glycolysis to the citric acid cycle (How does glycolysis...2014) as shown in Figure 2 below.

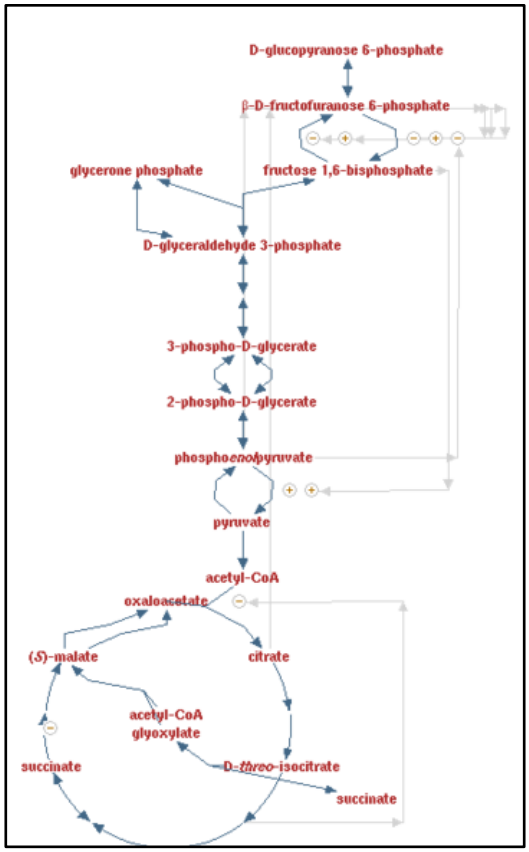



**Figure 2.** Pyruvate links glycolysis to the citric acid cycle. The end product of glycolysis is pyruvate, which is then converted to acetyl-CoA and used in the citric acid cycle. Image from: http://ecocyc.org/ECOLI/NEW-IMAGE?type=PATHWAY&object=GLYCOLYSIS-TCA-GLYOX-BYPASS.

Pyruvate is converted to acetyl-CoA via the pyruvate dehydrogenase complex (PDHC) which consists of three enzymes; E1, E2, and E3 (Eur. J. Biochem. 133, 155-162 (1983)). The specific genes that play the roles of E1, E2, and E3 in any given genome must be identified in order to determine if that organism's pyruvate metabolism pathway is similar to the reference pathway above. *E. coli* is a well-studied genome, and its pyruvate metabolism pathway is identical to that described above (CITATION). Its E1, E2, and E3 genes have been identified, making it a quality control group for studying the pyruvate metabolism pathway in other organisms that have not been so heavily studied. It is important to branch out and study other bacterial genomes because much of the knowledge about bacteria comes from a relatively small amount of species, and the rules observed from those few species do not always apply to other species. Studying *M. ruber*'s genome will contribute to the Tree of Life project currently being run by GEBA and expand humanity's understanding of bacteria (Phylogenetic...2015).

In *e. Coli*, E1, E2, and E3 are encoded by genes *aceE, aceF, and lpd,* respectively (Eur. J. Biochem. 133, 155-162 (1983)). These three genes are a part of an operon as shown in Figure 3.

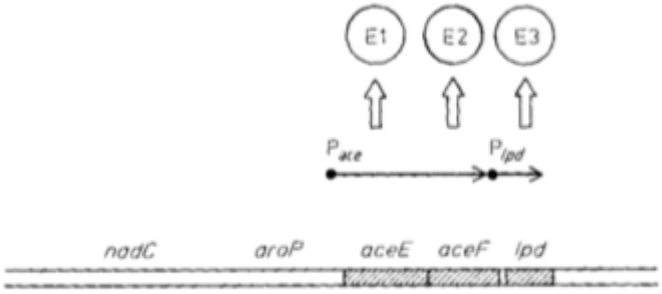



**Figure 3**. Organization of the PDHC operon in e. Coli . *AceE* (E1) is the most proximal gene, followed by *AceF* (E2), and finally *lpd* (E3) (Eur. J. Biochem. 133, 155-162 (1983).

The purpose of this experiment was to determine if mrub_0476, 1516, and 1517 are the E1 proteins in *M. ruber* by comparing it to the known E1 protein in *E. coli*, b0114/aceE, and to determine if mrub_0477 and mrub_2322 are the E2 proteins in *M. ruber* by comparing it to the known E2 protein in *E. coli*, b0115 (KEGG 2015). This knowledge will contribute to the overall understanding of how amino acid biosynthesis pathways in *M. ruber* compare to model organisms, such as *E. coli,* ultimately aiding in the endeavor to understand the evolutionary diversification of microbes (Phylogenetic...2015). In order to study the hypothesized E1 and E2 genes in *M. ruber* by comparing it to the known E1 and E2 genes in *E. coli*, a variety of bioinformatics tools were utilized. These tools included BLAST and databases such as Metacyc and Ecocyc. Ecocyc in particular served as a platform to data as it thoroughly described the genome, metabolic pathways, and regulatory network of *E. coli*, the control. Bioinformatics tools are not limited to studying bacteria, in fact, these same tools are used when diagnosing, treating, and researching disease in humans based on the human genome. The genome is the blueprint to the body and can reveal a great deal about abnormalities causing disease. Therefore, understanding how to use bioinformatics tools is highly useful for anyone studying a genome as it is a source for well-studied organisms (Shen 2014).

**Methods:**

At the beginning of the experiment, *E. coli* was BLASTed against *M. ruber* to determine if there was any gene that appeared similar to the E1 and E2 genes in *M. ruber.* Mrub_0476, 1516, and 1517 were identified as the possible E1 genes in *M. ruber* and Mrub_0477 and Mrub_2322 were identified as the possible E2 genes in *M. ruber.* The bioinformatics programs within the GENI-ACT lab notebook were used to compare the E1 and E2 genes from *E. coli* and *M. ruber*. The

descriptions for those tools can be found at the following link: http://www.geni-act.org/education/main/. From there, the rest of the protocol was followed, with the exception of a few minor deviations. Under the sequence based similarity module where T-coffee was used for *E. coli*, 20 sequences were used instead of 10. The first alignment sequences excluded escherichia and the second set of sequences excluded the phylum of proteobacteria. The paralog section of *E. coli* in the duplication and degradation was not done. Ecocyc was used for *E. coli* instead of metacyc.

**Results**

**Table 1.** b0114 and Mrub_0476, 1516, and 1517 are orthologs

| Description of Evidence | *E. coli b0114* | *Mrub_0476* | *Mrub_1516* | *M. ruber_1517* |
|---|---|---|---|---|
| Cellular Localization | Cytoplasm | Cytoplasm | Cytoplasm | Cytoplasm |
| BLAST *E. coli* against *M. ruber* | | Bit score: 892; E-value: 0.0 | Bit Score: 319; E-value: 1e-105 | Bit Score: 195; E-value: 6e-57 |
| KEGG Pathway | Pyruvate Metabolism | Pyruvate Metabolism | Pyruvate Metabolism | Pyruvate Metabolism |
| Pfam- protein family | P00456 Transketolase; E-value 1.2e-09 | PF00456 Transketolase; E-value 2e-08 | PF02779 Transketolase_pyrimidine; E-value 4.1e-43 | PF00676 Dehydrogenase E1 Component; E-value 1.5e-82 |
| TIGRfam family | TIGR00759 Pyruvate Dehyrogenase E1 Component; E-value 0.00 | TIGR00759 Pyruvate Dehydrogenase Component E1; E-value 0.00 | None | TIGR03181 PDH_E1; E-value 5.4e-182 |
| CCD (COG Category) | COG2609 Pyruvate dehydrogenase complex, (E1) component; E-value 0.00 | COG2609 Pyruvate dehydrogenase complex, (E1) component; E-value 0.00 | COG0022 AcoB; E-value 1.01e-178 | COG1071 AcoA; E-value 4.70e-139 |

| E.C. Number | 1.2.4.1 | 1.2.4.1 | 1.2.4.1 | 1.2.4.1 |
|---|---|---|---|---|
| PBD | 1W85 Pyruvate Dehydrogenase; E-value 8.54665e-63 | 1L8A Pyruvate Dehydrogenase; E-value 0.00 | 3DUF Pyruvate Dehydrogenase snapshots; E-value 0.00 | 1W85 Pyruvate Dehydrogenase; E-value 8.54665E-63 |

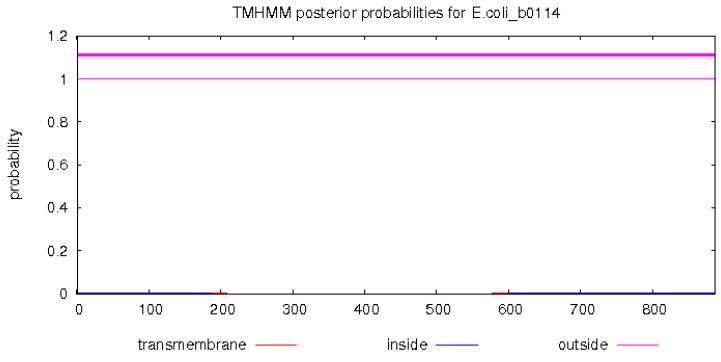
*Blast*

Table 1 shows the results of a protein BLAST comparison between E. coli b0014 and Mrub_0476, which produced an E-value of 0.00. This result is significant and indicates that sequence similarities are likely due to evolutionary alikeness as opposed to chance. However, BLAST comparison between b0014 and Mrub_1516 did not produce a significant e-value at 0.013, nor did b0014 against Mrub_1517, which showed no sequence similarities at all.

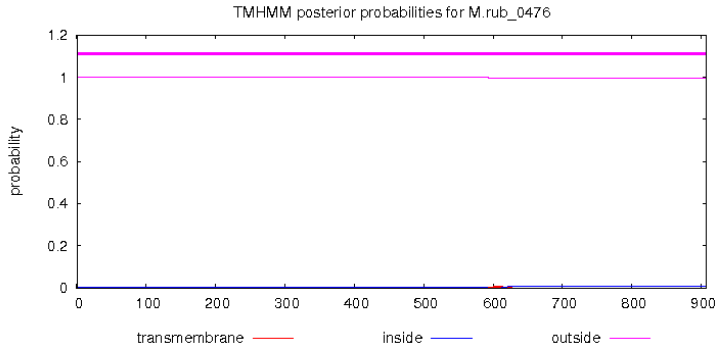
*Cellular Localization*

TMHMM analysis showed that *E. coli* b0014 (Figure 4 Panel A) has no transmembrane helices, indicating that it is not a membrane protein. Results were the same for Mrub_0476, Mrub_1516, and Mrub_1517, shown in Figure 4 Panel B, C, and D, respectively, below.

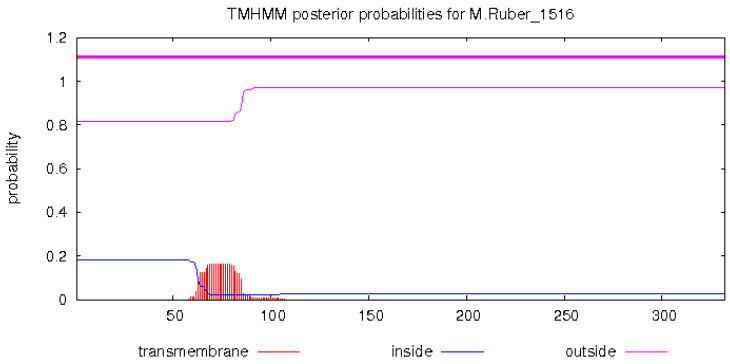
Panel A




Panel B




Panel C




Panel D

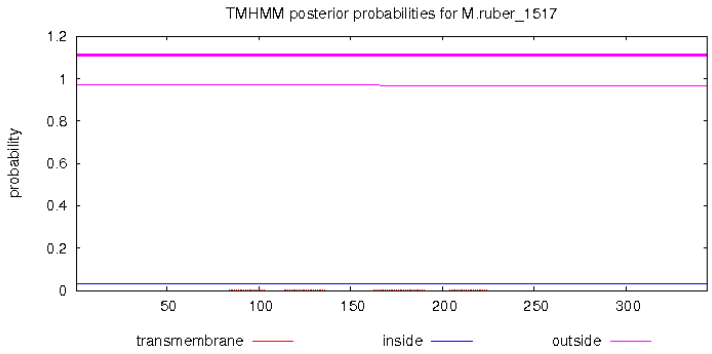
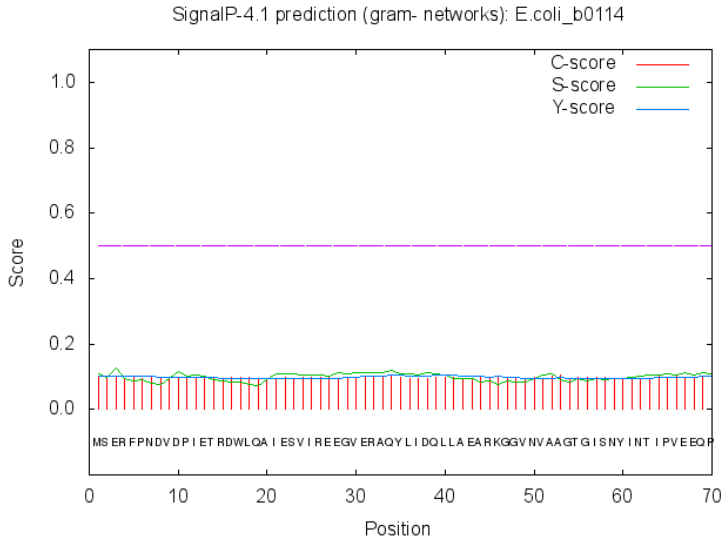



**Figure 4.** Mrub_0476, Mrub_1516, and Mrub_1517 and *E. coli* b0014 do not contain TMH regions. A cytoplasmic location is predicted. Panel A= *E. coli* b0014; Panel B= Mrub_0476; Panel C= Mrub_1516; Panel D=Mrub_1517. TMHMM server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) created this hydropathy plot.

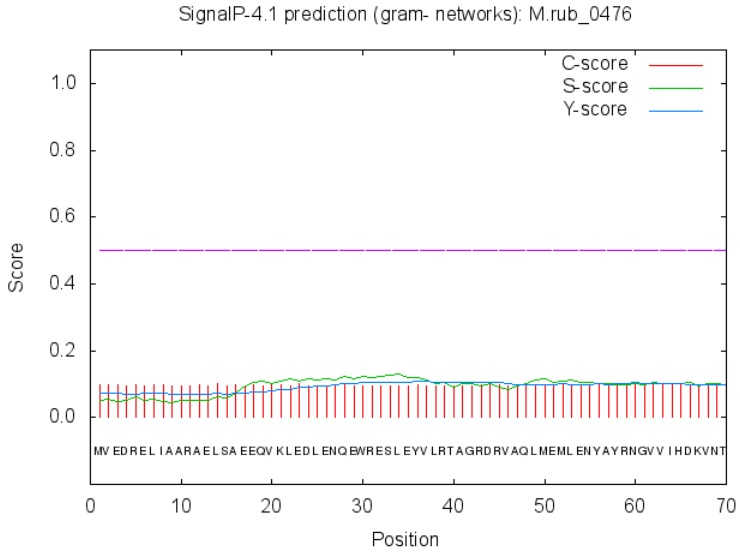
The Signal Peptide probability had a predicted value of 0.103 for *E. coli*. Since this value is far from 1, it indicates that *E. coli* has no N-terminal membrane helices. In addition, no signal peptides were predicted. Results are shown in Figure 3 Panel A. The predicted Signal Peptide

probability for Mrub_0476, Mrub_1516, and Mrub_1517, were also far less than 1.0 at 0.098, 0.133, and 0.097, respectively, indicating that they also contain no N-terminal membrane helices. No signal peptides were predicted for *M. ruber* either. Results for *M. ruber* are shown in Figure 3 Panel B, C, and D below.

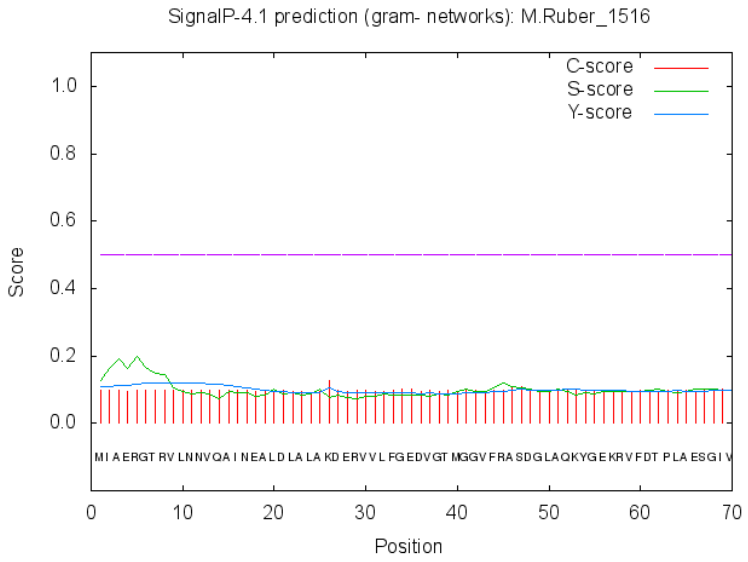
Panel A




Panel B




Panel C




Panel D

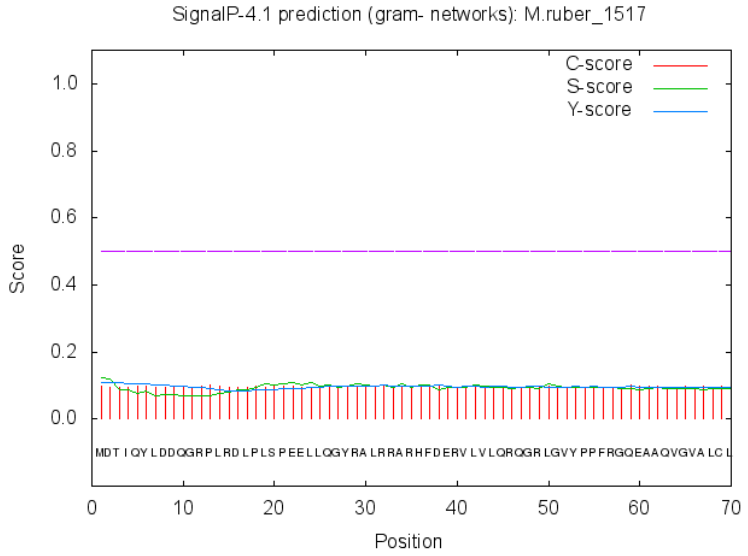



**Figure 5.** Mrub_0476, Mrub_1516, and Mrub_1517 and *E. coli* b0114 do not contain signal peptides. A cytoplasmic location is predicted. Panel A= *E. coli* b0114; Panel B=Mrub_0476; Panel C= Mrub_1516; Panel D= Mrub_1517. SignalP server v. 4.1 (http://www.cbs.dtu.dk/services/SignalP/) created this graph.

LipoP v1.0 (http://www.cbs.dtu.dk/services/LipoP/) did not predict any cleavage sites for *M. ruber* or *E. coli*, and its best cellular location prediction was the cytoplasm for all four genes. PSORT-B v3.0.2 (http://www.psort.org/psortb/) predicted all three *M. ruber* genes and *E. coli* to be located in the cytoplasm as well. It is known that *E. coli*'s b0114 gene is located in the cytoplasm. Similar cellular localization analyses suggest that Mrub_0476, Mrub_1516, and Mrub_1517 are also located in the cytoplasm.

*KEGG Pathway*

The Pyruvate Metabolism reference pathway indicates that *E. coli* b0114 performs the role of E1 in the pyruvate dehydrogenase complex in that it converts pyruvate to 2-hydroxy-ethyl-ThPP and 2-hydroxy-ethyl-ThPP to S-Acetyl-dihydrolipoamide-E. It is known that *E. coli* utilizes the b0114 gene, as shown in Figure 6 Panel A. The KEGG pathway (http://www.genome.jp/kegg/pathway.html) for *M. ruber* shown in Figure 4 Panel B below also indicates that it utilizes the proB gene (1.2.4.1).

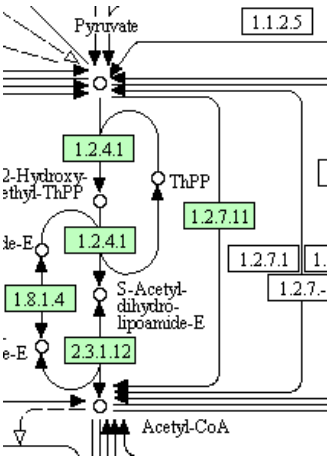



**Figure 6.** M. ruber and *E. coli* utilize the pyruvate dehydrogenase complex E1 protein in the synthesis of pyruvate when converting pyruvate to 2-hydroxy-ethyl-ThPP and 2-hydroxy-ethyl-ThPP to S-Acetyl-dihydrolipoamide-E. The pyruvate dehydrogenase complex E1 in each organism is notated as 1.2.4.1 in the pathway. The KEGG pathway is identical for both. The KEGG pathway database (http://www.genome.jp/kegg/pathway.html) was used to create the above image.

*Pfam- protein family:*

Pfam analysis identified one family within the *E. coli* b0114 sequence; Transketolase with an e-value of 1.2e09. The query gene, PDHC E1, is in fact a member of the transketolase protein family. The same Transketolase family was identified in Mrub_0476 with an e-value of 2e-08 as well. Pfam identified different families in Mrub_1516 and Mrub_1517. Mrub_1516 was in the family Transket_pyr, which is named Transketolase, pyrimidine binding domain. This had an e-value of 4.1e-43. Mrub_1517 was in the family E1_dh, a dehydrogenase E1 component with an e-value of 1.5e-82.

*TIGRfam- protein family:*

TIGRfam analysis (http://blast.jcvi.org/web-hmm/) predicts the name or function of the gene product. In this experiment, TIGRfam predicted aceE pyruvate dehydrogenase for the gene product of the query *E. coli* amino acid sequence. The e-value was significant at 0.00 with a bit score of 2573.5. TIGRfam also predicted aceE pyruvate dehydrogenase as the gene product of the query Mrub_0476 amino acid sequence. The e-value was significant at 0.00 with a bit score of 1931.1. The other two Mrub genes did not have the same predicted TIGERfam gene products. TIGRfam did not predict a gene product for Mrub_1516 at all. TIGRfam predicted PDH_E1: pyruvate dehydrogenase with an e-value of 5.4e-182 and bit score of 615.9.

*CCD (COG) Category:*

Conserved domains were located using CCD (http://www.ncbi.nlm.nih.gov/cdd/). Analysis of the b0114 sequence found two conserved domains, TktA1 and AceE. TktA1 had an e-value of 2.46e-15. AceE had an e-value of 0.00. The same domains were located when Mrub_0476 sequence was analyzed with significant e-values of 2.48e-14 and 0.00, respectively. Mrub_1516 had a conserved domain of AcoB, which is a dehydrogenase E1 component. Mrub_1517 had a conserved domain of AcoA, which is a TPP-dependent pyruvate or acetoin dehydrogenase.

*E.C. Number:*

The enzyme commission number for *E. coli* b0114 is 1.2.4.1 as identified on the KEGG pathway in Figure 4 Panel A. The enzyme commission number for Mrub_0476, Mrub_1516, and Mrub1517 is also 1.2.4.1.

*PBD:*

The protein bank database (PBD) identified crystal structure of pyruvate dehydrogenase E1 bound to the peripheral subunit binding domain of E2 as the protein when the *E. coli* b0114 sequence was analyzed. The E-value for this finding was 8.54665e-63, indicating that there is a low chance of sequence similarity are due to chance, rather, similarities are likely due to evolutionary relatedness. The PBD for Mrub_0476 identified *E. coli* pyruvate dehydrogenase as the protein, with an e-value of 0.00. PBD identified similar proteins for Mrub_1516 or Mrub_1517. PBD identified snapshots of catalysis in the E1 subunit of the pyruvate dehydrogenase multi-enzyme complex as the protein in Mrub_1516. This had a significant e-value of 0.00. PBD identified the crystal structure of pyruvate dehydrogenase E1 bound to the peripheral subunit binding domain of E2 as the protein in Mrub_1517, just as was identified in b0114. This also had a significant e-value of 8.54665E-63.

*Gene Context*

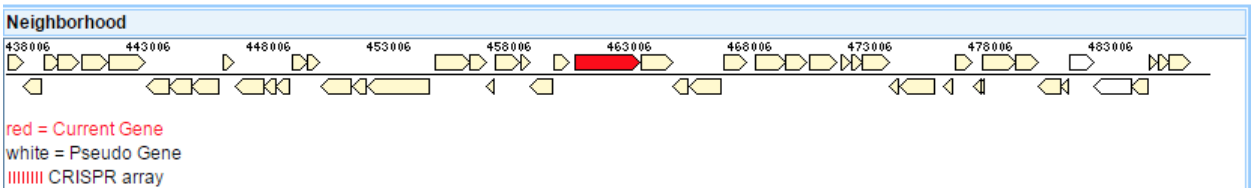



**Figure 7.** Mrub_0476 (red above) is part of an operon with mrub_0475 (left) and mrub_0477 (right). They are all on the same strand of DNA and going in the same direction. Image was obtained from http://img.jgi.doe.gov/cgi-bin/edu/main.cgi?section=FindGenes&page=geneSearch.

**Table 2.** b0115 and Mrub_0477 and Mrub_2322 are orthologs

| Description of Evidence | *E. coli b0115* | Mrub_0477 | Mrub_2322 |
|---|---|---|---|
| Cellular Localization | Cytoplasm | Cytoplasm | Cytoplasm |

| BLAST *E. coli* against *M. ruber* | | E-value 1e-99; bit score 302 | E-value 2e-70; bit score 226 |
|---|---|---|---|
| KEGG Pathway | Pyruvate Metabolism | Pyruvate Metabolism | Pyruvate Metabolism |
| Pfam- protein family | Biotin_lipoyl; e-value 2e-21 | Biotin_lipoyl; e-value 5.8e-20 | Biotin_lipoyl; e-value 2.7e-16 |
| TIGRfam family | TIGRfam01348 PDHac_trf_long; e-value 0.00 | TIGRfam01348 PDHac_trf_long; e-value 7.9e-155 | TIGRfam01348 PDHac_tr_long; e-value 5.6e-35 |
| CCD (COG Category) | COG0508 AceF; e-value 1.49e-139 | COG0508 AceF; e-value 6.79e-131 | COG0508 AceF; e-value 5.62e-156 |
| E.C. Number | 2.3.1.12 | 2.3.1.12 | 2.3.1.12 |
| PBD | 4N72 dihydrolipoamide acetyltransferase of pyruvate dehydrogenase; e-value 8.01065E-143 | 4N72 dihydrolipoamide acetyltransferase of pyruvate dehydrogenase; e-value 1.66494E-60 | 3DUF E1 subunit of pyruvate dehydrogenase complex; e-value 2.56059E-75 |

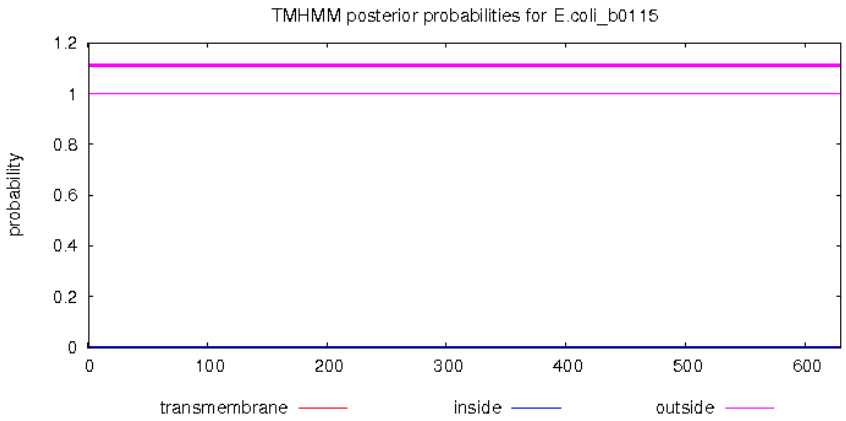
*Blast*

Table 2 shows the results of a protein BLAST comparison between E. coli b0015 and Mrub_0477, which produced an E-value of 1e-99 and bit score 302. This result is significant and indicates that sequence similarities are likely due to evolutionary relatedness as opposed to chance. BLAST comparison between b0015 and Mrub_2322 also produced a significant e-value at 2e-70 and bit score 226. This also indicates that sequence similarities are likely due to evolutionary relatedness.

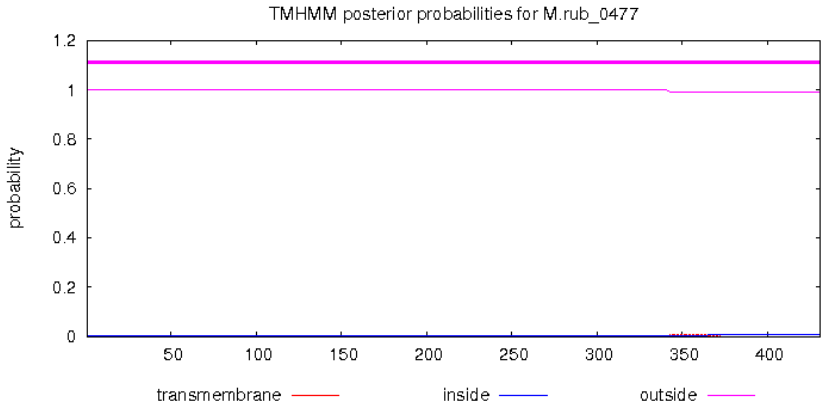
*Cellular Localization*

TMHMM analysis showed that *E. coli* b0015 (Figure 7 Panel A) has no transmembrane helices, indicating that it is not a membrane protein. Results were the same for Mrub_0477 and Mrub_2322 shown in Figure 4 Panel B and D, respectively, below.

Panel A




Panel B




Panel C

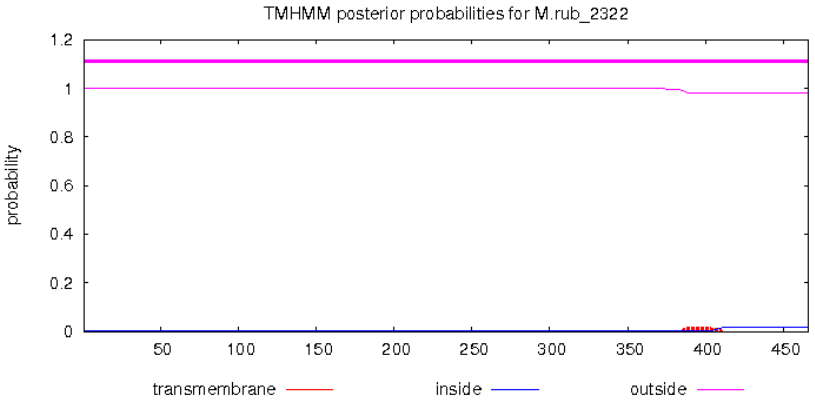
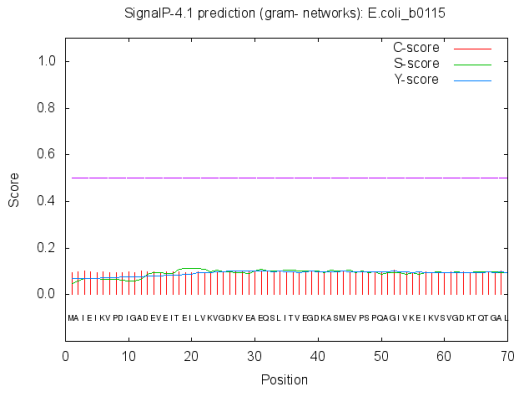


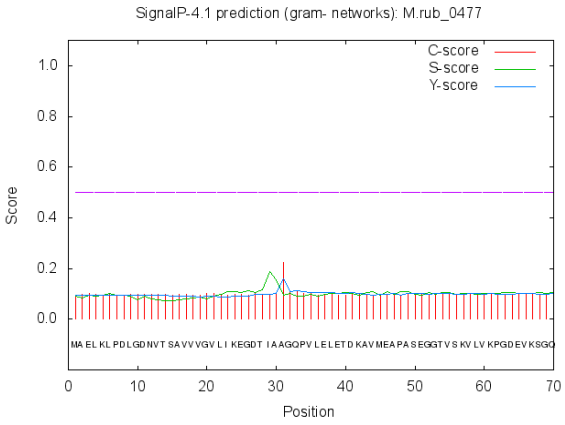

**Figure 10.** Mrub_0477, Mrub_2322, and *E. coli* b0015 do not contain TMH regions. A cytoplasmic location is predicted. Panel A= *E. coli* b0015; Panel B= Mrub_0477; Panel C= Mrub_2322. TMHMM server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) created this hydropathy plot.

The Signal Peptide probability had a predicted value of 0.097 for *E. coli*. Since this value is far from 1, it indicates that *E. coli* has no N-terminal membrane helices. In addition, no signal peptides were predicted. Results are shown in Figure 8 Panel A. The predicted Signal Peptide probability for Mrub_0477 and Mrub_2322, were also far less than 1.0 at 0.130 and 0.101, respectively, indicating that they also contain no N-terminal membrane helices. No signal peptides were predicted for *M. ruber* either. Results for *M. ruber* are shown in Figure 8 Panel B and C below.

Panel A




Panel B




Panel C

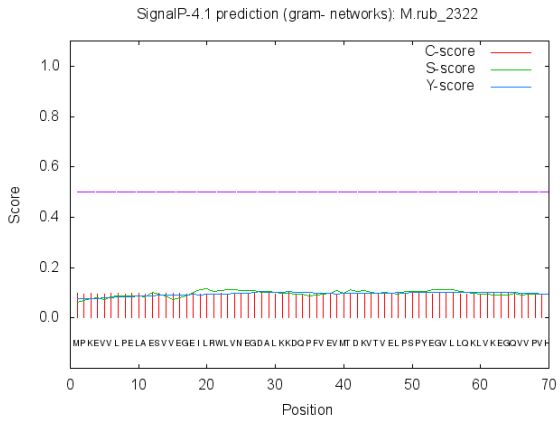


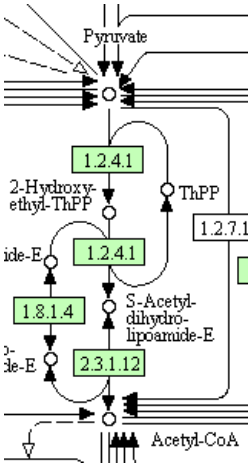

**Figure 11.** Mrub_0477, Mrub_2322 and *E. coli* b0115 do not contain signal peptides. A cytoplasmic location is predicted. Panel A= *E. coli* b0115; Panel B=Mrub_0477; Panel C= Mrub_2322. SignalP server v. 4.1 (http://www.cbs.dtu.dk/services/SignalP/) created this graph.

LipoP v1.0 (http://www.cbs.dtu.dk/services/LipoP/) did not predict any cleavage sites for *M. ruber* or *E. coli*, and its best cellular location prediction was the cytoplasm for all three genes. PSORT-B v3.0.2 (http://www.psort.org/psortb/) predicted both *M. ruber* genes and *E. coli* to be located in the cytoplasm as well. It is known that *E. coli*'s b0115 gene is located in the cytoplasm. Similar cellular localization analyses suggest that Mrub_0477 and Mrub_2322 are also located in the cytoplasm.

*KEGG Pathway*

The Pyruvate Metabolism reference pathway indicates that *E. coli* b0115 performs the role of E2 in the pyruvate dehydrogenase complex in that it converts S-Acetyl-dihydrolipoamide-E to Acetyl-CoA. It is known that *E. coli* utilizes the b0115 gene, as shown in Figure 9 Panel A. The KEGG pathway (http://www.genome.jp/kegg/pathway.html) for *M. ruber* shown in Figure 9 Panel B below also indicates that it utilizes the PDHC E2 gene (2.3.1.12).

Panel A




Panel B

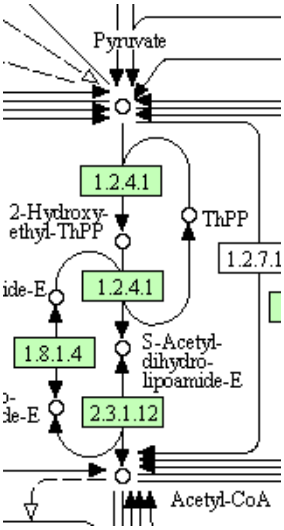



**Figure 12.** M. ruber and *E. coli* utilize the pyruvate dehydrogenase complex E2 protein in the synthesis of pyruvate when converting S-Acetyl-dihydrolipoamide-E to Acetyl-CoA. The pyruvate dehydrogenase complex E2 in each organism is notated as 2.3.1.12 in the pathway. Panel A= *E. coli* b0115; Panel B= Mrub_0477; Panel C= Mrub_2322. The KEGG pathway database (http://www.genome.jp/kegg/pathway.html) was used to create the above image.

*Pfam- protein family:*

Pfam analysis identified one family within the *E. coli* b0115 sequence; Biotin_lipoyl with an e-value of 2e-21. The query gene, PDHC E2, is in fact a member of the Biotin_lipoyl family.

The same Biotin_lipoyl family was identified in Mrub_0477 and Mrub_2322 with e-values of 5.8e-20 and 2.7e-16, respectively.

*TIGRfam- protein family:*

TIGRfam analysis (http://blast.jcvi.org/web-hmm/) predicts the name or function of the gene product. In this experiment, TIGRfam predicted pyruvate dehydrogenase complex dihydrolipoamide acetyltransferase, long form for the gene product of the query *E. coli* amino acid sequence. The sequence matched in three different places with three different e-values. Codons 4-73; e-value 2e-21; bit score 75.5, Codons 107-176; e-value 6.3e-23; bit score 80.3, and Codons: 207-277; e-value 4.7e-22; bit score 77.5. TIGRfam also predicted pyruvate dehydrogenase complex dihydrolipoamide acetyltransferase, long form as the gene product of the query Mrub_0477 amino acid sequence. The e-value was significant at 7.9e-155 with a bit score of 525.6. TIGRfam predicted pyruvate dehydrogenase complex dihydrolipoamide acetyltransferase, long form as the gene product of the query Mrub_2322 amino acid sequence as. The e-value was significant at 5.6e-35 with a bit score of 127.5.

*CCD (COG) Category:*

Conserved domains were located using CCD (http://www.ncbi.nlm.nih.gov/cdd/). Analysis of the b0115 sequence found one conserved domain, aceF, which is pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component. This had a significant e-value of 1.49e-139. The same domain was located when Mrub_0477 and Mrub_2322 sequences were analyzed with significant e-values of 6.79e-131 and 5.62e-156, respectively.

*E.C. Number:*

The enzyme commission number for *E. coli* b0115 is 2.3.1.12 as identified by the KEGG pathway. The enzyme commission number for Mrub_0477 and Mrub_2322 is also 2.3.1.12 as identified by the KEGG pathway.

*PBD:*

The protein bank database (PBD) identified dihydrolipoamide acetyltransferase of pyruvate dehydrogenase as the protein when the *E. coli* b0115 sequence was analyzed. The E-value for this finding was 8.01065E-143, indicating that there is a low chance of sequence similarity are due to chance, rather, similarities are likely due to evolutionary relatedness. The PBD for Mrub_0477 and Mrub_2322 also identified dihydrolipoamide acetyltransferase of pyruvate dehydrogenase as the protein, with significant e-values of 1.66494E-60 and 5.2862E-41, respectively.

*Gene Context*

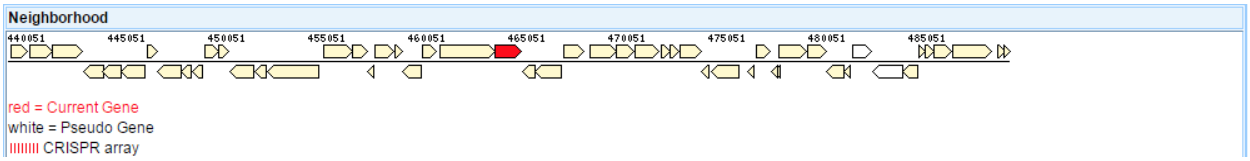



**Figure 13.** Mrub_0477 (red above) is part of an operon with mrub_0475 (far left) and mrub_0477 (middle). They are all on the same strand of DNA and going in the same direction. Image was obtained from http://img.jgi.doe.gov/cgi-bin/edu/main.cgi?section=FindGenes&page=geneSearch.

**Discussion**

Mrub_0476 and *E. coli* b0114 share enough structural and functional qualities to support the hypothesis that Mrub_0476 and b0114 are orthologous genes. Both are found in the cytoplasm. TIGRfam identified that both Mrub_0476 and b0114 were both a part of the Pyruvate Dehydrogenase E1 Component family. PBD identified both the Mrub_0476 and b0114 sequences as the pyruvate dehydrogenase E1 component. According to Pfam, both

Mrub_0476 and b0114 belong to the transketolase family. Both genes have the same E.C. number, indicating that they catalyze the same reaction. Finally, KEGG identified both as the E1 enzyme in the pyruvate dehydrogenase complex utilized in pyruvate metabolism. Since b0014 did not share enough similarities with either Mrub_1516 or Mrub_1517 to be convincing for evolutionary similarities, two separate BLAST searches were performed between Mrub_1516 and b0114 (Figure 8 Panel A) and Mrub_1517 and b0114 (Figure 8 Panel B). Neither had significant e-values, indicating that neither are orthologs with *E. coli* PDHC E1.

Panel A:

*Mrub_1516 against b0114*

**Range 1: 231 to 323** Graphics ·Next Match ·Previous Match

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 23.9 bits(50) | 0.013 | Compositional matrix adjust. | 27/95(28%) | 39/95(41%) | 24/95(25%) |

```
Query  766  ARDGQDCERWNM--LHPLETPRVPYIAQVMNDAPAVASTDYMKL----------FAEQVR  813
            AR+G + E   ++   L PL+TP +  +A V      AV    + M+            AE+
Sbjct  231  AREGVELEVVDLETLIPLDTPTI--LASVQKTGRAVVVYEAMRTGGFGAEIAARIAEEAL  288

Query  814  TYVPADDYRVLGTDG----------FGRSDSRENL  838
             Y+ A    RV G D            F R D+R  L
Sbjct  289  DYLQAPILRVAGWDAPYPPFSAVENFYRPDARRVL  323
```

**Range 2: 48 to 58** Graphics ·Next Match ·Previous Match ·First Match

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 18.9 bits(37) | 0.47 | Compositional matrix adjust. | 7/11(64%) | 9/11(81%) | 0/11(0%) |

```
Query  673  DGLERMYGEKQ  683
            DGL + YGEK+
Sbjct  48   DGLAQKYGEKR  58
```

**Range 3: 13 to 45** Graphics ·Next Match ·Previous Match ·First Match

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 18.5 bits(36) | 0.74 | Compositional matrix adjust. | 12/35(34%) | 18/35(51%) | 2/35(5%) |

```
Query  500  VRALNVMLKNKSIKDRLVPIIADEARTFGMEGLFR  534
            V+A+N  L     KD  V +  ++  T G  G+FR
Sbjct  13   VQAINEALDLALAKDERVVLFGEDVGTMG--GVFR  45
```

**Range 4: 69 to 170** Graphics ·Next Match ·Previous Match ·First Match

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 18.1 bits(35) | 0.83 | Compositional matrix adjust. | 26/109(24%) | 45/109(41%) | 26/109(23%) |

```
Query  392  IKGYGMGDAAEGKNIAHQVK----------KMNMDGVRH-IRDRFNVPVSDADIEKLP  438
            I G+G+G A  G     +++             ++  +RH  R RF +P+    + + P
Sbjct  69   IVGPGIGLAMAGLRPVAEIQFAGFLYPALDQILSHLGRMRHRTRGRFTIPM----VIRAP  124

Query  439  Y---ITFPE---GSEEHTYLHAQRQKLHGYLPSRQPNFTEKLELPSLQD  481
            Y   +  PE    S E    H    K+   +PS  P   + L L +++D
Sbjct  125  YGGGVKTPEQHADSPEAILAHVPGVKM--VIPS-SPERAKGLLLAAIED  170
```

**Range 5: 212 to 232** Graphics ·Next Match ·Previous Match ·First Match

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 17.3 bits(33) | 1.4 | Compositional matrix adjust. | 6/21(29%) | 12/21(57%) | 0/21(0%) |

```
Query  601  SMFGFQRIGDLCWAAGDQQAR  621
            S+F +  + ++C  A +  AR
Sbjct  212  SLFCYGGMVEVCLKAAEVAAR  232
```

**Range 6: 231 to 245** Graphics ·Next Match ·Previous Match ·First Match

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 15.4 bits(28) | 5.4 | Compositional matrix adjust. | 6/15(40%) | 9/15(60%) | 0/15(0%) |

```
Query  706  AEEGIRKGIYKLETI  720
            A EG+   +  LET+
Sbjct  231  AREGVELEVVDLETL  245
```

Panel B:

*Mrub_1517 against b0114*

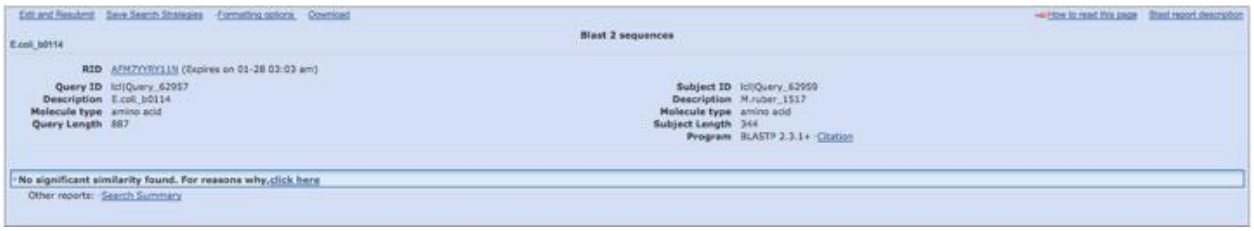



**Figure 8.** Panel A= mrub_1516 BLAST against b0114 was not significant with top hit e-value 0.013 and bit score 23.9. Panel B= mrub_1517 BLAST against b0114 found no significant sequence similarities. Image was obtained from http://blast.ncbi.nlm.nih.gov/Blast.cgi.

The mrub_0476 sequence was BLASTed against its own organism, Meiothermus ruber, to search for possible paralogs. No significant paralogs were identified.

Mrub_1516 and Mrub_1517 were BLASTed against the entire *E. coli* genome to unveil their role in *E. coli*. Mrub_1516 against *E. coli* showed that the first hit is 2-oxoisovalerate dehydrogenase with a significant e-value of 1e-105 and bit score 319 (Figure 9 Panel A). 2-oxoisovalerate dehydrogenase catalyzes the following reaction (Uniprot 2012):

3-methyl-2-oxobutanoate + CoA + NAD+ $\rightleftharpoons$ 2-methylpropanoyl-CoA + CO2 + NADH

During pyruvate metabolism, NAD+ is indeed converted to NADH (Henning 2012), so perhaps mrub_1516 is responsible for the above reaction in *M. ruber.*

Mrub_1517 against *E. coli* showed that the first hit is pyruvate dehydrogenase, partial. This protein belongs to the TPP_enzymes family, which is a group of enzymes that are cofactors for many different reactions in the body, where TPP stands for thiamine pyrophosphate. The first reaction of pyruvate metabolism, decarboxylation of pyruvate, is catalyzed by E1 as discussed above, but it requires a cofactor, TPP. Perhaps mrub_1517 encodes that TPP cofactor used in pyruvate dehydrogenase (Henning 1997).

Mrub_0477, Mrub_2322 and *E. coli* b0115 share enough structural and functional qualities to support the hypothesis that Mrub_0477 and Mrub_2322 are orthologous to b0115.

Both are found in the cytoplasm. TIGRfam identified that Mrub_0477, Mrub 2322 and b0115 were all a part of the pyruvate dehydrogenase complex dihydrolipoamide acetyltransferase, long form family. PBD identified all three sequences as the dihydrolipoamide acetyltransferase of pyruvate dehydrogenase. According to Pfam, the three belong to the biotin lipoyl family. All three genes have the same E.C. number, 2.3.1.12, indicating that they catalyze the same reaction. Finally, KEGG identified both as the E2 enzyme in the pyruvate dehydrogenase complex utilized in pyruvate metabolism.

**Conclusion**

Mrub_0476 and b0114 are orthologs. Neither Mrub_1516 nor Mrub_1517 are orthologous to b0114. The protein product of Mrub_1516 is predicted to be 2-oxoisovalerate dehydrogenase, which catalyzes the reaction that converts NAD+ to NADH. Mrub_1517 might catalyze the E1 reaction in the PDHC that decarboxylates pyruvate. Therefore, both Mrub_1516 and Mrub_1517 are involved with the PDHC.

Mrub_0477 and Mrub_2322 are paralogs and both are orthologous to b0115.

References

AAMC. (2008). Topics for Biological Science Section of the MCAT. [PDF Brochure].

Available Modules. [Internet]. 2015. Geni-Act; [2015 Dec 16]. Available from: http://www.geni-act.org/education/main/.

Berg JM, Tymoczko JL, Stryer L. Biochemistry. 5th edition. New York: W H Freeman; 2002. Chapter 17, The Citric Acid Cycle. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21163/.

Conserved Domains. [Internet]. NCBI; [2015 Dec 16]. Available from http://www.ncbi.nlm.nih.gov/cdd/.

Escherichia coli K-12 substr. MG1655 Pathway: superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glyoxylate bypass. [Internet]. 2015. SRI International: Ecocyc; [2015 Jan 28]. Available from http://ecocyc.org/ECOLI/NEW-IMAGE?type=PATHWAY&object=GLYCOLYSIS-TCA-GLYOX-BYPASS.

How does glycolysis connect to the citrate cycle. [Internet]. 2014 Sept 19. The Adapa Project: BioBook; [2015 Jan 28]. Available from: https://adapaproject.org/bbk_temp/tiki-index.php?page=Leaf%3A+How+does+glycolysis+connect+to+the+citrate+cycle%3F.

KEGG Pathway Database. [Internet]. 2015 Dec 2015. KEGG; [2016 Jan 13]. Available from: http://www.genome.jp/kegg-bin/show_pathway?org_name=mrb&mapno=00620&mapscale=&show_description=show.

Lodish, H., Berk, A., Kaiser, C., Kreiger, M., Scott, M. P., Bretscher. A., & Ploegh, H. (2008). Molecular Cell Biology (6th ed.). New York: W. H. Freeman and Company.

Molecular Mechanism of Regulation of the Pyruvate Dehydrogenase Complex from E. coli

Jana Hennig, Gunther Kern, Holger Neef, Michael Spinka, Hans Bisswanger,‖ and, and Gerhard Hübner. Biochemistry 1997 36 (50), 15772-15779

Phylogenetic Diversity. [Internet]. 2015. Joint Genome Institute; [2015 Dec 7].

Shen, B. 2014. Bioinformatics for Diagnosis, Prognosis, and Treatment of Complex Diseases. New York: Springer.

Uniprot. [Internet]. 2012. EMBL-EBI; [2016 Jan 20]. Available from http://www.uniprot.org/uniprot/P12694.

Prediction of lipoprotein signal peptides in Gram-negative bacteria.

A. S. Juncker, H. Willenbrock, G. von Heijne, H. Nielsen, S. Brunak and A. Krogh.

Protein Sci. 12(8):1652-62, 2003. Available from http://www.cbs.dtu.dk/services/LipoP/.

Protein Database. [Internet]. 2015. RCSB PBD; [2015 Dec 16]. Available from: http://www.rcsb.org/pdb/search/searchSequence.do.

Phylogenetic Diversity. [Internet]. 2015. Joint Genome Institute; [2015 Dec 7]. Available from http://jgi.doe.gov/our-science/science-programs/microbial-genomics/phylogenetic-diversity/.

Search Pfam. [Internet]. 2015. EMBL-EBI; [2015 Dec 16]. Available from: http://pfam.xfam.org/search.

SignalP 4.0: discriminating signal peptides from transmembrane regions

Petersen TN., Brunak S., von Heijne G. & Nielsen H.

Nature Methods, 8:785-786, 2011. Available from http://www.cbs.dtu.dk/services/SignalP/.

TMHMM 2.0: prediction of transmembrane helices in proteins. Krogh A. & Rapacki K.

CBS. 4 Aug 2015. Available from http://www.cbs.dtu.dk/services/TMHMM/.

TIGR fam proteins. [Internet]. 2015. J. Craig Venter Institute; [2015 Dec 16]. Available from: http://blast.jcvi.org/web-hmm/.