

5-2016

Mrub_2874 is homologous to b3386 and
Mrub_1349 is homologous to b2914, but
Mrub_1349 is not homologous to b4090

Samantha Murad

Augustana College, Rock Island Illinois

Dr. Lori Scott

Augustana College, Rock Island Illinois

Follow this and additional works at: <http://digitalcommons.augustana.edu/biolmruber>

 Part of the [Bioinformatics Commons](#), [Biology Commons](#), [Genetics Commons](#), [Genomics Commons](#), and the [Molecular Genetics Commons](#)

Augustana Digital Commons Citation

Murad, Samantha and Scott, Dr. Lori. "Mrub_2874 is homologous to b3386 and Mrub_1349 is homologous to b2914, but Mrub_1349 is not homologous to b4090" (2016). *Meiothermus ruber Genome Analysis Project*.
<http://digitalcommons.augustana.edu/biolmruber/17>

This Published Article 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.

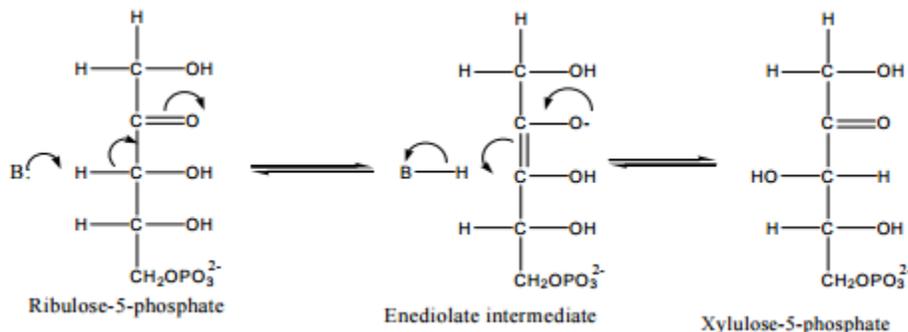
Mrub_2874 is homologous to b3386 and Mrub_1349 is homologous to b2914, but Mrub_1349 is not homologous to b4090

Introduction:

Meiothermus ruber (*M. ruber*) is a Gram-negative eubacteria that is nonmotile, nonspore-forming, red-pigmented, and rod-shaped. *M. ruber* belongs to the Deinococcus-Thermus phylum and grows between 35-70°C under highly aerated conditions (Loginova et al. 1984). While there are species within the Deinococcus-Thermus phylum that are well studied, such as *Thermus aquaticus*, not much is known about *M. ruber*. The Genomic Encyclopedia of Bacteria and Archaea (GEBA) project is designed to expand upon current knowledge of many microorganisms including *M. ruber* (Phylogenetic Diversity 2013). The sequencing of *M. ruber* through the GEBA project allows for further genomic analysis within the *M. ruber* genome.

Bioinformatics can be used to determine the relationship between two genes from two different species. The use of bioinformatics is beneficial since many different organisms can be used to find evolutionary similarities of genes or proteins. This method is useful to the scientific community since these databases would better support bio-sources for information and scientific inquiry (Gentleman et al. 2004). In addition, bioinformatics has the potential to affect all of society as more organisms are studied.

Since *Escherichia coli* (*E. coli*) is a well-studied species, *E. coli* will be used as a positive control to find evolutionary similarities to *M. ruber* through bioinformatics. Specifically, the pentose phosphate pathway will be investigated in *M. ruber* by comparing genes in *M. ruber* to genes in *E. coli*. The pentose phosphate pathway is essential to eubacteria because it generates NADPH and ribose-5-phosphate, a precursor to the biosynthesis of nucleotides (Kruger and von Schaewen 2003). Ribose-5-phosphate can be converted from xylulose-5-phosphate in a two-step process which involves two enzymes: ribulose-5-phosphate 3-epimerase and ribose-5-phosphate isomerase (Kruger and von Schaewen 2003). For the first step, xylulose-5-phosphate accepts a proton to form an enediolate intermediate and eventually is converted to ribulose-5-phosphate (Figure 1). This step is catalyzed by ribulose-5-phosphate 3-epimerase (Miles 2003). The second step involves the conversion of ribulose-5-phosphate to ribose-5-phosphate by forming an enediol intermediate (Figure 1). This second step is catalyzed by ribose-5-phosphate isomerase (Miles 2003). Once ribose-5-phosphate is formed, this substrate can be converted to other structures, such as phosphoribosyl pyrophosphate, in order to synthesize nucleotides (Kruger and von Schaewen 2003).



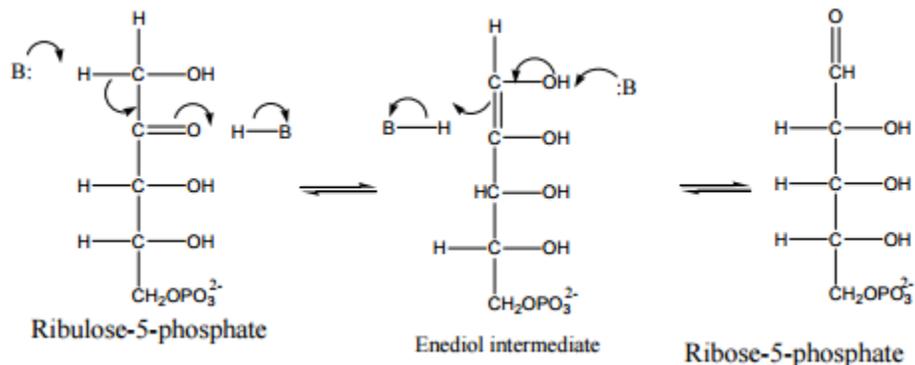


Figure 1. Xylulose-5-phosphate is converted to ribulose-5-phosphate and ribulose-5-phosphate is converted to ribose-5-phosphate. Panel A is the conversion of xylulose-5-phosphate to ribulose-5-phosphate through the ribulose-5-phosphate 3-epimerase enzyme (right to left). Panel B is the conversion of ribulose-5-phosphate to ribose-5-phosphate through the ribose-5-phosphate isomerase enzyme (left to right) (Miles 2003).

Functional evidence shows that the *rpe* gene in *E. coli* is likely to produce ribulose-5-phosphate 3-epimerase since *E. coli* cells with a mutation on the *rpe* gene did not produce the ribulose-5-phosphate 3-epimerase enzyme (Lyngstadaas et al. 1998). Another study was conducted comparing wild-type *E. coli* cells and *E. coli* cells with a mutation on *rpiA* gene. The mutated cells had significantly decreased the activity of the ribose-5-phosphate isomerase enzyme (Skinner and Cooper 1971). In addition, cloned DNA fragments with two open reading frames for the *rpiB* gene increased the enzymatic activity of ribose-5-phosphate isomerase (Sorensen and Hove-Jensen 1996). Thus, the *rpe* gene (b3386) in *E. coli* is hypothesized to code for ribulose-5-phosphate 3-epimerase, while *rpiA* (b2914) and *rpiB* (b4090) are hypothesized to code for ribose-5-phosphate isomerase in *E. coli*.

By using bioinformatics tools in the GENI-ACT lab notebook (<http://www.geni-act.org/>), the amino acid sequence of ribulose-5-phosphate 3-epimerase and ribose-5-phosphate isomerase of *M. ruber* will be compared to the amino acid sequence of ribulose-5-phosphate 3-epimerase and ribose-5-phosphate isomerase in *E. coli*. Mrub_2874 is the gene in *M. ruber* that is hypothesized to code for ribulose-5-phosphate 3-epimerase, while Mrub_1349 is hypothesized to code for ribose-5-phosphate isomerase in *M. ruber*. Comparison of these genes to *E. coli* through their amino acid sequences will determine if the protein from Mrub_2874 is homologous to the protein from b3386, if the protein from Mrub_1349 is homologous to the protein from b2914, and if the protein from Mrub_1349 is homologous to the protein from b4090.

Methods:

Bioinformatics programs were performed under the instructions within the GENI-ACT notebook (<http://www.geni-act.org/education/main/>) and the programs are described on the GENI-ACT website (<http://geni-act.org/>). Deviations from these instructions included an additional BLAST search of *E. coli* against *M. ruber* for each comparison that was saved in a separate document and not included in the GENI-ACT lab notebook. T-coffee was conducted using 20 different sequences, rather than the suggested 10 sequences for each gene. The Gene Context of the Horizontal Gene Transfer module was annotated with KEGG. In addition, EcoCyc was used to illustrate the pathway map rather than MetaCyc.

Results:

Mrub_2874 compared against b3386

A KEGG pathway map for the “Pentose Phosphate Pathway” displayed similarities between *E. coli* and *M. ruber* for the conversion of D-Xylulose-5P to D-Ribose-5P. The KEGG pathway ID was found to be 00030 for both organisms within the “Pentose Phosphate Pathway” map. The E.C. numbers of ribulose-5-phosphate 3-epimerase and ribose-5-phosphate isomerase are displayed in green for both organisms (5.1.3.1 and 5.3.1.6 respectively) and display the same protein name (Table 1). Table 1 displays the similarities and differences between these two genes (*Mrub_2874* and *b3386*) according to the various modules from the GENE-ACT lab notebook.

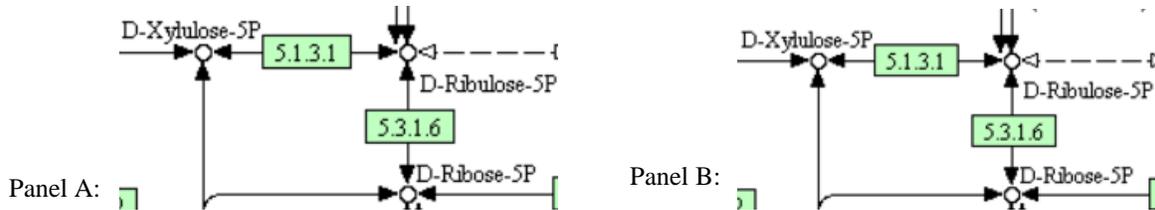


Figure 2. A KEGG pathway map displays ribulose-5-phosphate 3-epimerase and ribose-5-phosphate isomerase activities in both *E. coli* and *M. ruber*. Panel A displays the pathway map for *E. coli* where the green boxes display confirmed enzymatic activity of the respective proteins. Panel B displays the pathway map for *M. ruber* where green boxes also indicate confirmed enzymatic activities. The boxes designated 5.1.3.1 and 5.3.1.6 are the E.C. numbers which identify ribulose-5-phosphate 3-epimerase and ribose-5-phosphate isomerase respectively. KEGG (<http://www.genome.jp/kegg/pathway.html>) created this map.

Table 1: *E. Coli rpe* (b3386) and *Mrub_2874* are homologs

Description of evidence collected	<i>E. Coli</i>	<i>M. ruber</i>
Locus tag	b3386	<i>Mrub_2874</i>
KEGG pathway	Pathway ID: 00030 Pentose Phosphate Pathway	
E.C. number	5.1.3.1 ribulose-5-phosphate 3-epimerase	
BLAST <i>E. coli</i> against <i>M. ruber</i>	Score: 187 bits ; E-value: 4e-63	
CDD (COG category)	COG0036 Pentose-5-phosphate-3-epimerase [Carbohydrate transport and metabolism] E-value: 7.19e-123	
Cellular localization (Module 3)	E-value: 1.11e-103 Cytoplasmic	
TIGRfam - protein family	TIGR01163 <i>rpe</i> : ribulose-phosphate 3-epimerase Score: 504.4 E-value: 1.9e-148	
		Score: 438.4 E-value: 1.4e-128

Pfam - protein family	PF00834	
	Ribulose-phosphate 3 epimerase family Clan: TIM barrel (CL0036)	
PDB	Score: 306.0	Score: 268.7
	E-value: 7.5e-92	E-value: 2.7e-80
1TQJ		
Crystal structure of D-ribulose 5-phosphate 3-epimerase from Synechocystis to 1.6 angstrom resolution		
E-value: 1.5354E-59		E-value: 7.00934E-60

A BLAST search of both the amino acid sequence of Mrub_2874 in *M. ruber* and *rpe* in *E. coli* independently displayed the same protein, ribulose-5-phosphate 3-epimerase, as a top-hit in other organisms. A BLAST search of *E. coli* against *M. ruber* resulted in an alignment length of 224. The bit score was 187 and the E-value was 4e-63 (Figure 3). This high bit score and low E-value suggests that the similarities between *M. ruber* and *E. coli* are likely due to an evolutionary relationship and not due to chance.

Mrub_2874

Sequence ID: lcl|Query_88329 Length: 224 Number of Matches: 1

Range 1: 1 to 218 [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
187 bits(475)	4e-63	Compositional matrix adjust.	108/223(48%)	149/223(66%)	6/223(2%)
Query 1	MKQYLIAPSILSADFARLGEDTAKALAAGADVHFDVMDNHVVPNLTIGPMVLKSLRNYG				60
	M + L+APSIL+ADFARLGE +A AAG D +H DVM D +VNPNT GP+V++++R				
Sbjct 1	MPKLLVAPSILTADFARLGEQIREAEAAGVDWIHLVMDGRFVNPNTFGPLVVEAIRK--				58
Query 61	ITA-PIDVHLMVKPVDRIVPDFAAAGASITFHPEASEHVDRTLQLIKENGCKAGLVFNP				119
	+TA P+DVHLM+ +R + DFA AGA IT H EA+ H R +Q IKE G KAGL NP				
Sbjct 59	VTALPLDVHLMIVPEPERYIADFARAGADWITVHHEATPHAHRAVQQIKELGKKAGLAINP				118
Query 120	ATPLSYLDYVMDKLDVILLMSVNPFGGGQSFIPQTLDKLREVRRIIDESGFDIRLEVDGG				179
	TPL L ++ +LD+ LLMSVNPFGGGQ +IP + +++R + + ++VDGG				
Sbjct 119	GTPLEALLPLLPELDLALLMSVNPFGGGQKYIPASTERIRRLAALRGQMPACLIQVDGG				178
Query 180	VKVN NIGEIAAAGADMVAGSAIFDQDPDYKKVIDEMRSELAKV		222		
	+K N+ E+ AGAD+ V GSA+F+ + V + M+ L +V				
Sbjct 179	IKPVNVAEVYRAGADV VVGSALFNN---RPVAENMKKLLGEV		218		

Figure 3. A protein BLAST search of the *rpe* (b3386) of *E. coli* against Mrub_2874 of *M. ruber* indicates a likely evolutionary relationship. This alignment shows a length of 224. The bit score was 187 and the E-value was 4e-63. The “Query” sequence is the sequence of *E. coli* and “Sbjct” displays the sequence from *M. ruber*. BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) was used for the protein blast.

The CDD search identified one conserved protein domain for each of the query sequences (b3386 and Mrub_2874). Both sequences resulted in a COG number of COG0036, a COG name of “Pentose-5-phosphate-3-epimerase [Carbohydrate transport and metabolism],” and significant E-values (Table 1). The E-value of the *E. coli* search was determined to be 7.19e-123 while the *M. ruber* search was 1.11e-103 (Table 1). Since the domain name is the same as the query, the CDD tool suggests that these protein domains share an evolutionary relationship. Both the BLAST search and the CDD are indicative of a sequence-based similarity between the Rpe amino acid sequence of *E. coli* and the Mrub_2874 amino acid sequence of *M. ruber*.

To determine the cellular localization of the protein in question, TMHMM was used (Figure 4). Since no peaks were found on the transmembrane topology graph for either *E. coli* or *M. ruber*, the proteins for both organisms do not have transmembrane helices. In addition, the lack of transmembrane helices suggests that this protein is localized to the cytoplasm because it does not pass through or insert into a membrane.

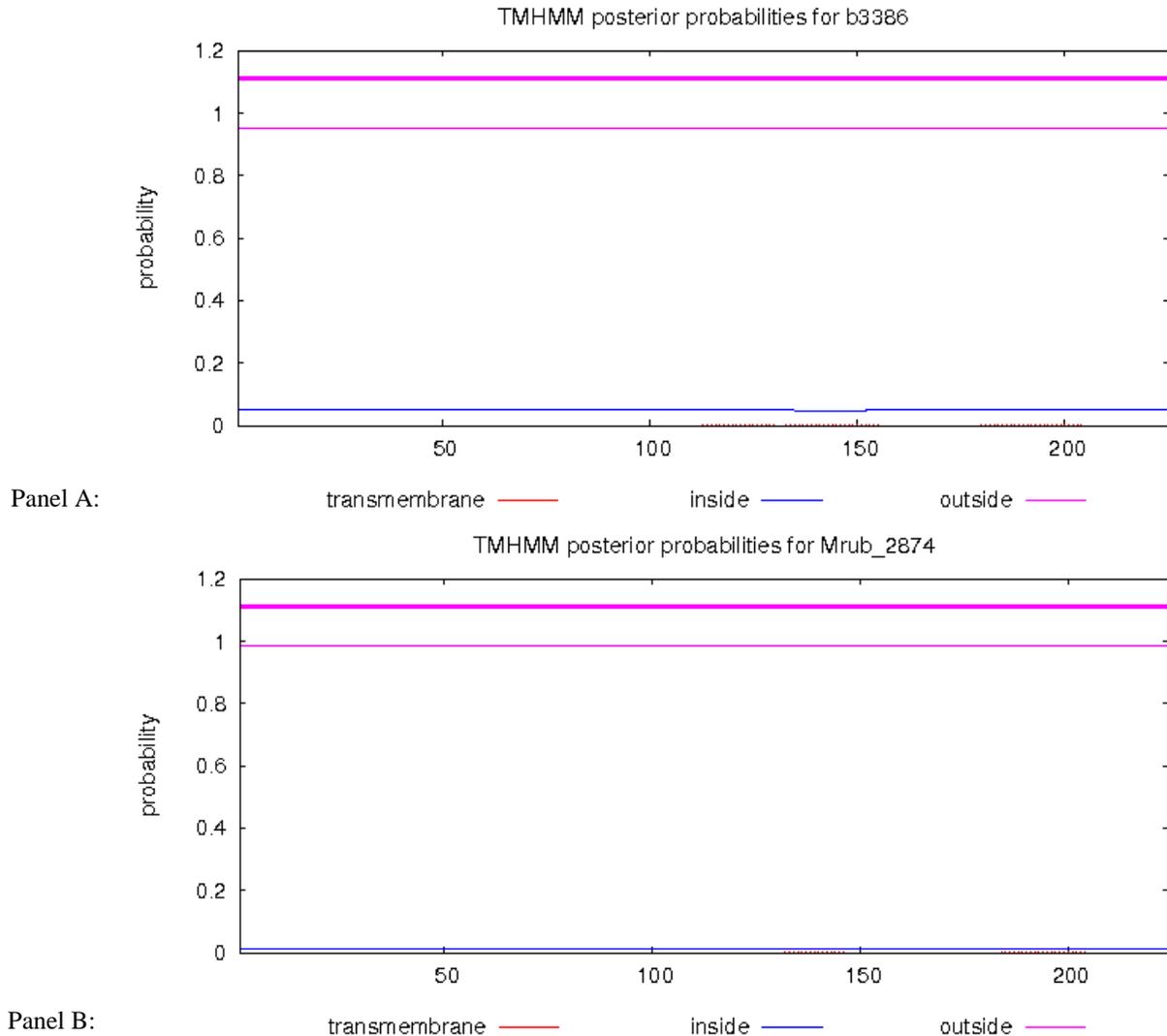
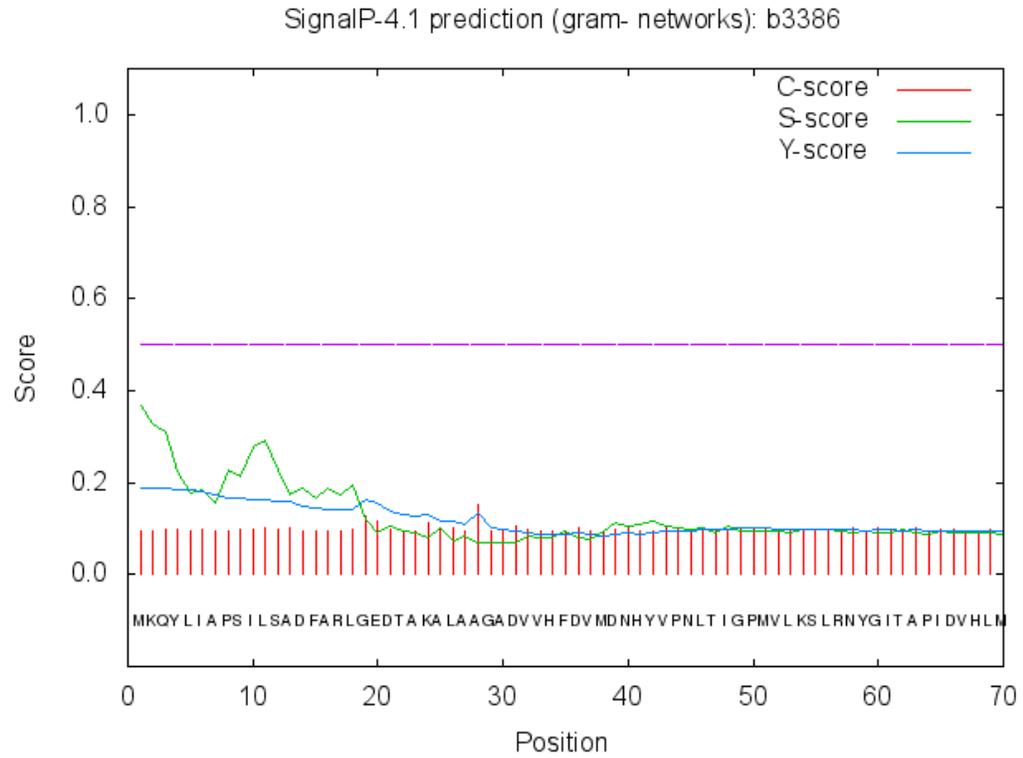
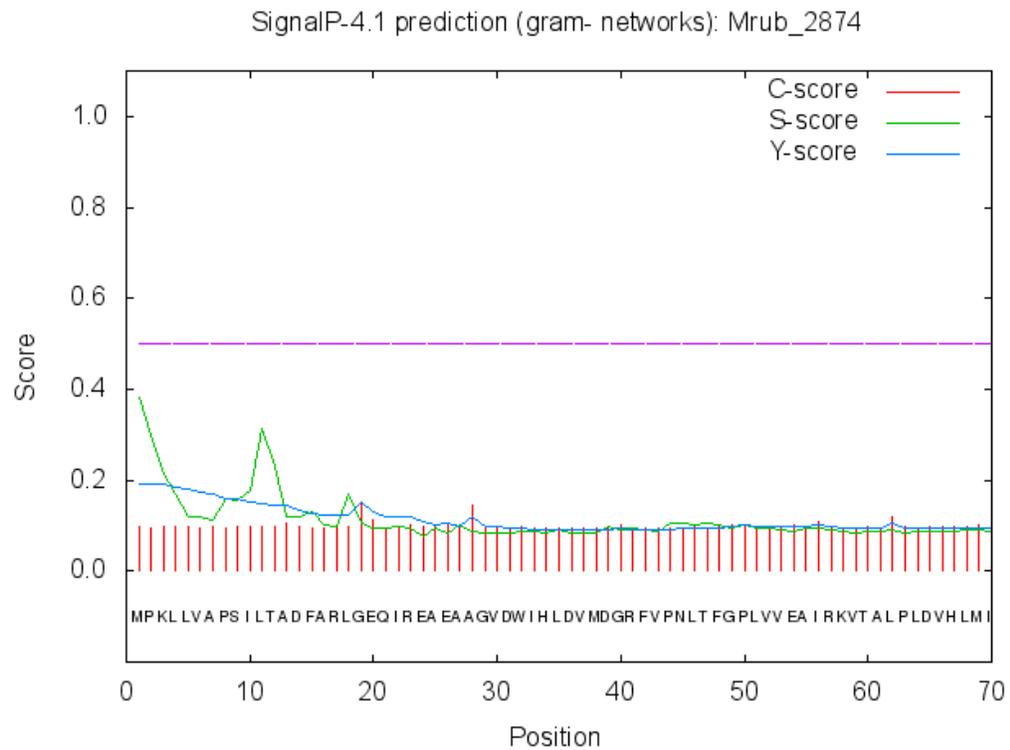


Figure 4. *E. coli rpe* and Mrub_2874 do not contain transmembrane helix regions and is likely in the cytoplasm. Panel A indicates the *E. coli rpe* query and Panel B indicates the Mrub_2874 query. TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) was used for the topology graph.

Signal peptides were predicted through SignalP (Figure 5). The signal peptide probability (D-score) of *E. coli rpe* and Mrub_2874 was 0.202 and 0.164 respectively. Due to the lack of peaks produced from Figure 5, there are no signal peptides predicted for either protein (no likely cleavage sites). The D-scores of these queries are considered low since they are not close to 1 (cut off value of 0.570 for both proteins) and also indicate no signal peptides located at the N-terminus of the protein sequence.



Panel A:



Panel B:

Figure 5. *E. coli rpe* and Mrub_2874 are not predicted to have signal peptides and are likely to be cytoplasmic proteins. Panel A is *E. coli rpe* and Panel B is Mrub_2874. SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP>) created this signal peptide graph.

The Mrub_2874 and b3386 sequences were also compared using the Phobius program. There are no significant vertical lines in gray that would indicate transmembrane helices (Figure 6). There is a slight spike in the signal peptide line (red) at the N-terminus of Mrub_2874 but all other cellular location modules (TMHMM, SignalP, and PSORT-B) do not support a signal peptide for Mrub_2874.

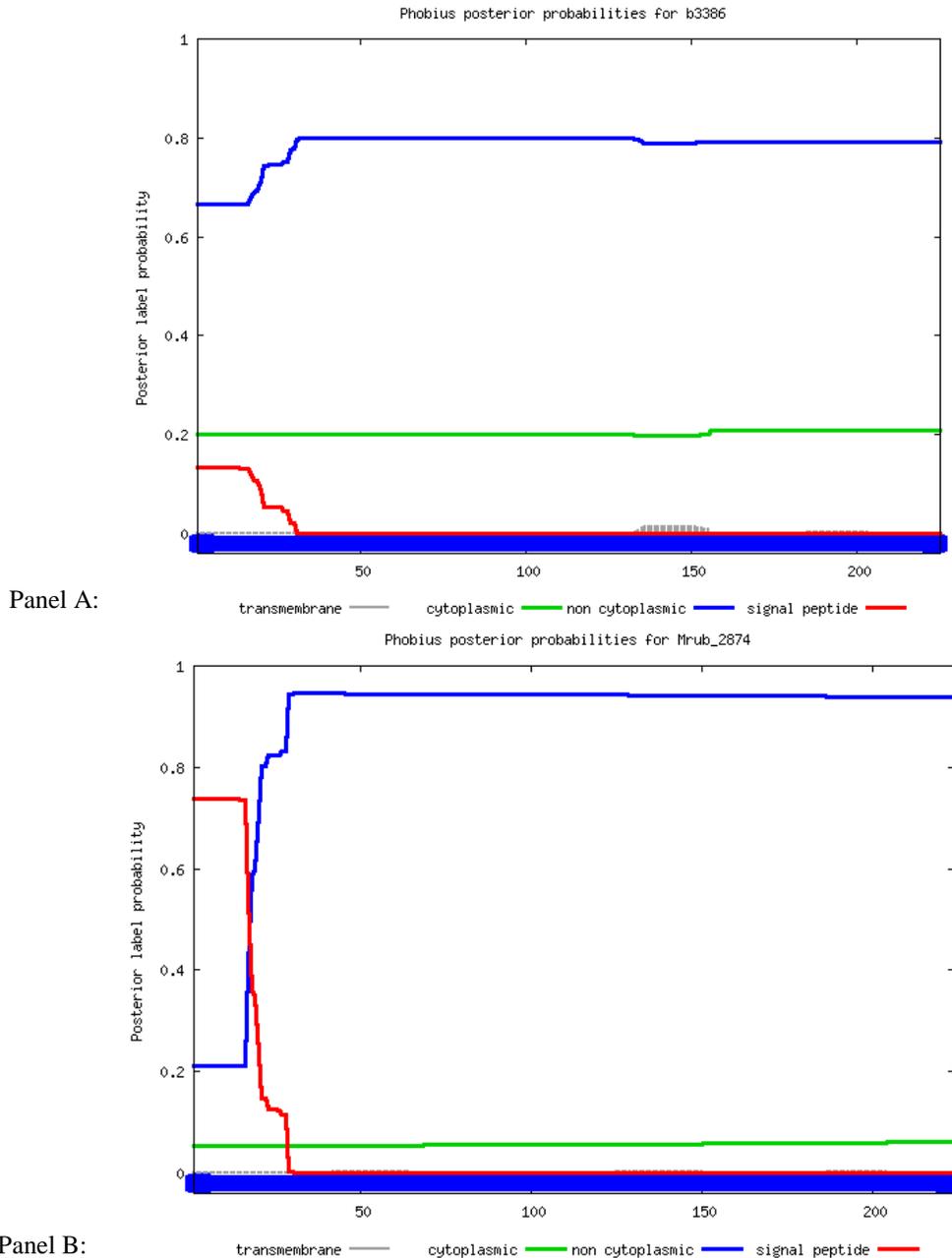


Figure 6. Mrub_2874 and *E. coli rpe* do not have transmembrane helices. No gray vertical lines are indicated which demonstrates a lack of hydrophobicity. Although there is a spike in the signal peptide line (red) by the N-terminus region for Mrub_2874, other data including SignalP suggests no signal peptides for both b3386 and Mrub_2874. Panel A is *E. coli* b3386 and Panel B is Mrub_2874. Phobius (<http://phobius.sbc.su.se/>) created these graphs.

Two more tools are used to determine the cellular localization of the protein query. LipoP suggested cytoplasm (Table 1) and the PSORT-B tool predicted a cytoplasmic score of 9.97 for *E. coli rpe* with a score of 0.01 for CytoplasmicMembrane and Periplasmic score. PSORT-B predicted the same cytoplasmic score of 9.97 for Mrub_2874 with a CytoplasmicMembrane and Periplasmic score of 0.01. Since the final prediction of the b3386 and Mrub_2874 cellular locations only included the cytoplasm, the 0.01 scores for CytoplasmicMembrane and Periplasmic are not significant. According to these tools, ribulose-5-phosphate 3-epimerase for *E. coli* and *M. ruber* is predicted to be in the cytoplasm which is reflective of the EcoCyc's prediction for the cellular location of ribulose-5-phosphate 3-epimerase.

Several tools were used to determine the structural similarities between *rpe* and Mrub_2874. TIGRFAM, which determines protein homology, resulted in identical TIGRFAM numbers: TIGR01163 (Table 1). The family name for this TIGRFAM number is “*rpe*: ribulose-phosphate 3-epimerase.” Both *E. coli* and *M. ruber* searches produced significant scores and E-values: the scores for *rpe* and Mrub_2874 were 504.4 and 438.4 respectively, while the E-values for *rpe* and Mrub_2874 were 1.9e-148 and 1.4e-128 respectively (Table 1). A Pfam search resulted in the same number and name (PF00834, Ribulose-phosphate 3 epimerase family) (Table 1). A pairwise alignment of both *E. coli rpe* and Mrub_2874 against the conserved sequence of ribulose-5-phosphate 3-epimerase displays how well the amino acid sequences relate to the consensus sequence (Figure 7). The scores for *rpe* and Mrub_2874 were 306.0 and 268.7 respectively, while the E-values for *rpe* and Mrub_2874 were 7.5e-92 and 2.7e-80 respectively (Table 1). Key functional groups for *E. coli rpe* were determined to be S9, H34, D36, D39, N45, H67, L68, H91, E93, G109, P142, G146, Q147, F149, D176, G177, G178, V196; while the key functional groups for *M. ruber* Mrub_2874 were determined to be S5, H30, D32, D35, N41, H63, L64, H87, E89, G105, P132, G142, Q143, D172, G173, G174, V192. Although they do not share the exact same amino acid positions, most of the key functional amino acids in *E. coli* are the same key amino acids in *M. ruber*. These results indicate similar protein domains and they resemble the CDD output domain. The scores, E-values, and key functional groups from Pfam also indicate significance in the conserved amino acid sequences between *E. coli* and *M. ruber*.

Panel A:

```
#HMM iapSllsadfarylgeevkaveeagadllHvDvmDghfvpNltigplvvealrklt.telpldvHLMveevdrivpefaeagasiisvHaEalehl
#MATCH iapS+lsadfarylge+ +++ +agad+++H+DvmD+h+vpNltigp+v+++lr++ +++p+dvHLMv++vdrivp+fa agasii++H+Ea+eh+
#PP g*****
#SEQ IAPSILSADFARLGEDTAKALAAAGADVHFDVMDNHYVPNLTIGPMVLKSLRNYGITAPIDVHLMVKPVDRIVPDFAAAGASIIIFHPEASEHV
krtiqeikeagakaglvlnpatplealeylllekldlvlmsvPgfGQkFipevlekikvrklidekeldkievDGGvklDnieqiaeagadvlvagsavf
+rt+q ike+G+kaglv+npatpl++l+y+++kld++llmsv+PgfGQ+Fi++l+k+++vr++ide+++d+++evDGGvk++ni +ia+agad++vagsa+f
*****
DRTLQLIKENGCKAGLVFNPATPLSYLDYVMDKLDVILLMSVNPFGFGQSFIPQTLDKLREVRRIIDESGFDIRLEVDGGVKVNNIGEIAAAGADMFVAGSAIF
```

Panel B:

```
#HMM iapSllsadfarylgeevkaveeagadllHvDvmDghfvpNltigplvvealrklt.telpldvHLMveevdrivpefaeagasiisvHaEalehl
#MATCH +apS+l+adfarlge+++++e+ag+d+++H+DvmDg+fvpNlt+gplvvea+rk+t lpldvHLM+ e++r++ +fa+aga++i+vH Ea++h +
#PP gg*****
#SEQ VAPSILTADFARLGEQIREAEEAGVDWIHLQVMDGRFVPLNLTGPLVVEAIRKVTALPLDVHLMVIEPERYIADFARAGADNITVHHEATPHAF
rtiqeikeagakaglvlnpatplealeylllekldlvlmsvPgfGQkFipevlekikvrklidekeldkievDGGvklDnieqiaeagadvlvagsavf
r +q+ike+G+kaglv+np tpleal ll++ldl llmsv+PgfGQk+ip+ +e+i+++ +l + + +i+vDGG+k n++++ +agad+v+gsa+f
*****
RAVQQIKELGKKAGLAINPGTLEALLPLPELDLALLMSVNPFGFGQKYIPASTERIRRLAALRGQNNPAQLIQVDGGIKPVNVAEVYRAGADVVVGSALF
```

Figure 7. *E. coli rpe* and Mrub_2874 have similar domain structures according to the consensus sequence of ribulose-5-phosphate 3-epimerase. Panel A is the *E. coli rpe* query (#SEQ) against the consensus sequence (#HMM). Panel B is the Mrub_2874 query (#SEQ) against the consensus sequence (#HMM). Pfam (<http://pfam.sanger.ac.uk/search>) was used to create the pairwise alignment.

In addition to TIGRFAM and Pfam, PDB suggested the similarities in the 3-D structure of the amino acid sequence. For both organisms, the PDB code was 1TQJ and was named “Crystal structure of D-ribulose 5-phosphate 3-epimerase from *Synechocystis* to 1.6 angstrom resolution.” The alignment length was 218 for b3386 and 209 for Mrub_2874. The E-values were 1.5354E-59 and 7.00934E-60 for b3386 and Mrub_2874 respectively. Since both queries resulted in the same protein name, the structures are similar. The significant E-values also indicate an evolutionary relationship between the two structures. All structure-based data suggest that the similarities in b3386 and Mrub_2874 domain and 3-D structures are due to their evolutionary relationship.

The IMG/EDU Gene finder was used to find the ortholog neighborhood region of *E. coli* *rpe* and Mrub_2874 with the color annotation of KEGG (Figure 8). The results for *rpe* shows that *rpe* (Panel A, underlined in red) is immediately next to the *gph* gene. Since these adjacent genes are aligned in the same direction, this gene (*rpe*) is likely to be an operon with *aroK*, *aroB*, *damX*, *dam*, *rpe*, *gph*, and *trpS*. Further research shows evidence that all these proteins act as an operon in *E. coli* (Lyngstadaas et al. 1999). These proteins are described as being a part of the “*dam*-containing operon” that uses seven genes for various functions in *E. coli*. On the other hand, the results for Mrub_2874 show slightly difference characteristics to the results for *rpe*. Mrub_2874 (Panel B, underline in red) is immediately next to the Mrub_2873 and Mrub_2872 genes but the *M. ruber* equivalent of the *dam* gene is not immediately next to Mrub_2874 in the same direction. However, all the other genes align in a similar manner to the *E. coli* ortholog neighborhood which may indicate that Mrub_2874 is part of an operon. Further functional evidence would be necessary to confirm if Mrub_2874 is part of an operon with Mrub_2873 and Mrub_2872.

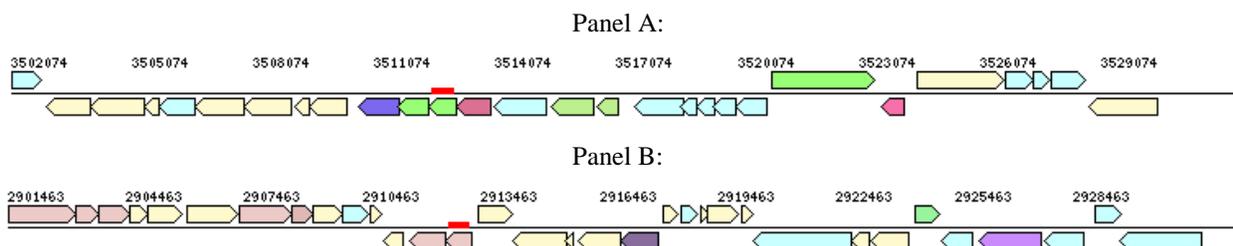


Figure 8. *E. coli* *rpe* and Mrub_2874 are likely part of an operon. Panel A is the visualization of the *E. coli* *rpe* ortholog neighborhood and Panel B is the Mrub_2874 ortholog neighborhood. The red lines indicate the location of either *rpe* (Panel A) or Mrub_2874 (Panel B). IMG/EDU (<http://img.jgi.doe.gov/cgi-bin/edu/main.cgi?section=FindGenes&page=geneSearch>) was used to illustrate the ortholog neighborhood regions.

Mrub_1349 compared against b2914

According to the KEGG pathway map for the “Pentose Phosphate Pathway” (Figure 2), Mrub_1349 and b2914 share the same E.C. number (5.3.1.6) (Table 2). Table 2 displays the similarities and differences between these two genes (Mrub_1349 and b2914) according to the various modules from the GENE-ACT lab notebook.

Table 2: *E. Coli rpiA* (b2914) and Mrub_1349 are homologs

Description of evidence collected	<i>E. Coli</i>	<i>M. ruber</i>
Locus tag	b2914	Mrub_1349
KEGG pathway	Pathway ID: 00030 Pentose Phosphate Pathway	
E.C. number	5.3.1.6 ribose-5-phosphate isomerase	
BLAST <i>E. coli</i> against <i>M. ruber</i>	Score: 115 bits ; E-value: 2e-35	
CDD (COG category)	COG0120 Ribose 5-phosphate isomerase [Carbohydrate transport and metabolism] E-value: 1.57e-112	
Cellular localization (Module 3)	Cytoplasmic	
TIGRfam - protein family	TIGR00021 <i>rpiA</i> : ribose 5-phosphate isomerase A Score: 406.6 E-value: 5.4e-119	
Pfam - protein family	PF06026 Ribose 5-phosphate isomerase A (phosphoriboisomerase A) Clan: ISOCOT_Fold (CL0246) Score: 197.8 E-value: 8.6e-59	
PDB	1O8B Structure of <i>Escherichia coli</i> ribose-5-phosphate isomerase, RpiA, complexed with arabinose-5-phosphate. E-value: 1.08859E-123	

A BLAST search of both the amino acid sequence of Mrub_1349 in *M. ruber* and *rpiA* in *E. coli* independently displayed the same protein, ribose-5-phosphate isomerase A, as a top-hit in other organisms. A BLAST search of *E. coli* against *M. ruber* resulted in an alignment length of 228. The bit score was 115 and the E-value was 2e-35 (Figure 9). This high bit score and low E-value suggests that the similarities between *M. ruber* and *E. coli* are likely due to an evolutionary relationship and not due to chance.

Mrub_1349

Sequence ID: Icl|Query_55811 Length: 228 Number of Matches: 1

Range 1: 4 to 223 [Graphics](#)

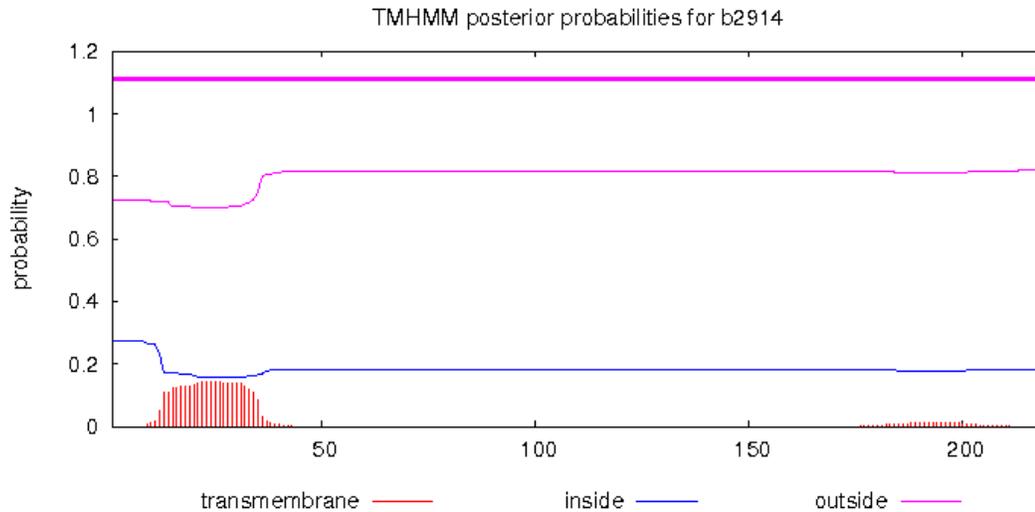
▼ Next Match ▲ Previous M

Score	Expect	Method	Identities	Positives	Gaps
115 bits(287)	2e-35	Compositional matrix adjust.	79/222(36%)	127/222(57%)	9/222(4%)
Query 3		QDELKKA V G W A A L Q Y V Q P G T I V G V G T G S T A A H F I D A L G T M - - K G Q I E G - - A V S S S D A S T E			58
Sbjct 4		QENFKQQAAL EAVK FV R S G M V V G L G T G S T A R Y A V L E L G R R L R E G E L S D V V A V P T S E A T A R			63
Query 59		K L K S L G I H V F D L N E V D S L G I Y V D G A D E I N G H M Q M I K G G G A A L T R E K I I A S V A E K F I C I A D			118
Sbjct 64		LAQSQGIPLVEL - E A R G V D L A I D G A D E I A P D L S L I K G L G G A L L R E K I V E A N A R Q F I V I A D			122
Query 119		A S K Q V D I L G K F P L P V E V I P M A R S A V A R Q L V K L G G - - - R P E Y R Q G V V T D N G N V I L D V H G M E			175
Sbjct 123		H S K K V G Q L G R G V V P V E I V R F G Y R A T L Q A L G R M G K A T L R M N G E E F F Y S D G G N L I A D I D F G P			182
Query 176		I L D P I A M E N A I N A I P G V V T V G L F A N R G A D V A L I G T P D G V K T I	217		
Sbjct 183		I P + E + I P G V V G L F A + A + + P + G + + +			223

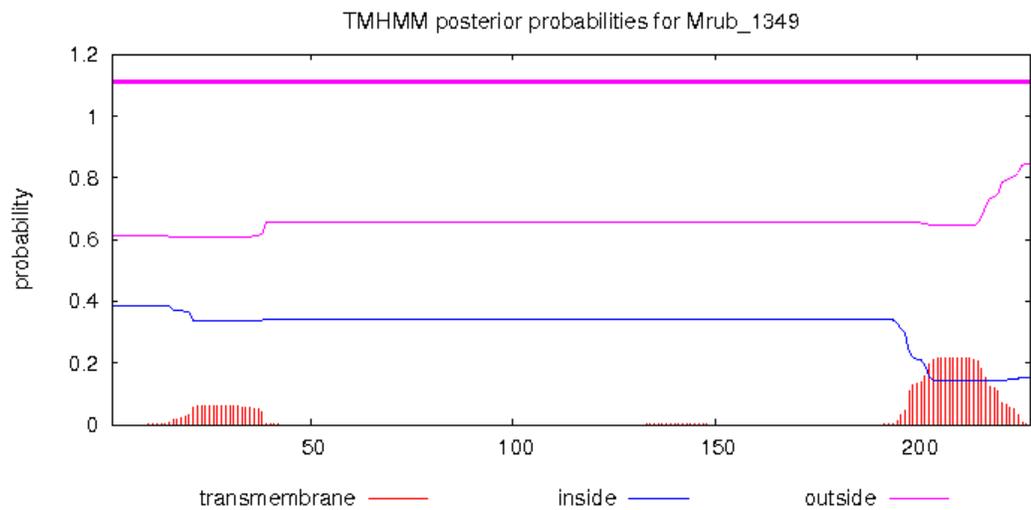
Figure 9. A protein BLAST search of the *rpiA* (b2914) of *E. coli* against Mrub_1349 of *M. ruber* indicates a likely evolutionary relationship. This alignment shows a length of 228. The bit score was 115 and the E-value was $2e-35$. The “Query” sequence is the sequence of *E. coli* and “Sbjct” displays the sequence from *M. ruber*. BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) was used for the protein blast.

The CDD search identified one conserved protein domain for each of the query sequences (b2914 and Mrub_1349). Both sequences resulted in a COG number of COG0120, a COG name of “Ribose 5-phosphate isomerase [Carbohydrate transport and metabolism],” and significant E-values (Table 2). The E-value of the *E. coli* search was determined to be $1.57e-112$ while the *M. ruber* search was $1.08e-88$ (Table 2). Since the domain name is the same as the query, the CDD tool suggests that these protein domains share an evolutionary relationship. Both the BLAST search and the CDD are indicative of a sequence-based similarity between the RpiA amino acid sequence of *E. coli* and the Mrub_1349 amino acid sequence of *M. ruber*.

To determine the cellular localization of the protein in question, TMHMM was used (Figure 10). Since no significant peaks were found on the transmembrane topology graph for either *E. coli* or *M. ruber*, the proteins for both organisms do not have transmembrane helices. In addition, the lack of transmembrane helices suggests that this protein is localized to the cytoplasm because it does not pass through or insert into a membrane.



Panel A:

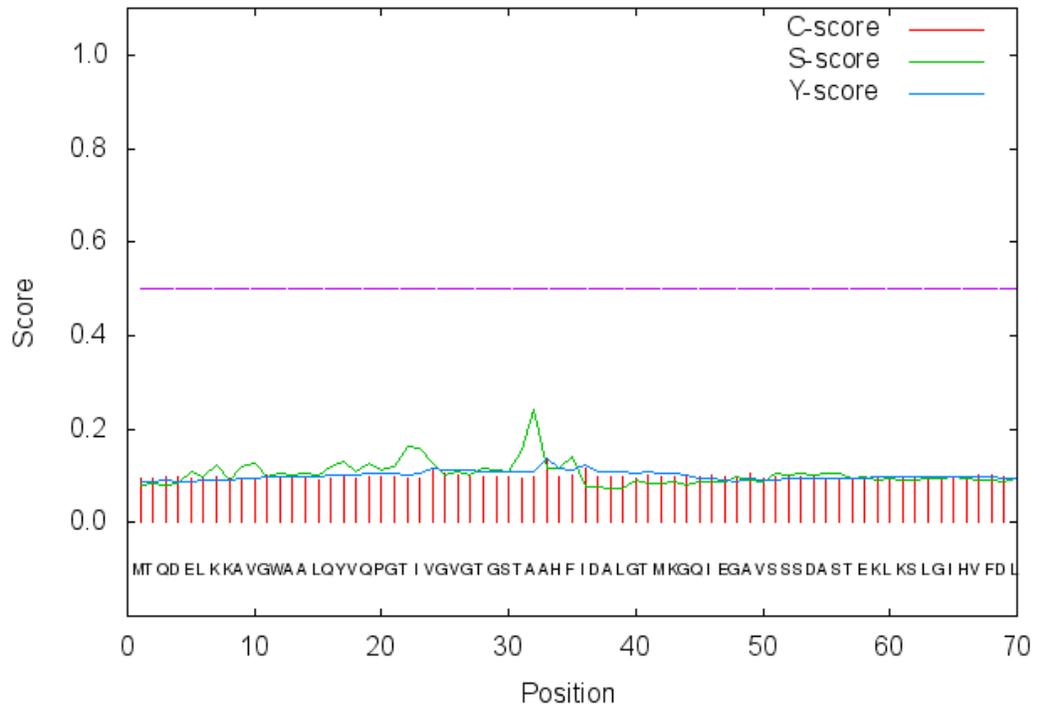


Panel B:

Figure 10. *E. coli rpiA* and Mrub_1349 do not contain transmembrane helix regions and is likely in the cytoplasm. Panel A indicates the *E. coli rpiA* query and Panel B indicates the Mrub_1349 query. TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) was used for the topology graph.

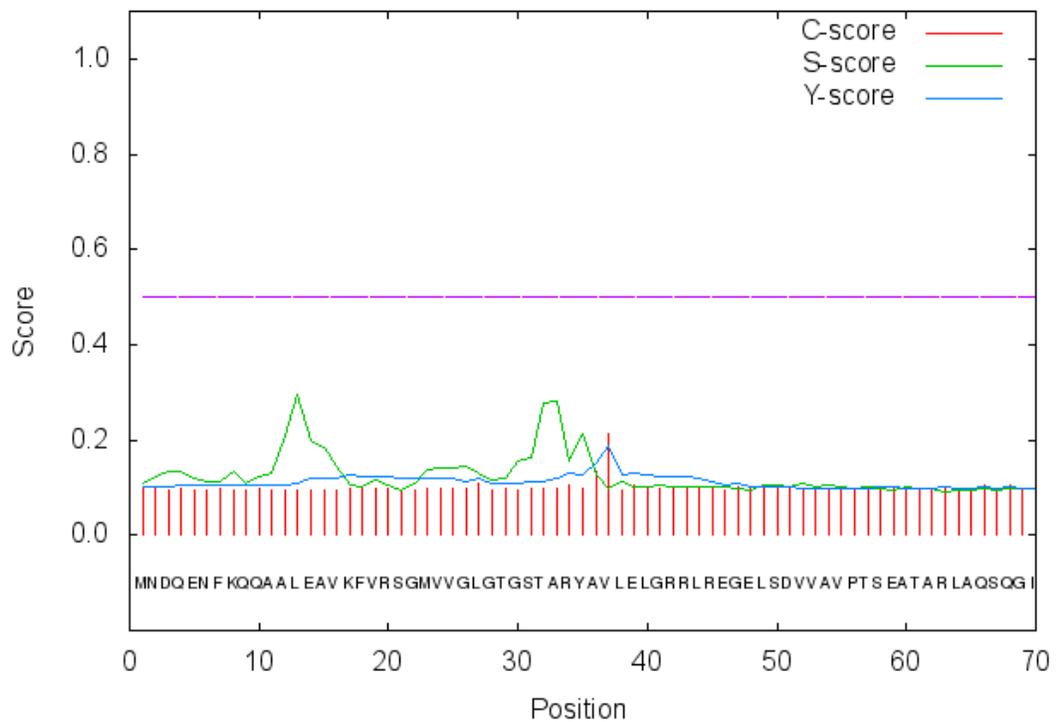
Signal peptides were predicted through SignalP (Figure 11). The signal peptide probability (D-score) of *E. coli rpiA* and Mrub_1349 was 0.127 and 0.170 respectively. Due to the lack of peaks produced from Figure 11, there are no signal peptides predicted for either protein (no likely cleavage sites). The D-scores of these queries are considered low since they are not close to 1 (cut off value of 0.570 for both proteins) and also indicate no signal peptides located at the N-terminus of the protein sequence.

SignalP-4.1 prediction (gram- networks): b2914



Panel A:

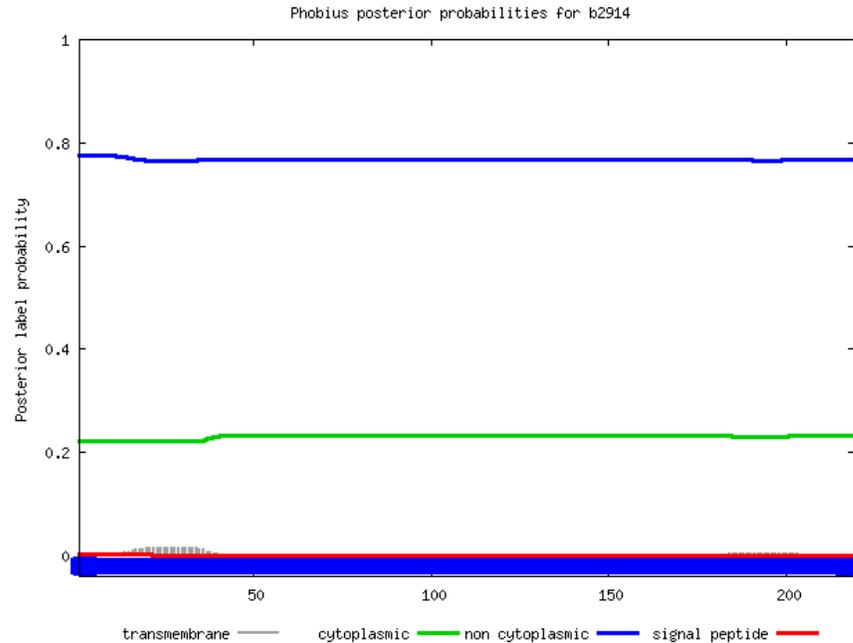
SignalP-4.1 prediction (gram- networks): Mrub_1349



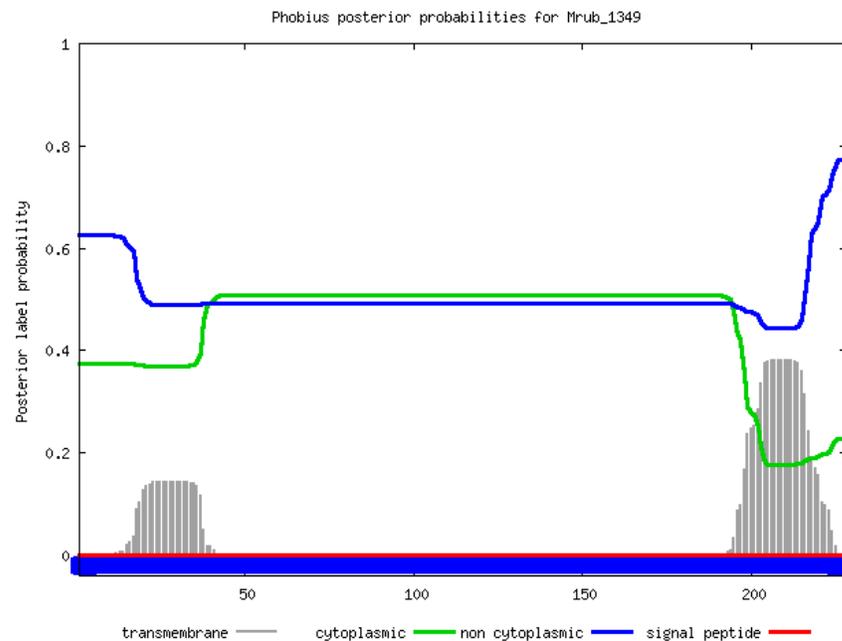
Panel B:

Figure 11. *E. coli rpiA* and Mrub_1349 are not predicted to have signal peptides and are likely to be cytoplasmic proteins. Panel A is *E. coli rpiA* and Panel B is Mrub_1349. SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP>) created this signal peptide graph.

The Mrub_1349 and b2914 sequences were also compared using the Phobius program. There are no significant vertical lines in gray that would indicate transmembrane helices (Figure 12). There are slight peaks (gray) at the N-terminus and C-terminus of Mrub_1349, but all other cellular location modules (TMHMM, SignalP, and PSORT-B) do not support a signal peptide for Mrub_1349. In addition, the peaks in the Mrub_1349 graph (Figure 12) do not surpass the non-cytoplasmic mark line (blue) which suggest that the peaks are insignificant.



Panel A:



Panel B:

Figure 12. Mrub_1349 and *E. coli rpiA* do not have transmembrane helices. Although there are two peaks by the N-terminus and C-terminus regions for Mrub_1349, they are insignificant and are still suggested to be cytoplasmic. Panel A is *E. coli* b2914 and Panel B is Mrub_1349. Phobius (<http://phobius.sbc.su.se/>) created these graphs.

Two more tools are used to determine the cellular localization of the protein query. LipoP suggested cytoplasm (Table 2) for both genes. However, the PSORT-B tool predicted a cytoplasmic score of 8.96 for Mrub_1349 with a CytoplasmicMembrane score of 0.51, a Periplasmic score of 0.26, an OuterMembrane score of 0.01, and an Extracellular score of 0.26. PSORT-B was inconclusive for *E. coli rpiA*. Since the final prediction of the Mrub_1349 cellular location was only cytoplasm, the scores for CytoplasmicMembrane, Periplasmic, OuterMembrane, and Extracellular are not significant. According to these tools, ribose-5-phosphate isomerase for *E. coli* and *M. ruber* is predicted to be in the cytoplasm which is reflective of the EcoCyc's prediction for the cellular location of ribose-5-phosphate isomerase.

Several tools were used to determine the structural similarities between *rpiA* and Mrub_1349. TIGRFAM, which determines protein homology, resulted in identical TIGRFAM numbers: TIGR00021 (Table 2). The family name for this TIGRFAM number is “*rpiA*: ribose 5-phosphate isomerase A.” Both *E. coli* and *M. ruber* searches produced significant scores and E-values: the scores for *rpiA* and Mrub_1349 were 406.6 and 349.6 respectively, while the E-values for *rpiA* and Mrub_1349 were 5.4e-119 and 7.6e-102 respectively (Table 2). A Pfam search resulted in the same number and name (PF06026, Ribose 5-phosphate isomerase A (phosphoriboisomerase A)) (Table 2). A pairwise alignment of both *E. coli rpiA* and Mrub_1349 against the conserved sequence of ribose-5-phosphate isomerase displays how well the amino acid sequences relate to the consensus sequence (Figure 13). The scores for *rpiA* and Mrub_1349 were 197.8 and 192.5 respectively, while the E-values for *rpiA* and Mrub_1349 were 8.6e-59 and 3.8e-57 respectively (Table 2). Key functional groups for *E. coli rpiA* were determined to be S54, D81, G82, D84, K94, G95, G97, E103, K104, K121, P132, V135, R157, T167, D175, G197, V198, G202, F204; while the key functional groups for *M. ruber* Mrub_1349 were determined to be S59, D86, G87, D89, K99, G100, G102, E108, K109, K126, P137, R162, D180, G202, V203, G207, F209. Although they do not share the exact same amino acid positions, most of the key functional amino acids in *E. coli* are the same key amino acids in *M. ruber*. These results indicate similar protein domains and they resemble the CDD output domain. The scores, E-values, and key functional groups from Pfam also indicate significance in the conserved amino acid sequences between *E. coli* and *M. ruber*.

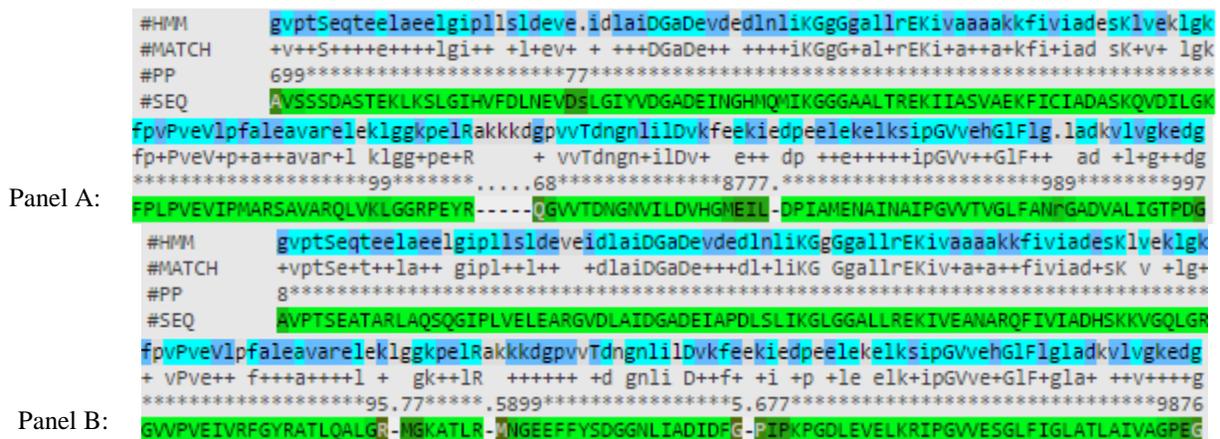


Figure 13. *E. coli rpiA* and Mrub_1349 have similar domain structures according to the consensus sequence of ribose-5-phosphate isomerase. Panel A is the *E. coli rpiA* query (#SEQ) against the consensus sequence (#HMM). Panel B is the Mrub_1349 query (#SEQ) against the consensus sequence (#HMM). Pfam (<http://pfam.sanger.ac.uk/search>) was used to create the pairwise alignment.

In addition to TIGRFAM and Pfam, PDB suggested the similarities in the 3-D structure of the amino acid sequence. For both organisms, the PDB code was 1O8B and was named “Structure of *Escherichia coli* ribose-5-phosphate isomerase, RpiA, complexed with arabinose-5-phosphate.” The alignment length was 219 for b2914 and 222 for Mrub_1349. The E-values were 1.08859E-123 and 2.68771E-24 for b2914 and Mrub_1349 respectively. Since both queries resulted in the same protein name, the structures are similar. The significant E-values also indicate an evolutionary relationship between the two structures. All structure-based data suggest that the similarities in b2914 and Mrub_1349 domain and 3-D structures are due to their evolutionary relationship.

The IMG/EDU Gene finder was used to find the ortholog neighborhood region of *E. coli* *rpiA* and Mrub_1349 with the color annotation of KEGG (Figure 14). The results for *rpiA* shows that *rpiA* (Panel A, underlined in red) is immediately next to the *serA* gene. Although these adjacent genes are aligned in the same direction, further research shows that the *serA* gene is promoted independently of the *rpiA* gene in *E. coli* (Yang et al. 2002). The *yqfE* gene that is also adjacent to the *rpiA* gene does not appear to be an operon with *rpiA* since there is no evidence that indicates that they are an operon. The *rpiA* gene is also able to be promoted independently of these adjacent genes (Hove-Jensen and Maigaard 1993). Therefore, *rpiA* is not likely to be part of an operon. On the other hand, the Mrub_1349 gene is surrounded by other adjacent genes. These adjacent genes do not appear to be equivalents to *serA* or the *yqfE* genes. Although these genes are indicated by KEGG to be involved in other pathways, further evidence would be necessary to indicate if Mrub_1349 is part of an operon. The ortholog neighborhood region of Mrub_1349 does not appear to resemble the ortholog neighborhood region of *rpiA*.

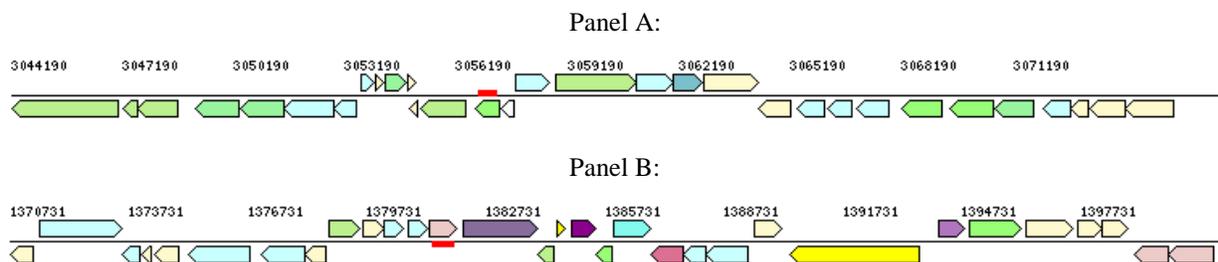


Figure 14. *E. coli* *rpiA* and Mrub_1349 are likely not part of an operon. Panel A is the visualization of the *E. coli* *rpiA* ortholog neighborhood and Panel B is the Mrub_1349 ortholog neighborhood. The red lines indicate the location of either *rpiA* (Panel A) or Mrub_1349 (Panel B). IMG/EDU (<http://img.jgi.doe.gov/cgi-bin/edu/main.cgi?section=FindGenes&page=geneSearch>) was used to illustrate the ortholog neighborhood regions.

***Mrub_1349* compared against b4090**

According to the KEGG pathway map for the “Pentose Phosphate Pathway” (Figure 2), Mrub_1349 and b4090 share the same E.C. number (5.3.1.6) (Table 3). Table 3 displays the similarities and differences between these two genes (Mrub_1349 and b4090) according to the various modules from the GENE-ACT lab notebook.

Table 3: *E. Coli rpiB* (b4090) and Mrub_1349 are not homologs

Description of evidence collected	<i>E. Coli</i>	<i>M. ruber</i>
Locus tag	b4090	Mrub_1349
KEGG pathway	Pathway ID: 00030 Pentose Phosphate Pathway	
E.C. number	5.3.1.6 ribose-5-phosphate isomerase	
BLAST <i>E. coli</i> against <i>M. ruber</i>	Length: 228 ; Score: 12.3 bits ; E-value: 5.7	
CDD (COG category)	COG0698 Ribose 5-phosphate isomerase RpiB [Carbohydrate transport and metabolism] E-value: 5.74e-70	COG0120 Ribose 5-phosphate isomerase [Carbohydrate transport and metabolism] E-value: 1.08e-88
Cellular localization (Module 3)	Cytoplasmic	
TIGRfam - protein family	TIGR01120 <i>rpiB</i> : ribose 5-phosphate isomerase B Score: 392.2 E-value: 1.1e-114	TIGR00021 <i>rpiB</i> : ribose 5-phosphate isomerase A Score: 349.6 E-value: 7.6e-102
Pfam - protein family	PF02502 Ribose/Galactose Isomerase Score: 182.6 E-value: 3.1e-54	PF06026 Ribose 5-phosphate isomerase A (phosphoriboisomerase A) Clan: ISOCOT_Fold (CL0246) Score: 192.5 E-value: 3.8e-57
PDB	1NN4 Structural Genomics, RpiB/AlsB Length: 149 E-value: 2.19079E-83	1O8B Structure of <i>Escherichia coli</i> ribose-5-phosphate isomerase, RpiB, complexed with arabinose-5-phosphate. Length: 222 E-value: 2.68771E-24

A BLAST search of both the amino acid sequence of Mrub_1349 in *M. ruber* and *rpiB* in *E. coli* independently did not display the same protein. Mrub_1349 showed ribose-5-phosphate isomerase A as a top-hit in other organisms, while the b4090 search showed ribose-5-phosphate isomerase B as a top-hit in other organisms. A BLAST search of *E. coli* against *M. ruber* resulted in an alignment length of 228. The bit score was 12.3 and the E-value was 5.7 (Figure 15). This low bit score and high E-value suggests that the similarities between *M. ruber* and *E. coli* are not likely due to an evolutionary relationship and likely due to chance.

Mrub_1349

Sequence ID: lcl|Query_6291 Length: 228 Number of Matches: 1

Range 1: 77 to 154 [Graphics](#)

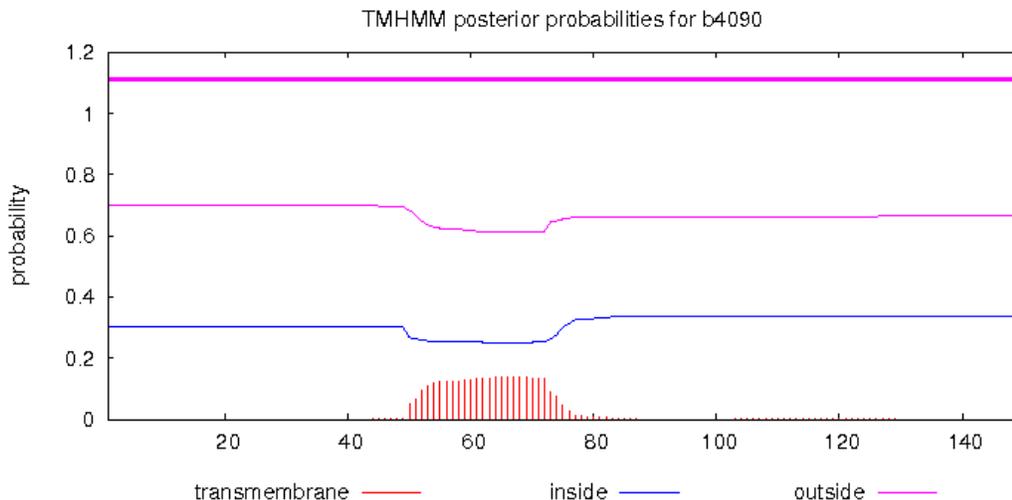
▼ Next Match ▲ Previous Match

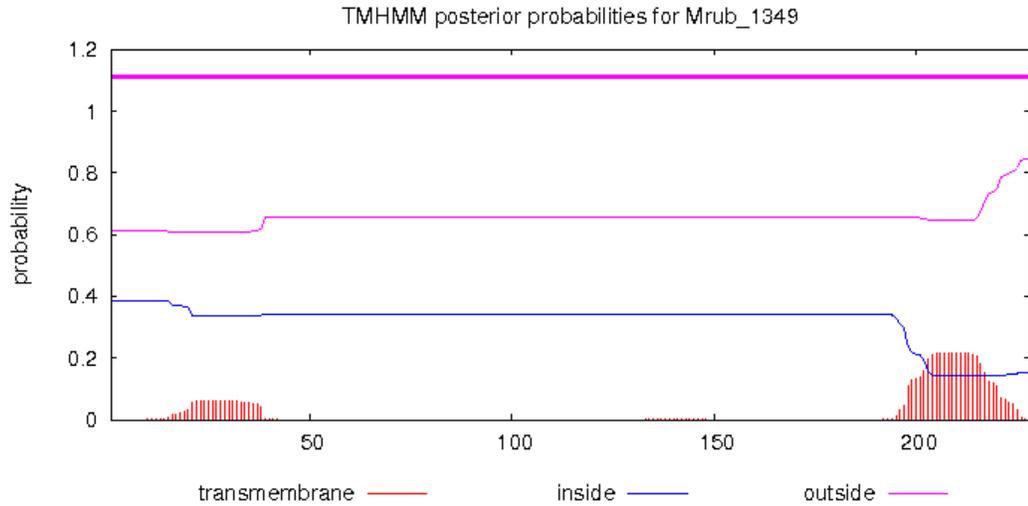
Score	Expect	Method	Identities	Positives	Gaps
12.3 bits(20)	5.7	Compositional matrix adjust.	20/78(26%)	30/78(38%)	5/78(6%)
Query 47	ASQVALAVAGGEVDGGILICGTGVGISIAANKF--AGIRAVVCSEPYSD---AQLSRQHND				101
	A V LA+ G + L G+G ++ K A R + +S QL R				
Sbjct 77	ARGVDLAIDGADEIAPDLSLIKGLGGALLREKIVEANARQFIVIADHSKKVQGLGRGVVP				136
Query 102	TNVLAFGSRVVGLELAKM				119
	++ FG R L +M				
Sbjct 137	VEIVRFGYRATLQALGRM				154

Figure 15. A protein BLAST search of the *rpiB* (b4090) of *E. coli* against Mrub_1349 of *M. ruber* indicates an unlikely evolutionary relationship. This alignment shows a length of 228. The bit score was 12.3 and the E-value was 5.7. The “Query” sequence is the sequence of *E. coli* and “Sbjct” displays the sequence from *M. ruber*. BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) was used for the protein blast.

The CDD search identified one conserved protein domain for each of the query sequences (b4090 and Mrub_1349). The Mrub_1349 sequence resulted in a COG number of COG0120, a COG name of “Ribose 5-phosphate isomerase [Carbohydrate transport and metabolism],” and a significant E-value (Table 3). The b4090 sequence resulted in a COG number of COG0698, a COG name of “Ribose 5-phosphate isomerase RpiB [Carbohydrate transport and metabolism],” and a significant E-value (Table 3). The E-value of the *E. coli* search was determined to be $5.74e-70$ while the *M. ruber* search was $1.08e-88$ (Table 3). Both the BLAST search and the CDD are indicative of that the RpiB amino acid sequence of *E. coli* and the Mrub_1349 amino acid sequence of *M. ruber* are not similar by sequence.

To determine the cellular localization of the protein in question, TMHMM was used (Figure 16). Since no significant peaks were found on the transmembrane topology graph for either *E. coli* or *M. ruber*, the proteins for both organisms do not have transmembrane helices. In addition, the lack of transmembrane helices suggests that this protein is localized to the cytoplasm because it does not pass through or insert into a membrane.

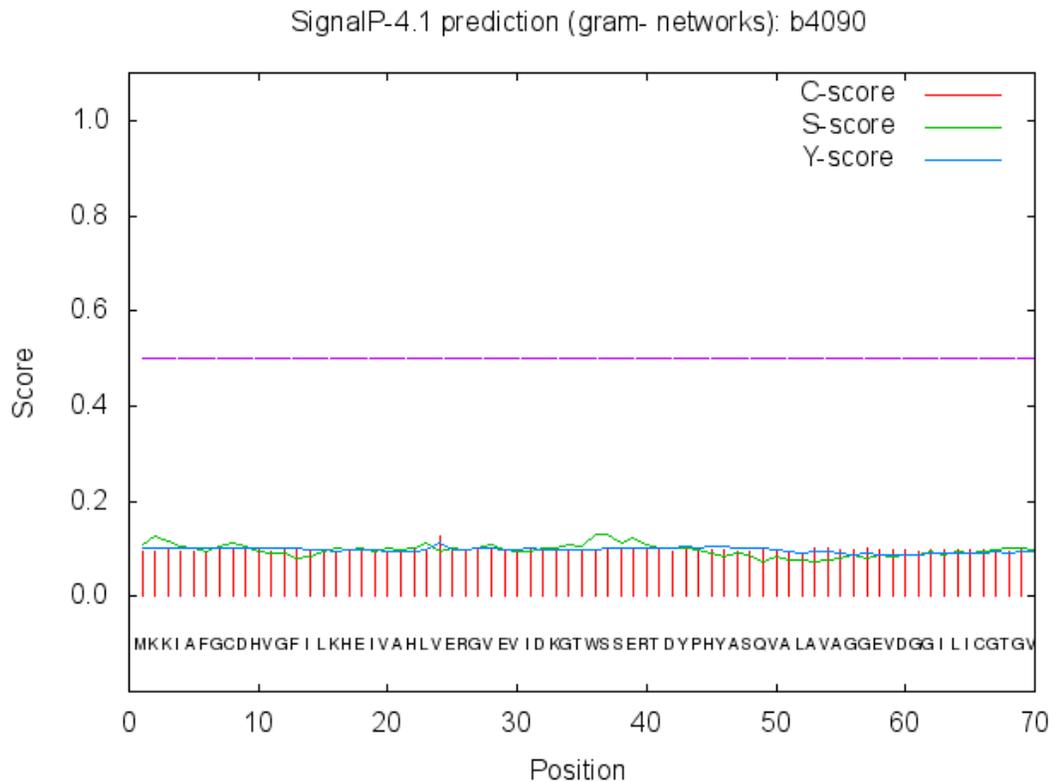




Panel B: — transmembrane — inside — outside

Figure 16. *E. coli rpiB* and Mrub_1349 do not contain transmembrane helix regions and is likely in the cytoplasm. Panel A indicates the *E. coli rpiB* query and Panel B indicates the Mrub_1349 query. TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) was used for the topology graph.

Signal peptides were predicted through SignalP (Figure 17). The signal peptide probability (D-score) of *E. coli rpiB* and Mrub_1349 was 0.106 and 0.170 respectively. Due to the lack of peaks produced from Figure 17, there are no signal peptides predicted for either protein (no likely cleavage sites). The D-scores of these queries are considered low since they are not close to 1 (cut off value of 0.570 for both proteins) and also indicate no signal peptides located at the N-terminus of the protein sequence.



Panel A:

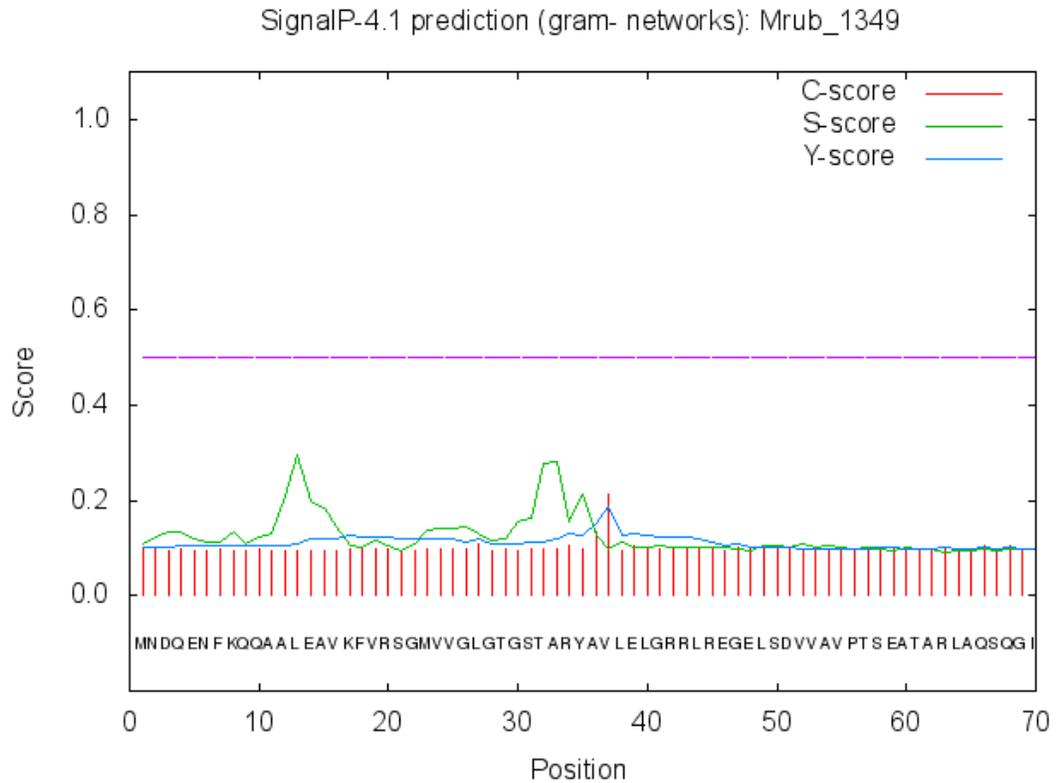
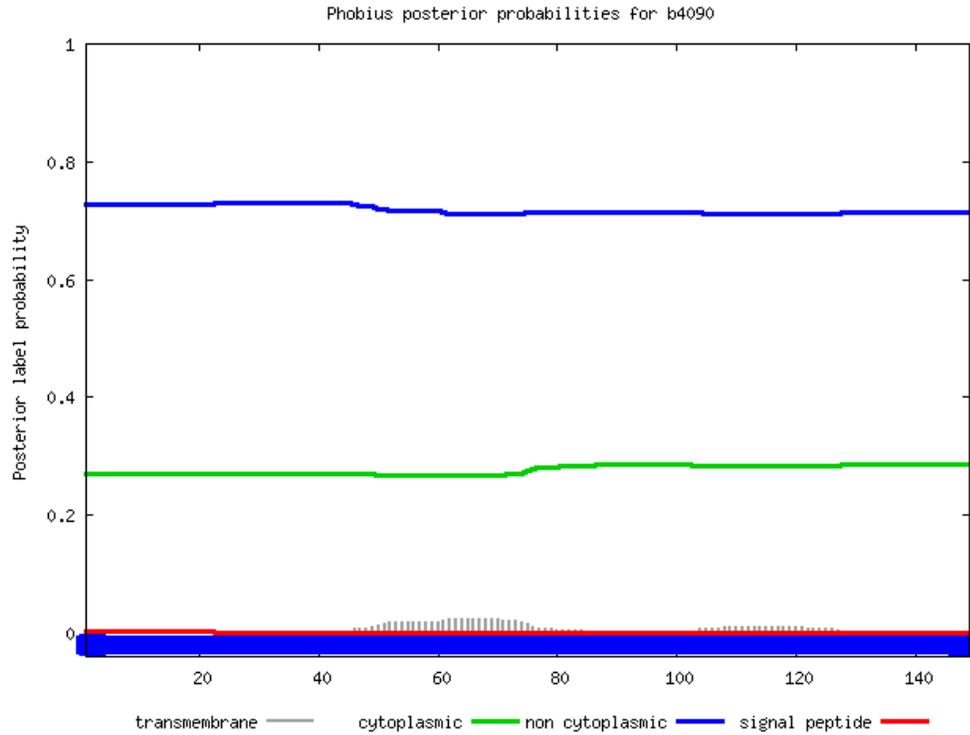
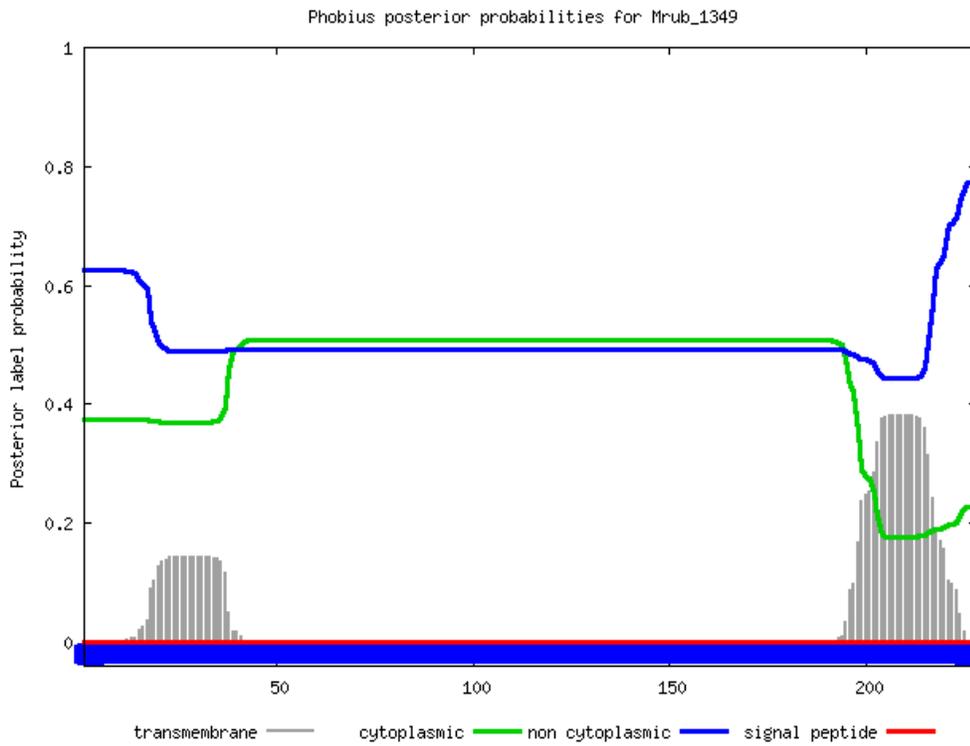


Figure 17. *E. coli rpiB* and Mrub_1349 are not predicted to have signal peptides and are likely to be cytoplasmic proteins. Panel A is *E. coli rpiB* and Panel B is Mrub_1349. SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP>) created this signal peptide graph.

The Mrub_1349 and b4090 sequences were also compared using the Phobius program. There are no significant vertical lines in gray that would indicate transmembrane helices (Figure 18). As stated previously, there are slight peaks (gray) at the N-terminus and C-terminus of Mrub_1349, but all other cellular location modules (TMHMM, SignalP, and PSORT-B) do not support a signal peptide for Mrub_1349. In addition, the peaks in the Mrub_1349 graph (Figure 18) do not surpass the non-cytoplasmic mark line (blue) which suggest that the peaks are insignificant.



Panel A:



Panel B:

Figure 18. Mrub_1349 and *E. coli rpiB* do not have transmembrane helices. Although there are two peaks by the N-terminus and C-terminus regions for Mrub_1349, they are insignificant and are still suggested to be cytoplasmic. Panel A is *E. coli* b4090 and Panel B is Mrub_1349. Phobius (<http://phobius.sbc.su.se/>) created these graphs.

Two more tools are used to determine the cellular localization of the protein query. LipoP suggested cytoplasm (Table 3) for both genes. However, the PSORT-B tool predicted a cytoplasmic score of 8.96 for Mrub_1349 with a CytoplasmicMembrane score of 0.51, a Periplasmic score of 0.26, an OuterMembrane score of 0.01, and an Extracellular score of 0.26. PSORT-B was inconclusive for *E. coli rpiB*. Since the final prediction of the Mrub_1349 cellular location was only cytoplasm, the scores for CytoplasmicMembrane, Periplasmic, OuterMembrane, and Extracellular are not significant. According to these tools, ribose-5-phosphate isomerase for *E. coli* and *M. ruber* is predicted to be in the cytoplasm which is reflective of the EcoCyc's prediction for the cellular location of ribose-5-phosphate isomerase.

Several tools were used to determine the structural similarities between *rpiB* and Mrub_1349. TIGRFAM, which determines protein homology, resulted in different TIGRFAM numbers: TIGR01120 for b4090 and TIGR00021 for Mrub_1349 (Table 3). The family name for this TIGRFAM number is “*rpiB*: ribose 5-phosphate isomerase B” for b4090 and “*rpiB*: ribose 5-phosphate isomerase A” for Mrub_1349. Both *E. coli* and *M. ruber* searches produced significant scores and E-values: the scores for *rpiB* and Mrub_1349 were 392.2 and 349.6 respectively, while the E-values for *rpiB* and Mrub_1349 were 1.1e-114 and 7.6e-102 respectively (Table 3). A Pfam search resulted in the different numbers and names for b4090 (PF02502, Ribose/Galactose Isomerase) and Mrub_1349 (PF06026, Ribose 5-phosphate isomerase A (phosphoriboisomerase A)) (Table 3). A pairwise alignment of both *E. coli rpiB* and Mrub_1349 against the conserved sequence of ribose-5-phosphate isomerase displays how well the amino acid sequences relate to the consensus sequence (Figure 19). The scores for *rpiB* and Mrub_1349 were 182.6 and 192.5 respectively, while the E-values for *rpiB* and Mrub_1349 were 3.1e-54 and 3.8e-57 respectively (Table 3). Key functional groups for *E. coli rpiB* were determined to be D9, L23, Y43, C66, G69, G71, A93, G108; while the key functional groups for *M. ruber* Mrub_1349 were determined to be S59, D86, G87, D89, K99, G100, G102, E108, K109, K126, P137, R162, D180, G202, V203, G207, F209. These sequences do not share the same key functional residues which indicate different protein structures.



Figure 19. *E. coli rpiB* and Mrub_1349 do not have similar domain structures according to the consensus sequence of ribose-5-phosphate isomerase. Panel A is the *E. coli rpiB* query (#SEQ) against the consensus sequence (#HMM). Panel B is the Mrub_1349 query (#SEQ) against the consensus sequence (#HMM). Pfam (<http://pfam.sanger.ac.uk/search>) was used to create the pairwise alignment.

In addition to TIGRFAM and Pfam, PDB suggested the differences in the 3-D structure of the amino acid sequence. For Mrub_1349, the PDB code was 1O8B and was named “Structure of

Escherichia coli ribose-5-phosphate isomerase, RpiB, complexed with arabinose-5-phosphate.” For b4090, the PDB code was 1NN4 and was named “Structural Genomics, RpiB/AlsB.” The alignment length was 149 for b4090 and 222 for Mrub_1349. The E-values were 2.19079E-83 and 2.68771E-24 for b4090 and Mrub_1349 respectively. Since the queries resulted in the different protein name, the structures are not similar. The significant E-values indicate an evolutionary relationship to the suggested PDB structure but they are not comparable between b4090 and Mrub_1349 since they produced different structures.

The IMG/EDU Gene finder was used to find the ortholog neighborhood region of *E. coli* *rpiB* and Mrub_1349 with the color annotation of KEGG (Figure 20). The results for *rpiB* shows that *rpiB* (Panel A, underlined in red) is immediately next to the *yjdP* gene. Although these adjacent genes are aligned in the same direction, further research shows that the *rpiB* gene is able to be transcribed independently of this adjacent gene (Sorensen and Hove-Jensen 1996). Therefore, *rpiB* is not likely to be part of an operon. On the other hand, the Mrub_1349 gene is surrounded by other adjacent genes. These adjacent genes do not appear to be equivalents to the *yjdP* gene. Although these genes are indicated by KEGG to be involved in other pathways, further evidence would be necessary to indicate if Mrub_1349 is part of an operon. The ortholog neighborhood region of Mrub_1349 does not appear to resemble the ortholog neighborhood region of *rpiB*.

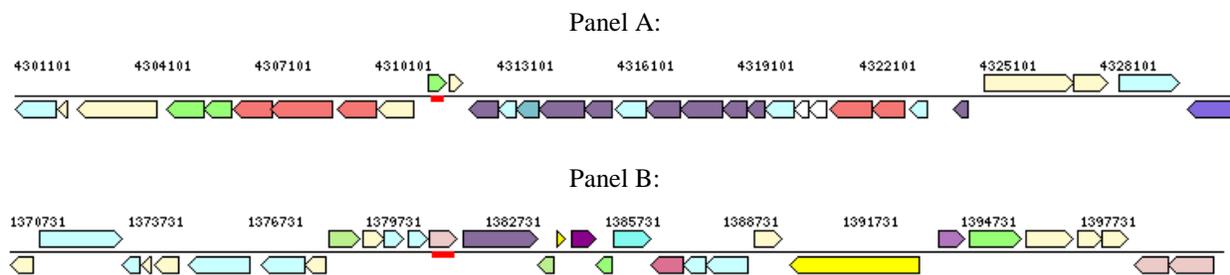


Figure 20. *E. coli* *rpiB* and Mrub_1349 are likely not part of an operon. Panel A is the visualization of the *E. coli* *rpiB* ortholog neighborhood and Panel B is the Mrub_1349 ortholog neighborhood. The red lines indicate the location of either *rpiB* (Panel A) or Mrub_1349 (Panel B). IMG/EDU (<http://img.jgi.doe.gov/cgi-bin/edu/main.cgi?section=FindGenes&page=geneSearch>) was used to illustrate the ortholog neighborhood regions.

Conclusion:

All tools included in the comparison of *rpe* and Mrub_2874 (Table 1) or *rpiA* and Mrub_1349 (Table 2) indicate similarities in sequence, evolutionary relationships, identical cellular locations, and structure. However, most tools included in the comparison of *rpiB* and Mrub_1349 (Table 3) indicate significant differences in sequence, evolutionary relationships, and structure. The similarities within the ortholog neighborhood regions also indicate similarities in function through operons for *rpe* and Mrub_1349. The *rpe* and Mrub_2874 sequences also shared significant similarities with the consensus sequence of ribulose-5-phosphate 3-epimerase, the protein from the *rpe* gene. The RpiA and Mrub_1349 sequences shared significant similarities with the consensus sequence of ribose-5-phosphate isomerase, the protein from the *rpiA* gene. However, the RpiB and Mrub_1349 sequences did not share significant similarities with the consensus sequence. Most of these bioinformatics tools indicate an evolutionary relationship due to the consistently low E-values for each comparison. As a result of these

findings, Mrub_2874 is the equivalent of b3386 (*rpe*) and Mrub_1349 is the equivalent of b2914 (*rpiA*), but Mrub_1349 is not the equivalent of b4090 (*rpiB*).

References:

- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* 5(10):R80.
- Hove-Jensen B and Maigaard M. 1993. *Escherichia coli rpiA* gene encoding ribose phosphate isomerase A. *J Bacteriol* 175(17):5628-35.
- Kruger NJ and von Schaewen A. 2003. The oxidative pentose phosphate pathway: Structure and organization. *Curr Opin Plant Biol* 6(3):236-46.
- Loginova L, Egorova L, Golovacheva R, Seregina L. 1984. *Thermus ruber* sp. nov., nom. rev. *Int J Syst Evol Microbiol* 34(4):498-9.
- Lyngstadaas A, Løbner-Olesen A, Grelland E, Boye E. 1999. The gene for 2-phosphoglycolate phosphatase (*gph*) in *Escherichia coli* is located in the same operon as *dam* and at least five other diverse genes. *Biochimica Et Biophysica Acta (BBA)-General Subjects* 1472(1):376-84.
- Lyngstadaas A, Sprenger GA, Boye E. 1998. Impaired growth of an *Escherichia coli rpe* mutant lacking ribulose-5-phosphate epimerase activity. *Biochimica Et Biophysica Acta (BBA)-General Subjects* 1381(3):319-30
- Miles B. 2003. Pentose Phosphate Pathway Aka the Hexose Monophosphate Shunt. [Internet]. Texas A&M University; [cited 2016 Feb 28]. Available from <https://www.tamu.edu/faculty/bmiles/lectures/Pentose%20Phosphate%20Pathway.pdf>
- Phylogenetic Diversity [Internet]. California: Joint Genome Institute United States Department of Energy; c2015 [cited 2016 Feb 28]. Available from: <http://jgi.doe.gov/our-science/science-programs/microbial-genomics/phylogenetic-diversity/>.
- Skinner A and Cooper R. 1971. The regulation of ribose-5-phosphate isomerization in *Escherichia coli* K12. *FEBS Lett* 12(5):293-6
- Sorensen KI and Hove-Jensen B. 1996. Ribose catabolism of *Escherichia coli*: Characterization of the *rpiB* gene encoding ribose phosphate isomerase B and of the *rpiR* gene, which is involved in regulation of *rpiB* expression. *J Bacteriol* 178(4):1003-11.
- Yang L, Lin RT, Newman EB. 2002. Structure of the Lrp-regulated *serA* promoter of *Escherichia coli* K-12. *Mol Microbiol* 43(2):323-33.