2-2016

Comparing *Meiothermus ruber* and *Myxococcus xanthus* in the Purine Metabolism Pathway

Linnea J. Ritchie
Augustana College - Rock Island

Dr. Lori Scott
Augustana College, Rock Island Illinois

Comparing *Meiothermus ruber* and *Myxococcus xanthus* in the Purine Metabolism Pathway

Linnea Ritchie

Bio-375 Molecular Genetics (Dr. Lori Scott)

Background

The purine metabolism pathway is an essential part of an organism's ability to make nucleotides. It is through this pathway that adenine and guanine are made, these molecules later become the bases of nucleotides, which are a key component in DNA (Westby 1974). There are two different routes for purine synthesis: the de novo pathway and the salvage pathway (Berg 2002). During the de novo pathway the purine molecules are essentially built from scratch. While this route uses comparatively simple molecules and amino acids there is a high energy requirement which is why at times the salvage pathway is used instead. While the de novo pathway requires hydrolysis of ATP or GTP on 5 of 12 steps of the pathway, the salvage pathway only requires energy input on one step (Berg et al. 2002). The salvage pathway takes free purine bases either from degraded RNA or DNA and recycles them to make new RNA or DNA. These free purine bases are broken down into free guanine monophosphate (GMP) or adenosine monophosphate (AMP) (Koonin 2003). This allows the system to save resources. If more nucleic acids than needed are present in the system they can be degraded and excreted.

While the purine metabolism pathway as a whole is fascinating, the gene sequences I am concerned with are involved in the first two steps of the process, before the system differentiates between de novo and salvage pathways. As shown in Figure 1, the Purine metabolism pathway begins with Ribose-5-phosphate which is obtained from the pentose phosphate pathway (Berg et al. 2002). Ribose-5-phosphate is then activated into Phosphoribose-1-pyrophosphate (PRPP) via

the combination of ATP and PRPP synthetase (PRS). It is within PRS that we find four of the six gene sequences of interest. Mxan_7156, Mrub_1053, Mxan_5075, Mrub_2281 all code for (or are expected to code for) PRS (Kanehisa 2000).

In the next step of the purine metabolism pathway the PRPP interacts with phosphoribosyl pyrophosphate (PRPP) amidotransferase as well as Glutamine and ATP to form 5-phosphoribosylamine (PRA) (Figure 1). I have been attempting to prove via bioinformatics that the *Meiothermus ruber* gene Mrub_2299 does in fact code for PRPP amidotransferase as does Mxan_1103 (Westby 1974). The purine pathway will take either the de novo synthesis or salvage synthesis pathway after this step depending on the abundance of PRPP. This step's dependence on the concentration of PRPP makes it the rate determining step in the reaction. This paper will focus primarily on the first two steps of the purine synthesis pathway, before a distinction is made between de novo and salvage.

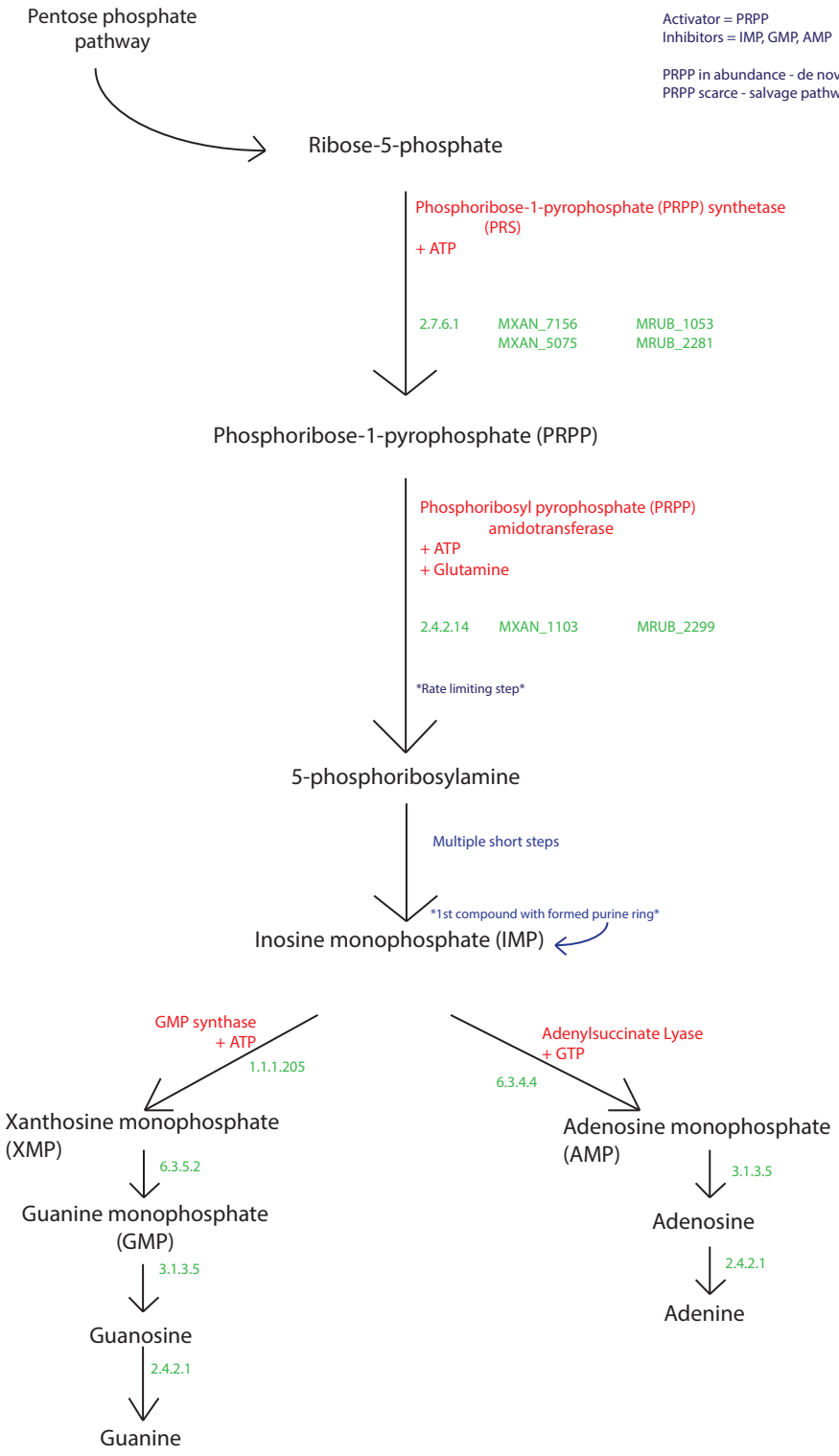


Figure 1: Purine Metabolism Pathway as it Relates to *M. xanthus* and *M. ruber*

In researching the relationship between these two bacteria I encountered problems finding a useable visual representation of the pathway. Most figures were either much to complicated for the purpose of this paper or they did not progress past Inosine monophosphate. I decided to create my own figure to show the pathway and the location of the gene sequences in question.

As shown in Figure 1, PRPP amidotransferase is the enzyme that catalyzes PRPP into 5-phosphoribosylamine (PRA). After the synthesis of PRA in the de novo pathway seven smaller steps build the purine ring directly on the ribose, which is annotated as 'ribotde' in Figure 2 (Berg et al. 2002). As explained earlier some steps in this pathway require energy input in order to perform their duty and others do not. The end product shown in Figure 2 has a completely formed purine ring and is called Inosine monophosphate (IMP). From this step the pathway branches off to form either guanine or adenine (Figure 1).

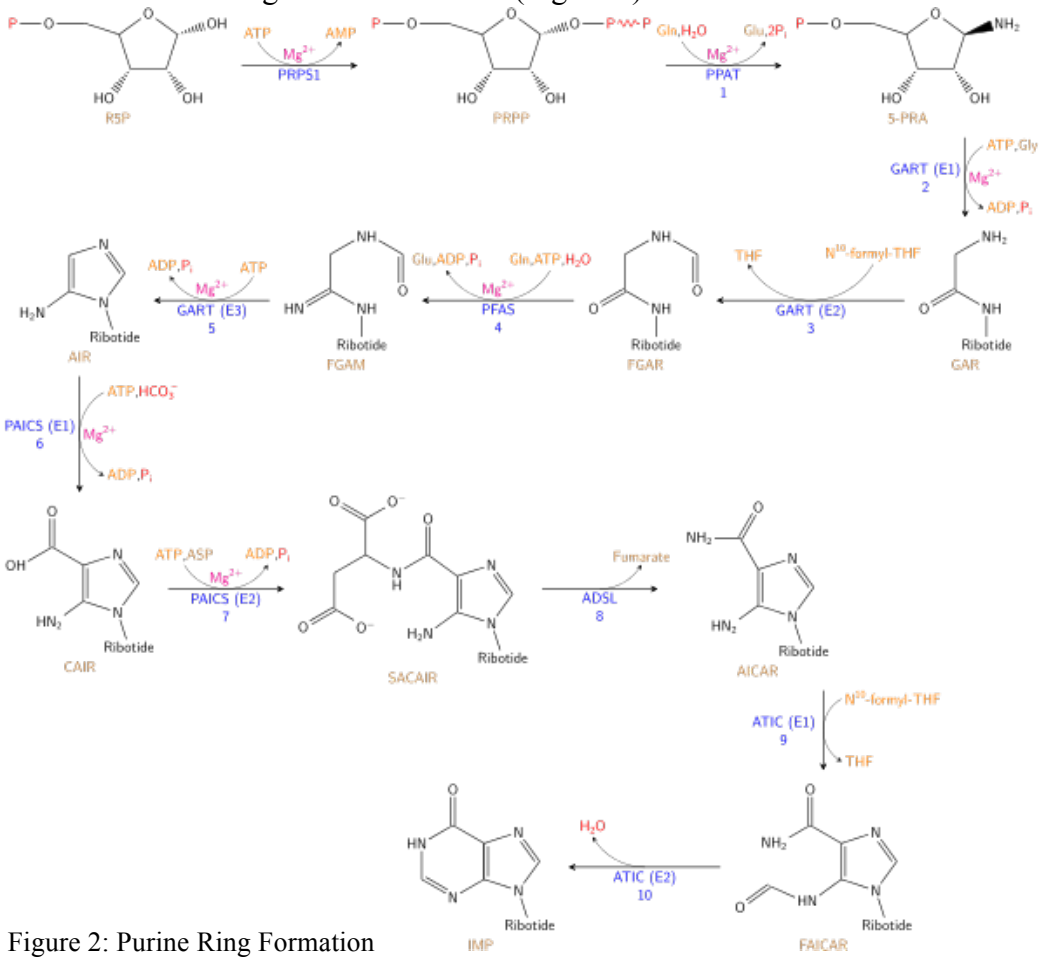


Figure 2: Purine Ring Formation

This figure shows a more chemical explanation of the events leading up to the production of Inosine monophosphate (IMP). Each new component of the purine ring (starting with step 2) is added directly to the ribose. The resulting IMP molecule has a fully formed purine ring and will differentiate into either adenine or guanine (Image from: https://upload.wikimedia.org/wikipedia/commons/thumb/2/25/Nucleotides_syn1.svg/600px-Nucleotides_syn1.svg.png)

My decision to use the bacteria *Myxococcus xanthus* as my reference bacteria in this project rose in part from my own knowledge of the organism having worked with it extensively for over a year. *M. xanthus* purine metabolism has also been studied extensively as myxospore formation can sometimes be triggered by the start or finish of some reactions in purine metabolism (Westby et al. 1974). In 1974 Westby and Tsai monitored purine metabolism in well nourished *M. xanthus* cells as well as metabolically dormant cells which were exposed to starvation conditions. Not only did this study confirm the locus tags of the *M. xanthus* gene sequences involved in purine synthesis but it also confirmed that the purine metabolism is constantly functioning, even when cells are dormant due to starvation. Because myxospores and their subsequent biofilms are so heavily studied there are many experiments explaining the function of the gene sequences in question (Westby 1974, Westby 1978, Hanson 1974).

Mrub_2299, Mxan_1103, Mrub_2281, Mxan_5075, Mrub_1053, and Mxan_7156 have all been identified by KEGG as being part of the same pathway. Their presence on the same steps of the same pathway was a good preliminary indicator of similarity. Using multiple bioinformatics tools, I have been able to confirm the similarity between these gene sequences not only in their amino acid structure but also in their function within the purine synthesis pathway.

I focused on the purine metabolism pathway because I am very interested in DNA in general and I wanted to know more about the process of building it. I was also curious if, since the structure of DNA remains fairly similar through most organisms, the process for acquiring the many building blocks would be different. In searching for pathways shared by *M. xanthus* and *M. ruber* I found that they are in many ways very different bacteria. *M. xanthus* is an extremely mobile bacterium while *M. ruber* is mostly sedentary. I found that while they seem to perform similar tasks and have similar pathways they often use different routes to reach the same

product. The purine pathway caught my attention partially because it was one of few pathways where they seemed to be somewhat similar.

An extensive knowledge of bioinformatics tools was necessary in order to complete my analysis. Not only did this speed up the process of comparing multiple sequences but it provided visual representation of the data which facilitated analysis. As stated earlier I focused on gene sequences associated with the first two steps (shown in Figure 1) of the purine metabolism pathway. Mxan_7156 and Mxan_5075 have been shown to code for the enzyme PRS (PRPP synthetase) and I compared these sequences to Mrub_1053 and Mrub_2281 in hopes of finding a similar function (Westby 1978). Mxan_1103 has been proven to code for PRPP amidotransferase and I believe that Mrub_2299 does the same (Westby 1978).

Methods

For this project I used the bioinformatics tools found in the GENI-ACT system (which can be found here: http://www.geni-act.org/student/view_assignment/find/6fa2ae446a0244ad/88178c02e87e4060/). I found very few instances where I was required to deviate from the instructions given on the website. Before I began, I used NCBI BLAST technology to confirm that there was a comparable gene sequence present in each genome. Having assured myself that I was indeed looking at gene sequences which had the potential to have similar functions I began characterizing each gene in turn.

I deviated slightly from the instructions while using T-Coffee (Tree-based Consistency Objective Function For Alignment Evaluation). Here the instructions simply instruct for ten sequences to be aligned to highlight conserved gene sequences. For consistency's sake I decided to make six genes in each *M. ruber* sequence set from the same genus (*Meiothermus*). This allowed me to keep a more consistent pool of gene sequences between annotations. Similarly, *M.*

xanthus was always compared to three other bacteria in the *Myxococcus* genus. The difference in numbers is simply because only three other *Myxococcus* bacteria were available on the NCBI BLAST database.

While Metacyc was useful during the beginning stages of my research I did not find it useful in looking for previous research indicating the function of the *M. xanthus* gene sequences in question. Since *M. xanthus* is not a model organism like E. coli websites like Metacyc do not focus on it. Because of this I instead used the NCBI database to look for previous research involving the purine metabolism pathway. Luckily there are studies regarding proline synthesis and its relation to sporulation and biofilms (Westby 1974 & 1978). So while I had to think outside of the box in order to indicate the function of the *M. xanthus* genes I was still able to prove that my model organism had the genes necessary for this project.

Results

Interestingly, there are two gene sequences in both the *M. ruber* and *M. xanthus* genome which code for PRPP synthetase. This could possibly indicate the occurrence of gene duplication. This potential duplication is interesting because it suggests that these sequences which are very similar to each other (see Table 1) perform different tasks and are present in different pathways. For instance, Mxan_5075 is also annotated as being involved in the pentose phosphate pathway (the product of which begins the purine metabolism pathway) and it is also involved in biosynthesis of secondary metabolites as well as carbon metabolism. My hypothesis of gene duplication is further supported by the fact that these genes are not located anywhere near each other on the genome. The numbers after their bacteria identifier (ex. Mxan_####) indicate their location on the genome. The sequences which I believe have been duplicated

(Mxan_7156 and Mxan_5075 as well as Mrub_1053 and Mrub_2288) are not near each other on the genome, which could indicate that they have more than one task within the bacteria.

I have several reasons to be confident in my assertion that the *Meiothermus ruber* gene sequences in question have functions similar to those of *Myxococcus xanthus* in the purine metabolism pathway. To begin, my preliminary BLAST searches where I compared the *M. xanthus* gene sequences against their (hopeful) counterparts in the *M. ruber* genome came back with very encouraging indicators. There appear to be many homologous protein-coding regions in *M. ruber* and *M. xanthus*. The highest E-value I encountered was 5 e (-63) (see Table 1) which is still very statistically significant. This indicates that these gene sequences have a similar amino acid sequence which is unlikely to be simply due to chance. In the case of the four sequences associated with EC number 2.7.6.1, when I compared sequences with the other gene sequence in the category (not the one they were paired with by KEGG) I got very similar data. I also compared Mxan_7156 & Mxan_5075 and Mrub_2281 & Mrub_1053. The *M. xanthus* genes matched with a score of 188 bits and an E-value of 4 e (-61) and the *M. ruber* genes matched with a score of 239 bits with an E-value of 2 e (-80). This would seem to further support my gene duplication theory. All of these gene sequences which code for the same enzyme seem to be very similar to each other.

Table 1: Evidence indicating similarity

Description of evidence collected	Mrub_2281	Mxan_5075	Mrub_1053	Mxan_7156	Mrub_2299	Mxan_1103
Cellular Localization	CYTOPLASM					
NCBI BLAST*	Score: 247 bits(630) E-Value:9 e (-84)		Score: 226 bits(575) E-Value: 2 e (-75)		Score: 427 bits(1090) E-Value: 2 e (-149)	
	See Mxan_7156 column	Score: 272 bits(489) E-Value:2 e (-93)		Score: 192 bits(489) E-Value: 5 e (-63) **Mxan_7156 vs Mrub_2281**		
KEGG Pathway	PURINE METABOLISM PATHWAY					
Pfam – protein family	PF13793 N-terminal domain of ribose phosphate pyrophosphokinase				PF13537 Glutamine amidotransferase domain	
	E-Value: 2.1 e (-40)	E-Value: 1.2 e (-51)	E-Value: 1 e (-45)	E-Value: 1.9 e (-31)	E-Value: 2.4 e (-22)	E-Value: 9.8 e (-24)
CDD (COG category)	COG0462 phosphoribosylpyrophosphate synthetase				COG0503 Adenine/guanine phosphoribosyltransferase	
TIGRfam – protein family	TIGR1251 Ribose-phosphate diphosphokinase				TIGR01134 amidophosphoribosyltransferase	
	E-Value: 4.4 e (-117)	E-Value: 6.3 e (-131)	E-Value: 1.2 e (-105)	E-Value: 1.3 e (-79)	E-Value: 6.8 e (-240)	E-Value: 1.2 e (-219)
E.C. number	2.7.6.1				2.4.2.14	
PDB	1DKR Crystal structures of *Bacillus subtilis* Phosphoribosylpyrophosphate synthetase				1ECB *Escherichia coli* Glutamine phosphoribosylpyrophosphate (PRPP) amidotransferase	
	E-Value: 1.04586 e (-77)	E-Value: 2.18684 e (-86)	E-Value: 6.96043 e (-73)	E-Value: 5.42333 e (-51)	E-Value: 1.47261 e (-74)	E-Value: 7.43275 e (-82)

* All BLAST results are the result of the amino acids of the two indicated gene sequences being compared (Ex. Mrub_2299 compared to Mxan_1103 and Mrub_2281 compared to Mxan_5075)

Table 1 also shows the results of my Conserved Domain Database (CDD) search which placed both Mrub_2299 and Mxan_1103 in the same category; Adenine/guanine phosphoribosyltransferase. The database does stipulate that they could also code for another PRPP binding protein, but since both the *M. ruber* and *M. xanthus* gene sequences were placed in this category I still believe that this information supports my hypothesis that they perform the same step in the purine synthesis pathway within their respective organisms. This search also positioned both sequences within the same cluster of orthologous groups (COG0503). Mrub_2281, Mrub_1053, Mxan_5075 and Mxan_7156 were also placed in the phosphoribosylpyrophosphate synthetase category (which was the expected result based on Figure 1). These four sequences were also all placed in COG0462. This means that all genes are likely to belong in a similar set of orthologs to those that were aligned to build the COG model (Marchler-Bauer et al. 2011). This is further proof that my hypothesis is correct and Mrub_2299 and Mxan_1103 are orthologs, as well as Mrub_2281 and Mxan_5075 & Mrub_1053 and Mxan_7156.

Furthermore, the Pfam database indicates that these six gene sequences (as paired above in Table 1) have the same domain; N-terminal domain of ribose phosphate pyrophosphokinase and Glutamine amidotransferase domain respectively. Their pairwise alignments (Figure 3A and 3B) also indicate a significant amount of conserved amino acids between the compared sequences. Similarly, all three sets of genes were consistent in their TIGRfam designations. The gene sequences coding for PRPP synthetase (Mrub_2281, Mrub_1053, Mxan_5075 and Mxan_7156) were all placed within TIGR1251 (Ribose-phosphate diphosphokinase). Mxan_1103 and Mrub_2299 were both annotated as part of TIGR01134 (amidophosphoribosyltransferase). Both of these names are listed as alternative names for the

enzymes they code for (Kanehisa et al. 2015). Unlike Pfam, TIGRfam uses full-length protein sequences with well understood functions. While the data gained from both databases is the same, the algorithms and parameters used during the search are different. This lends even more credence to my hypothesis since these gene segments are being connected repeatedly and through different search techniques.

3. 1

```
#HMM      lkifagssskeLaekiakrlgvelgkakvsrfadgEikveieesvrgkdvfiiqslsnpvndnlmelllmidalrrasakritaViPylgysrqdrk.ksrepisaklvaklleaaG
#MATCH    l++f+g+s+  La++ia++lgv+lgk+ ++rfa++   v++e+s+r+ d+fiiqsl++pv+d+l+ell+mida++ +sa+r+taViPy+ y+r+d+k ++r +i+a+lva+l++aaG
#PP       589*********************************************************************************************999**************99
#SEQ       MVFSGQSNLPLAQEIADNLGVPLGKSITERFANDCLFVRYEQSLREGDIFIIQSLTPPVQDHLVELLMMIDAAKSSSAARVTAVIPYYSYARSDKKGAPRISIAARLVADLIQAAG
```

3. 2

```
#HMM      kifagssskeLaekiakrlgvelgkakvsrfadgEikveieesvrgkdvfiiqslsnpvndnlmelllmidalrrasakritaViPylgysrqdrk.ksrepisaklvaklleaaG
#MATCH    k+fag+s++ La++i+++l+ +lgka+v+rf+dgEi vei e+vrg+dvfi+qs+++p+nd+lmell+m+dal+rasa +itaViPy+gy+rqdrk ++r+pi+akl+a+lle+aG
#PP       89*********************************************************************************************************99
#SEQ       KVFAGNSNPGLAHRICEYLKRPLGKAEVGRFSDGEIHVEIGENVRGHDVFILQSTCPPANDHLMELLIMCDALKRASAGSITAVIPYYGYARQDRKvAPRTPITAKLIADLLESAG
```

3. 3

```
#HMM      kifagssskeLaekiakrlgvelgkakvsrfadgEikveieesvrgkdvfiiqslsnpvndnlmelllmidalrrasakritaViPylgysrqdrk.ksrepisaklvaklleaaG
#MATCH    k+f g++++ Lae++a+ lg++lgka+v+rf+dgE++v++ es+rg+dv+++qs+++pvnd+lmell+++da+rr+sa ri+aViPy+gy+rqd++ + repi+a+lva lle++G
#PP       89*********************************************************************************************777**************998
#SEQ       KLFCGNANRPLAEAVARALGISLGKATVERFPDGEVQVRLLESIRGDDVYLLQSTAPPVNDHLMELLVLADAARRSSAGRINAVIPYFGYARQDKQtQGREPITARLVAGLLEHVG
```

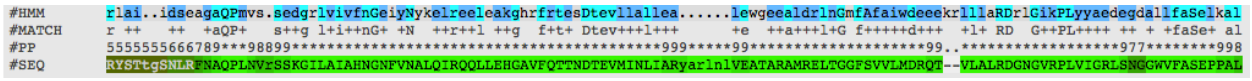
3. 4

```
#HMM      ifagssskeLaekiakrlgvelgkakvsrfadgEikveieesvrgkdvfiiqslsnpvndnlmelllmidalrrasakritaViPylgysrqdrk.ksrepisaklvaklleа
#MATCH    ++ g++s++L++++ + lgv  ++++ +rf+dgE+ ve+ e+vrg++v+++qs+++p+ ++l+elllm+da+ r  a+r  aV+Pylgy+rqdr+ ++ ep+  +lva+ll++
#PP       67899*********************************************************************************************99*************975
#SEQ       LVGTASPHLGRALGQALGVAPADCHFERFPDGEMHVEVPEQVRGRTVILVQSTTPPAGEHLLELLLMADACWRMGAARLEAVVPYLGYARQDRRaRPGEPLGGRLVADLLSQ
```

Figure 3A: Pairwise Alignment for PRPP synthetase shows similarity

3.1 = Mrub_2281, 3.2 = Mxan_5075, 3.3 = Mrub_1053, 3.4 = Mxan_7156. Each Pairwise alignment shows a high degree of similarity to ribose-phosphate diphosphokinase. Each panel shows the similarities in amino acid sequence between the gene in question (bottom line in green) and the selected HMM sequence (top line in blue), in this case to ribose-phosphate diphosphokinase. All of the E-Values in this case were significant, meaning that there is a strong possibility that these sequences are matches with Pfam13793 by more than random chance. Pfam (http://pfam.xfam.org/search) created these alignments.

3.5



3. 6

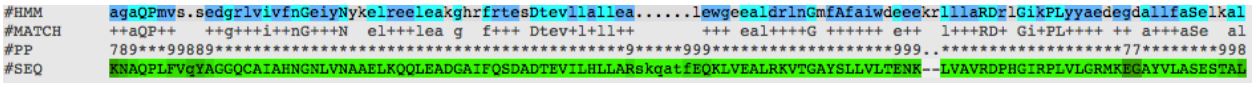


Figure 3B: Pairwise Alignment for PRPP amidotransferase confirms similarity

3.5 = Mrub_2299, 3.6 = Mxan_1103. Each pairwise alignment shows a high degree of similarity to amidophosphoribosyltransferase. As in Figure 3A, the green line indicates the gene sequence in question and the blue line above it represents the HMM sequence of the TIGRfam group. The large amount of conserved amino acids in conjunction with the significant E-Values indicates a strong probability of similar function.

My hypothesis was further supported by the KEGG (Kyoto Encyclopedia of Genes and Genomes) purine metabolism pathway (Figure 4). Since these genes were placed not only in the same pathway but assigned the same EC number (2.7.6.1 and 2.4.2.14 respectively) within that pathway this gave strong evidence for their shared functionality. Enzymes in different organisms which catalyze similar reactions are given the same EC number, since the number refers not to a specific enzyme but to the reaction catalyzed by the enzyme (Kanehisa et al. 2015). Their similarity was further supported when the bioinformatics tool ExPASy used the EC numbers to assign the same names to the enzymes as was expected at these two steps in the purine metabolism pathway (Figure 1).

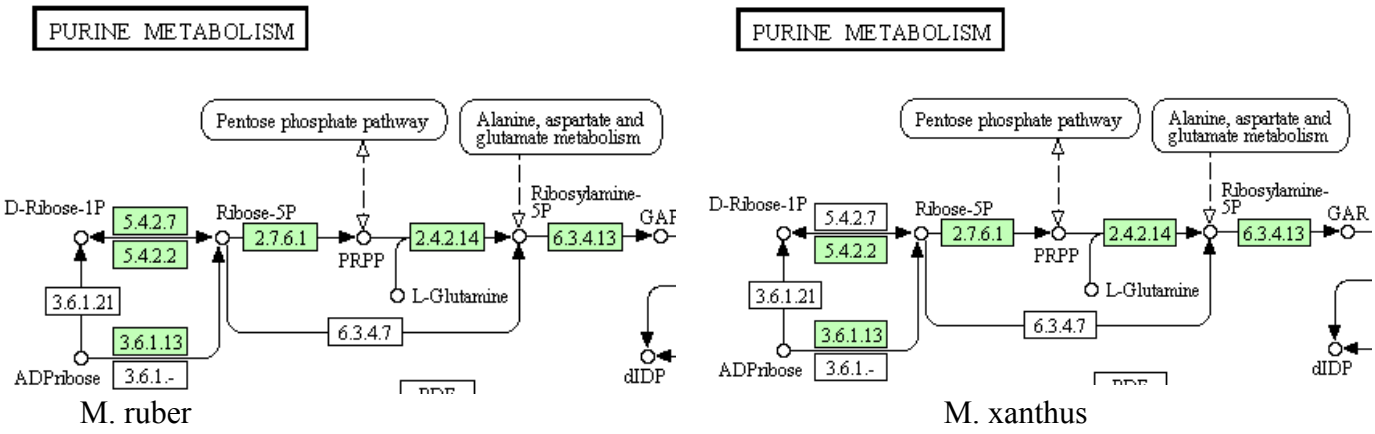


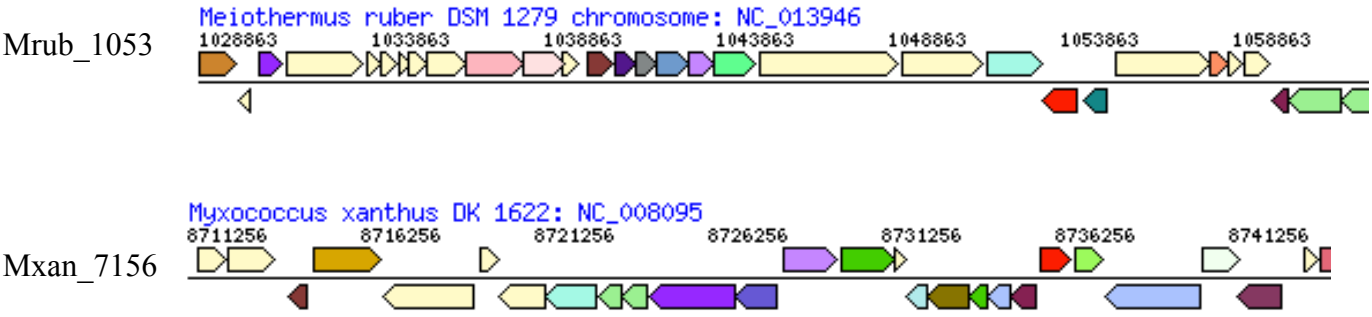
Figure 4: KEGG pathway indicates similarities between *M. ruber* and *M. xanthus*

This zoomed in shot of the purine metabolism pathway as provided by KEGG shows that both bacteria follow a similar pathway (with similar EC numbers) in the preliminary steps of the process of purine synthesis. While this similarity remains constant throughout the process only the first steps are shown because they related directly to the genes being annotated.

Using the Protein Data Bank (PDB) I searched for sequence-based similarity using all of the gene sequences in question. The benefit of utilizing this large bioinformatics tool is that PDB searches for matches to the query gene segment against gene sequences with solved structures. According to the PDB website, solved structures are protein sequences with at least 95% similarity grouped together and solved by multiple experimental methods such as X-Ray, NMR and EM (Berman et al. 2000). This database provides further evidence for the similarity between these gene sequences. Mrub_2299 and Mxan_1103 were placed in the group 1ECB (*Escherichia coli* Glutamine phosphoribosylpyrophosphate (PRPP) amidotransferase) and Mrub_2281, Mxan_5075, Mrub_1053 and Mxan_7156 were placed in the group 1DKR (Crystal structures of *Bacillus subtilis* Phosphoribosylpyrophosphate synthetase). Utilizing this database not only allowed me to further confirm each gene sequence's identity but it supported my hypothesis that the segments are evolutionarily related.

When the gene segments being considered in this paper are lined up in orthologous gene neighborhoods most seem to remain surrounded by the same gene segments, though these segments and their functions do not seem to be consistent between *M. ruber* and *M. xanthus* (Figure 5). While this would have further proven my hypothesis the lack of similar operons between the two does not rule out their being orthologs. These gene segments can still perform the same tasks while at different locations on their respective chromosomes. Interestingly, the *M. xanthus* genes of interest all seem to remain surrounded by the same genes, both within the *Myxococcus* genus and in unrelated species.

The fact that in most cases the gene order was maintained around the gene of interest is also more evidence of evolutionary relatedness and possibly horizontal gene transfer. Since these genes do not appear to have functions involved in the purine synthesis pathway it is safe to exclude the possibility of them forming an operon. The possibility of an operon cannot be so easily excluded in the case of *M. ruber*, indeed preliminary evidence suggests that Mrub_2281 and Mrub_2299 are indeed in an operon. Several genes between these two are annotated as being part of the purine synthesis pathway. The identity as an operon cannot be confirmed at this time because two genes in the potential operon are annotated as hypothetical, meaning that they may or may not be involved in the same pathway (Markowitz et al. 2014).



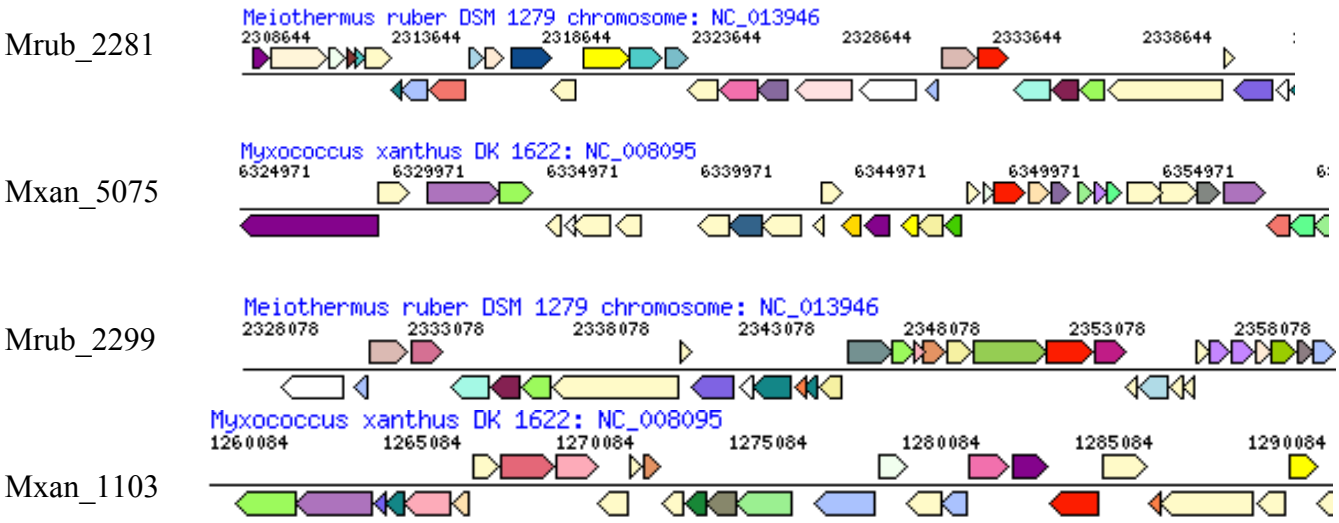
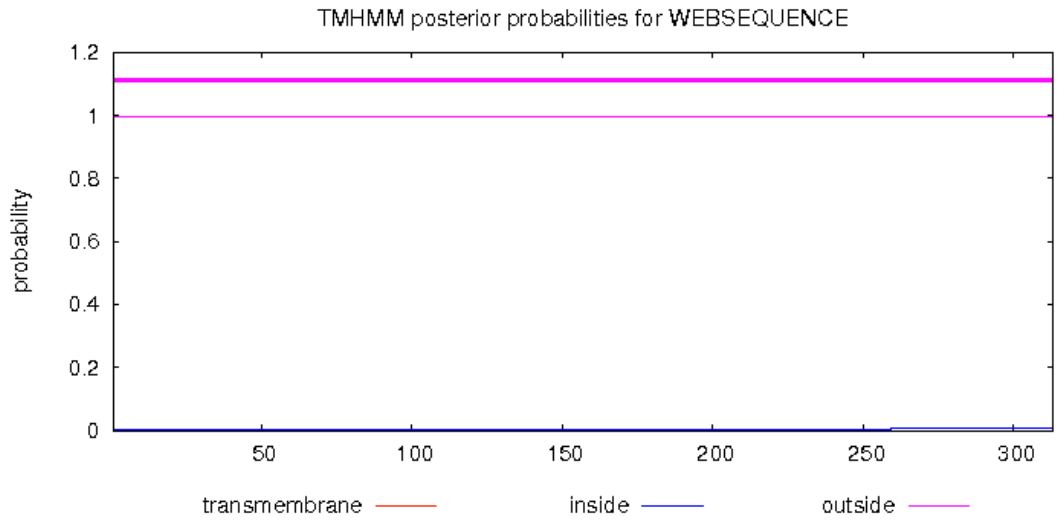


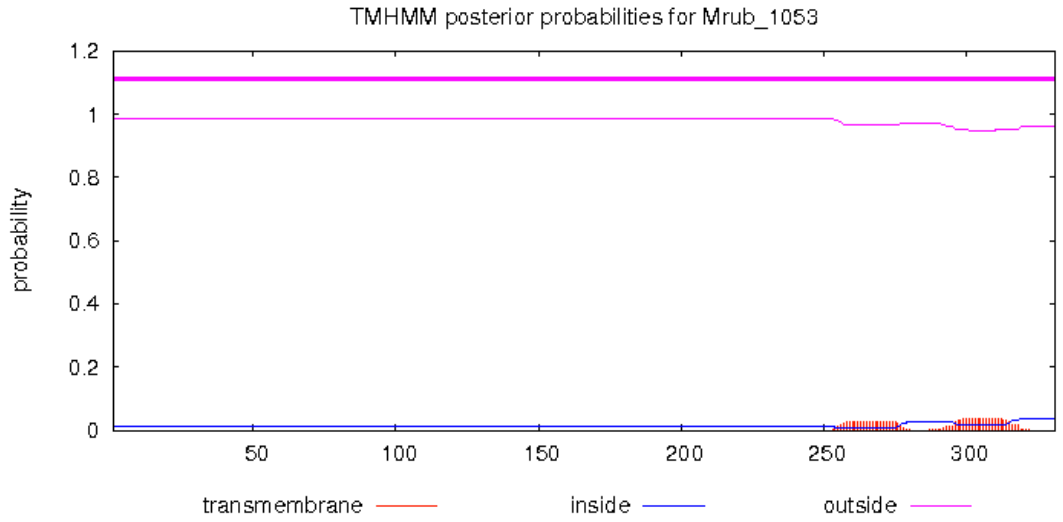
Figure 4: Orthologous gene neighborhoods reveal consistent gene placement
All gene sequences (Red) seemed to remain surrounded by the same gene sequences when compared to similar bacteria using JGI IMG/EDU. The one exception is Mxan_1103 which seems to be mostly solitary except in the case of *C. coralloides* which seems to have the same sequence as part of a group. The fact that the gene order is maintained across many different species presents the possibility that these theoretical operons are evolutionarily related.

Furthermore, none of the gene sequences in question indicated the presence of transmembrane helices. As shown below in the TMHMM topology graphs (Figure 6) while in some cases a slight amount of red is visible (indicating the possible presence of a transmembrane alpha helix) it remains below the level where it definitely can be identified as being transmembrane. This abnormality can be explained by large amounts of hydrophobic regions within the protein which can give the appearance of a transmembrane alpha helix.

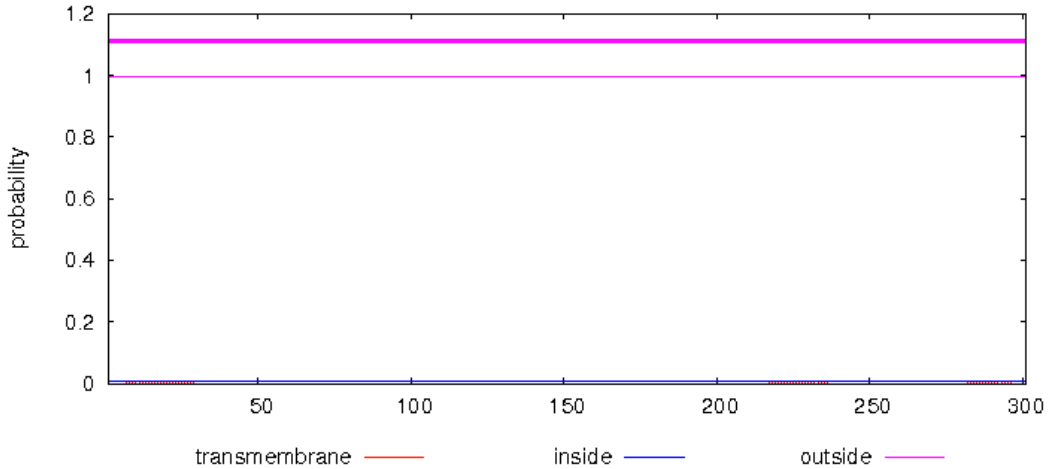
6.1



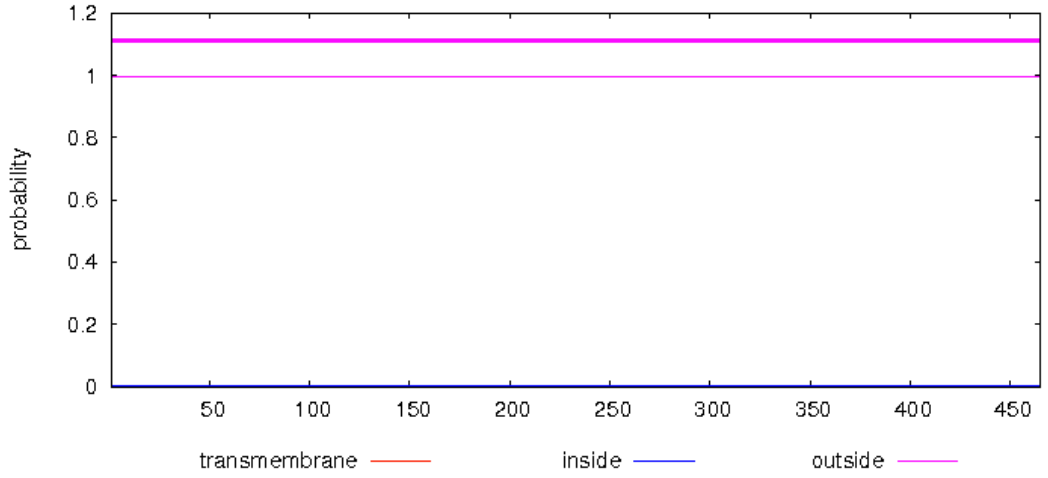
6.2


TMHMM posterior probabilities for Mrub_1053
probability
1.2
1
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300
transmembrane
inside
outside

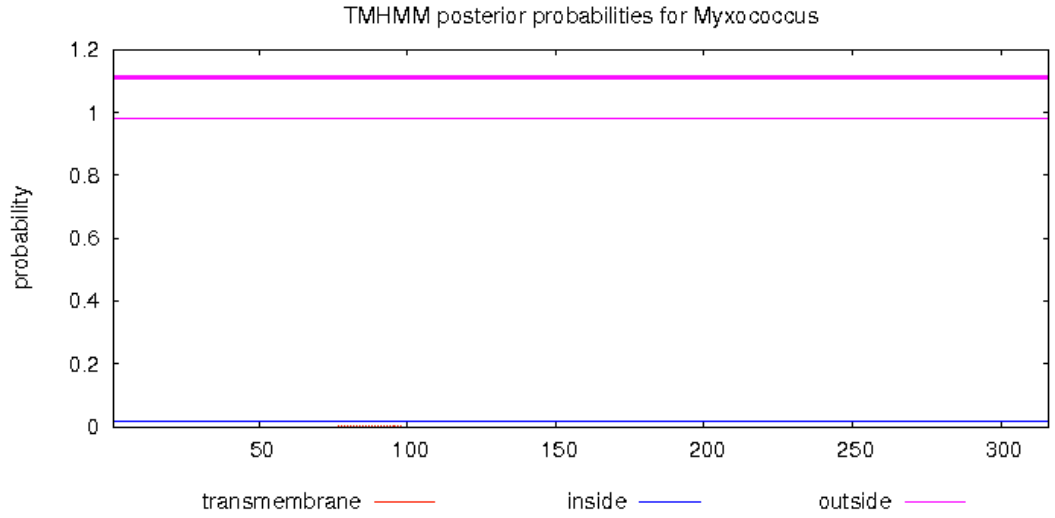
6.3


TMHMM posterior probabilities for Mxan_7156
probability
1.2
1
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300
transmembrane
inside
outside

6.4


TMHMM posterior probabilities for MRUB_2299
probability
1.2
1
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300
350
400
450
transmembrane
inside
outside

6.5



6.6

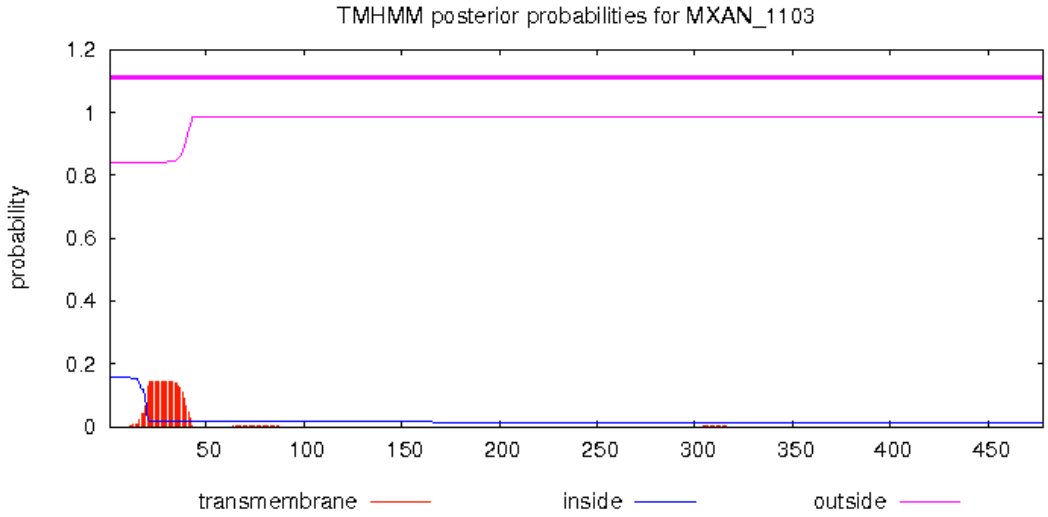


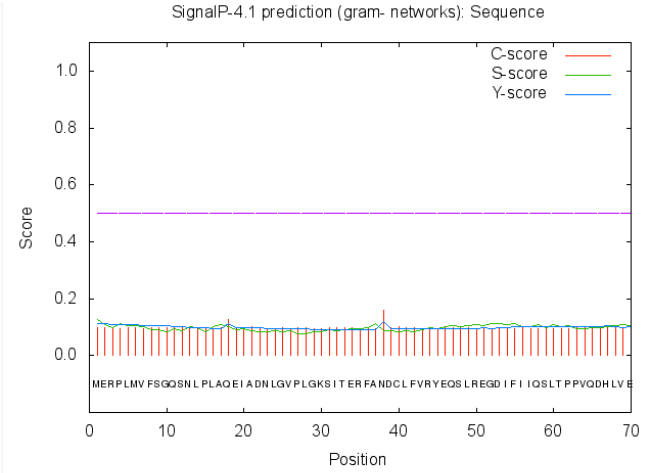
Figure 6: No transmembrane helices expected

6.1 = Mrub_2281, 6.2 = Mxan_5075, 6.3 = Mrub_1053, 6.4 = Mxan_7156, 6.5 = Mrub_2299, 6.6 = Mxan_1103. All amino acid sequences were compared to a database which specializes in the detection of transmembrane helices (TMHMM: http://www.cbs.dtu.dk/services/TMHMM/). The blue lines at the bottom of the graphs which are sometimes hard to see represent amino acids predicted to remain inside of the cell in cytoplasm. In graph 3.3 and 3.6 a red section of transmembrane proteins is depicted however since they do not reach the pink line across the top of the graph they are not considered to be transmembrane.

I used the bioinformatics tool LipoP to support the location of the gene segments within the cell. LipoP is used to predict the presence of any lipoproteins within the gene segment which were embedded in the membrane. This database is able to discriminate between lipoprotein signal peptides as well as other signal peptides and n-terminal membrane helices in gram-negative bacteria (Juncker et al. 2003). Using this resource, I got their best prediction for the location of my gene sequences (every single one was labeled as residing in the cytoplasm) as well as their predictions for any possible cleavage sites (none were predicted). Supporting LipoP's prediction of a cytoplasmic gene sequence, Psort-B predicted overwhelmingly that each gene sequence resided within the cytoplasm. Psort-B is able to predict a protein's location by searching the gene sequence for hydrophobic alpha helices which can be indicative of membrane bound regions (Yu et al. 2010). Each gene sequence was given a score which indicates the certainty of the gene segment's location. Most genes had a cytoplasmic score of 9.97 with no score dropping below 9.26. Psort also allowed me to confirm that, despite the slight presence of red in the TMHMM graph of Mxan_1103, there are no predicted regions of gene segment presence anywhere other than the cytoplasm (Mxan_1103 received a cytoplasmic score of 9.97).

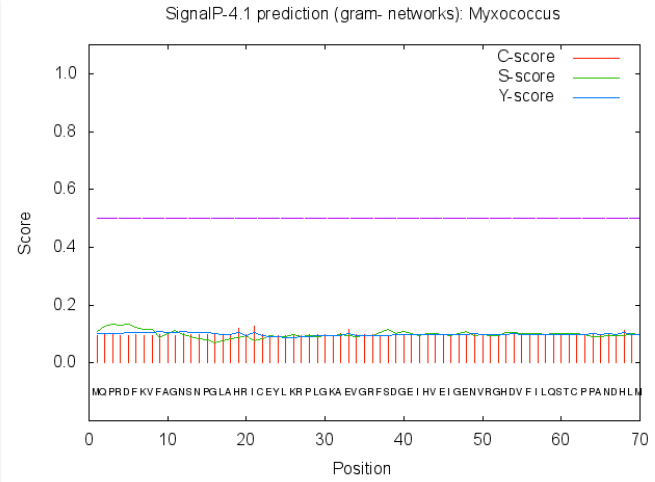
SignlP is used in a similar manner as LipoP to locate proteins whether inside or outside of the cell based on similar known peptides. These peptides, if present, direct the cell toward the cell membrane so that the segment can either adhere to the membrane or traverse it completely (SignalP). Since no gene segment annotated in this paper was predicted to go through the membrane I did not expect to see any areas of contact between the gene segment and the cell membrane, not any cleavage sites. As shown in Figure 7 no signal peptides or cleavage sites are predicted, although the Mxan_1103 graph does seem to indicate that a signal peptide (S-score) rises above the Y-score which is the geometric average of the raw cleavage score and the signal

peptide score (C and S scores). Just as with the TMHMM graphs in Figure 3 this very slight derivation from the expected result can again be explained by larger amounts of hydrophobic regions of DNA which could give the appearance of a signal peptide without actually traversing the membrane.



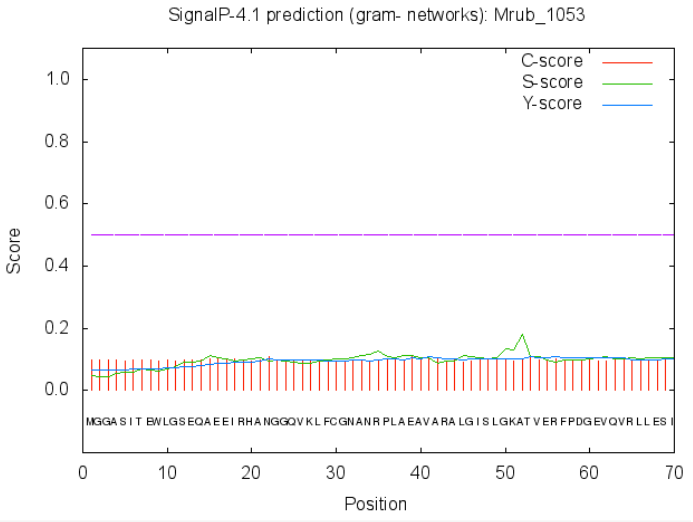

```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    38        0.160
  max. Y    38        0.120
  max. S     1        0.127
  mean S     1-37     0.094
       D     1-37     0.108   0.570   NO
Name=Sequence   SP='NO' D=0.108 D-cutoff=0.570 Networks=SignalP-noTM
```

Mrub_2281



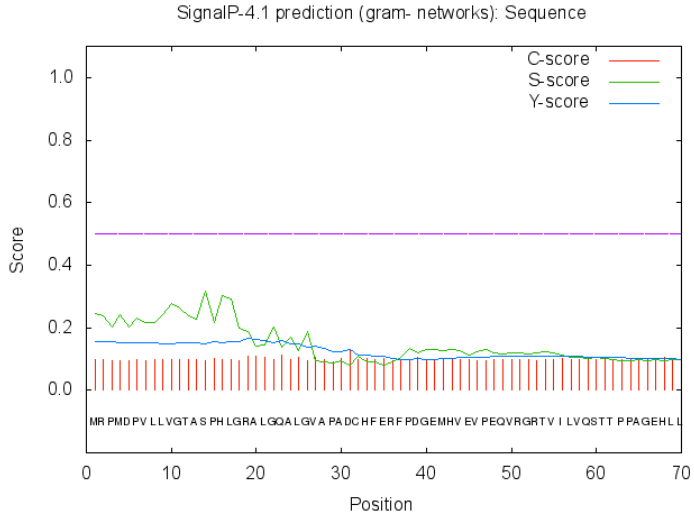

```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    21        0.127
  max. Y    12        0.108
  max. S     5        0.135
  mean S     1-11     0.117
       D     1-11     0.112   0.570   NO
Name=Myxococcus SP='NO' D=0.112 D-cutoff=0.570 Networks=SignalP-noTM
```

Mxan_5075



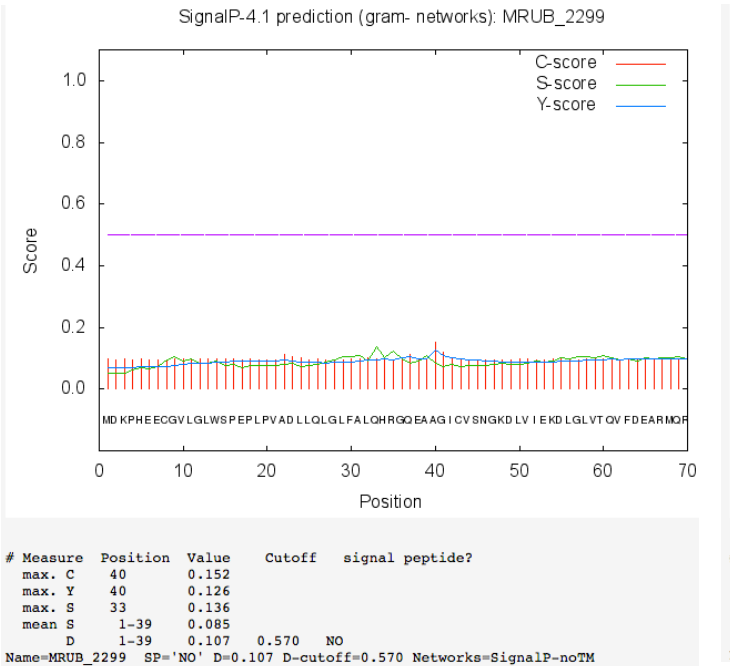

```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    22        0.109
  max. Y    53        0.110
  max. S    52        0.181
  mean S     1-52     0.095
       D     1-52     0.103   0.570   NO
Name=Mrub_1053  SP='NO' D=0.103 D-cutoff=0.570 Networks=SignalP-noTM
```

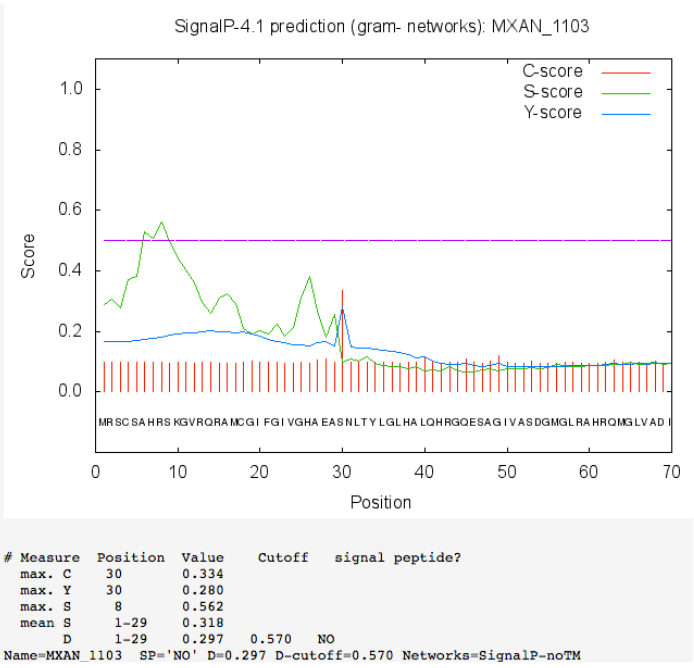
Mrub_1053




```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    31        0.125
  max. Y    19        0.165
  max. S    14        0.316
  mean S     1-18     0.243
       D     1-18     0.201   0.570   NO
Name=Sequence   SP='NO' D=0.201 D-cutoff=0.570 Networks=SignalP-noTM
```

Mxan_7156





Mrub_2299 Mxan_1103

Figure 7: No hydrophobic spikes indicates no membrane attachment.
While Mxan_1103 exhibits a small spike in the graphical representation no gene sequence is predicted to have signal peptides or cleavage sites. This is expected since all other bioinformatics tools predicted the gene sequences to reside entirely in the cytoplasm.

The last bioinformatic tool used to compare gene segment location within a cell was Phobius. This database can predict both TMHMM and signal peptide regions within a protein if given the correct conditions. Because none of the genes of interest seem to contain transmembrane helices nor signal peptides Phobius was unable to predict anything when given the amino acid sequences.

Conclusion

The evidence gained from multiple bioinformatics tools supports my hypothesis about the shared functionality of Mrub_2299 and Mxan_1103 as well as Mrub_1053, Mxan_7156, Mrub_2281 and Mxan_5075. Among many other reliable sources of information both sequences have the same EC number which indicates that they catalyze the same chemical reactions. This is

an excellent indication of their similarity. This combined with the fact that each database I searched placed all gene sequences in the same families or groups (Pfam: N-terminal domain of ribose phosphate pyrophosphokinase & Glutamine amidotransferase domain, COG Category: phosphoribosylpyrophosphate synthetase & Adenine/guanine phosphoribosyltransferase, TIGRfam: ribose-phosphate diphosphokinase & amidophosphoribosyltransferase, PDB: Phosphoribosylpyrophosphate synthetase & Glutamine phosphoribosylpyrophosphate (PRPP) amidotransferase) makes it very easy to deduce that my hypothesis is correct. Mrub_2281 and Mrub_1053 are similar in structure and function to Mxan_5075 and Mxan_7156 and likely code for the enzyme Phosphoribose-1-pyrophosphate synthetase. Mrub_2299 is similar in both structure and function to Mxan_1103 and likely is responsible for the production of Phosphoribosyl pyrophosphate amidotransferase.

Works Cited

Berg JM, Tymoczko JL, Stryer L. 2002. Section 25.2: Purine bases can be synthesized de novo or recycled by salvage pathways. In: Biochemistry 5th Edition. New York: W H Freeman.

Berman HM, Westbrook J, Zeng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. Nucleic Acids Research, 28: 235-242.

Hanson CW, Dworkin M. 1974. Intracellular and extracellular nucleotides and related compounds during the development of Myxococcus xanthus. Journal of Bacteriology [Internet]. 118(2): 486-496

Juncker AS1, Willenbrock H, Von Heijne G, Brunak S, Nielson H, Krogh A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. Aug; 12(8):1652-1662.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2015. KEGG as a reference resource for gene and protein annotation. [Internet]. [cited 2106 feb 5] 4:44: 457-462. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26476454

Kanehisa M. 2000. Post-genome informatics. Oxford University Press. KEGG (Kyoto Encyclopedia of Genes and Genomes) database at http://www.genome.jp/kegg/pathway.html

Koonin EV, Galperin MY. 2003. Sequence - Evolution - Function: Computational Approaches in Comparative Genomics. Boston: Kluwer Academic. Chapter 7.

Marchler-Bauer A et al. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acid Res. 43(D)222-6.

Markowitz V, Chen I, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A Ivanova N, and

Kyrpides N. 2014. IMG 4 version of the integrated microbial genomes comparative analysissystem. 42:D560-D567.

SignalP 4.0: discriminating signal peptides from transmembrane regions. Petersen TN., Brunak S., von Heijne G. & Nielsen H. Nature Methods, 8:785-786, 2011

Westby CA, Tsai WC. 1974. De Novo Purine Synthesis in Vegetative Cells and Myxospores of Myxococcus Xanthus. Journal of Bacteriology [Internet]. [cited 2016 Jan 21] 117(3): 1099–1107. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC246589/

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, and Brinkman FSL. 2010. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, Bioinformatics 26(13): 1608-1615