View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Augustana College, Illinois: Augustana Digital Commons

Augustana College
Augustana Digital Commons

Meiothermus ruber Genome Analysis Project

Biology

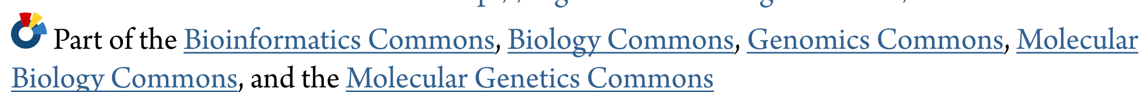
Winter 2-2016

The *Meiothermus ruber* Mrub_2572 gene is an ortholog of the *Escherichia coli* pyrE b3642 gene and the *Meiothermus ruber* Mrub_2071 gene is an ortholog of the *Escherichia coli* pyrF b1281 gene

Cale J. McCormick
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, Genomics Commons, Molecular Biology Commons, and the Molecular Genetics Commons

Recommended Citation
McCormick, Cale J. and Scott, Dr. Lori. "The *Meiothermus ruber* Mrub_2572 gene is an ortholog of the *Escherichia coli* pyrE b3642 gene and the *Meiothermus ruber* Mrub_2071 gene is an ortholog of the *Escherichia coli* pyrF b1281 gene" (2016). *Meiothermus ruber Genome Analysis Project.*
http://digitalcommons.augustana.edu/biolmruber/1

This Student Paper is brought to you for free and open access by the Biology at Augustana Digital Commons. It has been accepted for inclusion in Meiothermus ruber Genome Analysis Project by an authorized administrator of Augustana Digital Commons. For more information, please contact digitalcommons@augustana.edu.

The *Meiothermus ruber* Mrub_2572 gene is an ortholog of the *Escherichia coli* pyrE b3642 gene and the *Meiothermus ruber* Mrub_2071 gene is an ortholog of the *Escherichia coli* pyrF b1281 gene

Authors: Cale McCormick and Dr. Lori Scott

Published: February 2016

Background/ Introduction:

Nucleotides are the building blocks of DNA in cells and are fundamentally essential to life. It is believed that nucleotides and RNA molecules had evolved long before DNA and proteins had evolved (Joyce, 2002). This early evolution of nucleotides and their cellular importance ultimately lead to the conclusion that nucleotide metabolism is a vital part to life. This is why nucleotide de novo synthesis is highly conserved among almost all living organisms seen today (Lehninger *et al.*, 1994) (Kafer *et al*., 2004). And due to this circumstance, pyrimidine synthesis serves great importance to all organisms. The availability of uridine is critical to the synthesis of RNA and cell membranes alike in that the overall growth and function of cells is dependent on the generation of pyrimidine nucleotide–lipid conjugates (Connolly and Duley, 1999). Nucleotides can be separated into two categories, pyrimidines and purines. As far as pyrimidine structure goes, pyrimidines can be identified as being heterocyclic aromatic compounds that are six membered rings accommodating two nitrogen's located at positions 1 and 3 (Sharma *et* al., 2014). Pyrimidines are composed of three molecules called uracil (6), thymine (7), and cytosine (8) as seen in figure 1.

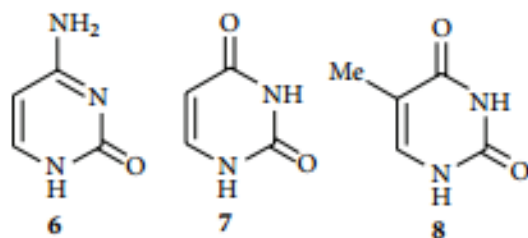


Figure 1: Pyrimidines and their molecular structures- 6: Uracil, 7: thymine, 8: Cytosine.

Image came from Figure 2b (Sharma *et* al., 2014) (http://www.hindawi.com/journals/ijmc/2014/202784/ref/)

Correspondingly, pyrimidine rings are synthesized from bicarbonate, an aspartic acid, and ammonia from the hydrolysis of a glutamine side chain (Berg *et al*., 2002). Pyrimidines are assembled as free bases, meaning they are not built upon one specific molecule within the cytosol. With that being said, pyrimidines are instead synthesized from 5-phospho-α-D-ribose 1-diphosphate (PRPP) and orotate alike (King). In the synthesis of these pyrimidines, there are two proteins that are particularly noticeable to the overall reaction. These two proteins are orotate phosphoribosyltransferase and orotidine-5'-phosphate decarboxylase. These two proteins are the final two proteins needed for the catalysis of uridine monophosphate (UMP), which is the first pyrimidine nucleotide catalyzed in pyrimidine synthesis and the precursor to other pyrimidines (Moffatt and Ashihara, 2002). Orotate phosphoribosyltransferase is dependent on the precursors catalyzed by the carbamoyl biosynthesis pathway and the PRPP synthesis pathway, as seen in Figure 2. The carbamoyl biosynthesis pathway is responsible for the formation of orotate (Makoff and Radford, 1978). The PRPP biosynthesis pathway is responsible for the other precursor 5-phospho-α-D-ribose 1-diphosphate (PRPP) (Wolucka. 2008). When both precursors are synthesized, 5-phospho-α-D-ribose 1-diphosphate (PRPP) and orotate are catalyzed together by orotate phosphoribosyltransferase to form a pyrophosphate and orotidine 5′-monophosphate

(Henriksen et al., 1995) (Christopherson and Finch, 1978). After orotidine 5′-monophosphate has been catalyzed by the orotate phosphoribosyltransferase, there is one remaining step until the formation of UMP is complete. In this final step, orotidine-5'-phosphate decarboxylase removes the carboxyl group of the orotidine-5′-monophosphate to form the uridine monophosphate. This reaction is the sixth and final step to the formation of UMP, which is also the precursor to all of the other pyrimidines (Turnbough et al., 1987).

In further analysis of orotate phosphoribosyltransferase and orotidine-5'-phosphate decarboxylase and their enzymatic function, these genes were identified to be a part of the UMP biosynthesis pathway (Figure 2). *pyrE* codes for the protein orotate phosphoribosyltransferase (EC 2.4.2.10), it has been shown through genetic analysis when pyrE is mutated increases of orotic acid levels rise as a result of this proteins mutation (Shimosaka et al., 1984). It is also seen that exogenous adenine will in turn increase the accumulation of precursors of orotate phosphoribosyltransferase by lowering the intracellular content of PRPP. As a result, the decrease in the synthesis of orotate to orotidine-5′-monophosphate consequently decreases the use of orotate for synthesis of pyrimidine nucleotides. (Bagnara and Finch, 1974). Through even further investigation, *pyrE* has also been found in the *E. coli* linkage map and is not linked to any other genes in pyrimidine synthesis (O'Donovan and Neuhard, 1970). Also seen in Figure 2, *pyrF* codes for the protein orotidine-5'-phosphate decarboxylase (EC 4.1.1.23), and is genetically shown that mutations in the *pyrF* gene increase the excretion of orotic acid (Womack and O'Donovan, 1978). Due to mutations of this gene and its importance to pyrimidine synthesis, all mutants of *pyrF* have an absolute requirement for pyrimidines in order to survive (O'Donovan and Neuhard, 1970). *pyrF* has also been located in the *E. coli* linkage map suggesting that it is not linked to any other genes in pyrimidine synthesis. Whereas *pyrF* is not linked to any other

gene, recent studies have suggested the likelihood of a *pyrF* operon downstream of the gene (Turnbough et al., 1987). Here we see that the *orfF* initiation site overlaps the *pyrF* termination site suggesting translational coupling between the two genes.

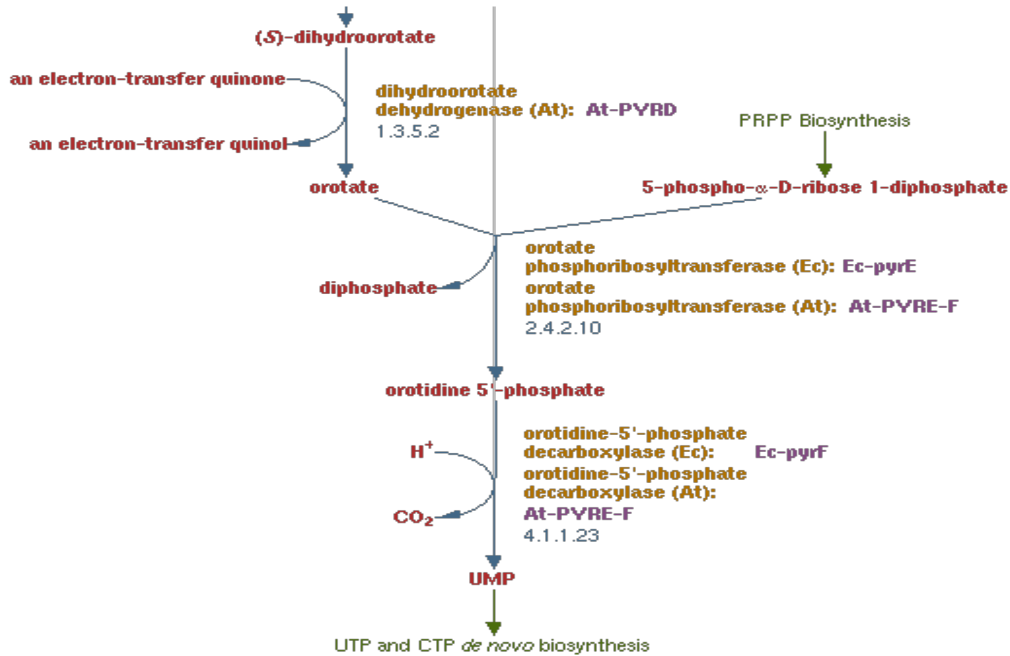


Figure 2: The *E. coli* b3642 and b1281 genes, as well as the *M. ruber Mrub*_2572 and Mrub_2071 genes, both have enzymatic function in the UMP biosynthesis pathway. Image came from (http://metacyc.org/)

The UMP biosynthesis pathway is of interest for a number of reasons, much of which has to do with its importance that it can provide. The Genomic Encyclopedia of Bacteria and Archaea (GEBA) project has sequenced nearly 200 genomic sequences in hopes of using those genomes to reach groundbreaking discoveries (DOE Joint Genome Institute, 2015). By analyzing these genomic sequences, researchers hope to find new breakthroughs that might lead to new

healthcare methods, energy productions, or bioremediations (DOE Joint Genome Institute, 2015). *M. ruber* is part of poorly studied section of the Tree of Life that has not be analyzed as in depth as other parts of the Tree of Life. So by analyzing *M. ruber's* genomic sequence in further detail, that section of the Tree of Life will hopefully be further understood in more ways than before and thus providing purpose to this study (DOE Joint Genome Institute, 2015). In regards to pyrimidine synthesis, further research into uridine synthesis and its nucleotides will increase our understanding of certain diseases. From this, new drug discoveries can be made and developments will be promoted, many of which can be therapeutic (Connolly and Duley, 1999). At a microbiological level, microbes have been a serious global problem due to their ability to adapt and live in varying conditions. Pyrimidine experimentation with pyrimidine derivatives might ultimately help the development of more potent procedures for the production of effective antimicrobial agents (Sharma *et al*., 2014). Advancements in these areas could provide new opportunities for research and treatments and thus opening the door to new discoveries.

Lastly, the Mrub_2572 and b3642 genes and the Mrub_2071 and b1281 genes will be analyzed with numerous bioinformatics tools. Bioinformatics tools will use the genomic sequences of these four genes to show their similarities or differences with E-scores and bit scores. The E-values will show us if the genes are relevant enough to each other to classify them as orthologs or not and will show the likelihood of the two sequences being similar by chance. A low E-value may suggest sequence similarity, function similarity, or the possibility of the two organisms being evolutionary related. Bioinformatics tools will look at the DNA sections for gene regulations, DNA motifs, and structure/function of the coded protein among other things (Pujari, 2014). Nevertheless, it is hypothesized that the *Meiothermus ruber* Mrub_2572 gene and

Escherichia coli b3642 are orthologs of each other. It is also hypothesized that the *Meiothermus ruber* Mrub_2071 gene and *Escherichia coli* b1281 are orthologs of each other as well.

Methods:

In order to distinguish whether or not the Mrub_2572 and b3642 genes and the Mrub_2071 and b1281 genes are orthologs of each other, the use of the bioinformatics programs located in the GENI-ACT lab notebook system were applied to the selected genes. These bioinformatics programs descriptions are located here: http://www.geni-act.org/education/main/. Although most of these bioinformatics programs were used directly as stated, some programs had slight deviations in their procedures. The first deviation was at the beginning of the gene annotation process; the *Meiothermus ruber* genes were put into BLAST to make sure there were *Escherichia coli* sequences that could be used for annotation and comparison. A second deviation from the GENI-ACT system came during the use of the bioinformatics program Tcoffee. This program calls for the use of ten sequences to run a multiple sequence alignment with *E. coli*, but fifteen sequences were used instead. For the first alignment sequence, *E. coli* was excluded from the search to provide room for other organisms that have sequence similarity. The next deviation was located in the pathways. During the use of the KEGG map pathway, we highlighted the enzymatic pathway facilitated by the annotated gene to find what pathways *E. coli* and *M. ruber* used for pyrimidine synthesis.

Results:

The bioinformatics tools described in the methods were used to analyze the two genes *M. ruber* Mrub_2572 *and E. coli* b3642.

Table 1: *E.coli* pyrE gene b3642 and *Mrub*_2572 are orthologs

Evidence collected during analysis	*E. coli* pyrE gene b3642	*M. ruber Mrub*_2572 gene
Cellular localization	Cytoplasmic	
KEGG Pathway	UMP biosynthesis pathway	
E.C. number	E.C. 2.4.2.10 Orotate phosphoribosyltransferase	
BLAST *E. coli* against *M. ruber*	Score: 53.5 E-value= 7e-14 32% identity	
CDD (COG category)	COG0461 (E=2e-211) Orotate phosphoribosyltransferase	COG0461 (E= 1.09e-41) Orotate phosphoribosyltransferase
TIGRfam – protein family	TIGR00336 (E= 3.8e-111) pyrE Orotate phosphoribosyltransferase	TIGR01367 (E= 3e-80) pyrE Orotate phosphoribosyltransferase
Pfam – protein family	PF00156 (E= 1.3e-11) Pribosyltran	PF00156 (E= 2e-17) Pribosyltran
PDB (Protein Database)	1ORO (E= 3.04621e-122) A FLEXIBLE LOOP AT THE DIMER INTERFACE IS A PART OF THE ACTIVE SITE OF THE ADJACENT MONOMER OF ESCHERICHIA COLI OROTATE PHOSPHORIBOSYLTRANSFERASE (E.C. 2.4.2.10)	4PAW (E= 4.80105e-26) Hypothetical protein HP1257 (E.C. 2.4.2.10)

Family and Domains:

Both genomic sequences for the *E.coli* b3642 gene and *Mrub*_2572 gene had several important results in consideration to their BLAST. The BLAST for the *E. coli* b3642 genome sequence to the *Mrub*_2572 genome sequence revealed a bit score of 53.5. This BLAST also had

an E-value of 7e-14 with a 32% identity. *E. coli* b3642 and *Mrub*_2572 both have similarities in their families and domains as well. Both genes showed evidence of having the same CDD domain with a COG number COG0461, indicating both genes to be an Orotate phosphoribosyltransferase (Table 1). *E. coli* b3642 yielded an E-value of 4e-211 and *Mrub*_2572 yielded an E-value of 1.09e-41, both of which suggest they are similar not due to chance. The next domain tested was the TIGRfam of *E.coli* b3642 and *Mrub*_2572. Both genes showed a TIGRfam number of TIGR00336 in the protein family pyrE Orotate phosphoribosyltransferase (Table 1). *E. coli* b3642 yielded an E-value of 3.8e-111 and *Mrub*_2572 yielded an E-value of 3e-80, both of which suggest they are similar not due to chance. As far as the protein database (PDB) of the selected genes goes, things weren't as concise. The conclusion of this test for the *E. coli* b3642 gene suggests that the protein is 1ORO - A FLEXIBLE LOOP AT THE DIMER INTERFACE IS A PART OF THE ACTIVE SITE OF THE ADJACENT MONOMER OF ESCHERICHIA COLI OROTATE PHOSPHORIBOSYLTRANSFERASE with an E-value of 3.04621e-12 (Table 1). The conclusion of this test for the *M. ruber Mrub*_2572 gene suggests that the protein is 4PAW – Structure of hypothetical protein HP1257 with an E-value of 4.80105e-26 (Table 1). Both proteins identified had an enzyme commission number of E.C. 2.4.2.10. When viewing these proteins on the PDB, both proteins were Orotate phosphoribosyltransferase.

Secondly, the Pfam of *E.coli* b3642 and *Mrub*_2572 came up to be PF00156 named Pribosyltran (Table 1). *E. coli* b3642 yielded an E-value of 1.3e-11 and *Mrub*_2572 yielded an E-value of 2e-17, both of which suggest they are similar not due to chance This Pribosyltran domain also showed similar conserved amino acids. The *E.coli* b3642 pairwise alignment had important amino acids 75G, 86L, 121V, 124V, 125D, 126D, and 132G. Mrub_2572 had similar

conserved amino acids as well at 68G, 79L, 111V, 113D, and 118G. The *E.coli* b3642 sequence starts at amino acid 48 and the *Mrub*_2572 sequence starts are amino acid 33 (Figure 3).

A

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites
				Start	End	Start	End	From	To				
Pribosyltran	Phosphoribosyl transferase domain	Domain	CL0533	38	184	48	159	13	129	159	44.2	1.3e-11	n/a

```
#HMM    davarlaaeileeyngkpdvvvgiargGvpfakiladeL......evplayvrkvkyrpdgseeaargssdesrlledpagkkVliVDDildtGgTllaalelLkekgpeevriavlvdkpat
#MATCH     +++a+ ++++ + + d+++g+a++G+p+a+ +a +L      ++p+++ rk+ ++++ +++ +              ++     +V +VDD+++ G+++++ +e++ + g++  ++  ++d+ ++
#PP     4566777788888.***********************************************99997..........4556789***************************9999999988776
#SEQ    LLGRFYAEALVDS-GIEFDLLFGPAYKGIPIATTTAVALaehhdlDLPYCFNRKEAKDHGEGGNLV----------GSALQGRVMLVDDVITAGTAIRESMEIIQANGATLAGVLISLDRQER
```

B

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
Pribosyltran	Phosphoribosyl transferase domain	Domain	CL0533	31	174	33	165	3	146	159	63.0	2e-17	n/a	Hide

```
#HMM    sellenpaildavarlaaeileeyngkpdvvvgiargGvpfakiladeLevplayvrkvkyrpdgseeaargssdesrlledpagkk
#MATCH  ++ll++p  ++av +++ e+++++    + d v+g+a gGv++a  +a++L+v+  + +k     + +  e          l++  g++
#PP     68999******************..****************************9999988888887777..........57999***
#SEQ    TTLLQHPLYAEAVGQAMGELFDT--LELDFVIGPAMGGVVLAFVTAKALGVRALFAEKDGQGGMRVREG----------LTIHPGER

VliVDDildtGgTllaalelLkekgpeevriavlvdkpateeevdknakrvfsls.pd
 l V+D+++tGg+++ a+++ ++kg+ +v + ++vd++a++ e ++ ++++++l  p+
*************************************************999998776
FLAVEDVVTTGGSVQQAIRAAEAKGARCVAVGAIVDRSAGRAEFSAPYRSLVQLDfPT
```

Figure 3: The *E.coli* b3642 and *Mrub*_2572 have conserved amino acid sequences that line up with each other. Lane A- *E. coli* b3642. Lane B- *Mrub*_2572. #match- match between consensus sequence. The pairwise alignments came from (http://pfam.xfam.org/search).

Cellular localization:

To begin, the *E.coli* b3642 and *Mrub*_2572 genes displayed no indications of transmembrane helices (Figure 4 and Figure 5). The *E. coli* b3642 transmembrane topology graph clearly indicates that there are no transmembrane helices due to the lack of helices reaching the outside of the cell. The *Mrub*_2572 transmembrane topology graph does not have

any transmembrane helices either, but might have a section of amino acids that is hydrophobic as seen in Figure 5. This provides evidence that the two proteins are cytoplasmic.

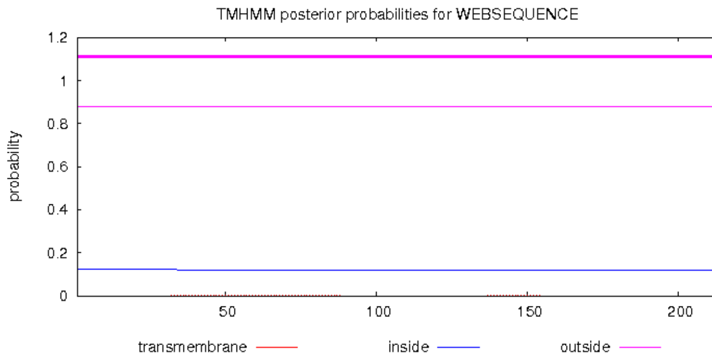


Figure 4: *E. coli* b3642 does not contain TMH regions; it is predicted to be a cytoplasmic. TMHMM (http://www.cbs.dtu.dk/services/TMHMM) created this plot.

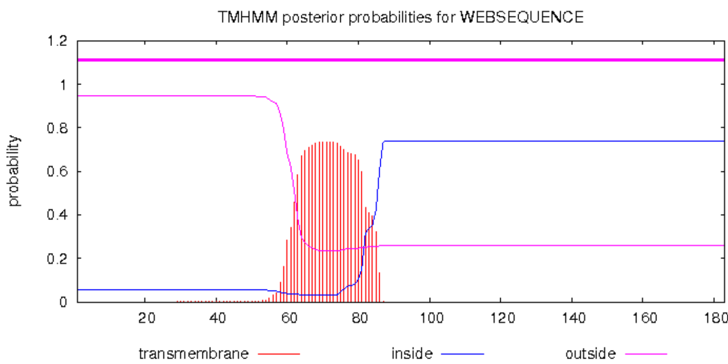


Figure 5: *Mrub*_2572 does not contain TMH regions; it is predicted to be a cytoplasmic. TMHMM (http://www.cbs.dtu.dk/services/TMHMM) created this plot.

To further understand the cellular localization of *E. coli* b3642 and *Mrub*_2572, the PSORT-B computational program was also put to use to predict the protein localization of these genes. Out of a score of 10, the *E.coli* b3642 gene had a cytoplasmic score of 9.97 and the *Mrub*_2572 gene had a cytoplasmic score of 8.96. In conclusion; both the transmembrane topography graph and P-SORT for *E. coli* b3642 and *Mrub*_2572 highly suggest that these proteins are located in the cytoplasm (Table 1).

Pathways and E.C.:

Next, it is seen that both the *E.coli* b3642 and *Mrub*_2572 genes have an enzyme commission number of E.C. 2.4.2.10. By knowing that the *E.coli* b3642 gene and *Mrub*_2572 gene have the same E.C. number, it can be applied to the KEGG pathway to find its enzymatic pathway. And by doing so, it is seen that the *E.coli* b3642 and *Mrub*_2572 gene use the UMP biosynthesis pathway (Figure 2 and Figure 6). Both the *E.coli* b3642 and *Mrub*_2572 genes have a KEGG pathway ID number of 00240.

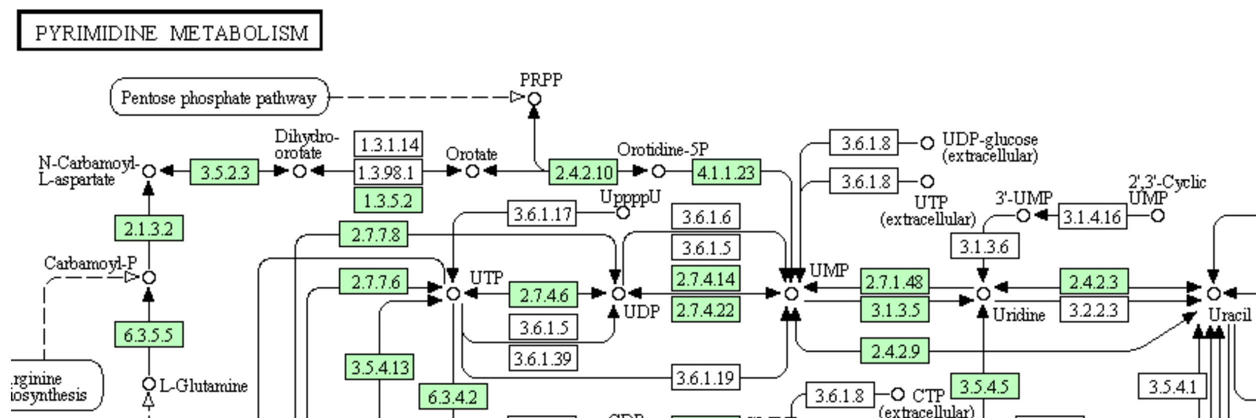
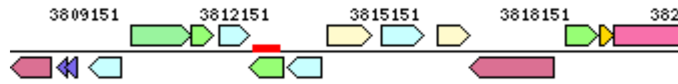


Figure 6: E.C. 2.4.2.10 is part of the UMP biosynthesis pathway. KEGG (http://geni-act.org/lab_notebook/student/isolate_genome_gene/6fa2ae446a0244ad/045600a4ae1a41a4/25eff9dbb2834e92/) created this pathway.

The *E.coli* b3642 and *Mrub*_2572 genes are not part of an operon (Figure 7 and Figure 8)



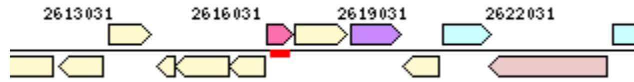


Figure 7: *E.coli* b3642 is not part of an operon.

Figure 8: *Mrub*_2572 is not part of an operon

(http://img.jgi.doe.gov/cgi-bin/edu/main.cgi?section=FindGenes&page=geneSearch) create these pathways.

Results:

The bioinformatics tools described in the methods were used to analyze the two genes *M. ruber* Mrub_2071 *and E. coli* b1281.

Table 2: *E.coli* pyrF gene b1281 and *Mrub*_2071 are orthologs

Evidence collected during analysis	*E. coli* pyrF gene b1281	*M. ruber Mrub*_2071 gene
Cellular localization	Cytoplasmic; protein does have any certain localization- assumed to be cytoplasmic.	
KEGG Pathway	UMP biosynthesis pathway	
E.C. number	E.C. 4.1.1.23 Orotidine-5'-phosphate decarboxylase	
Blast *E. coli* against *M. ruber*	Score: 53.5 E-value= 7e-14 32% identity	
CDD (COG category)	COG0284 (E=5.20e-98) Orotidine-5'-phosphate decarboxylase	COG0284 (E= 2.71e-24) Orotidine-5'-phosphate decarboxylase
TIGRfam – protein family	TIGR01740 (E= 2.5e-67) pyrF: orotidine 5'-phosphate decarboxylase	TIGR02127 (E= 1.6e-69) pyrF_sub2: orotidine 5'-phosphate decarboxy
Pfam – protein family	PF00215 (E= 1.5e-54) OMPdecase	PF00215 (E= 1.6e-24) OMPdecase
PDB (Protein Database)	1EIX (E= 5.70819e-136) STRUCTURE OF OROTIDINE 5'-MONOPHOSPHATE DECARBOXYLASE FROM E. COLI, CO-CRYSTALLISED WITH THE INHIBITOR BMP (E.C. 4.1.1.23)	3V75 (E= 6.43827e-23) Crystal structure of putative orotidine 5'-phosphate decarboxylase from Streptomyces avermitilis ma-4680 (E.C. 4.1.1.23)

Family and Domains:

Both genomic sequences for the *E.coli* b1281 gene and *Mrub*_2071 gene had several important results in consideration to their BLAST. The BLAST for the *E. coli* b1281 genome

sequence to the *Mrub*_2071 genome sequence revealed a bit score of 23.1. This blast also had an E-value of 0.004 with a 30% identity. *E. coli* b1281 and *Mrub*_2071 both have similarities in their families and domains as well. Both genes showed evidence of having the same CDD domain showing a COG number COG0284, indicating both genes to be an Orotidine-5'-phosphate decarboxylase (Table 2). *E. coli* b1281 yielded an E-value of 5.20e-98 and *Mrub*_2071 yielded an E-value of 2.71e-24, both of which suggest they are similar not due to chance. The next domain tested was the TIGRfam of *E.coli* b1281 and *Mrub*_2071. *E.coli* b1281 showed a TIGRfam number of TIGR01740 in the protein family pyrF: orotidine-5'-phosphate decarboxylase with an E-value of 2.5e-67 (Table 2). The *Mrub*_2071 showed a TIGRfam number of TIGR02127 in the protein family pyrF_sub2: orotidine 5'-phosphate decarboxy with an E-value of 1.6e-69 (Table 2). Both genes suggest they are similar not due to chance. As far as the protein database (PDB) of the selected genes goes, differing results were noted. The conclusion of this test for the *E. coli* b1281gene suggests that the protein is 1EIX - STRUCTURE OF OROTIDINE 5'-MONOPHOSPHATE DECARBOXYLASE FROM E. COLI, CO-CRYSTALLISED WITH THE INHIBITOR BMP with an E-value of 5.70819e-136 (Table 2). The conclusion of this test for the *M. ruber Mrub*_2071 gene suggests that the protein is 3V75 – Crystal structure of putative orotidine-5'-phosphate decarboxylase from Streptomyces avermitilis ma-4680 with an E-value of 6.43827e-23 (Table 2). Both proteins identified had an enzyme commission number of E.C. 4.1.1.23. When viewing these proteins on the PDB, both proteins were orotidine-5'-phosphate decarboxylase.

Secondly, the Pfam of *E.coli* b1281 and *Mrub*_2071 came up to be PF00215 named OMPdecase (Table 2). *E. coli* b1281 yielded an E-value of 1.5e-54 and *Mrub*_2071 yielded an E-value of 1.6e-24, both of which suggest they are similar not due to chance. This OMPdecase

domain also showed similar conserved amino acids. The *E.coli* b1281 pairwise alignment had important amino acids 22D, 44K, 71D, 73K, 76D, 77I, 103G, 165G, 221G, and 222R. Mrub_2071 had similar conserved amino acids at 13L, 18D, 51K, 85D, 87K, 90D, 91I, 92G, 118G, and 189G. The *E. coli* b1281 sequence starts at amino acid 15 and the *Mrub*_2572 sequence starts are amino acid 12 (Figure 9).

A

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
OMPdecase	Orotidine 5'-phosphate decarboxylase / H ...	Domain	CL0036	15	237	15	237	1	226	226	184.9	1.5e-54	n/a	Hide

```
#HMM     snLclalDvetarelleladkvgpsivvlKtgydlveawgfetlaklvallrkhgflifeDrKfaDIGntvekqakagsarikewadivtvhavaGkasvkalaeaaekwg..nrgllilaqmsskgsldlqelk.
#MATCH   s++++alD+++++++l+++dk++p  + lK+g ++++ +g     ++v++l+++gf if+D+Kf+DI+nt++++++   a++++++++v+vha  G  +++a+ ea+  +g  ++ l+++++++s+++ dl +l+
#PP      5799*************************************....********************************7...899******************************9988899*****************777
#SEQ     SPVVVALDYHNRDDALAFVDKIDPRDCRLKVGKEMFTLFG----PQFVRELQQRGFDIFLDLKFHDIPNTAAHAVA---AAADLGVWMVNVHASGGARMMTAAREALVPFGkdAPLLIAVTVLTSMEASDLVDLGm

.kgytqecveaarehkaedfvmGfvaqatln.ateed.addfihltpgigleqggdgggqqyntpealigiagadiiivGRgIlkagdpveeaery
  +  + +e++++ +++++++G+v++a+++ + ++  +++f+ +tpgi++ qg+++g+q++ ++++++  ag d++++GR+++++ dp+++++++
799****************************86666689************.*****************************************99875
mTLSPADYAERLAALTQKCGLDGVVCSAQEAvRFKQVfGQEFKLVTPGIRP-QGSEAGDQRRIMTPEQALSAGVDYMVIGRPVTQSVDPAQTLKAI
```

B

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
OMPdecase	Orotidine 5'-phosphate decarboxylase / H ...	Domain	CL0036	11	226	12	220	2	187	226	86.6	1.6e-24	n/a	Hide

```
#HMM     nLclalDveta...........relleladkvgpsivvlKtgydlveawgfe...tlaklvallrkhgflifeDrKfaDIGntvekqakagsarikewadivtvhavaGkasvka
#MATCH   +L+l++D++ +           r  lel++ +   i+ +K + +++ea g +   ++ +l+  +r    ++++D+K++DIG+t+e++ak  ++   +     ++tv++++G +++++
#PP      6899999999999999999999*****************************99**9999*******************************9..56689999*****************
#SEQ     PLVLGVDPRPElhgpnplahirRYTLELLEALAGRICAVKFQAAFFEALGPQgftLMHELIVGARVLSLPVIVDAKRGDIGSTAEAYAK--AYLQAYPGSALTVNPYLGRDALEP

laeaaekwgnrgllilaqmsskgsldlqe..lk..kgytqecveaarehkae.........dfvmGfvaqatlnateedaddfihltpgigleqggdg
+ +aa+++g   +++l+++s++gs  +q+  l+   + +q++++ +++++++         + v+G+  +++  ++++   ++ + l pg g+ qgg++
********8.8*************99999643387677788888877776699*****************996677789**********9.88875
FFNAAQASG-GAVFVLVKTSNPGSGLFQDlpLEqgGRLYQAVADYLAQQAEQqrvgewsrvAAVVGVTYPEQVTEIRQRLPHSLLLLPGLGA-QGGQA
```

Figure 9: The *E.coli* b1281 and *Mrub*_2071 have conserved amino acid sequences that line up with each other. Lane A- *E. coli* b1281. Lane B- *Mrub*_2071. #match- match between consensus sequence. The pairwise alignments came from (http://pfam.xfam.org/search).

Cellular localization:

To begin, both the *E.coli* b1281 and *Mrub*_2071 displayed no indications of having any transmembrane helices (Figure 10 and Figure 11). The *E. coli* b1281 transmembrane topology graph clearly indicates that there are no transmembrane helices due to the lack of helices reaching the outside of the cell. The *Mrub*_2071 transmembrane topology graph clearly demonstrates that there are no transmembrane helices due to the lack of helices reaching the outside of the cell as well. This provides evidence that the two proteins are cytoplasmic.

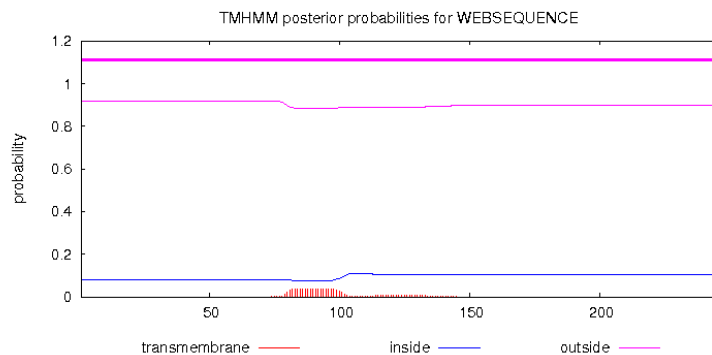


Figure 10: *E. coli* b1281 does not contain TMH regions; it is predicted to be a cytoplasmic. TMHMM (http://www.cbs.dtu.dk/services/TMHMM) created this plot.

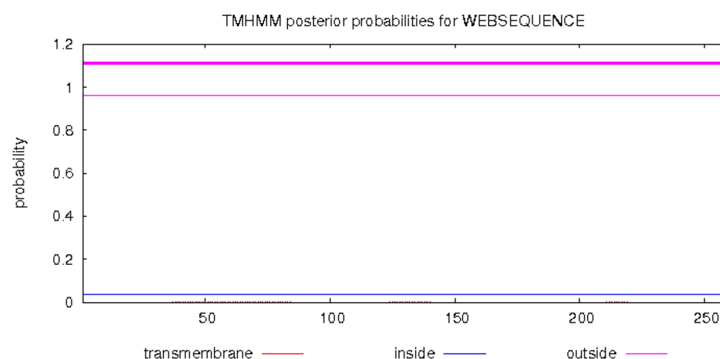


Figure 11: *Mrub*_2071 does not contain TMH regions; it is predicted to be a cytoplasmic. TMHMM (http://www.cbs.dtu.dk/services/TMHMM) created this plot.

To further understand the cellular localization of *E. coli* b1281 and *Mrub*_2071, the PSORT-B computational program was also put to use to predict the protein localization of these genes. Out

of a score of 10 once again, the *E.coli* b1281 gene had a cytoplasmic score of 9.26 and thus suggesting that this protein is very likely to be cytoplasmic. On the other hand, the *Mrub*_2071 gene had scores of 2.00 in all of the categories except for the cell wall. This data suggests that the computer program PSORT-B does not know what to do with the amino sequence provided. Due to this insufficiency, both genes were tested for signal peptides for further localization evidence. The SignalP for both *E. coli* b1281 and *Mrub*_2071 exhibited numbers below the cutoff line of 0.450 suggesting that these genes are not signal peptides (Figure 12 and Figure 13). This evidence suggests that these proteins are not localized to any part of the cell. In conclusion; both the transmembrane topography graphs and SignalP for *E. coli* b3642 and *Mrub*_2572 highly suggest that these proteins are located in the cytoplasm and do not localize the protein to a specific location. (Table 2).

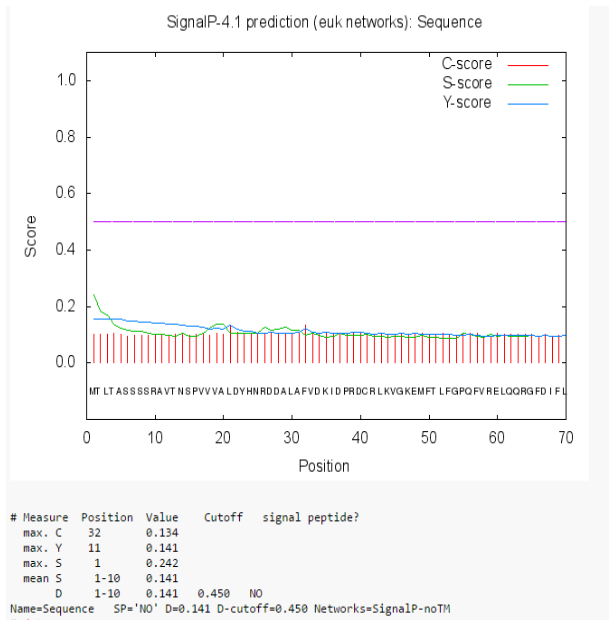


Figure 12: *E. coli* b1281 does not contain any Signal Peptides and is under the cutoff line of 0.450. (http://www.cbs.dtu.dk/services/SignalP) created this plot.

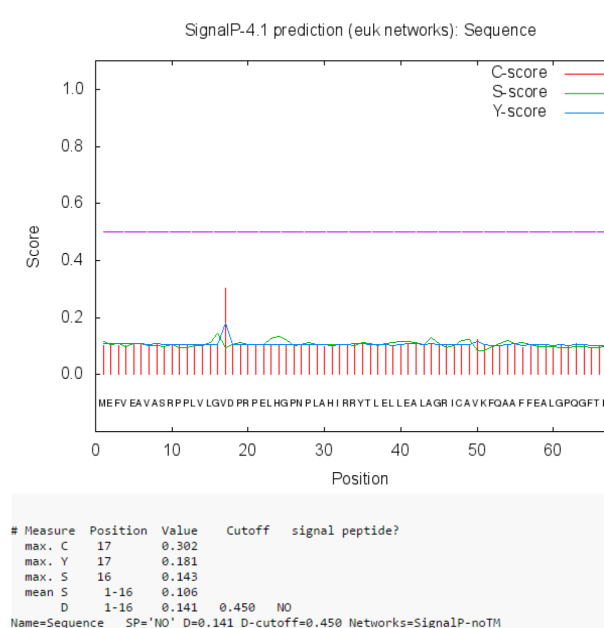


Figure 13: *Mrub*_2071 does not contain any Signal Peptides and is under the cutoff line of 0.450. (http://www.cbs.dtu.dk/services/SignalP) created this plot.

Pathways and E.C.:

Next, it is seen that both the *E.coli* b1281 and *Mrub*_2071 genes have an enzyme commission number of E.C. 4.1.1.23. By utilizing this E.C. number, it can be applied to the KEGG pathway to find its enzymatic pathway. By doing so, it is seen that the *E.coli* b1281 and *Mrub*_2071 gene use the UMP biosynthesis pathway (Figure 2 and Figure 14). Both the *E.coli* b1281 and *Mrub*_2071 genes have a KEGG pathway ID number of 00240.

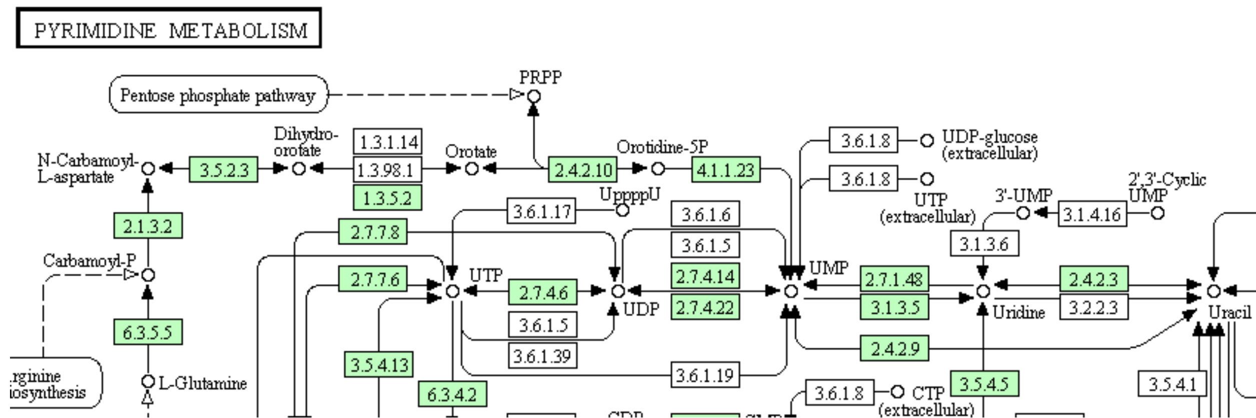


Figure 14: E.C. 4.1.1.23 is part of the UMP biosynthesis pathway. KEGG (http://geni-act.org/lab_notebook/student/isolate_genome_gene/6fa2ae446a0244ad/045600a4ae1a41a4/25eff9dbb2834e92/) created this pathway.

The *E.coli* b1281 and *Mrub*_2071 genes are not part of an operon (Figure 15 and Figure 16).

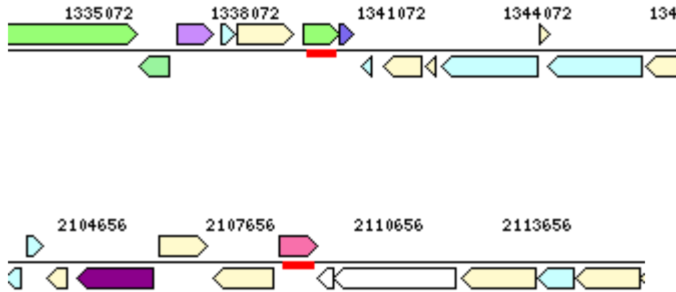


Figure 15: *E.coli* b1281 is not part of an operon.

Figure 16: *Mrub*_2071 is not part of an operon

(http://img.jgi.doe.gov/cgi-bin/edu/main.cgi?section=FindGenes&page=geneSearch) create these pathways.

Conclusion:

In conclusion, the cellular localization of the Mrub_2572 and b3642 genes is shown to be cytoplasmic with no transmembrane helices. The cellular localization of the Mrub_2071 and b1281 genes is also shown to be cytoplasmic with no transmembrane helices and no certain protein localization points. All four genes have the same enzymatic function and pathways in the UMP biosynthesis pathways. *E.coli* b3642 and *Mrub*_2572 showed compelling sequence similarity of the *pyrE* from BLAST with a significant E-value. For the *E.coli* b1281 and *Mrub*_2071 genes, the E-value was not significant enough to completely suggest the *pyrF* sequence similarity. This lower E-value means additional information is needed for a better conclusion. When considering the families and domains of both *E.coli* b3642 and *Mrub*_2572, orotate phosphoribosyltransferase was the protein identified and was verified by significant E-values. In regards to the *E.coli* b1281 and *Mrub*_2071, Orotidine-5'-phosphate decarboxylase was the protein identified and was also verified by significant E-values. Likewise, the Pfam

shows support for both sets of genes being in their respected domains Pribosyltran and OMPcase. When viewing the pairwise alignments, both *E.coli* b3642 and *Mrub*_2572 displayed similar conserved amino acid sequences to each other. The same was seen for the pairwise alignments of b1281 and Mrub_2071. The last evidence utilized came form gene context and whether or not these genes were part of an operon. It was concluded that *E.coli* b3642 and *Mrub*_2572 as well as b1281 and Mrub_2071 are not part of an operon. This conclusion was seen by the color trends highlighted by the KEGG pathway. To conclude, all of this evidence supports the hypothesis that the *Meiothermus ruber Mrub*_2572 gene is indeed an ortholog of the *Escherichia coli* pyrE b3642 gene. It can also be concluded that all of this evidence supports my hypothesis that the *Meiothermus ruber Mrub*_2071 gene is indeed an ortholog of the *Escherichia coli* pyrF b1281 gene.

Literature Cited

Bagnara, A. S. & Finch, L. R. (1974) *Eur. J. Biochem. 41*, 421–430

Berg J, Tymoczko J, Stryer L. 2002. In de Novo Synthesis, the Pyrimidine Ring Is Assembled from Bicarbonate, Aspartate, and Glutamine. W H Freeman. [accessed 2016 Feb 5]. http://www.ncbi.nlm.nih.gov/books/NBK22447/

Christopherson R, Finch L. 1978. Response of the Pyrimidine Pathway of Escherichia coli K12 to Exogenous Adenine and Uracil. Eur J Biochem 90:347-358. [accessed 2016 Feb 5]

Connolly G, Duley J. 1999. Uridine and its nucleotides: biological actions, therapeutic potentials. Trends in Pharmacological Sciences 20:218-225. [accessed 2016 Feb 5]

DOE Joint Genome Institute. 2015. Phylogenetic Diversity - DOE Joint Genome Institute. [accessed 2015 Dec 16]. http://jgi.doe.gov/our-science/science-programs/microbial-genomics/phylogenetic-diversity/

GENI. 2015. Meiothermus ruber Genome Analysis Project. [accessed 2015 Dec 16]. http://geni-science.org/secure/projects/view/

Henriksen A, Aghajari N, Jensen K, Gajhede M. 1995. A Flexible Loop at the Dimer Interface is a Part of the Active Site of the Adjacent Monomer of Escherichia coli Orotate Phosphoribosyltransferase, - Biochemistry (ACS Publications). Pubs.acs.org. [accessed 2016 Feb 5]. http://pubs.acs.org/doi/pdf/10.1021/bi952226y

Joyce, G.F. (2002). The antiquity of RNA-based evolution. Nature 418:214-22

Kafer, C., Zhou, L., Santoso, D., Guirgis, A., Weers, B., Park, S., and Thornburg, R. (2004). Regulation of pyrimidine metabolism in plants. Front. Biosci. 9: 1611-162

King M. Nucleotide Metabolism: Nucleic Acid Synthesis. Themedicalbiochemistrypage.org. [accessed 2016 Feb 5]. http://themedicalbiochemistrypage.org/nucleotide-metabolism.php

Lehninger, A.L., Nelson, D.L., and Cox, M.M. (1994). Prinzipien der Biochemie, H. Tschesche, ed (Heidelberg, Berlin, Oxford: Spektrum Akademischer Verlag

Makoff A, Radford A. 1978. Genetics and biochemistry of carbamoyl phosphate biosynthesis and its utilization in the pyrimidine biosynthetic pathway. Microbiological Reviews 42:307. [accessed 2016 Feb 5]. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC281432/

Moffatt B, Ashihara H. 2002. Purine and Pyrimidine Nucleotide Synthesis and Metabolism. The Arabidopsis Book. The American Society of Plant Biologists.

O'Donovan G, Neuhard J. 1970. Pyrimidine Metabolism in Microorganisms. BACTERIOLOGICAL REVIEWS 34. [accessed 2016 Feb 5]

Pujari S. 2014. Bioinformatics: An Useful Essay on Bioinformatics | Biotechnology. YourArticleLibrary.com. [accessed 2015 Dec 16]. http://www.yourarticlelibrary.com/essay/bioinformatics-an-useful-essay-on-bioinformatics-biotechnology/29374/

Sharma V, Chitranshi N, Agarwal A. 2014. Significance and Biological Importance of Pyrimidine in the Microbial World. International Journal of Medicinal Chemistry 2014:1-31. [accessed 2016 Feb 5]

Shimosaka M, Fukuda Y, Murata K, Kimura A. 1984. Purine-Mediated Growth Inhibition Caused by a pyrE Mutation in Escherichia coli K-12. JOURNAL OF BACTERIOLOGY 160:1101-1104. [accessed 2016 Feb 5]

Turnbough C, Kerr K, Funderburg W, Donahue J, Powell F. 1987. Nucleotide sequence and characterization of the pyrF operon of Escherichia coli K12. Journal of Biological Chemistry 262:10239-10245. [accessed 2016 Feb 5]. http://www.jbc.org/content/262/21/10239.long

Wolucka B. 2008. Biosynthesis of D-arabinose in mycobacteria - a novel bacterial pathway with implications for antimycobacterial therapy. FEBS Journal 275:2691-2711. [accessed 2016 Feb 5]

Womack J, O'Donovan G. 1978. Orotic Acid Excretion in Some Wild-Type Strains of Escherichia coli K-12. JOURNAL OF BACTERIOLOGY, 136:825-827. [accessed 2016 Feb 5]