2-2016

Meiothermus ruber Mrub_0976 and Mrub_1641 share the same functions as *Escherichia coli* b3940 and b3433 in the biosynthesis of homoserine

Cody Stephans
Augustana College, Rock Island Illinois

Dr. Lori Scott
Augustana College, Rock Island Illinois

Cody Stephans

BIOL 375 – Scott

06 February 2015

Meiothermus ruber Mrub_0976 and Mrub_1641 share the same functions as *Escherichia coli* b3940 and b3433 in the biosynthesis of homoserine

Introduction: *Meiothermus ruber* is an organism that is not very well studied. It was originally thought to come from the genus *Thermus* in 1975 (Tindall *et al*. 2010). But it was discovered later (1996) that *M. ruber* should be re-categorized under the *Meiothermus* genus. *M. ruber* prefers warm temperatures (60°C), but slightly less warm than those of *Thermus* (Tindall *et al*. 2010).

Bioinformatics tools were the main source of information for this research project. Bioinformatics is a branch of science in which biological data is gathered, analyzed, and compared to solve complex problems in life sciences (Pujari 2014). Recent growth in molecular genetics data such as: complete genomes of biological species, protein sequences, metabolic pathways, and protein structures have made biotechnology a hot topic for research (Pujari 2014). For this particular project, bioinformatics tools were used to compare the *Meiothermus ruber* genes *Mrub*_0976 and *Mrub*_1641 to the *E. coli* genes b3940 and b3433 to determine if the functions were similar. Bit scores and E-values were used often as a way to measure the similarity between the two genes (Pujari 2014). A bit score is a measure of how close the alignment of amino acids is to each other. A higher bit score indicates a better alignment (Pujari 2014). E-values are a way to measure the likelihood of sequences aligning just by random chance. A very low E-value (E<1.0e-5) is needed in order to conclude that the similarity in sequences is not due to random chance (Pujari 2014).

One of the main functions of *E. coli* b3940 and b3433 is participating in the biosynthesis of homoserine. Homoserine is an amino acid that differs from serine by an additional (-CH_2-) in its backbone (Cambell 2012). Homoserine is not a proteinogenic amino acid; meaning it is not one of the 'common' amino acids encoded by DNA. However, it is still considered an amino acid. Homoserine is an intermediate in the biosynthesis of two essential amino acids: methionine and threonine (Chassagnole *et al*. 2001). An additional two reactions are required to produce threonine from homoserine. Methionine requires an additional 4.

Threonine is an essential amino acid for birds and mammals (Chassagnole *et al*. 2001). Our bodies cannot produce threonine so we need to consume threonine through our diets (Cambell 2012). Threonine is important for the formation of tooth enamel, collagen, and the production of antibodies. It also metabolizes fat and prevents the buildup of fat in the liver. Threonine also plays a role in maintaining the normal function of our central nervous system, cardiovascular system, liver, and immune system.

The name of the protein that *E.coli* b3940 codes for is aspartate kinase (Chassagnole *et al*. 2001). This enzyme is 810 amino acids long. It catalyzes the reaction of aspartate and ATP → aspartyl-4-phopsphate. Its cellular location is in the cytoplasmic membrane. *E.coli* b3940 is located between DNA coordinates 4129835-4132267. This gene is not part of an operon in *E.coli*. According to the chromosome viewer colored by KEGG, this gene and the one next to it go in the same direction but are different colors so they most likely perform different functions. The name of the protein that *E.coli* b3433 codes for is aspartate semialdehyde dehydrogenase (Chassagnole *et al*. 2001). This enzyme is 367 amino acids long. It catalyzes the reaction of aspartyl-4-phosphate and NADPH → aspartate semialdehyde (Chassagnole *et al*. 2001). Its cellular location is the cytoplasm. *E.coli* b3433 is located between DNA coordinates 3573775-

3574878. This gene is not part of an operon in *E.coli*. According to the chromosome viewer colored by KEGG, this gene and the ones next to it are either are colored differently or there is a gap in between them, so they most likely perform different functions.

By studying genes in *M. ruber*, a poorly-researched organism, we can gain more information on bacteria in general (Tindall *et al*. 2010). This will help us understand the functions of different bacteria and the metabolic pathways that occur in them. It could also give us a better understanding of diseases involving bacteria (Tindall *et al*. 2010). For this research project, I predict that if *Mrub*_1641 and *Mrub*_0976 are similar to *E. coli* b3433 and b3940 respectively, in structure, location, and protein families, then *Mrub*_1641 and *Mrub*_0976 are involved in homoserine biosynthesis.

Methods: All of the results were gathered using the bioinformatics programs within GENI-ACT. The results were recorded in the GENI-ACT lab notebook. This link geni-act.org/education/main identifies a list as well as a brief description of all of the bioinformatics programs used in GENI-ACT. However, there were several deviations from the GENI-ACT instructions. When analyzing *E.coli* with the T-coffee program, 20 sequences were analyzed instead of 10. Also, proteobacteria was excluded from the search to allow more variety in the results. No paralog was done for *E. coli* as it was already known that *E. coli* doesn't have any. A phylogenic tree was not produced for *E. coli* because it was already known that *E. coli* has not recently undergone horizontal gene transfer. Finally, a BLAST of *E.coli* was run against *M. ruber* to determine the similarity of their amino acid sequences.

Results:

***E. coli* b3433 and *Mrub*_1641**

A protein BLAST comparison between *E. coli* b3433 and *Mrub*_1641 produced an E-value of 2e-29 and a bit score of 114 (table 1). These numbers indicate that the amino acid sequences are very similar and this is not likely due to chance. Table 1 indicates that both of these genes are involved in homoserine and threonine metabolism. This pathway was found on KEGG. A Pfam search showed that both genes shared the semialdehyde dehydrogenase family PF01118 (table 1). Both E-values (4.6e-24 and 5.6e-28) indicated that the similarities were unlikely due to chance. The CDD program indicated that these genes share a common COG number (COG0136) with E-values of (1.77e-142 and 6.33e-144) (table 1). These results indicate that both *E.coli* b3433 and *Mrub*_1641 are also known as aspartate-semialdehyde dehydrogenase. Table 1 indicates that a TIGRfam search shows *E.coli* b3433 and *Mrub*_1641 have a protein family called aspartate-semialdehyde dehydrogenase (TIGR01296 and TIGR01745) and both of the E-values are significant (7.8e-283 and 7.6e-155). Furthermore, both *E.coli* b3433 and *Mrub*_1641 possess the E.C. number of 1.2.1.11 (table 1). Both these sets of results indicate that these genes are aspartate-semialdehyde dehydrogenase. Although the crystal structure of aspartate-semialdehyde dehydrogenase for *M.ruber* has not yet been determined, the PDB results in table 1 show that *E.coli* b3433 and *Mrub*_1641 share a PDB result of aspartate-semialdehyde dehydrogenase (2YV3 and 1BRM). Both E-values are significant for the PDB results (0 and 8.5e-127) which means the similarities in amino acid sequences were not due to chance.

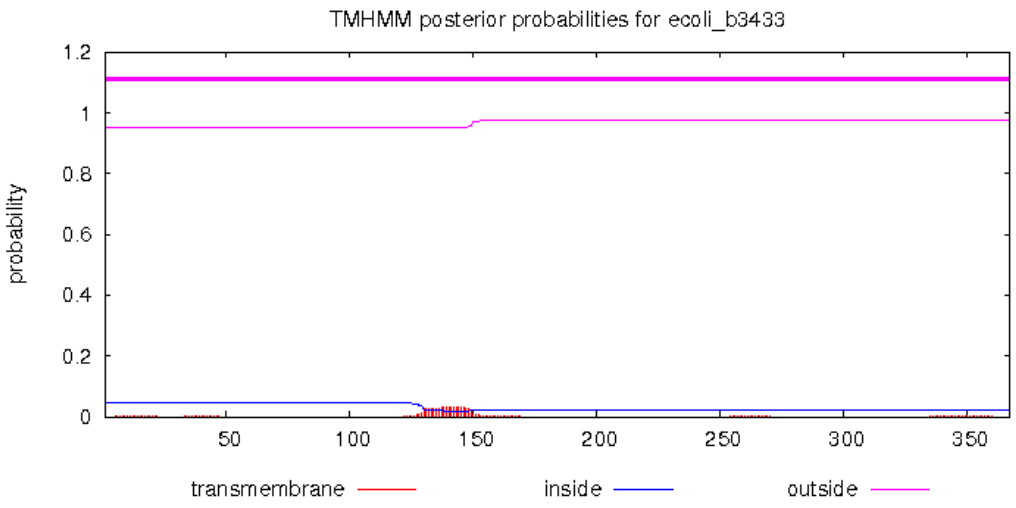
Table 1: *E. coli* b3433 and *Mrub*_1641 are orthologs

Description of evidence collected	*E. coli*	*M. ruber*
Cellular location	Cytoplasm	
BLAST *E. coli* against *M. ruber*	Score: 114 bits; E-value: 2e-29	
KEGG pathway	Homoserine and threonine metabolism	

Pfam- protein family	PF01118 Semialdehyde dehydrogenase family	
	E=4.6e-24	E=5.6e-28
CDD (COG category)	COG0136 Aspartate-semialdehyde dehydrogenase	
	E=1.77e-142	E=6.33e-144
TIGRfam – protein family	Aspartate-semialdehyde dehydrogenase	
	TIGR01296 E=7.8e-283	TIGR01745 E=7.6e-155
E.C. number	E.C.1.2.1.11 Aspartate-semialdehyde dehydrogenase	
PDB	Aspartate-semialdehyde dehydrogenase	
	1BRM E=0e+00	2YV3 E=8.5e-127
Gram stain	Gram-negative	

Both *E.coli* b3433 and *Mrub*_1641 showed no evidence of transmembrane helices (Figure 1). Figure 2 shows that both genes have no signal peptides, indicating a cytoplasmic location. *E.coli* b3433 and *Mrub*_1641 share the protein family PF01118. A logo of conserved sequences for this Pfam can be seen in figure 3. A comparison of KEGG pathways is shown in figure 4. Both *E.coli* b3433 and *Mrub*_1641 are involved in homoserine biosynthesis (aspartate → homoserine). A snapshot of the metabolic pathway for aspartate → homoserine is provided in figure 5. Figure 6 shows evidence that neither of these genes are part of an operon.

Panel A



Panel B

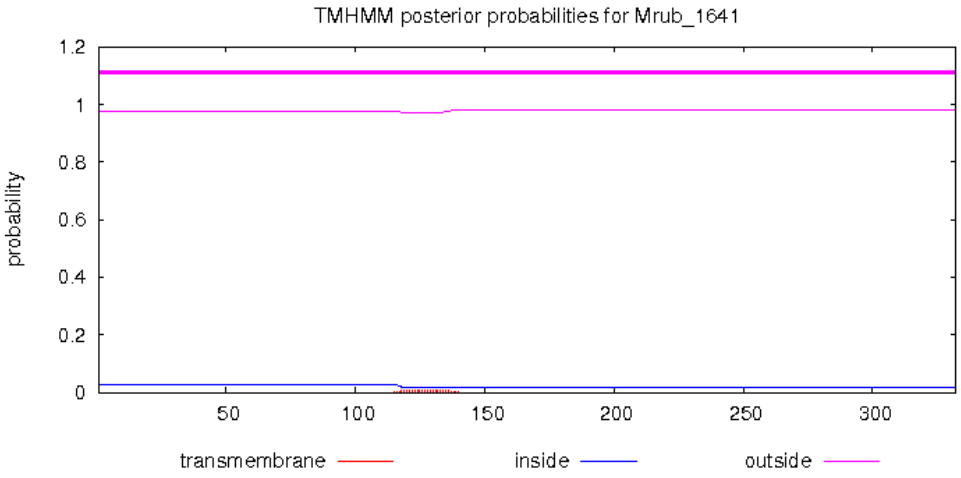
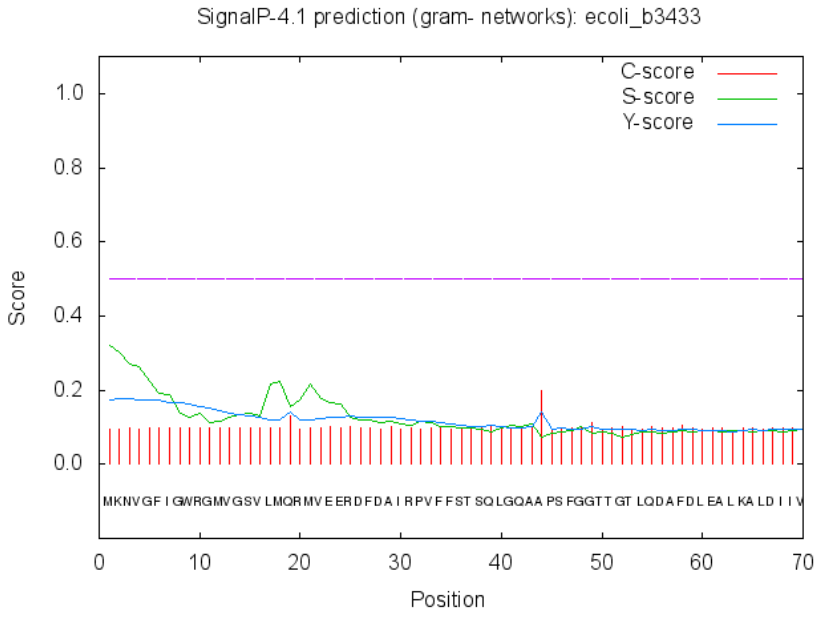


Figure 1: *E. coli* b3433and *Mrub*_1641 do not contain any transmembrane helices; a cytoplasmic location is predicted. Panel A = *E. coli* b3433; Panel B = *Mrub*_1641. TMHMM Server v. 2.0 created this hydropathy plot. http://www.cbs.dtu.dk/services/TMHMM

Panel A



Panel B

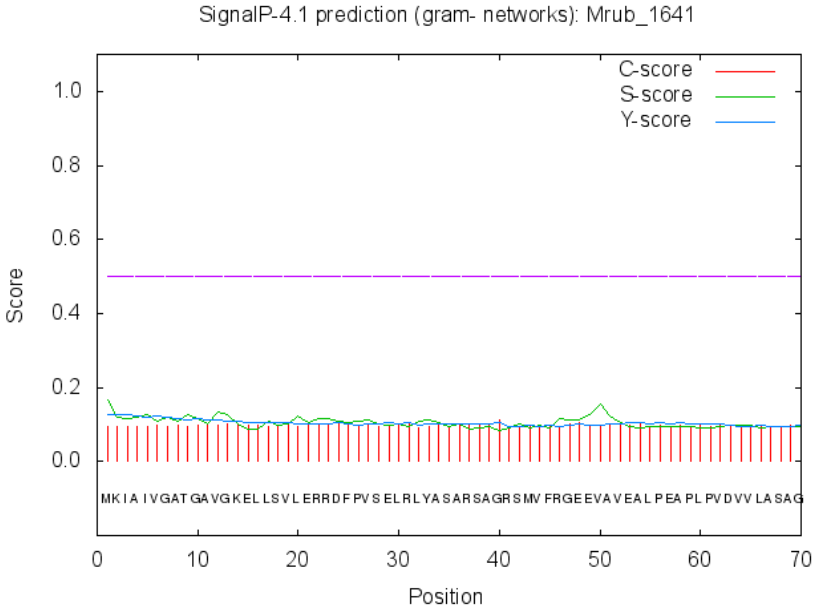


Figure 2: *E. coli* b3433 and *Mrub*_1641 do not contain signal peptides. A cytoplasmic location is predicted. Panel A = *E. coli* b3433; Panel B = *Mrub*_1641. SignalP 4.1 Server created this plot. http://www.cbs.dtu.dk/services/SignalP

Panel A

Panel B

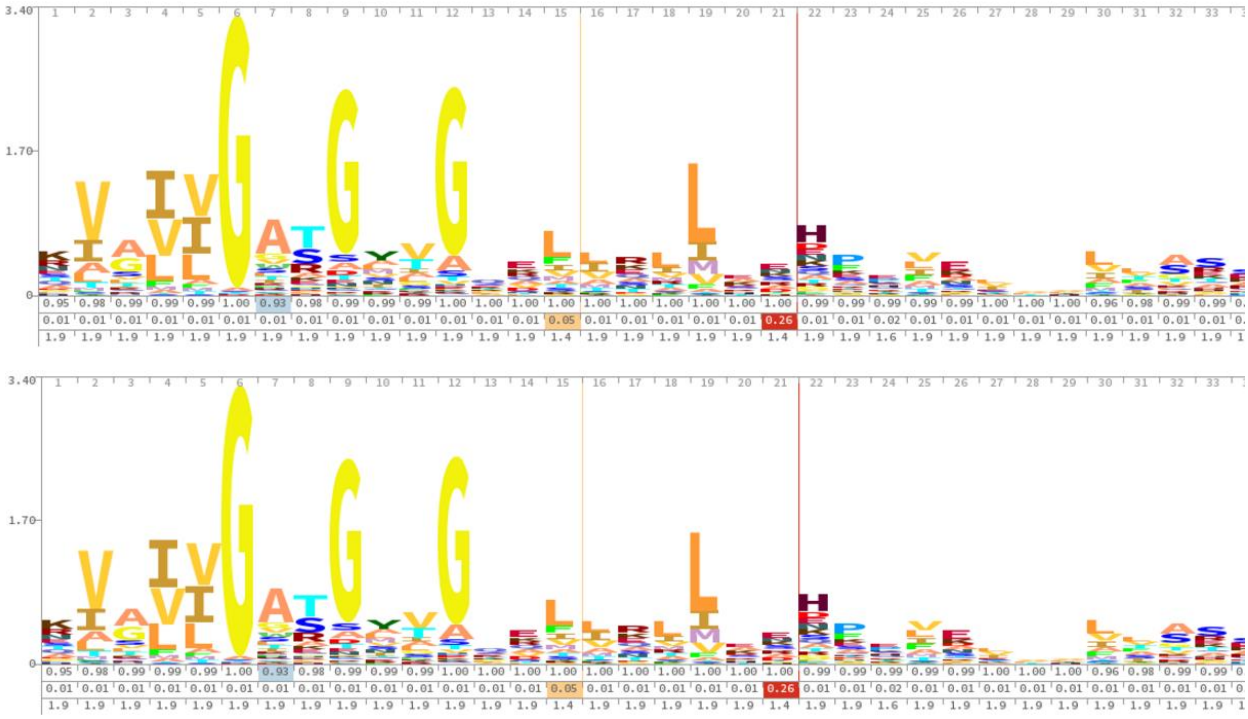


Figure 3: *E.coli* b3433 and *Mrub*_1641 have identical conserved Pfam HMM logos. Both *E.coli* b3433 and *Mrub*_1641 share the protein family PF01118 and therefore the HMM logos of conserved amino acid sequences are identical. Panel A= *E.coli* b3433; Panel B= *Mrub*_1641. Pfam HMM logo created this logo. http://pfam.sanger.ac.uk/search

Panel A

Panel B

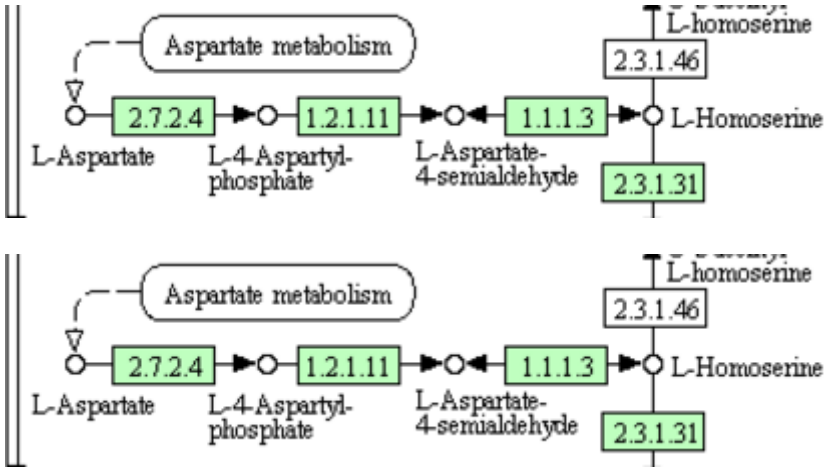
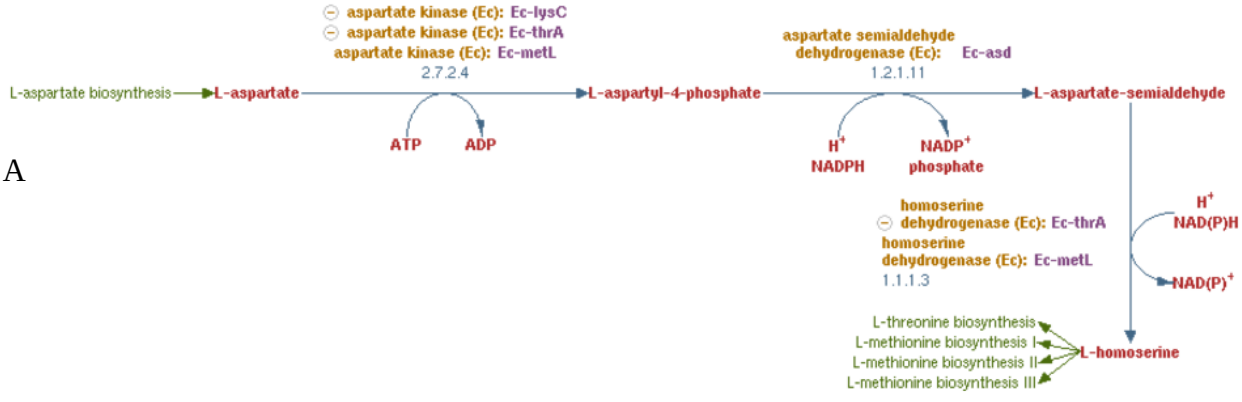


Figure 4: KEGG pathway showing the biosynthesis of homoserine. *E.coli* b3433 and *Mrub*_1641 are both identified by the E.C. number 1.2.1.11 (aspartate-semialdehyde dehydrogenase). Panel A = *E.coli* b3433; Panel B = *Mrub*_1641. Pathway provided by KEGG Pathway Database. http://www.genome.jp/kegg/pathway.html

Panel A



Panel B

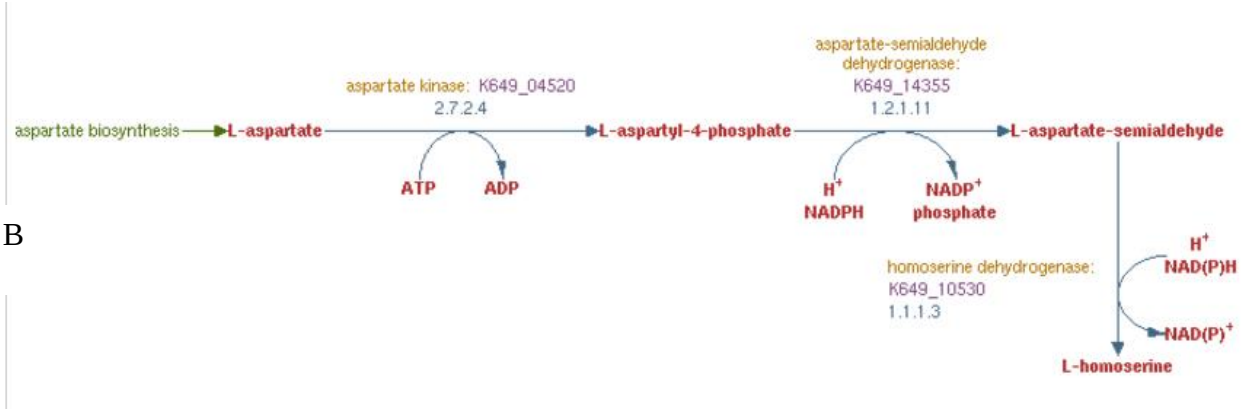


Figure 5: Metabolic pathways for aspartate → homoserine. *E.coli* b3433 and *Mrub*_1641 are both identified by the E.C. number 1.2.1.11 (aspartate-semialdehyde dehydrogenase). Panel A = *E.coli* b3433; Panel B = *Mrub*_1641. Panel A provided by EcoCyc E. Coli Database http://ecocyc.org; Panel B provided by Metacyc Metabolic Pathway Database. http://metacyc.org

Panel A

Panel B

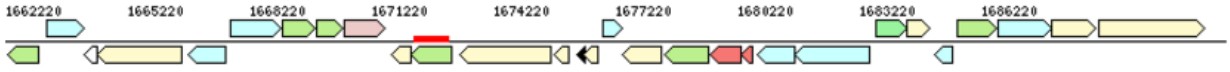


Figure 6: Neither *E.coli* b3433 nor *Mrub*_1641 are part of an operon. The genes directly next to *E.coli* b3433 or *Mrub*_1641 are either colored differently or there is a gap between them. Therefore, the genes likely perform different functions and are not part of an operon. Panel A= *E.coli* b3433; Panel B= *Mrub*_1641. Diagram provided by IMG/EDU gene finder http://img.jgi.doe.gov/cgi-bin/edu/main.cgi?section=FindGenes&page=geneSearch

***E. coli* b3940 and *Mrub*_0976**

A protein BLAST comparison between *E. coli* b3940 and *Mrub*_0976 produced an E-value of 1e-25 and a bit score of 105 (table 2). These numbers indicate that the amino acid sequences are

very similar and this is not likely due to chance. Table 2 indicates that both of these genes are involved in homoserine and threonine metabolism. This pathway was found on KEGG. A Pfam search showed that both genes shared the amino acid kinase family PF00696 (table 2). Both E-values (4.9e-44 and 7.6e-51) indicated that the similarities were unlikely due to chance. The CDD program indicated that these genes share a common COG number (COG0527) with E-values of (1.93e-133and 5.63e-139) (table 2). These results indicate that both *E.coli* b3940 and *Mrub*_0976 are also known as aspartate kinase. Table 2 indicates that a TIGRfam search shows *E.coli* b3940 and *Mrub*_0976 have a protein family called aspartate kinase (TIGR00657) and both of the E-values are significant (3.6e-153 and 1.3e-40). Furthermore, both *E.coli* b3940 and *Mrub*_0976 possess the E.C. number of 2.7.2.4 (table 2). Both these sets of results indicate that these genes are aspartate kinase. Although the crystal structure of aspartate kinase for *M.ruber* has not yet been determined, the PDB results in table 2 show that *E.coli* b3940 and *Mrub*_0976 share a PDB result of aspartate kinase (3L76 and 1EBF). Both E-values are significant for the PDB results (5.43e-86 and 2.7e-43) which means that the similarities in amino acid sequences were not due to chance.

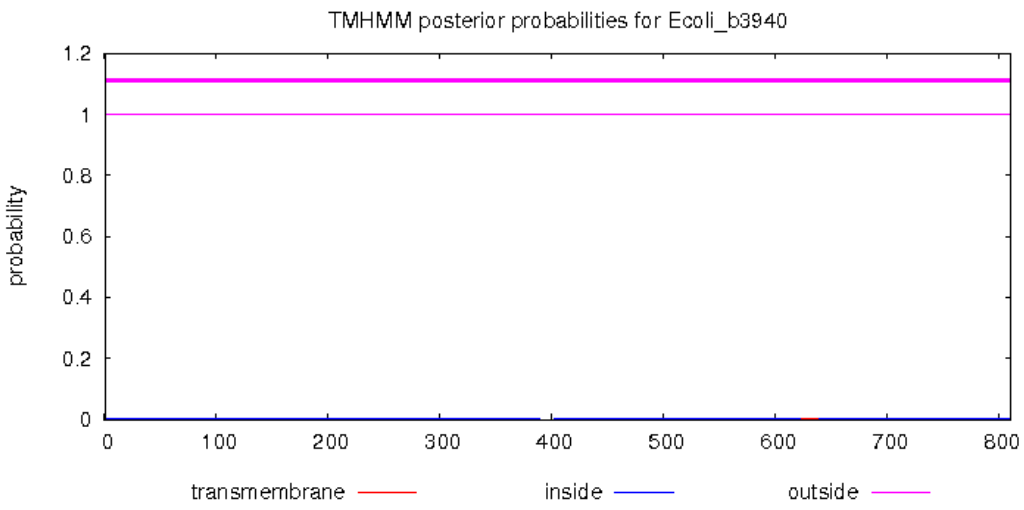
Table 2: *E. coli* b3940 and *Mrub*_0976 are orthologs

Description of evidence collected	*E. coli*	*M. ruber*
Cellular location	Cytoplasmic membrane	cytoplasm
BLAST *E. coli* against *M. ruber*	Score: 105 bits; E-value: 1e-25	
KEGG pathway	Homoserine and threonine metabolism	
Pfam- protein family	PF00696 Amino acid kinase family	
	E=4.9e-44	E=7.6e-51
CDD (COG category)	COG0527 Aspartate kinase	

	E=1.93e-133	E=5.63e-139
TIGRfam – protein family	TIGR00657 Aspartate kinase	
	E=3.6e-153	E=1.3e-40
E.C. number	E.C.2.7.2.4 Aspartate kinase	
PDB	Aspartate kinase	
	1EBF E=5.43e-86	3L76 E=2.7e-43
Gram stain	Gram-negative	

Both *E.coli* b3940 and *Mrub*_0976 showed no evidence of transmembrane helices (Figure 7). Figure 8 shows that both genes have no signal peptides, indicating a cytoplasmic location. *E.coli* b3940 and *Mrub*_0976 share the protein family PF00696. A logo of conserved sequences for this Pfam can be seen in figure 9. A comparison of KEGG pathways is shown in figure 10. Both *E.coli* b3940 and *Mrub*_0976 are involved in homoserine biosynthesis (aspartate → homoserine). A snapshot of the metabolic pathway for aspartate → homoserine is provided in figure 11. Figure 12 shows evidence that neither of these genes are part of an operon.

Panel A



Panel B

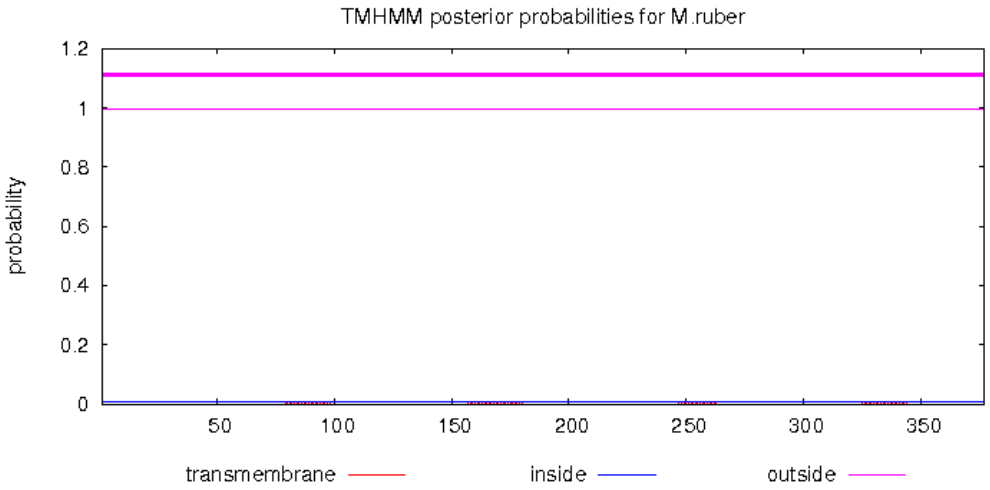
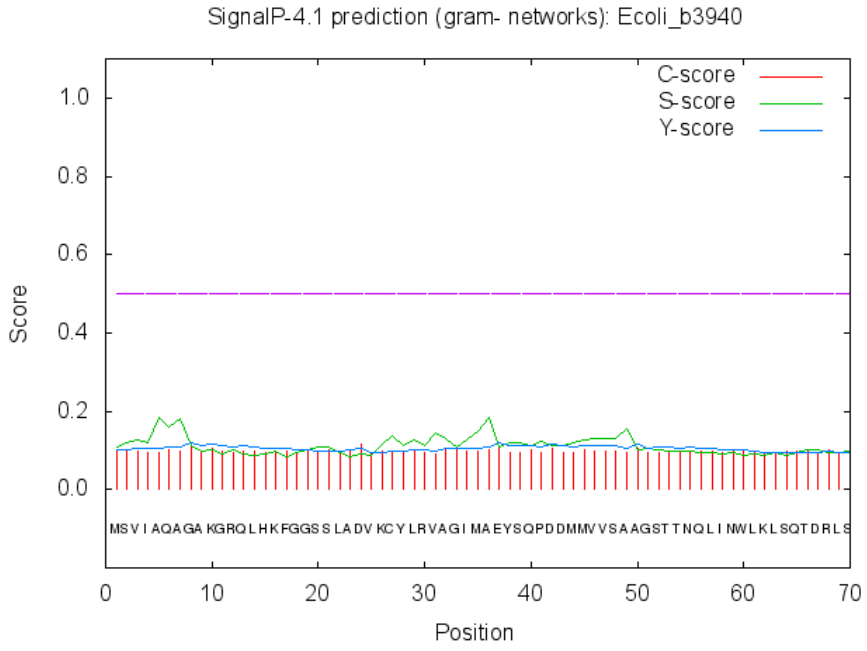


Figure 7: *E. coli* b3940 and *Mrub*_0976 do not contain any transmembrane helices; a cytoplasmic location is predicted. Panel A = *E. coli* b3940; Panel B = *Mrub*_0976. TMHMM Server v. 2.0 created this hydropathy plot. http://www.cbs.dtu.dk/services/TMHMM

Panel A



Panel B

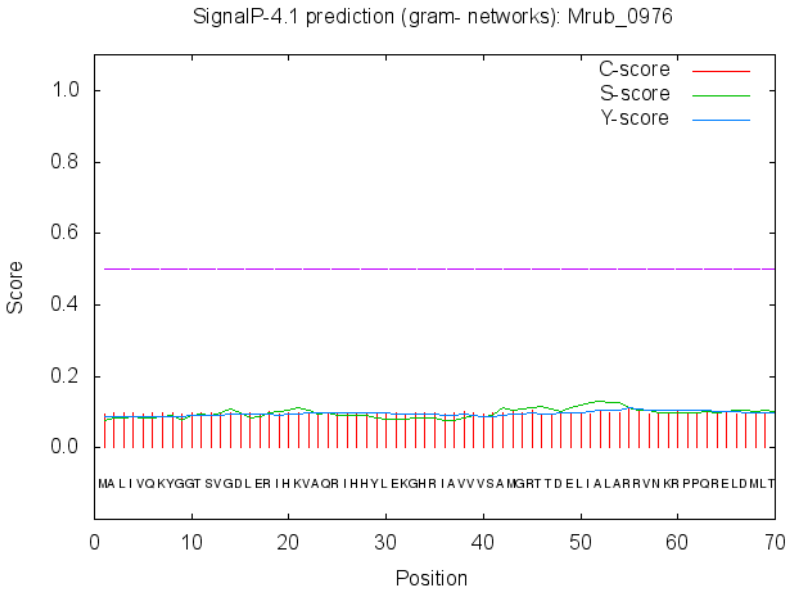


Figure 8: *E. coli* b3940 and *Mrub*_0976 do not contain signal peptides. A cytoplasmic location is predicted. Panel A = E. coli b3940; Panel B = Mrub_0976. SignalP 4.1 Server created this plot. http://www.cbs.dtu.dk/services/SignalP

Panel A

Panel B

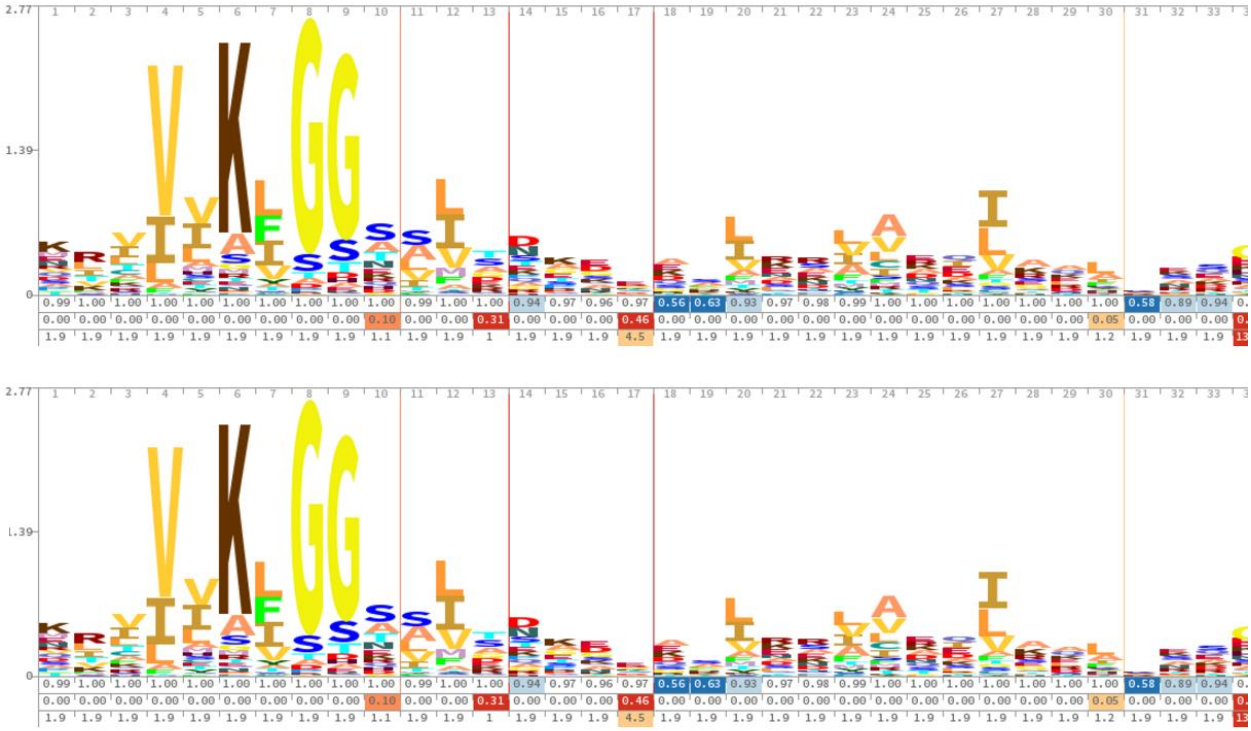


Figure 9: *E.coli* b3940 and *Mrub*_0976 have identical conserved Pfam HMM logos. Both *E.coli* b3940 and *Mrub*_0976 share the protein family PF00696 and therefore the HMM logos of conserved amino acid sequences are identical. Panel A= *E.coli* b3940; Panel B= *Mrub*_0976. Pfam HMM logo created this logo. http://pfam.sanger.ac.uk/search

Panel A

Panel B

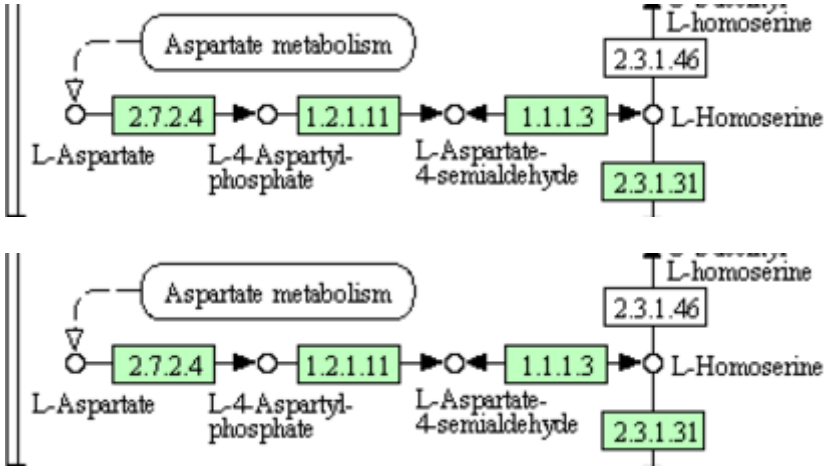
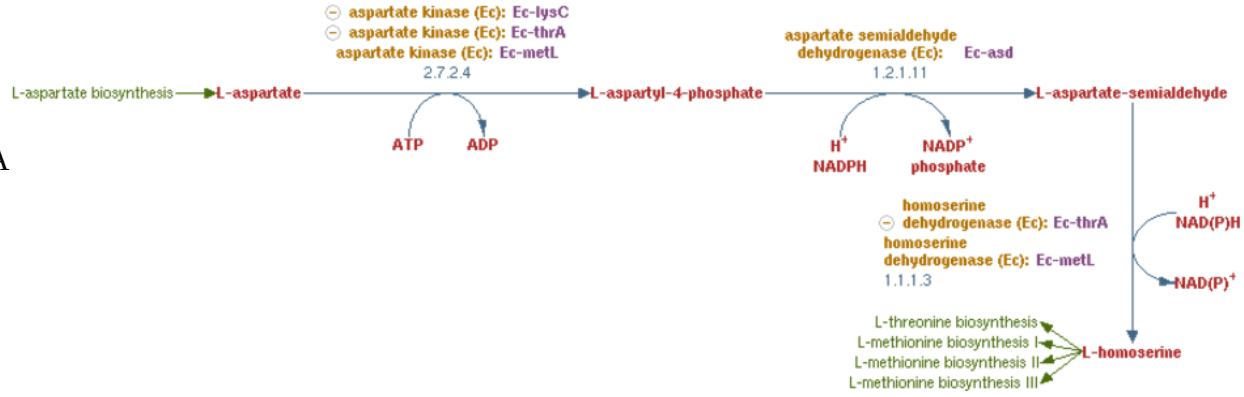


Figure 10: KEGG pathway showing the biosynthesis of homoserine. *E.coli* b3940 and *Mrub*_0976 are both identified by the E.C. number 2.7.2.4 (aspartate kinase). Panel A = *E.coli* b3940; Panel B = *Mrub*_0976. Pathway provided by KEGG Pathway Database. http://www.genome.jp/kegg/pathway.html

Panel A



Panel B

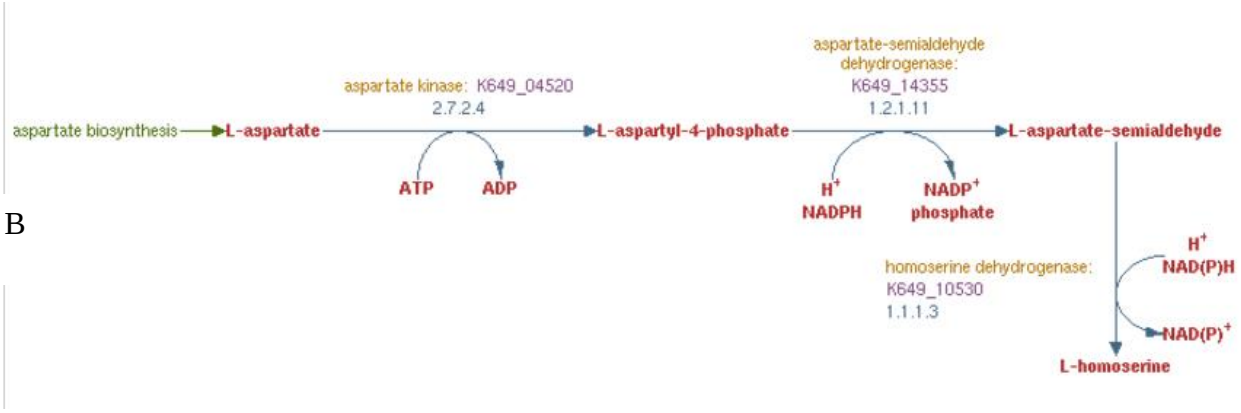
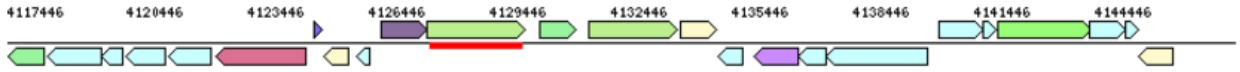


Figure 11: Metabolic pathways for aspartate → homoserine. *E.coli* b3940 and *Mrub*_0976 are both identified by the E.C. number 2.7.2.4 (aspartate kinase). Panel A = *E.coli* b3940; Panel B = *Mrub*_0976. Panel A provided by EcoCyc E. Coli Database http://ecocyc.org; Panel B provided by Metacyc Metabolic Pathway Database. http://metacyc.org

Panel A



Panel B

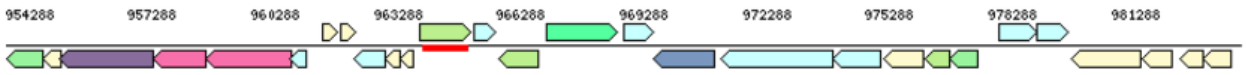


Figure 12: Neither *E.coli* b3940 nor *Mrub*_0976 are part of an operon. The genes directly next to *E.coli* b3940 or *Mrub*_0976 are either colored differently or there is a gap between them. Therefore, the genes likely perform different functions and are not part of an operon. Panel A= *E.coli* b3940; Panel B= *Mrub*_0976. Diagram provided by IMG/EDU gene finder http://img.jgi.doe.gov/cgi-bin/edu/main.cgi?section=FindGenes&page=geneSearch

Conclusion: My hypothesis for this project was: If *Mrub*_1641 and *Mrub*_0976 are similar to *E. coli* b3433 and b3940 respectively, in structure, location, and protein families, then *Mrub*_1641 and *Mrub*_0976 are involved in homoserine biosynthesis. This hypothesis was supported. By utilizing the bioinformatics tools in GENI-ACT, I compared the amino acid sequence of *E. coli* b3433 and *Mrub*_1641, and also *E.coli* b3940 and *Mrub*_0976. I also analyzed the results of many tests located in GENI-ACT. *E. coli* b3433 and *Mrub*_1641 are: located in the cytoplasm, involved in homoserine biosynthesis, have the same protein families, have the same COG numbers, and have the same EC number. *E.coli* b3940 and *Mrub*_0976 are: located in the cytoplasm, involved in homoserine biosynthesis, have the same protein families, have the same COG numbers, and have the same EC number. After analyzing all of this information, it is clear that *Mrub*_1641 and *Mrub*_0976 are involved in homoserine biosynthesis.

Work Cited

1. Chassagnole C, Rais B, Quentin E, Fell DA, Mazat JP. 2001. An integrated study of threonine-pathway enzyme kinetics in *Escherichia coli*. Biochem J 356:415-23
2. Pujari S. 2015. History of Biotechnology: Ancestors, Modern Applications, and Birth of Bioinformatics.
3. Tindall BJ, *et al*. 2010. Complete genome sequence of *Meiothermus ruber* type strain (21^T). Stand in Gen Sci 3:26-36.
4. Cambell W. 2012. Threonine: The essential amino acid that helps to maintain the proper protein balance in the body.
5. Alvarez E, Ramon F, Magan C, Diez E. 2004. L-cystine inhibits aspartate-beta-semialdehyde dehydrogenase by covalently binding to the essential 135Cys of the enzyme. Biochim Biophys Acta 1696: 23-29.