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

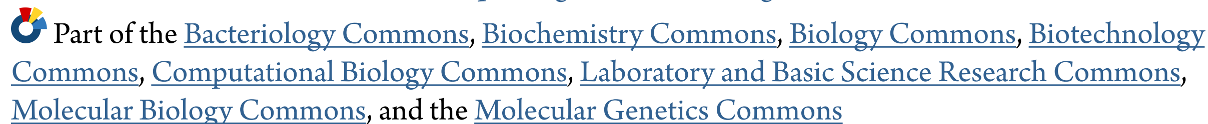
2017

Annotation and Identification of Several Glycerolipid Metabolic Related Ortholog Genes; Mrub_0437, Mrub_1813 and Mrub_2759 In The Organism *Meithermus Ruber* and Their Predicted Respective Orthologs b3926, b4042 and BO514 Found In *E.coli*.

Abdul Rahman Abdul Kader
Augustana College, Rock Island Illinois

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

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


Part of the Bacteriology Commons, Biochemistry Commons, Biology Commons, Biotechnology Commons, Computational Biology Commons, Laboratory and Basic Science Research Commons, Molecular Biology Commons, and the Molecular Genetics Commons

Augustana Digital Commons Citation
Abdul Kader, Abdul Rahman and Scott, Dr. Lori R.. "Annotation and Identification of Several Glycerolipid Metabolic Related Ortholog Genes; Mrub_0437, Mrub_1813 and Mrub_2759 In The Organism *Meithermus Ruber* and Their Predicted Respective Orthologs b3926, b4042 and BO514 Found In *E.coli*." (2017). *Meiothermus ruber Genome Analysis Project.*
http://digitalcommons.augustana.edu/biolmruber/29

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.

Annotation and Identification of Several Glycerolipid Metabolic Related Ortholog Genes; Mrub_0437, Mrub_1813 and Mrub_2759 In The Organism *Meithermus Ruber* and Their Predicted Respective Orthologs b3926, b4042 and BO514 Found In *E.coli*.

Abdul W. Kader
Dr. Lori R. – Molecular Genetics
Augustana College 639 38th Street, Rock Island, IL 61201

Introduction:

Significance of *M. Ruber*:

Meiothermus Ruber received its current name in 1996 when transferred from the genus Thermus into the then novel genus Meiothermus by Nobre et al. (1996). Currently, there are eight species placed in the genus Meiothermus (Euzéby 2007). The genus name derives from the Greek words 'meion' and 'thermos' meaning 'lesser' and 'hot' to indicate an organism in a less hot place (Nobre et al 1996, Euzéby 2007). The species epithet derives from the Latin word 'ruber' meaning red, to indicate the red cell pigmentation (Loginova, Egorova 1984, Euzéby 2007). Members of the genus Meiothermus were isolated from natural hot springs and artificial thermal environments (Nobre et al 1996) in Russia (Loginova, Egorova 1984), Central France (Albuquerque et al, 2007), both Northern and Central Portugal (Tenreiro et al 1995, Pires et al 2005), North-Eastern China (Zhang et al, 2010), Northern Taiwan (Chen et al, 2002) and Iceland (Chung et al, 1997) Interestingly, the genus Meiothermus is heterogeneous with respect to pigmentation. The yellow pigmented species also form a distinct group on the basis of the 16S rRNA gene sequence similarity, with the red/orange pigmented strains forming two groups, one comprising M. silvanus and the other the remaining species (Pires et al 2005, Zhang et al, 2010). The annotation of these organisms' genes can lead to further understanding and functionality of these proteins, and can potentially uncover interesting discoveries whether they be medical or industrially advantageous. For example if an understudied bacterial species could produce lipids at a significantly increased rate due to not containing certain proteins, then it could potentially be studied as a potential bio-lipid-synthetic production medium, discoveries like these are only made when we investigate more of these understudied organisms. Due to *M. Rubers* resilience to heat perhaps a method of ensuring cell health in high heat environments could be developed. Through the use of *E. coli* as a control, we can compare these understudied genes in *M. Ruber* to the very well-known and understood proteins in *E. coli.*

Significance of Glycerolipids:

Glycerolipids serve an array of significant roles in biological organisms they primarily through the degradation and synthesis of energy storage related substrates they accomplish this primarily through being an intermediary step in the synthesis of Triacylglycerol. Triglycerides are the main constituents of body fat in humans and other animals, as well as vegetable fat.(Nelson et al, 2000) They are also present in the blood to enable the bidirectional transference of adipose fat and blood glucose from the liver, and are a major component of human skin oils. (Lampe et al

1983) Further more they can feed into the Glycolytic pathway for energy purposes; this is accomplished by mrub_2759 which synthesizes 2-Phospho-D-glycerate which can directly go through to the glycolytic pathway, furthermore it is also related to Glycerophospholipid metabolism and fatty acid synthesis/degradation as seen in figure 1 below. (Kanehisa et al, 2000) Therefor the identification of orthologs in *M. Ruber* is vital in understanding how this organism processes these energy related proteins.

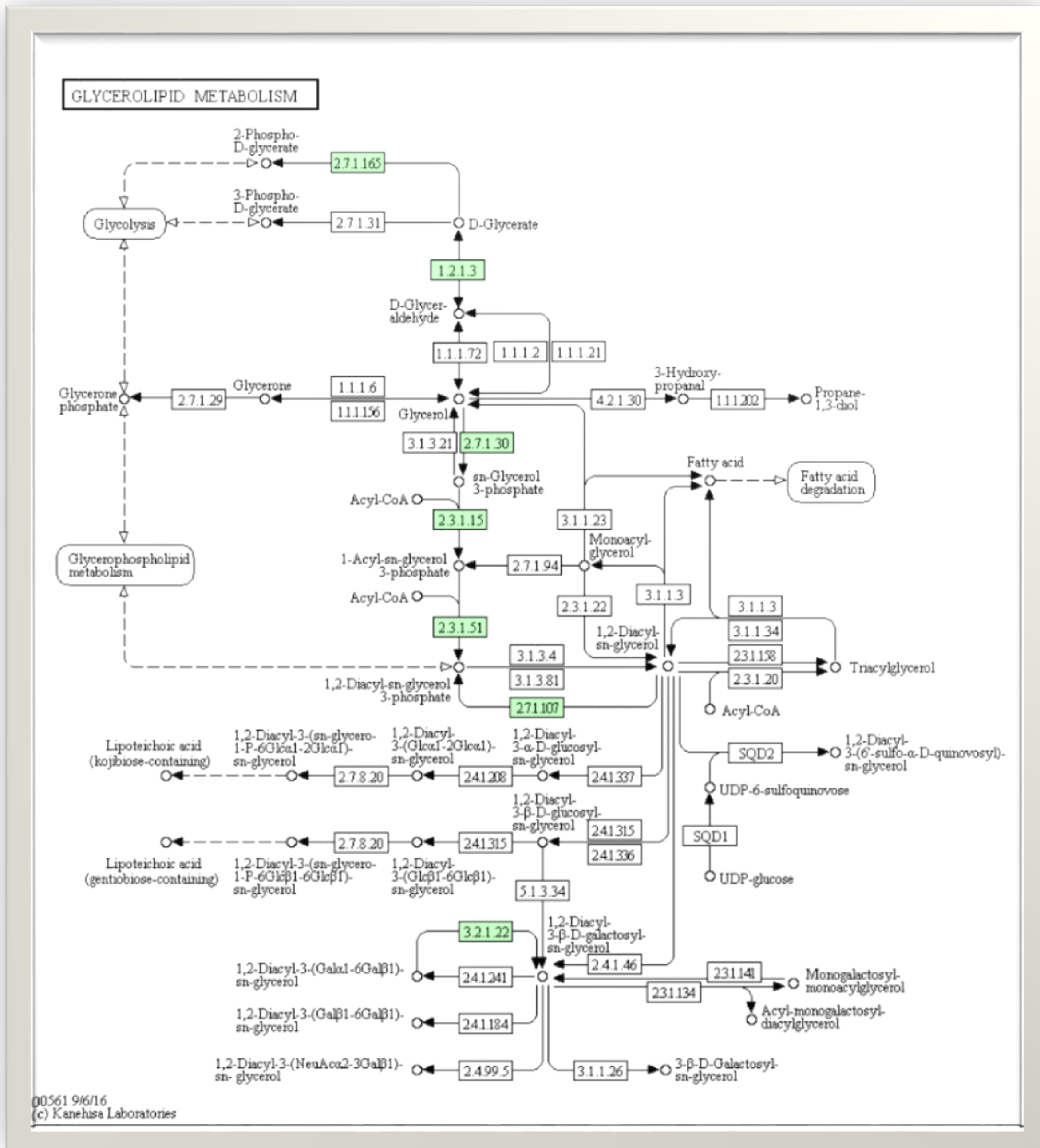


Figure 1: Lipoglycerol metabolism, courtesy of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database at http://www.genome.jp/kegg/pathway.html

Bioinformatic Database and Tools: Bioinformatics has become a crucial tool in modern biological research. In experimental molecular biology, bioinformatics allows for the analysis and comparison of large sets of raw data into useful results. In the field of genetics and genomics, it aids in sequencing and annotating genomes and their observed mutations. It plays a role in the text mining of biological literature and the development of biological and gene ontologies to organize and query biological data. It also plays a role in the analysis of gene and protein expression and regulation. Bioinformatics tools aid in the comparison of genetic and genomic data and more generally in the understanding of evolutionary aspects of molecular biology. At a more integrative level, it helps analyze and catalogue the biological pathways and networks that are an important part of systems biology. In structural biology, it aids in the simulation and modeling of DNA. (Sim et al, 2012) (Dawson et al, 2016) (Kmiecik et al, 2016) (Wong, 2016)

Significance of E.coli:

Because of their comparative simplicity, prokaryotic cells (bacteria) are ideal models for studying many fundamental aspects of biochemistry and molecular biology. The most thoroughly studied species of bacteria is E. coli, which has long been the favored organism for investigation of the basic mechanisms of molecular genetics. Most of our present concepts of molecular biology—including our understanding of DNA replication, the genetic code, gene expression, and protein synthesis—derive from studies of this humble bacterium. (Cooper 2000) Because this organism's full genome has been sequenced and because it is so well understood we will use it as a control, and fill in the gaps and make comparisons from our less studied M. Ruber organism.

Goal:

The aim of this experiment is to form an annotation and identification of several glycerolipid metabolic related ortholog Genes; Mrub_0437, Mrub_1813 and Mrub_2759 in the organism *Meithermus Ruber* and their predicted respective orthologs b3926, b4042 and BO514 Found In E.coli. This will be accomplished by comparing data that we collected from several bioinformatics websites and comparing it to the well-studied E. Coli strain as a control.

Methods:

We began by Identifying the top two hits related to the *E. coli* and *M. ruber* genes. In Module 2, a BLASTp using the amino acid sequence of each *E. coli* and *M. ruber* gene in their respective systems was preformed, assuming these sequences have the correct start codon. For the *E. coli* BLASTp, it is recommended that you "exclude" *E. coli* genes from the BLASTp output, because there are so many in the NCBI database. The top two hits for the *E. coli* or *M. ruber* genomes were recorded, as explained in the module.

In module 7: Duplication & Deletion the goal was to determine if there are paralogs of the *M. ruber* genes in the genome this step was done early to have an estimate of the amount of annotations that would need to be done before beginning the process

to assess time management. Multiple genes in the genome could produce confusing results in future modules if not identified in advance. Use the KEGG map to identify if duplicate genes are present in the genome; duplicate genes will be revealed as two or more locus tags for an enzyme/protein on a KEGG map. In Dr. Scott's class, all paralogs must be annotated along with the original set of genes. If paralogs are identified that show high sequence homology to the original gene, then these genes will also need to be annotated, several of our genes had paralogs on the KEGG path way but when blasted no results were found.

We then followed GENI-ACT protocol to collect bioinformatics data on the genes in *E. coli* and *M. ruber*, which includes Module 2/Sequence Similarity (CDD), Module 3/Cellular Localization, Module 5/Functional Similarity, Module 6/Metabolic Pathways, and the Gene Context section of Module 8/Horizontal Gene Transfer. The methods on how to use the various bioinformatics programs are included in the GENI-ACT modules. The next step was then to compare the annotation pages of both organisms to determine if they are actually orthologs. To determine if horizontal gene transfer occurred Module. Modified instructions were given by the primary investigator through the use of Phylogeny.fr, The resulting phylogentic tree to remove the branch numbers and use the "cladogram" format. Upload image into Module 8. Follow the GENI-ACT instructions for the Gene Context and Heat Map (%GC map). Upload the indicated information.

Overview of the different bioinformatics tools used in GENI-ACT

Module 1: Involves transferring sequence information from the Gene Details page into the lab notebook. Module 2: Involves comparing the amino acid sequence of the gene to proteins in a variety of databases to determine how similar they are. These comparisons will include identifying protein domain(s) within the protein, predicting possible active sites within the protein, and assigning the protein to a particular functional group called a COG. Module 3: Involves identifying signal peptides on the protein and predicting the location of the functional protein. It is not uncommon to get conflicting information between the various programs in this module. It was recommended that we research the literature prior to beginning module 3 as it gives a good reference on to where the protein should be located. look in the published literature to see where the protein is localized in the model organism, we then performed Module 3 to confirm the published work. Module 4 determines if the correct start codon was identified for the gene. This is probably a module that should be done prior to the other modules. Module 5 involves identifying orthologous proteins, protein family and domain(s) of the protein. This is a critical module for our

study, especially the Pfam and PDB databases. These are highly curated databases and a hit is very meaningful.

Module 6 involves putting the protein into the context of its metabolic pathway. Module 7 involves identifying if the gene has paralogs (duplicate genes in the *M. ruber* genome). Module 8 analyzes the phylogenetic history of the protein to determine if a gene was acquired by horizontal/lateral gene transfer, which includes creating a phylogenetic tree, analyzing the chromosome region for a potential operon and analyzing the chromosome region for %GC content. Studying the presence of a potential operon will take two approaches. First, compare the chromosome region between *E. coli* and *M. ruber* ("color by KEGG") to see if the GOI has the same adjacent genes in the two organisms. It is especially relevant if those flanking genes are in the same KEGG pathway. Second, follow the instructions in the GENI-ACT site to compare the gene context of *M. ruber* to other organisms. Bacterial chromosomes have undergone a high level of reorganization. If the gene order is maintained, even between related organisms, then this is highly suggestive of a coordinately expressed transcription unit (*i.e.*, operon). The the same Gene Context box to upload two different sets of data (*E. coli* vs *M. ruber* chromosome region and *M. ruber* vs other orthologs). (The methods sections has its own citation page)

Results:

Bioinformatics tool used	*E. coli* b3926	M. ruber Mrub_0537
BLAST *E. coli* against M. ruber	Score: 608 E-value 0	
CDD Data (COG category)	COG Number: COG0554	
	E-value: 0	E-value: 0
Cellular Localization	Cytoplasmic	
TIGRfam – protein family	TIGR TIGR01311	
	E-value: 0	E-Value: 0
Pfam – protein family	family of carbohydrate kinases, N-terminal domain	
	E-values 4.1e-96	E-values: 1.3e-80
Protein Database	1BO5	2DPN
	E-value: 0	E-value: 1.86687E-160
Enzyme commission number	EC 2.7.1.30	
KEGG pathway map	map00561	

As seen by the data summarized above we can conclude that these organisms are definitely orthologues. Most conclusively is that they are localized in the same location, code for the same protein match the same Tigrfam and Cog indicating similar functions. The low E values are very conclusive in making these assumptions. One thing to note was that the PDB top hit for both genes was different but both these PDBs coded for phosphate kinase domains so it is not a source of refuting evidence. Lastly due to both having accurate GC %'s we can confirm that the gene did not come to be as a result of HGT.

Bioinformatics tool used	***E. coli*** b4042	**M. ruber** Mrub_1813
BLAST *E. coli* against M. ruber	**Score:** 75.5 bits **E-value:** 2e-23	
CDD Data (COG category)	**COG Number:** COG0818	
	E-value: 3.42e-51	**E-value:** 6.86e-20
Cellular Localization	Cytoplasmic Membrane	
TIGRfam – protein family	**TIGR** N/A	
	E-value: N/A	**E-Value:** N/A
Pfam – protein family	PFK-like superfamily	
	E-values: 1.1E-5	**E-values:** 6.3e-33
Protein Database	Crystal structure of the integral membrane diacylglycerol kinase - wild-type	NMR Solution Structure of E. coli diacylglycerol kinase (DAGK) in DPC micelles
Enzyme commission number	2.7.1.107	
KEGG pathway map	map00561	

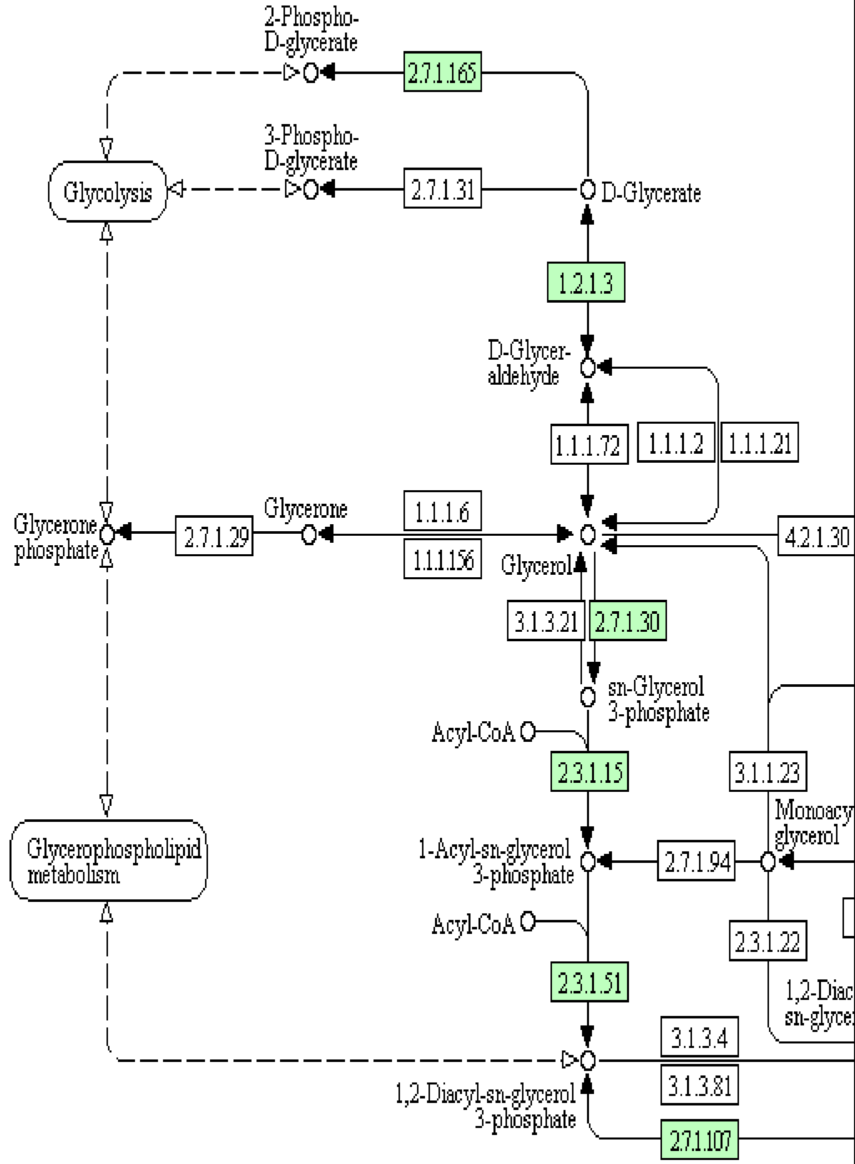
As seen by the data summarized above we can conclude that these organisms are definitely orthologues. Most conclusively is that they are localized in the cytoplasmic membrane and contain 3 helixes, confirmation of similar translation of the same protein cab be seen by the matching Cogs indicating similar functions. The low E values are very conclusive in making these assumptions. One thing to note was that the PDB top hit for both genes was different but both these PDBs coded for diacylglycerol phosphate kinase domains so it is not a source of refuting evidence. Lastly due to both having accurate GC %'s we can confirm that the gene did not come to be as a result of HGT. One thing to note is that no TIGRFAM hits were found but the PFAM is substantial enough in its confirmation of function.

Bioinformatics tool used	*E. coli* b0514	M. ruber Mrub_2759
BLAST *E. coli* against M. ruber	Score: 317 E-value: 0.0	
CDD Data (COG category)	COG Number: COG1929	
	E-value: 0e+00	E-value: 4.06e-137
Cellular Localization	Cytoplasmic	
TIGRfam – protein family	TIGR No Hit	
	E-value: N/A	E-Value: N/A
Pfam – protein family	Glycerate kinase family	Domain of unknown function (DUF4147)
	E-values: 1.3e-145	E-values: 1.9e-48
Protein Database	Crystal structure of putative glycerate kinase 2 from Salmonella typhimurium	*Crystal structure of Glycerate kinase (EC 2.7.1.31) (tm1585) from THERMOTOGA*
	E-value: 7.53881E-111	E-value: 1.55879E-51
Enzyme commission number	2.7.1.165	
KEGG pathway map	map00561	

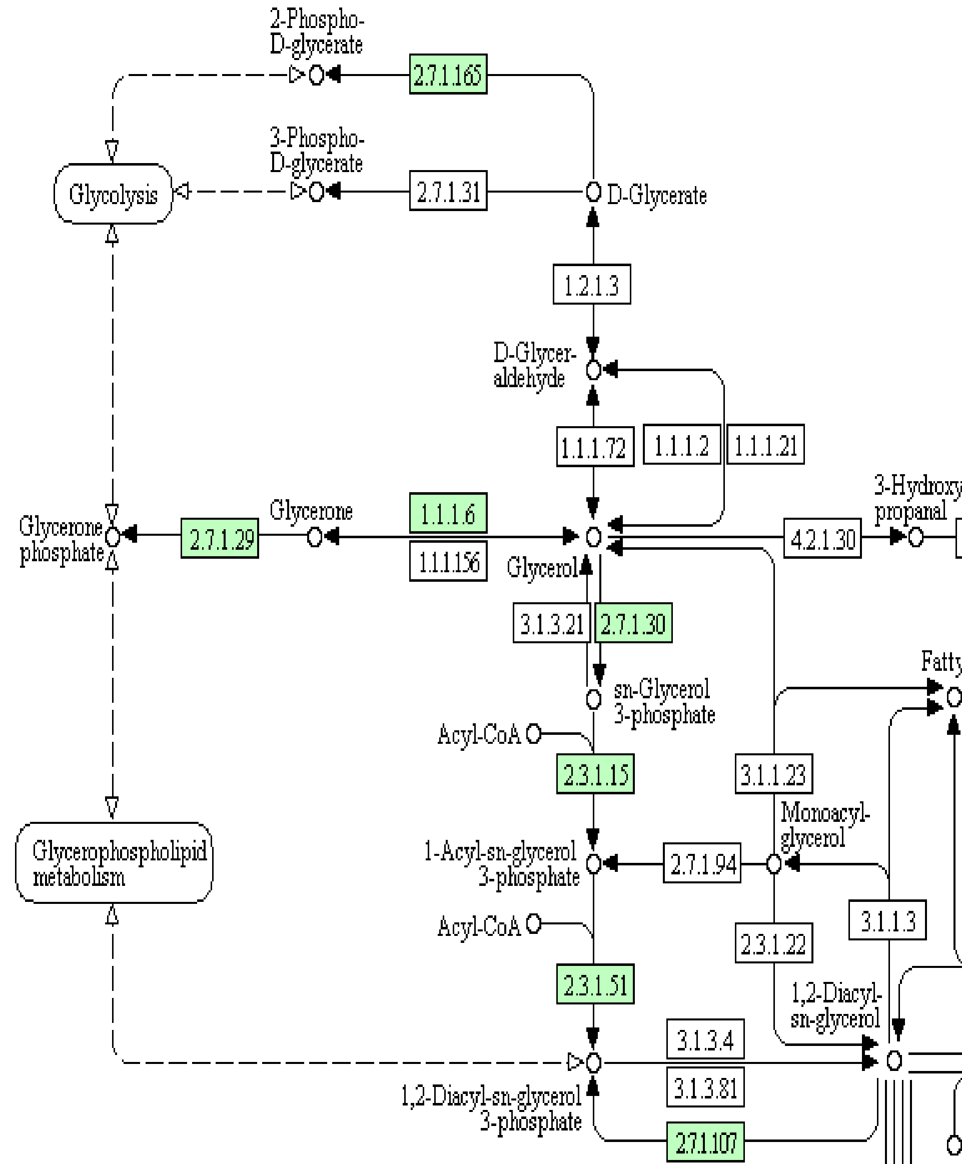
As seen by the data summarized above we can conclude that these organisms are definitely orthologues as they have an e value of 0 when blasted against each other. Furthermore is that they are localized in the cytoplasm, but what is interesting is that there is a high degree of non-polar residues observed early in the sequence that almost lead to a false positive membrane helix conclusion, this is potential an avenue for further research into the structure/function of the

protein. The matching Cogs indicating similar functions. One thing to note was that the PDB top hit for both genes was different but both these PDBs coded for glycerate kinase domains so it is not a source of refuting evidence. Lastly due to both having accurate GC %'s we can confirm that the gene did not come to be as a result of HGT. One thing to note is that no TIGRFAM hits were found and the Pfam of ruber only came up with an unknown domain, therefor there is a source of refuting evidence, but due to the homology of the sequence and the PDB hits it is still fair to assume that these are orthologues.

As seen by the respective keg maps in the image below all the E.C values are hits on both organism's pathways indicating that both have a similar biochemical pathway, and that both



Meiothermus ruber DSM 1279
Go
GLYCEROLIPID METABOLISM
2-Phospho-D-glycerate
2.7.1.165
3-Phospho-D-glycerate
2.7.1.31
Glycolysis
D-Glycerate
1.2.1.3
D-Glycer-aldehyde
1.1.1.72
1.1.1.2
1.1.1.21
Glycerone phosphate
2.7.1.29
Glycerone
1.1.1.6
1.1.1.156
Glycerol
4.2.1.30
3.1.3.21
2.7.1.30
sn-Glycerol 3-phosphate
Acyl-CoA
2.3.1.15
3.1.1.23
Glycerophospholipid metabolism
1-Acyl-sn-glycerol 3-phosphate
2.7.1.94
Acyl-CoA
2.3.1.22
2.3.1.51
3.1.3.4
1,2-Diacyl-sn-glycerol 3-phosphate
3.1.3.81
2.7.1.107



Escherichia coli K-12 MG1655
Go
GLYCEROLIPID METABOLISM
2-Phospho-D-glycerate
2.7.1.165
3-Phospho-D-glycerate
2.7.1.31
Glycolysis
D-Glycerate
1.2.1.3
D-Glycer-aldehyde
1.1.1.72
1.1.1.2
1.1.1.21
Glycerone phosphate
2.7.1.29
Glycerone
1.1.1.6
1.1.1.156
Glycerol
4.2.1.30
3.1.3.21
2.7.1.30
sn-Glycerol 3-phosphate
Acyl-CoA
2.3.1.15
3.1.1.23
Monoacyl-glycerol
Glycerophospholipid metabolism
1-Acyl-sn-glycerol 3-phosphate
2.7.1.94
3.1.1.3
Acyl-CoA
2.3.1.22
2.3.1.51
1,2-Diacyl-sn-glycerol
3.1.3.4
1,2-Diacyl-sn-glycerol 3-phosphate
3.1.3.81
2.7.1.107

CONCLUSION: In conclusion through the use of bioinformatics tools like BLAST we have confirmed the presence of 3 pairs of glycolipid metabolizing ortholog genes, due to the low e values and respective PDB hits and for cases when PFAM and TIGR supported these conclusions (not always). Furthermore, due to background research and confirmation through bioinformatics tools such as TMH, SignalP, LipoP, and PSORT-B we have analyzed the location of these proteins without needing to step into the lab, this really demonstrates just how resourceful and cost effective these free bioinformatics databases are. Potential follow up avenues of research can be done on conserved residues as they are potential active sites that could shed a light on the functionality of these proteins a little better, also due to allow of these genes being present on homologous orthologue neighborhoods analysis of their operons could be of interest. This project could not have been accomplished without these essential bioinformatics tools that are essential to resourceful and accurate research.

Nobre MF, Trüper HG, Da Costa MS. Transfer of Thermus ruber (Loginova et al. 1984), Thermus silvanus (Tenreiro et al. 1995), and Thermus chliarophilus (Tenreiro et al. 1995) to Meiothermus gen. nov. as Meiothermus ruber comb. nov., Meiothermus silvanus comb. nov., and Meiothermus chliarophilus comb. nov., respectively, and emendation of the genus Thermus. Int J Syst Bacteriol 1996; 46:604-606 10.1099/00207713-46-2-604 [Cross Ref]

Loginova LG, Egorova LA. Obligate thermophilic-bacterium Thermus ruber in hot springs of Kamchatka.Mikrobiologiya 1975; 44:661-665 [PubMed]

Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. Int J Syst Bacteriol 1980; 30:225-420 10.1099/00207713-30-1-225 [Cross Ref]

Lapage SP, Sneath PHA, Lessel EF, Skerman VBD, Seeliger HPR, Clark WA. 1992. International Code of Nomenclature of Bacteria (1990 Revision). Bacteriological Code. American Society for Microbiology, Washington, DC.

Loginova LG, Egorova LA, Golovacheva RS, Seregina LM. Thermus ruber sp. nov., nom. rev. Int J Syst Bacteriol 1984; 34:498-499 10.1099/00207713-34-4-498 [Cross Ref]

Euzéby JP. List of bacterial names with standing in nomenclature: A folder available on the Internet. Int J Syst Bacteriol 1997; 47:590-592 10.1099/00207713-47-2-590 [PubMed] [Cross Ref]

NCBI BLAST [Internet] National Center for Biotechnology Information; [2015 Dec 16]

Available from

http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=F

AQ#expect

Persidis A. 1999. Bioinformatics. Nature Biotechnology. 17(8): 828-830.

Phylogenetic Diversity. [Internet] U.S. Department of Energy Joint Genome Institute; [2015 Dec

16]. Available from http://jgi.doe.gov/our-science/science-programs/microbialgenomics/phylogenetic-diversity/

Scott LR. Meiothermus ruber Genome Analysis Project. [Internet]. GENI-ACT; [2015 Dec 16].

Available from: http://geni-science.org/secure/projects/view/

Tindall et al. 2010. Complete genome sequence of Meiothermus ruber type strain. Stand

Genomic Sci 3(1): 26-36.

Albuquerque L, Ferreira C, Tomaz D, Tiago I, Veríssimo A, da Costa MS, Nobre MF. Meiothermus rufus sp. nov., a new slightly thermophilic red-pigmented species and emended description of the genus Meiothermus. Syst Appl Microbiol 2009; 32:306-313 10.1016/j.syapm.2009.05.002 [PubMed] [Cross Ref]

Tenreiro S, Nobre MF, da Costa MS. Thermus silvanus sp. nov. and Thermus chliarophilus sp. nov., two new species related to Thermus ruber but with lower growth temperatures. Int J Syst Bacteriol 1995; 45:633-639 10.1099/00207713-45-4-633 [PubMed] [Cross Ref]

Pires AL, Albuquerque L, Tiago I, Nobre MF, Empadinhas N, Veríssimo A, da Costa MS. Meiothermus timidus sp. nov., a new slightly thermophilic yellow-pigmented species. FEMS Microbiol Lett 2005; 245:39-45 10.1016/j.femsle.2005.02.011 [PubMed] [Cross Ref]

Zhang XQ, Zhang WJ, Wei BP, Xu XW, Zhu XF, Wu M. Meiothermus cateniformans sp. nov., a slightly thermophilic species from north-eastern China. Int J Syst Evol Microbiol 2010; 60:840-844 10.1099/ijs.0.007914-0 [PubMed] [Cross Ref]

Chen MY, Lin GH, Lin YT, Tsay SS. Meiothermus taiwanensis sp. nov., a novel filamentous, thermophilic species isolated in Taiwan. Int J Syst Evol Microbiol 2002; 52:1647-1654 10.1099/ijs.0.02189-0 [PubMed] [Cross Ref]

Chung AP, Rainey F, Nobre MF, Burghardt J, Costa MSD. Meiothermus cerbereus sp. nov., a new slightly thermophilic species with high levels of 3-hydroxy fatty acids. Int J Syst Bacteriol 1997; 47:1225-1230 10.1099/00207713-47-4-1225 [PubMed] [Cross Ref]

Dawson, Wayne K.; Maciejczyk, Maciej; Jankowska, Elzbieta J.; Bujnicki, Janusz M. (2016-07-01). "Coarse-grained modeling of RNA 3D structure". Methods. Advances in RNA Structure Determination. 103: 138–156. doi:10.1016/j.ymeth.2016.04.026.

Kmiecik, Sebastian; Gront, Dominik; Kolinski, Michal; Wieteska, Lukasz; Dawid, Aleksandra Elzbieta; Kolinski, Andrzej (2016-06-22). "Coarse-Grained Protein Models and Their Applications". Chemical Reviews. 116: 7898–936. doi:10.1021/acs.chemrev.6b00163. ISSN 0009-2665. PMID 27333362.

Wong, KC (2016). Computational Biology and Bioinformatics: Gene Regulation. CRC Press (Taylor & Francis Group). ISBN 9781498724975.

Lee SY (Mar 1996). "High cell-density culture of Escherichia coli". Trends in Biotechnology. 14 (3): 98–105. doi:10.1016/0167-7799(96)80930-9. PMID 8867291

Nelson, D. L.; Cox, M. M. (2000). Lehninger, Principles of Biochemistry (3rd ed.). New York: Worth Publishing. ISBN 1-57259-153-6.

Lampe, M. A.; Burlingame, A. L.; Whitney, J.; Williams, M. L.; Brown, B. E.; Roitman, E.; Elias, M. (1983). "Human stratum corneum lipids: characterization and regional variations". J. Lipid Res. 24 (2): 120–130. PMID 6833889

Kanehisa, Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K.; KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45, D353-D361 (2017). [pubmed] [pdf]

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44, D457-D462 (2016). [pubmed] [pdf]

Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27-30 (2000).

Sim, Adelene YL; Minary, Peter; Levitt, Michael (2012-06-01). "Modeling nucleic acids". Current Opinion in Structural Biology. Nucleic acids/Sequences and topology. 22 (3): 273–278. doi:10.1016/j.sbi.2012.03.012. PMC 4028509 . PMID 22538125

Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. Cells As Experimental Models. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9917/

Annotations for the methods:

1U.S. Department of Energy Joint Genome Institute. [Internet: http://jgi.doe.gov/our-science/science-programs/microbial-genomics/phylogenetic-diversity/ accessed 7-29-14]

2Wu, D et al. 2009. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature 462(7276):1056-60.

3Loginova, LG et al. 1984. Thermus ruber sp. nov., nom. rev. Int J Syst Bacteriol 34:498-499.

4Loginova LG,, Egorova LA. 1975. Obligate thermophilic-bacterium Thermus ruber in hot springs of Kamchatka.Mikrobiologiya 44:661-665

5Ekman J, et al. 2007. Detection and quantitation of colored deposit-forming Meiothermus spp. in paper industry processes and end products. J Ind Microbiol Biotechnol. 34(3):203-11.

6Albuquerque L et al. 2009. Meiothermus rufus sp. nov., a new slightly thermophilic red-pigmented species and emended description of the genus Meiothermus. Syst Appl Microbiol 32:306-313.

7Tindall, B et al. 2010. Complete genome sequence of Meiothermus ruber type strain (21T). Standards in Genomic Sciences 3:26-36; and Department of Energy, Joint Genome Institute Division; http://img.jgi.doe.gov/cgi-bin/geba/main.cgi?section=TaxonDetail&page=taxonDetail&taxon_oid=2500395348

8Da Costa et al. 2006 The genus Thermus and relatives. In Prokaryote, eds: E. Rosenberg, E. Stackebrandt, F. Thompson, S. Lory, E. DeLong. 7:797-812.

9U.S. Department of Energy, Joint Genome Institute division [accessed July 31, 2014 http://jgi.doe.gov/our-science/science-programs/microbial-genomics/phylogenetic-diversity/

10Krieger, C et al. 2004. MetaCyc: a multiorganism database of metabolic pathways and enzymes Nucleic Acids Res. 32 (Database issue): D438–D442.

11Markowitz, V. et al. 2014. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucl. Acids Res. 42 (D1):D560-D567.

12Kanehisa M, Goto S. 2000. "KEGG: Kyoto Encyclopedia of Genes and Genomes". Nucleic Acids Res 28 (1): 27–30.

13Kanehisa M. 1997. "A database for post-genome analysis". Trends Genet 13(9):375–6.

14Kanehisa M, et al. 2006. "From genomics to chemical genomics: new developments in KEGG". Nucleic Acids Res 34 (Database issue):D354–7.

15Nishida, H. and Nishiyama, M. 2012. Evolution of Lysine Biosynthesis in the Phylum Deinococcus-Thermus. Intern. J. Evol. Bio. 2012: Article ID 745931, 6 pages

16Altschul, SF et al. 1990. Basic local alignment search tool.J Mol Biol. 215(3):403-10. http://blast.ncbi.nlm.nih.gov/Blast.cgi

17Ashburner, M et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 25(1):25-9. List of functional genomics tools accessed on the Internet: http://www.geneontology.org/book/export/html/801 (CLICK HERE)

18Inquiry Wheel image taken from the Internet at: http://cires.colorado.edu/education/outreach/rescipe/collection/inquirystandards.html [Accessed July 31, 2014].

19Reiff, R. et al. 2002. Proceedings of the 2002 Annual International Conference of the Association for the Education of Teachers in Science, eds. P. Rubba, J. Rye, W. Di Biase, B. Crawford. ERIC Document Reproduction Service No. ED (465 602).