Analysis of the Role of Methyl-Accepting Chemotaxis Proteins and the GGDEF Motif in the Signal Transduction Chain of Photoactive Yellow Protein

Saylor F. Hampton

Department of Microbiology and Molecular Genetics,

Oklahoma State University, Stillwater OK

Author Note

Parts of this work were carried out in the Microscopy Laboratory, Oklahoma State University, which received funds for purchasing the equipment from the NSF MRI program.

Table of Contents

Abstract

Many bacterial genomes encode photoreceptors which detect illumination and send signal cascades to response-regulated proteins to illicit physiological responses. The Photoactive yellow protein (PYP) is a known photoreceptor which has been characterized to illicit a range of biological responses in different bacteria upon blue light illumination indicating diversity in its downstream signal transduction chain. Such responses include photoregulation of photo-protective pigment biosynthesis, photoregulation of biofilm formation, and negative phototaxis. We performed bioinformatics analyses of PYP homologs to identify both recurring genes within the same predicted operon as PYP and protein domains occurring in multidomain PYPs to further understand possible biological functions of PYP and its signal transduction pathways. Two identified genes of interest include the PYP-MCP fusion protein and the GGDEF motif as it relates to the signal transduction chain of PYP.

The domain structure of the PYP-MCP fusion protein suggests that it is sensitive to blue light illumination via its PYP domain and that it controls motility via its methyl-accepting chemotaxis protein (MCP) moiety, leading to the hypothesis that this protein triggers negative phototaxis. To further investigate the PYP-MCP fusion protein and its signal transduction chain, we experimentally analyzed the *Nitrincola alkalilacustris sp.*, which contains a multidomain PYP, to test the prediction that *N. alkalilacustris sp.* containing the MCP-PYP fusion protein will respond to a stimulus of blue-light and undergo negative phototactic movement due to the downstream signaling from the PYP to the methyl accepting chemotaxis protein which controls the MCP moiety.

Using prior studies as well as bioinformatics research the GGDEF domain is suggested to be a key player in regulating biofilm formation upon light illumination triggered by the *pyp gene.* The GGDEF proteins have been characterized to convert GTP into cyclic-di-GMP which is needed to synthesize exopolysaccharides in biofilm formation. Therefore, the domain structure of tis GGDEF motif coupled with the PYP suggests that it is sensitive to light illumination via its PYP and it controls the regulation of biofilm formation via its GGDEF protein, leading to the hypothesis that this multi-domain protein triggers the regulation of biofilm formation. To investigate this GGDEF motif as a multi-domain PYP we experimentally analyzed *Massilia albidiflava* and *Massilia plicata* species containing the GGDEF motif in a multi-domain PYP using *Idiomarina loihiensis* as an experimental control as it has been described to have biofilm formation regulation via they PYP to test the prediction that, *M. albidiflava* and *M. pilcata* containing the PYP-GGDEF motif will respond to the blue-light stimulus and undergo photo-regulation of biofilm formation due to downstream signaling from the photoactive yellow protein to the GGDEF motif in the multidomain PYP.

Since PYP functional photocycle is initiated by photoisomerization of its p-coumaric acid (pCA) chromophore, we used the addition of pCA to indicate the blue-light illumination response to be triggered by the PYP signaling due to PYP being the only photoreceptor to use pCA as its key component in photoisomerization of its chromophore to initiate an active response. Upon the addition of pCA within the media used to test phototaxis within *N. alkalilacustris sp,* there appears to be a restoration in positive phototaxis, which needs further biological replicates to be a conclusive result. Furthermore, for the analysis of biofilm formation there were inconclusive results for *M. albidiflava* while the addition of pCA did not appear to have an effect on *M. plicata.* Upon further investigation of *M. plicata* there does however appear

to be a light effect which causes a decrease in biofilm formation upon light illumination, again further biological replicates are needed to confirm these results. The addition of a trans-locked pCA chromophore is also needed to determine this result is due to the PYP.

Keywords: Photoactive yellow protein, *Nitrincola alkalilacustris, Massilia albidiflava, Massilia plicata,*

Acknowledgements

I would like to thank Dr. Wouter Hoff for providing guidance and wisdom as my primary reader, thesis director, and professor during my time at Oklahoma State University. Dr. Hoff is a master in his field who has provided me with the knowledge and tools to accelerate my academic growth and personal attributes as he challenged me to think critically and scientifically through my time in his undergraduate laboratory.

I would like to thank Dr. Jeff Hadwiger for acting as my second reader and providing vital input to heighten my writing and scientific approach during my thesis journey. Dr. Hadwiger challenged me to think beyond my research for practical application and approaches to broaden my understanding of life and the world around us.

I would like to give a great thanks to Dr. Matthew Cabeen for his input on understanding motility as well as providing me with the light source used within this experiment. Dr. Cabeen further added in the understanding and protocols behind evaluating for biofilm formation and without his support and helping had this study would not have operated with such swift motion.

I wish to thank Dr. Randy Morganstein for his current and future collaborations and aid in describing and understanding our bacterium cell characteristics and understanding of cell imaging process. Without his help our motility assays as well as discoveries on our bacterium's would not have been possible.

I further like to thank Dr. Paul Kirchberger for his aid in understanding viruses and their interactions within bacteria as this information proved vital to understanding under characterized bacteria and the analysis of viruses which could be present within our species.

Dr. Rolf Prade as well as his laboratory deserve heartfelt thanks for allowing me to perform the analysis of my biofilm results using their 96-well plate reader as this allowed for efficient and effective analysis of our biofilm results. I would also like to thank them for their input in enhancing our studies and for comments during my thesis project.

I cannot thank Rosalie Dohmen enough for being my primary graduate supervisor and her aid in every aspect of my research. Rosalie is an inspiration and mentor for research studies but also for life. She no doubt led by example and provided uplifting correction and advice that kept be pursuing my passion for research. Furthermore, she was my go-to assistant and collaborator for day-to-day experiments, and I would not have been able to finish my work without her.

I would like to thank Gunnar Hoogerwerf for his programming expertise as his work helped us to analyze and present my research. Gunnars programs made bioinformatic comparisons and research far less daunting and this work could not have been done without him.

I would further like to give my gratitude to my research contributors and lab mates Catalina Bradley, Clarice Huffman, Scout Powell, and Sarah Teeman who helped with bioinformatic research and provided vital input during laboratory meetings each week.

Without the help and guidance of Brent Johnson and the Oklahoma State University Microscopy Laboratory obtaining TEM images would not have been possible and thus motility assays would have been far more difficult so for that I wish to thank them for their contribution to this study.

Lastly, I would like to thank the Oklahoma State University Honors College for allowing me to participate in their curriculum and honors thesis program as it has no doubt prepared me for my future endeavors. With this I would like to add and thank Sam Morse for his outstanding job as an honors advisor and for pushing me to pursue my thesis. Thank you all.

1. Introduction

Prokaryotes containing proteins responsible for sensory detection including but not limited to chemo-, mechano-, thermo-, and or photoreceptors allow the organism to detect and respond to its environmental and physiological alterations. Organisms use photoreceptors to detect illuminations through chromoprotein photoreceptors which send a biochemical signal cascade to response-regulated proteins for physiological action [1]. There is a limited number of documented photoreceptors used for detection of environmental illuminations and each photoreceptor can be generally classified into six photosensor families[2]: rhodopsin's[3, 4], phytochromes[5], xanthopsin's[6], cryptochromes[7], phytotropins[8], and BLUF proteins[9]. It is important to note that about 25% of all bacterial genomes (also from chemotrophs) encode photoreceptors [10]. This leads us to question what biological responses are triggered by these photoreceptors, which we are studying within photoactive yellow protein. Photoactive yellow protein (PYP) is a model protein proposed to be part of the xanthopsin family of photoreceptors which is analogous to the rhodopsin family[6] responsible for variable biological responses. PYP is considered to be a part of the xanthopsin family due to its covalently bound 4-hydroxycinnamic acid which is the key deciding factor for classification within the xanthopsin family.[11]

PYP is a blue-light regulated photoreceptor that was initially discovered in *Halorhodospira halophila [12, 13]* and is identified as the structural model for the three-dimensional fold found in PAS (Per-Arnt-Sim) domain [14]. Within this PAS domain there is a protein group LOV (light, oxygen, or voltage) which senses inputs from the environment and alters physiological responses[15, 16]. When PYP is introduced to blue-light illuminations it undergoes a

photoisomerization of the vinyl bond within its *p*-coumaric acid (*p*CA) chromophore resulting in the trans to cis confirmational change from its initial pG state to a short-lived, red-shifted pR state where *p*CA protonation occurs by Glu46 forming the pB' state[17, 18]. These cascading changes result in the pB state which is blue-shifted and the assumed signaling state of the PYP[17]. While PYP is a novel protein known for its blue-light regulation for structural conformational change, PYP has been shown to illicit a range of biological responses in an assortment of bacterial organisms. Such biological responses include but are not limited to phototaxis, particularly negative phototaxis[19], photo-regulation of photoprotective pigments[20], and photo-regulation of biofilm formation[21]. Notwithstanding the fact that a lot is known about the biophysics of isolated PYP from *H. halophila,* very little information is available on the downstream signal transduction chain(s) to which PYP supplies its signal. The approach that we are using here is to first perform a bioinformatics analysis of *pyp* operons and PYP fusion proteins to generate hypotheses both about possible PYP signal transduction chains and possible functions triggered by PYP, and then to test experimentally these hypotheses for selected microorganisms.

When looking at the relationship between phototaxis and PYP we need to understand the basis of which the research stemmed from. It is known that *Escherichia coli* undergoes chemotactic movement when exposed to attractant and repellant chemicals as it is regulated by four chemoreceptors in the cytoplasmic membrane of the organism; however, *E. coli* also alters its swimming motility upon exposure to blue light of high intensity[22-24]. Such results and studies on phototaxis in *E. coli, Halobacterium halobium, and Natronobacterium pharaonis* led to the understanding of the sensory rhodopsin I (SRI) and sensory rhodopsin II (SRII) as key components to such phototactic responses[19, 25-27]. This ultimately propelled the study of

Ectothiorhodospira halophila (now known as *Halorhodospira halophila*) and its phototactic response to blue-light illumination as it had photophysical similarities to the rhodopsin's known for phototaxis [19]. The results of this study showed that *H. halophila* exhibited negative phototactic motility to blue-light of 400-500 nm wavelength, making it the only motile purple bacteria to respond in such a way thus far[19, 28]. However, the signal transduction chain causing such a response is unknown.

For multiple different types of photoreceptors we found genome bioinformatic support for a link with photoregulation of biofilm formation [29]leading to further inquiry. When analyzing the association between PYP and biofilm formation bioinformatic studies showed chemotrophic bacteria including *Idiomarina loihiensis* contained many known photoreceptor protein families[21]. The genome of *I. loihiensis* contains encoded proteins that produce and secrete polysaccharides associated with known biofilm formation as well as a homolog protein to PYP as its photoreceptor protein [30, 31]. Furthermore, the genome of *I. loihiensis* indicates that the PYP is within close proximity to the diguanylate cyclase (c-di-GMP) gene which is the bacterial signaling molecule and regulator for the synthesis of exopolysaccharides in biofilm formation[32-35]. The GGDEF domain converts GTP into cyclic-di-GMP which is needed to produce the exopolysaccharides for biofilm formation. For analysis of PYP regulation of biofilm formation trans-locked *p*CA was used to prevent the PYP from entering into the pB state needed for signaling and cis-locked *p*CA to keep the PYP in the active signaling pB state[21]. The trans-locked PYP resulted in significantly decreased biofilm formation while the cis-locked increased the formation of biofilm, concluding that the PYP was the photoreceptor responsible for photo-regulation of biofilm formation[21].

Such variation and diversity in biological responses such as negative phototaxis and biofilm formation associated with organisms containing PYP leads to question of how the downstream signaling and transduction chain of PYP is conducted. To further analyze the downstream signaling and signal transduction chain of PYP we used two bioinformatic analysis techniques (1) we identified recurring genes with the same predicted operon as PYP and (2) identified protein domains occurring in multidomain PYPs to further understand possible biological functions of PYP. From there we identified genes of interest such as PYP-Methyl accepting chemotaxis protein (MCP) fusions and PYP-GGDEF domains, as well as organisms from which contain such genes of interest; *Nitrincola alkalilacustris* for negative phototaxis containing PYP-MCP, and *Massilia albidiflava* and *Massilia pilcata* for photo-regulation of biofilm formation containing PYP-GGDEF motif. Using said organisms we hypothesize that *N. alkalilacustris sp.* containing the MCP-PYP fusion protein will respond to a stimulus of blue-light and undergo negative phototactic movement due to the downstream signaling from the PYP to the methyl accepting chemotaxis protein which controls the MCP moiety; and *M. albidiflava* and *M. pilcata* containing the PYP-GGDEF motif will respond to the blue-light stimulus and undergo photo-regulation of biofilm formation due to downstream signaling from the photoactive yellow protein to the GGDEF motif in the multidomain PYP. This biofilm regulation assumption is concluded on assumption that *M. albidiflava* and *M. plicata* will function under the same mechanistic nature of *I. loihiensis*, as this species was shown to have its GGDEF motif regulated by PYP to cause alterations in biofilm formation. Since PYPs functional photocycle is initiated by photoisomerization of its *p*CA chromophore, we plan to use a *trans*-locked pCA analog that prevents the photocycle to attribute the observed phototaxis and biofilm responses to PYP photoexcitation. These findings will provide novel information regarding the PYP signal

transduction chain and provide a basis for future mechanistic studies on how PYP relays its signal to downstream signaling partners. Furthermore, understanding the signal cascade of biological functions such as biofilm formation will aid in the fight against antibiotic resistant bacteria due to biofilms and lead to further studies on maintaining medical advances in biochemical therapies.

2. Methods

2.1 Bioinformatics

Bioinformatics research began with the original PYP1 from *Halorhodospira halophila* where a genomic BLAST was performed using NCBI tools on the known PYP where an analysis of the generated results provided a list of organisms containing similar genomic sequences. For BLAST results from NCBI we retrieved the full-length sequence of the proteins and used ClAlign to remove the proteins with extensive gaps in the sequence. The multiple sequence alignment (MSA) with PYP homologs were analyzed with selected queries such as their statistical significance, percent of sequence similarities, and diversity in taxonomic classification of the organism for both single and multi-domain PYPs. All organisms containing a sequence that was validated as being likely to encode a functional PYP based on the presence of functionally important residues were organized into an excel spreadsheet with the organism's name, ascension number, and PYP amino acid sequence.

We used the KEGG database and its genome viewer to identify and examine possible *pyp* operons. As a first step, members of the PYP family encoded in genomes in the KEGG database were identified. Each PYP amino acid sequence identified using BLAST was then put through a BLAST within the KEGG database and populated organisms were compared using sequence

similarity. Sequence similarity was compared using Clustal Omega using a similarity cut off of 95%. Once a very high degree of similarity was established, we used the KEGG genome mapping tool to identify the *pyp* gene and its flanking genes for single domain PYP's. Each flanking gene was analyzed for their function pathway as annotated in the KEGG database, motif as analyzed for conserved residues in the SMART database, and gene identity for sequence identity using their closes annotated BLAST his in the KEGG and NCBI Databases, each was performed using their respective amino acid sequences. Analysis of the amino acid sequences involved performing a BLASTP using the NCBI Protein BLAST tool for other possible functional annotations and pathways related to the amino acid sequence. Of the multi-domain PYPs a domain analysis was performed using the SMART and Pfam database, and resultant statistics and identified conserved domains within the multi-domain PYPs were analyzed to provide a list of possible protein functions and biological functions associated with the PYP family. For KEGG BLAST hits, these results were manually inspected for candidate PYP operons and possible functional signal pathways. Associated genes were then analyzed using BLAST and resultant proteins associated with the PYP and then were analyzed for possible protein and biological function using prior research and other similar sequence hits with known functional pathways.

After Identification of proteins and domains of interest organisms containing such genes were identified for further analysis using experimental approaches. It should be noted that for future identical bioinformatics searches, when performing these bioinformatic analyses, queries may contain alternate results due to updates and modifications in both NCBI and KEGG.

Bioinformatics analyses of the complete genomes of *N. alkalilacustris, M. albidiflava,* and *M. plicata* were conducted to determine the presence of the tyrosine ammonia lyase (TAL) that

converts tyrosine into *p*-coumaric acid and *p*-coumaryl-coA ligase (*p*CL) that uses ATP to attach *p*-coumaric acid to coenzyme A, which then serves as the *p*-coumarly donor for attaching *p*-coumaric acid to Cys69 in apoPYP to convert it to PYP. This analysis was performed by using the *p*CL and TAL amino acid sequence for these two protein from the KEGG database for *H. halophila (Hhal)* and performing a protein BLAST searches within NCBI for each respective organism. The KEGG identifier for Hhal TAL is Hhal_1820 and for *p*CL it is Hhal_1819. The organism identifiers used within NCBI for the protein BLAST of each organism is as follows; *N. alkalilacustris* NCBI tax ID:1571224, *M. albidiflava* DSM:17472, and *M. plicata* DSM:17505. Each respective generated hit was analyzed for percent amino acid sequence identity to the Hhal sequence and then further analyzed by inserting the amino acid sequence for BLAST in KEGG. The generated result in KEGG was examined for similarity to the original sequence and functional pathway.

We also used a genome bioinformatics approach to explore if *N. alkalilacustris* encodes components of the bacterial chemotaxis signaling pathway. Using the amino acid sequence from each respective protein from *Escherichia coli K-12 MG1655* within the KEGG database, *N. alkalilacustris* was analyzed for components needed for MCP flagellar signaling such as CheA, CheY, and CheW chemotaxis proteins. The amino acid sequence for each protein was used for a protein BLAST within NCBI and the top generated result with the highest percent identity to the original sequence was used for further analysis. The result was then used in a KEGG BLAST for which the top generated result was observed for pathway and identification of the chemotaxis proteins in relation to the original *Escherichia coli K-12 MG1655* species proteins.

2.2 Growth Conditions

We selected our bacteria for further experimental studies. *Nitrincola alkalilacustris sp.* contains key components of interest including PYP and a methyl accepting chemotactic protein in its multidomain structure. *Massilia albidiflava* and *Massilia plicata* contain genes of interest, namely a PYP and a downstream gene encoding a protein containing GGDEF motif in a likely *pyp* operon. *Idiomarina loihiensis* serves as a positive control, as it has been reported to show reduced biofilm formation upon PYP photo-excitation [21].

Upon receiving each bacterium as a dried culture from Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures the samples were rehydrated in sterile media and inoculated into their respective growth media per the manufacturer's specifications located at www.dsmz.de/cultivation-microbes. Each bacterium was grown using the conditions listed in *Table 1.*

Additional images were obtained for *N. alkalilacustris* using transmission electron microscopy (performed at the OSU microscopy Laboratory), since the morphology of this organism has not yet been reported in literature. *N. alkalilacustris* will undergo further analysis for cell chape and size variation under the direction of Dr. Morgenstein upon a later date.

Organism	Culture Medium	Temperature (°C)	Incubation Period	Sample Type/ Isolate location	Reference
Nitrincola alkalilacustris sp.	Marine Broth	25	1-2 days	The alkaline water of Zab-szek soda pan	[36-39]
Idiomarina loihiensis	Marine Broth	30	1-2 days	Hydrothermal vent fluids mixed with ambient seawater	[36, 37, 39, 40]
Massilia albidiflava	R2A Medium Broth	30	3 days	Soil from Jiangsu province	[36, 37]
Massilia plicata	R2A Medium Broth	30	1-2 days	Heavy-metal-polluted farm soil	[36, 37]

Table 1: Growth conditions and strain type per species of inoculate.

2.3 gDNA Extraction, PCR, and Gel-Electrophoresis

Extraction of genomic DNA (gDNA), polymerase chain reaction (PCR), and gel-electrophoresis were performed on all obtained samples to ensure correct sample identity as well as verification of the presences of the genes of interest. DNA extraction was performed using the protocol provided in Zyma Research Quick DNA Fungal Bacterial Mini Prep Kit (cat:6005). All steps were followed according to the protocol with modification to step 10, where we added 30 µL of DNA elution buffer instead of the minimum 35 µL. Extracted gDNA was then quantified using NanoDrop standard protocol with purity measure the 260/280 column, 260/230 column, and concentration given in ng/µL. Samples were then stored at -80°C until thawed for PCR and gel-electrophoresis.

Primers were selected using the genome sequence for each organism as given in NCBI which was then imported into the SNAP gene database to visually find primers with a GC content between 40-60% and a melting temperature within 5°C between two paired primers, the SNAP gene database allowed us to easily visualize the needed region of which to select primers for. Each primer was checked for self-dimers, hetero-dimers, hairpins, and annealing temperatures than 10°C, DNA primers were ordered from Integrated DNA Technologies. Once each primer was received, each respective oligo was diluted to a concentration of 100µM and placed in a -80°C freezer. See *Table 2* for primer names, related organism and gene of interest, melting temperature, and nucleotide sequences. Three different PCR reactions were performed on *Nitrincola alkalilacustris sp.,*: one sample contained forward primer PYP_F_N.A and reverse primer genomic1_R_N.A to identify the *pyp* gene, one sample with forward primer genomic1_F_N.A and reverse primer MCP-PYP_R_N.A to identify the gene encoding the MCP, and one sample containing the forward primer genomic1_F_N.A and reverse primer genomic1_R_N.A to identify the gene encoding the entire PYP-MCP fusion protein. For *Idiomarina loihiensis* only the presence of the *pyp* gene was confirmed using PCR as there are limitations to PCR due to the size of the GGDEF domain and the positional gap between the GGDEF and PYP as the two are not within the same protein.

PCR reactions were performed using the recipe as shown in *Table 3* and performed using the thermocycler settings as shown in *Table 4*. Standard protocol was used for thermocycler settings and primer annealing temperature was obtained using lowest melting temperature +3°C from the primer sets within each set up. *Nitrincola alkalilacustris sp* was run separately with an annealing temperature of 62.6°C and *Idiomarina loihiensis*, *Massilia albidiflava*, and *Massilia plicata* were

run together with an annealing temperature of 69.6°C. Upon completion of the PCR process, samples were stored at 4°C until they were used for gel-electrophoresis.

Gel-electrophoresis casting gels were made using a 1% agarose gel using 0.3g agarose to 30 mL of 1xTAE Buffer following the specifications from the RunOne™ System Vers. 022108 allowing the gel to solidify for 30 minutes. Each DNA sample was mixed with 1µL of Purple Loading Dye 6X (B7024A) from BioLabs and then loaded into wells as shown in *Table 5* for *N. alkalilacustris sp,* and *Table 6* for *I. loihiensis, M. albidiflava,* and *M. Plicata*. Each gel also contained a DNA ladder GeneRuler 1kb #SM0311 for identification of DNA size that contained 1µL of the ladder, 1µL of loading dye, and 4µL of NF H_2O. The system was run for five minutes on 25V and then twenty minutes on 100V. The gel was then removed and gently rotated in 9µL of Gel Green nuclei acid gel stain from Biotium (catalog no. 41005) in 30mL of 1x TAE buffer for twenty minutes shielded from light before images were obtained. The results observed in the gel picture were compared to a simulated gel created using the SnapGene database.

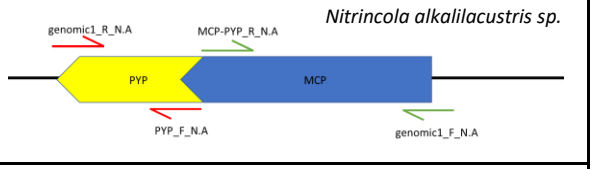
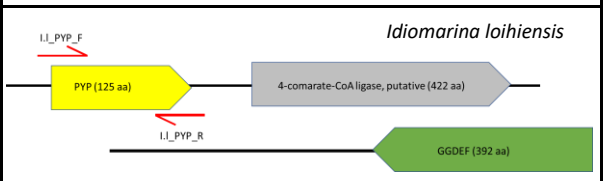
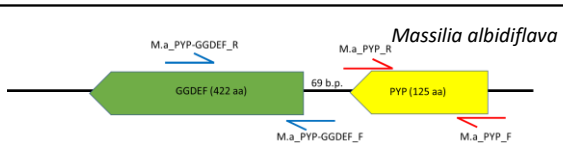
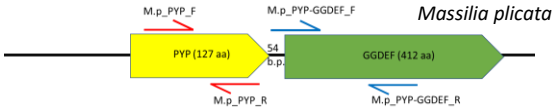
Primer name	T_m	Sequence	Notes
PYP_F_N.A	59.6	5′- TAA GCA GGA TAT CGA CAG TCT GTC AGC -3′	
MCP-PYP_R_N.A	63.9	5′- GAG TAG CCT GTT GTG TGT TGG ATG CCA GTT -3′	
Genomic1_F_N.A	59.6	5′- GAT AAA CAT TAC AAG ACT ACT GAC TAA CGC CCT G -3′	
Genomic1_R_N.A	61.9	5′- GTC GCT GTA TCG TCA GTC GAT GCA TGA -3′	
I.l_PYP_F	60.1	5′- CTT AGA GGT AAC ACA ATG GAG ATT GTT CAA TTC GG -3′	
I.l_PYP_R	59.4	5′- AAG TGA GTC TGT CAC TTT ATA GTC GCT TAA CG -3′	
M.a_PYP_F	67.2	5′- TAC CTG GAG TTT CGA TGA CCG CAC TTG CCT TCG A -3′	
M.a_PYP_R	67.6	5′- GGT ATT CAC GGC TGA CGC TGC ACC AGC A -3′	
M.a_PYP-GGDEF_F	69.1	5′- TGC CGT TGC GGA TGC GGC CCA CCA A -3′	
M.a_PYP-GGDEF_R	68.7	5′- ATA ATC GGC GAC ACT GGC CAG CAG TGC GTT CA -3′	
M.p_PYP_F	66.6	5′- GGC AGT CAA CGT GGA GCG TTT CAT GAG TGC ACT TA -3′	
M.p_PYP_R	69.2	5′- AGC TCT GTC ACA GCC GCC GCT GCA -3′	
M.p_PYP-GGDEF_F	69.2	5′- TAC GTC CTG ACG CTG CGG ATG AAG CCG GT -3′	
M.p_PYP-GGDEF_R	68.6	5′- CGC ATC CAG TTT CAC CAG CGA CAA CGC GAG GAT -3′	

Table 2: PCR reactions used to confirm the presence of genes of interest in bacterial strain used for physiological studies. N.A denotes *Nitrincola alkalilacustris sp.*, I.l denotes *Idiomarina loihiensis*, M.a denotes *Massilia albidiflava*, M.p. denotes *Massilia plicata.* The Notes section shows the domain architecture of the encoded protein of interest in each organism in relation to the primer location and direction, each species shows the number of base pairs (bp) between each gene if indicated. T_m indicates melting temperature in °C of each primer.

Component	**Amount µL**	
NF H2O:	31.5	
5xHF Buffer	10	
10mM dNTPs:	1	
10mM R primer:	2.5	
10mM F primer:	2.5	
Phusion pol:	0.5	
Total volume	**48**	
Split master mix of total volume	24	24
gDNA template	1	0
NF H2O	0	1
Total Volume of each sample for PCR	**25**	**25 (Negative control)**

Table 3: PCR Recipe Master Mix created and then divided into two even samples, where one sample contains the template gDNA and one contains no gDNA as the negative control, resulting in two samples of 25µL.

Number of cycles	1	30			1	1
Temperature (°C)	98	98	Primer Dependent	72	72	4
Time (s)	30	20	30	30	180	∞

Table 4: Summary of PCR conditions used. Note: *Nitrincola alkalilacustris sp.* primer sets dependent temperature was 62.6°C. *Idiomarina loihiensis, Massilia albidiflava*, and *Massilia plicata* primer dependent temperature was 69.6°C.

Well	1	2	3	4	5	6	7	8
Component	DNA Ladder	PYP	MCP	PYP-MCP fusion	empty	PYP Negative Control	MCP Negative Control	PYP-MCP Fusion Negative Control

Table 5: Nitrincola alkalilacustris sp. DNA gel pipetting scheme. PYP contains PYP_F_N.A, Genomic1_R_N.A, and gDNA, MCP contains Genomic1_F_N.A, MCP-PYP_R_N.A, and gDNA, and PYP-MCP fusion contains Genomic1_F_N.A, Genomic1_R_N.A, and gDNA. The negative controls contain the same components as their respective sample type without the gDNA.

Well	1	2	3	4	5	6	7	8
Component	Negative Control	blank	DNA Ladder	I.l	M.a PYP	M.a GGDEF	M.p PYP	M.p GGDEF

Table 6: DNA gel pipetting scheme for PCR products obtained using *I. loihiensis* and *Massilia* DNA as template. I.l denotes *Idiomarina loihiensis* and contains I.l_PYP_F, I.l_PYP_R, and gDNA. M.a denotes *Massilia albidiflava* and M.p denotes *Massilia plicata.* M.a PYP contains M.a_PYP_F, M.a_PYP_R, and gDNA, M.a GGDEF contains M.a_PYP-GGDEF_F, M.a_PYP-GGDEF_R, and gDNA. M.p PYP contains M.p_PYP_F, M.p_PYP_R, and gDNA, and M.p GGDEF contains M.p_PYP-GGDEF_F, M.p_PYP-GGDEF_R, and gDNA. Negative control used was for I.I and contained all components of I.l except the gDNA template which was substituted for water.

2.4 Motility

Motility is needed for phototactic movement to be achievable, and therefore understanding the type of motility performed by *N. alkalilacustris* is relevant. Motility testing was performed on *Nitrincola alkalilacustris sp.* using macroscopic altered concentration semi-

solid agar assays following the qualitative semi-solid concentration cut offs as described in research as seen in *Table 7*. Marine broth to agar concentrations were followed using the medium ratios as described in *Table 8*. Each plate of altered concentration of semi-solid agar medium was inoculated with *N. alkalilacustris* using a sterile loop into the center of the plate and incubated at 30°C for 72hr under white light and observed for motility.

As described in research, swarming motility has been reported to often results in a greater number of flagella than swimming cells, and thus microscopic studies were performed to observe the presence and number of flagella within *N. alkalilacustris[41].* Further microscopic motility studies were performed through visual observation of flagella using transmission electron microscopy (TEM). To prepare our sample for TEM our culture was grown for 24 hours, and was then centrifuged to obtain a pellet, while the supernatant was discarded. The bacterial cell sample was then fixed using 2% buffered glutaraldehyde (pH 7.0) by slowly adding the fixative to the tube not to disturb or dislodge the pellet and allowed to fix for 15 minutes at room temperature. The pellet was then released and allowed to continue to remain in the fixative until specimen was transported to Microscopy Laboratory at Oklahoma State University. Upon arrival at the Microscopy Laboratory, negative staining was performed with the sample placed on a formvar coated grid and allowed to sit at room temperature for one minute before excess liquid was removed from the grid and a drop of 2% uranyl acetate was added and allowed to sit for one minute. All excess liquid was removed from the grid using filter paper and allowed to dry before observing using TEM.

Published information indicates that for swarming type motility a surfactant is produced and excreted by bacteria to aid in the multicellular motion and cell-cell interactions. To examine

the possible presence of secreted surfactants on agar plates, a drop-collapse assay was performed on *N. alkalilacustris* to aid in the examination of motility[42]. The drop-collapse assay was performed by placing a drop of water on each agar plate to observe the appearance of the destabilization of the liquid drop by surfactants.

Motility Type	**Agar Concentration**	**Reference**
Swimming	~0.3%	[43, 44]
Swarming (temperate)	0.5-0.8%	[43, 45]
Swarming (robust)	>1.5%	[43, 45]
Twitching (inoculated at bottom of media)	1%	[43, 46]
Sliding	0.3-0.4% or 1-2%	[43, 47, 48]
Gliding	≤7% in Myxococcus xanthus	[43, 49]

Table 7: Literature overview of information on different motility types as described in research for semi-solid agar assays using altered concentrations of agar.

Desired Agar %	Agar(g): Marine Broth(g) per 1000mL water	Agar(g):Marine broth (g) per 500mL Water	Agar(g):Marine broth (g) per 250mL Water	Agar(g):Marine broth (g) per 125mL Water
1.5	15 : 37.4	7.5 : 18.7	3.75 : 9.35	1.875 : 4.675
1.0	10 : 37.4	5.0 : 18.7	2.5 : 9.35	1.2 5 : 4.675
0.5	5 : 37.4	2.5 : 18.7	1.25 : 9.35	0.625 : 4.675
0.4	4 : 37.4	2.0 : 18.7	1.0 : 9.35	0.5 : 4.675
0.3	3 : 37.4	1.5 : 18.7	0.75 : 9.35	0.375 : 4.675
0.2	2 : 37.4	1.0 : 18.7	0.5 : 9.35	0.25 : 4.675

Table 8 : Summary of the agar plates used here to study motility. Agar to marine broth medium ratios for desired agar percent concentrations for *N. alkalilacustris sp.* using Agar, Granulated BP1423-500 (*https://www.fishersci.com/shop/products/fisher-bioreagents-powdered-agar-2/BP1423500*) and Marine Broth 2216 (*https://www.sigmaaldrich.com/US/en/product/sial/76448*)

2.5 Phototactic Motility

Each inoculated plate of 0.3% and 0.2% concentrated agar was placed under white light for 48 hours at 30°C, with each concentration of plate containing a negative control completely shielded from light using aluminum foil, a positive control completely exposed to the white light, and a test plate that was partially covered with aluminum foil to allow one half of the plate to be exposed to white light while the other half was shielded from light. We used the Fenix UC35 V2.0 1000 Lumen Rechargeable Tactical Flashlight with LumenTac Organizer on the 150 lumen setting (see *Figure 1* for further information on the experimental setup). To further test motility we also illuminated the plates using blue light of ~360nm-580nm wavelength "blue circle" colored glass filter using the same techniques and light source as described above. See *Figure 2* for filter absorbance for "blue circle" filter and *Figure 3* for the experimental setup. Absorbance spectra were measured using an Agilent Cary 8454 UV-Vis Spectroscopy System.

To examine the possibility that some bacteria take up but do not synthesize the *p*-coumaric acid (*p*CA) chromophore, we tested the effect of adding *p*CA to the original growth media on phototactic motility. The binding of *p*CA as a chromophore is needed for the photoisomerization of PYP from it *trans*-to-*cis* state as shown in *H. halophila[50]*. The residues that aid in the binding of *p*CA chromophore to *H. halophila* are Cys69, Tyr42, and Glu46. Glu46 and Tyr42 bind to the phenolate oxygen of pCA whereas Cys69 forms a covalent thioester bond[51-53]. For calculating the total amount of *p*CA to be added to each media we used the molar solution formula $C = \frac{m}{V} \times \frac{1}{MW}$ to calculate the amount of solute to add to obtain a 0.5 mM solution. The *p*CA was dissolved in sterile water by adjusting the pH of the solution by adding 3M NaOH dropwise and agitating using a vortex until the solute was dissolved. We used SIGMA

C9008-5G *p*-coumaric acid with a molecular weight of 164.16. We then grew the plates using the same conditions as described above with a negative control completely covered with aluminum foil, a positive control completely exposed to the white light, and test plates partially covered with only half the plate exposed to the light. Again, these were grown in a controlled environment at 30°C, see *Figure 4.*

The exact same set up and experiment was then replicated with the only variable being altered: we switched the light source from white light to a blue light with a wavelength range 430-460nm. The light source emission spectrum was obtained by using an optical density filter with an absorption of 0.5 placed in front of the light detector of the Agilent Cary 8454 UV-Vis Spectroscopy System. A zero starting value was obtained using the measure of a blank with the optical density filter placed over the detector to obtain a baseline absorption of 0.5. Then the new LED light source 'Blue LED' was measured by angling the light source at the detector with the optical density filter still in place. The wavelength of the light source was then determined using the provided numerical value of the wavelength, *Figure 5* for graphical representation.

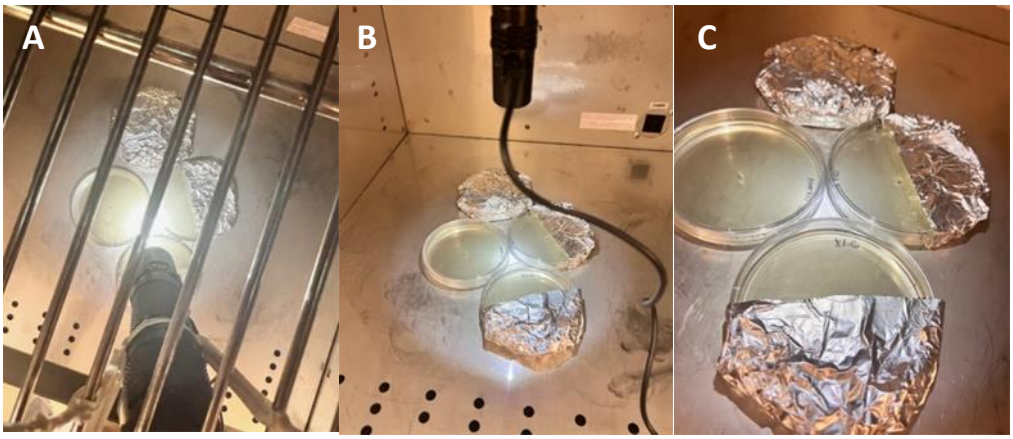


Figure 1 A is the overhead view of light position, B is the lateral view of light position, C is the direct view of light illumination on the plates.

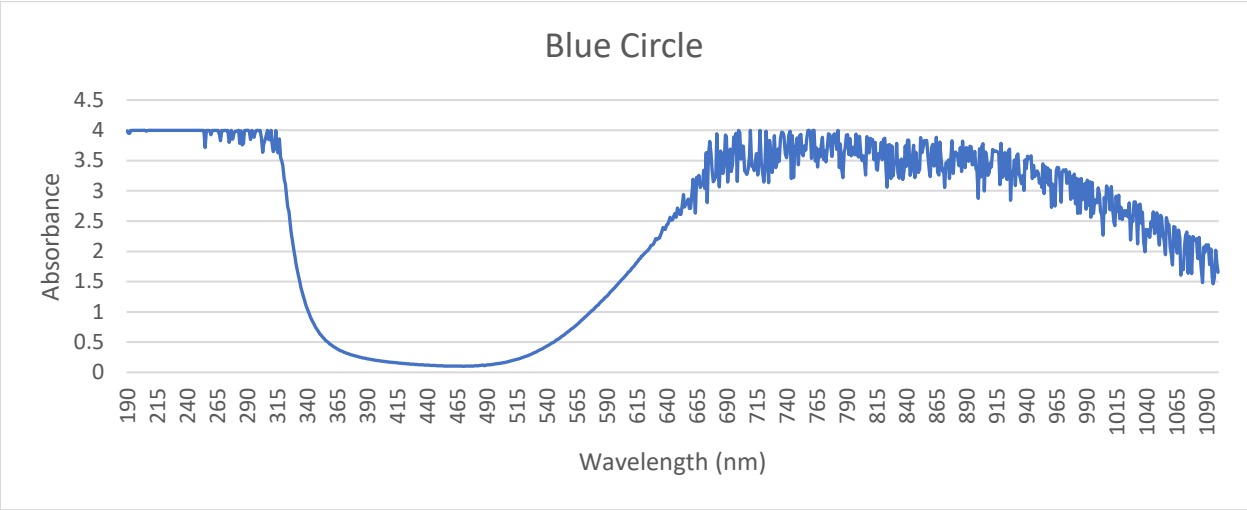


Figure 2 Absorbance spectrum obtained for blue circle filter using a spectrophotometer with the 'blue circle' filter placed over the detector after initial zeroed blank value was obtained.

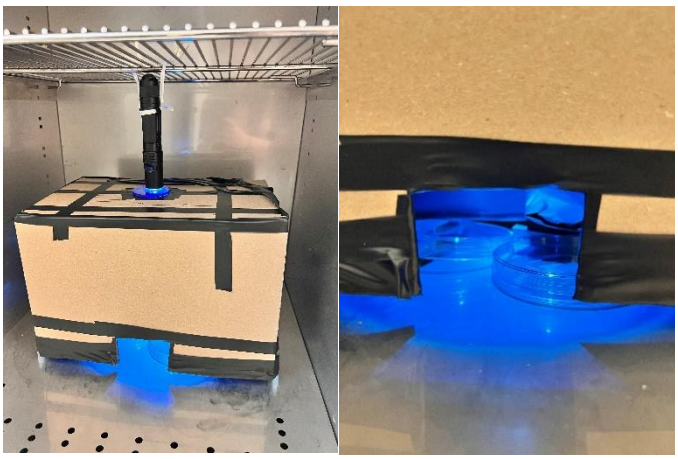


Figure 3 Phototactic movement set up for *N. alkalilacustris sp* using the 'blue circle' filter and *the Fenix UC35 V2.0 1000 Lumen Rechargeable Tactical Flashlight with LumenTac Organizer on the 150 lumen setting.* The box is lined with black paper to eliminate reflection of light outside the desired circular location.

Figure 4 Plate set up for phototactic assay under both white and blue light illumination. The upper level images of the Blue light set show the positive control fully exposed to blue light and the negative control wrapped in black paper to eliminate light exposure. The two test plates were place with half of the plate exposed to the blue light with the other half covered in black paper and placed in the shadow of the box to elicit partial illumination. The white light set up as shown in the below image shows the positive control in the upper left hand corner and the negative control in the bottom right hand corner. The two test plates are also partially covered with black paper.

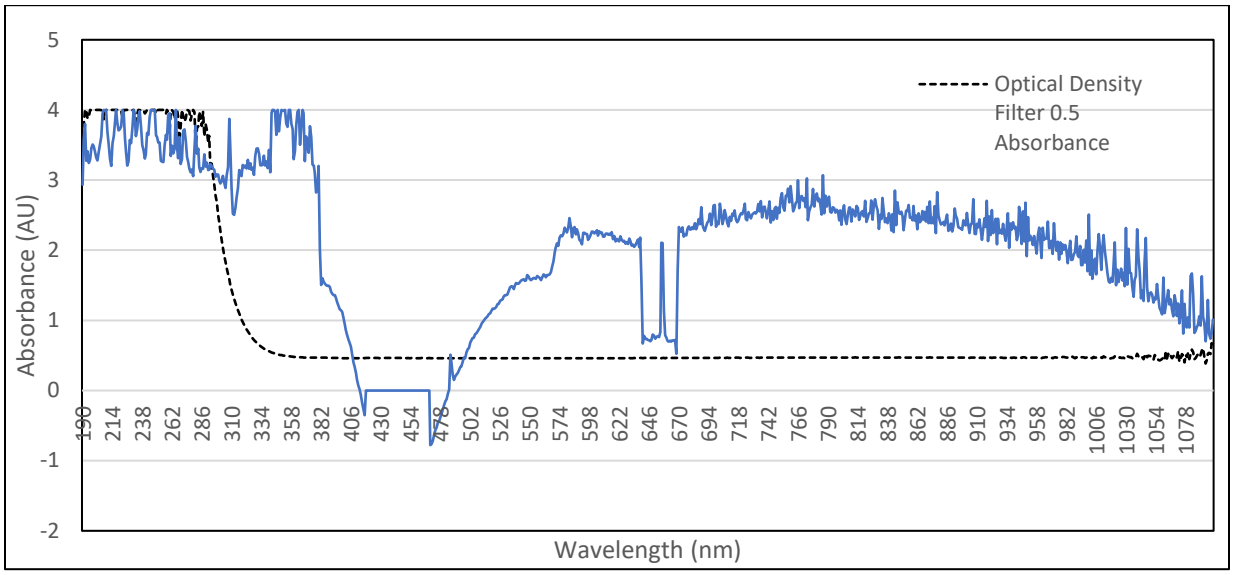


Figure 5 Filter absorbance and wavelength of the LED blue light source obtained using an optical density filter with an absorbance value of 0.5 as the baseline blank value. The blue light LED wavelength depicted as the values below the 0.5 absorbance baseline as shown as the optical density filter 0.5 absorbance. These data show that the blue LED source used here emits light in the region ~400 nm to ~500 nm, exactly in the region that will photoexcite PYP.

2.6 Biofilm formation

To test for biofilm formation in *Massilia albidiflava* and *Massilia plicata*, we inoculated cultures into tubes and allowed them to grow at 30°C for 48 hours. *Idiomarina loihiensis* was inoculated under the same conditions and acted as our positive control, as it has been reported that this organism has reduced biofilm formation triggered by PYP upon blue light illumination [21]. One experiment was conducted with each sample inoculated in their respective growth medium of 20mL volume and allowed to grow at 30°C for 72 hours with exposed to white light of 170 μE/m2/s intensity under static conditions. After growth for 72 hours each sample was visually inspected for biofilm formation.

The following experiment was conducted with the same parameters as stated above with the addition of a sample from each bacterium grown under the same conditions except for the light condition. Here the negative control sample added was covered in aluminum foil to obstruct the light. Again, all samples were observed for biofilm formation after 72 hours.

The next set of parameters for testing biofilm formation consisted of two samples of *I. loihiensis. Each* sample was grown in 15mL of Marine broth and inoculated with the bacterium. One sample was grown with exposure to white light while the other was completely covered in aluminum foil. Of the *Massilia species, M. plicata* and *M. albidiflava* were both inoculated into 2 samples of 15mL of R2A broth. Each sample medium contained 0.5mM solution of pCA. One of each sample was grown exposed to white light while the other was covered in aluminum foil. After 72 hours cultures were visually inspected for biofilm formation.

Another experiment to test biofilm formation included testing multiple conditions and its effect on biofilm formation. Each sample was grown at 30°C in a controlled environment. See *Table 9* for each sample condition. Each sample was grown for 72 hours and observed for biofilm formation.

After the initial biofilm tests were performed and optimal conditions were obtained, 96-well plate assays were performed to quantify biofilm formation using methods somewhat modified from those previously reported [21] and [54] as follows. Each culture was inoculated into 15mL of their respective growth medium with *M. albidiflava* and *M. plicata* and grown at 30°C with light rotation of the cultures for their respective incubation periods to their stationary phase. Each culture was then diluted 1:100 into their respective media and 250 µL of each culture was placed into the wells of the 96-well flat bottom PVC microtiter plates containing

250µL of media, each well contained 2.5µL of bacteria inoculate. One 96-well plate was placed under blue light illumination (with an intensity of 13 µE/m2/s) using the 'Blue LED' and one was placed under white light illumination (170 µE/m2/s), half of each plate was covered with black paper as to perform as a negative control, see *Figure 6* for set up and *Table 10* for well location of each inoculate. The plates were then incubated for 72 hours at 30°C. After incubation for 72 hours the plates were then washed with DI water three times to remove non-attached cells and attached cells were stained with 1% crystal violet for 20 minutes. The plates were then washed with DI water three times to remove non-bound crystal violet. 200µL of 30% acetic acid was then added to the wells to solubilize bound crystal violet and the plates were coved and allowed to incubate at room temperature for 15 minutes. The contents of each well were then gently mixed and 125µL of the solution was placed into a separate optically clear well and optical density was measured at a wavelength of 600 nm.

A second experiment was conducted using the same parameters as above except for the placement of the inoculated wells and the number and light illumination of the plate. One plate was placed under full white light illumination, one was placed under full blue light illumination, and one was completely shielded from light illuminations using black paper in a shaded area of the control environment. All bacteria were placed in the same controlled environment. The placement of the wells was optimized to contain an outer ring of wells consisting of sterile DI water to control moisture exposure and the border effect [55]. See *Figure 7* for the well placement of each bacterium.

I. loihiensis		***M. albidiflava***								***M. plicata***							
2 samples without pCA		4 samples with pCA				4 samples without pCA				4 samples with pCA				4 samples without pCA			
2 samples aerobic		2 samples aerobic		2 samples anaerobic		2 samples aerobic		2 samples anaerobic		2 samples aerobic		2 samples anaerobic		2 samples aerobic		2 samples anaerobic	
light	dark	light	dark	light	dark	light	dark	light	dark	light	dark	light	dark	light	dark	light	dark

Table 9: Species with each condition and number of samples under each condition. *I. loihiensis* has two samples with the only alternation being light and dark. *M. albidiflava* and *M. plicata* will have eight total samples with 4 samples grown with pCA four without pCA, each of those four two are grown aerobically (with the lid unscrewed) and two anaerobically (with the lid tightly screwed on), and of those two on in grown in full white light and one covered in aluminum foil.

Figure 6 Set up for 9-6 well plates A shows the white light illumination set up and B-C shows the blue light illumination set up. Each are contained within the same controlled environment and inoculated concurrently.

M.p (1wk)	M.p (1wk)	M.p (1wk)			M.p (1wk)	M.p (1wk)	M.p (1wk)
M.p (1d)	M.p (1d)	M.p (1d)			M.p (1d)	M.p (1d)	M.p (1d)
M.p pCA (1wk)	M.p pCA (1wk)	M.p pCA (1wk)			M.p pCA (1wk)	M.p pCA (1wk)	M.p pCA (1wk)
M.p pCA (1d)	M.p pCA (1d)	M.p pCA (1d)			M.p pCA (1d)	M.p pCA (1d)	M.p pCA (1d)
M. a (1wk)	M. a (1wk)	M. a (1wk)			M. a (1wk)	M. a (1wk)	M. a (1wk)
M. a (1d)	M. a (1d)	M. a (1d)			M. a (1d)	M. a (1d)	M. a (1d)
M.a pCA (1wk)	M.a pCA (1wk)	M.a pCA (1wk)			M.a pCA (1wk)	M.a pCA (1wk)	M.a pCA (1wk)
M.a pCA (1d)	M.a pCA (1d)	M.a pCA (1d)			M.a pCA (1d)	M.a pCA (1d)	M.a pCA (1d)
I.l (1wk)	I.l (1wk)	I.l (1wk)			I.l (1wk)	I.l (1wk)	I.l (1wk)
I.l (1d)	I.l (1d)	I.l (1d)			I.l (1d)	I.l (1d)	I.l (1d)

Table 10: Well placement for each species and type of culture. M.p represents the wells inoculated with *M. plicata*, M.a represents the wells inoculated with *M. albidiflava*, I.l represents the wells inoculated with *I. loihiensis.* Wells inoculated with week old cultures with previous biofilm formation denoted with (1wk), wells inoculated with 24-hour fresh culture denoted with (1d). Well containing 0.5mM solution by volume of pCA is denoted with annotation pCA all wells without this annotation do not contain pCA. The dark shaded(left) portion of the table represents the side of the 96-well plate covered with black paper. This set up was performed under white light and blue light illumination.

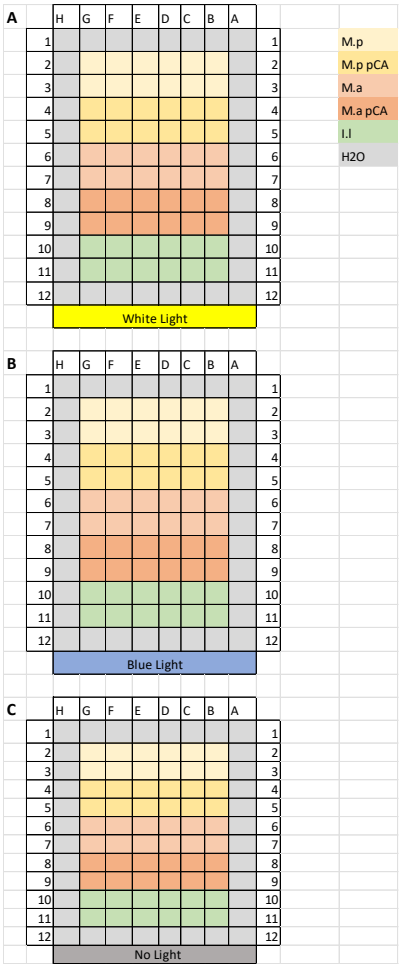


Figure 7 96-well plate assays performed on *M. plicata, M. albidaflava and I. loihiensis.*. Wells annotated and color coded for pCA contained a 0.5mM solution of pCA within the media. All samples were inoculated with week old cultures and allowed to grow statically in a controlled environment at 30°C. Each well contained 250µL of media or sterile DI water respectively. Plate optimized to contain the outer well of the plate with water to eliminate any outstanding factors or alterations due to the outside environment or moisture. Cultures were added to obtain a 1:100 dilution of inoculate to media. M.p= *M. plicata,* M.a= *M. albidiflava,* and I.l= *I. loihiensis.*

3. Results and Discussion

3.1 Bioinformatics

The genomic BLAST performed in NCBI using PYP1 from *H. halophila* query resulted in 1,120 homologs after manual curation of the hits based on the presence of highly conserved residues in the PYP family. To gain insights into the degree of sequence diversity in these 1,120 proteins, all pairwise percentages, sequence identify, and similarity for this set of proteins were computed and plotted as a histogram. *Figure 8* shows the vast diversity between the PYP sequences when analyzing the percent identity and similarity revealed by this approach. Single and multi-domain PYPs were found to be present in organisms belonging to a total of 8 different bacterial phyla. A phylogenetic tree for all 135 PYP homologs found within the KEGG database also showing the domain structure of these proteins is depicted in *Figure 9.* Approximately 23% of PYPs identified were multi-domain proteins while the remaining PYPs were single domain.

We aimed to use this set of PYP homologs to do further bioinformatics studies to identify additional candidate protein functionally associated with PYP. In this work we followed a two-pronged approach: (1) we examined the presence of conserved domains in the multi-domain PYPs; and (2) we examined the presence of genes that are part of predicted operons also containing the *pyp* gene. Published research regarding those conserved domains and genes then can help generate functional hypotheses on these proteins and how they relate to PYP function. Due to the limitations of the NCBI database in depicting the genome map of each species encoding a PYP, further studies were conducted using the KEGG database. The KEGG database has a more limited array of organisms, containing 135 of the PYP homologs identified in NCBI. Further studies using the full homolog query results from the NCBI database is underway.

Of the PYP homologs identified using the KEGG BLAST, using cut off of 95% sequence similarity compared to the sequences found in NCBI, the homologs were analyzed for flanking genes that appear to be part of a multi-domain PYP based on short inter-genic distances and being transcribed in the same direction. *Table 11* depicts the flanking genes of interest in the multi-domain PYPs identified using the KEGG genome map and *Figure 11* shows graphical representation of other key genes of interest related to PYP. Each flanking gene was then considered for possible biological functions using published results on the flanking genes.

Based on this bioinformatics dataset and considering the availability of strains for laboratory growth under aerobic conditions, we selected two aspects of PYP function for further study. First, we selected *N. alkalilacustris* because its genome encodes an MCP and possible relation to phototactic movement as seen by *E. coli* which contains the MCP and undergoes chemotaxis and shows motility upon exposure to chemical compounds, and archaea which use an MCP and the photoreceptor rhodopsin to trigger biological response [23, 24]. See *Figure 10* for the predicted structure of *N. alkalilacustris sp.* constructed using the known structure for *E. coli.* Second, *Massilia albidiflava* and *plicata* were selected because they contain genes of proteins containing a GGDEF motif in a predicted operon also containing the pyp gene *Figure 12* , leading to the prediction that this organism exhibits PYP-induced light-triggered photo-regulation of biofilm formation, similar to the response reported for *I. loihiensis[21].*

The genomes of *N. alkalilacustris, M. albidiflava,* and *M. plicata* were also analyzed for the presence of genes encoding the enzymes TAL and *p*CL (required for the biosynthesis of the pCA chromophore of PYP) using the amino acid sequence of the genes encoding these proteins in *H. halophila.* For TAL analysis *N. alkalilacustris* upon analysis using the NCBI protein

BLAST tools had two blast hits with accensions as follows; WP_151704889.1 and WP_240776273.1. Accension WP_151704889.1 had a 34.78 percent identity and WP_240776273.1 a 31.32 percent identity. *M. albidiflava* analysis resulted in three BLAST hits; WP_131148831.1, GGY46706.1, and WP_131144526.1. Each had a percent identity as follows, WP_131148831.1 34.4%, GGY46706.1 34.4%, and WP_131144526.1 34.56%. *M. plicata* resulted in three hits, GGZ08051.1 with 34.54% identity, WP_134385340.1 with 34.54%, and WP_134387103.1 with 31.58%. Each hit when performing a KEGG BLAST upon each FASTA amino acid sequence resulted in a histidine pathway with EC number 43.1.3. The original TAL within PYP1 of *H. halophila* has an EC number of 43.1.23. For pCL analysis *N. alkalilacustris* resulted in four results with accension and percent identity, WP_151703207.1 34.89%, WP_151703530.1 25.86%, WP_151705422.1 21.04%, and WP_151704733.1 24.53%. Each accension was then analyzed within KEGG for related sequence identity WP_151703207.1 resulted in a hypothetical protein in *Methylomonas koyamae,* WP_151703530.1 a putative acetyl-coenzyme a synthetase with an EC number 6.2.1.32 in *Azoarcus olearius BH72,* WP_151705422.1 a AMP-binding protein with an EC 6.2.1.- in *Nitrincola iocasae,* and WP_151704733.1 a acyl-CaA synthetase with a EC 6.2.1.44 in *Nitrincola iocasae. M. albidiflava* results included WP_131148053.1 with a percent identity of 20.91. This ascension classification in KEGG resulted in an amino acid adenylation domain-containing protein in *Pseudoduganella albidiflava (M. albidiflava). M. plicata* results as follows; WP_134387101.1 with 27.74%, WP_189568707.1 with 25.44%, and WP_134382918.1 with 31.36%. WP_134387101.1 within KEGG resulted in a long-chain fatty acid—CoA ligase in *Pseudoduganella plicata (M. plicata),* WP_189568707.1 an AMP-binding protein with EC 6.2.1.32 in *Telluria chitinolytica,* and WP_134382918.1 a non-ribosomal peptide synthetase with

EC 6.2.1.63 in *Pseudoduganella plicata.* The *p*CL within PYP1 is an AMP-binding protein in *H. halophila.*

The low percentage amino acid identity and non-matching EC numbers that do not correspond to the same gene of interest as identified in *H. halophila* indicates that *N. alkalilacustris*, *M. albidiflava*, and *M. plicata* do not contain clear homologs of the TAL or pCL genes. This is indicative of the plausible need for adding pCA to elicit the desired responses as these bacteria may obtain their pCA from their environment as TAL and pCL are needed to make the *p*CA and attach it to apoPYP, to elicit the active state of PYP, thus without the ability to produce *p*CA the bacteria would need to obtain it from the environment. In these considerations it should be noted that the annotation of substrate specificity of enzymes based on their amino acid sequence is difficult and often non conclusive.

Since we hypothesize that the MCP-PYP multidomain protein in *N. alkalilacustris* triggers a phototaxis response to blue light, we analyze the genome of this organism for the presence of components related to MCP flagellar signaling such as CheA, CheY, and CheW using the amino acid sequence of these proteins from *Escherichia coli K-12 MG1655* for each respective protein using NCBI BLASTP. CheA top populated results as follows; WP_151704063.1 for chemotaxis protein CheA with EC 2.7.13.3. CheY top result was WP_151702677.1 for chemotaxis protein CheY with KEGG KO3413. CheW; WP_151704062.1 for chemotaxis protein CheW with KO3408. *E. coli* KEGG results for CheA were chemotaxis protein CheA with EC 2.7.13.3, CheY chemotaxis protein CheY with KO3413, and CheW chemotaxis protein CheW with KO3408. These results indicate that *N. alkalilacustris* does

contain the needed components for signal transduction to occur from the MCP to the flagellar proteins in order to elicit tactic movement.

See supplemental documents for raw analysis of bioinformatics work regarding the TAL, pCL, and Chemotaxis proteins.

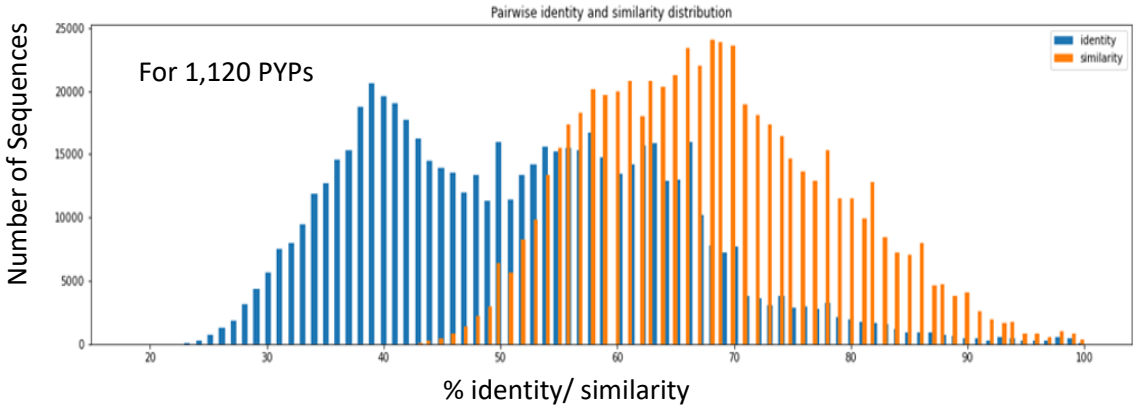


Figure 6 Percent identity and similarity distribution derived from PYP homologs found using NCBI BLASTP. The distribution was generated using a general protein pipeline coded by Gunnar Hoogerwerf.

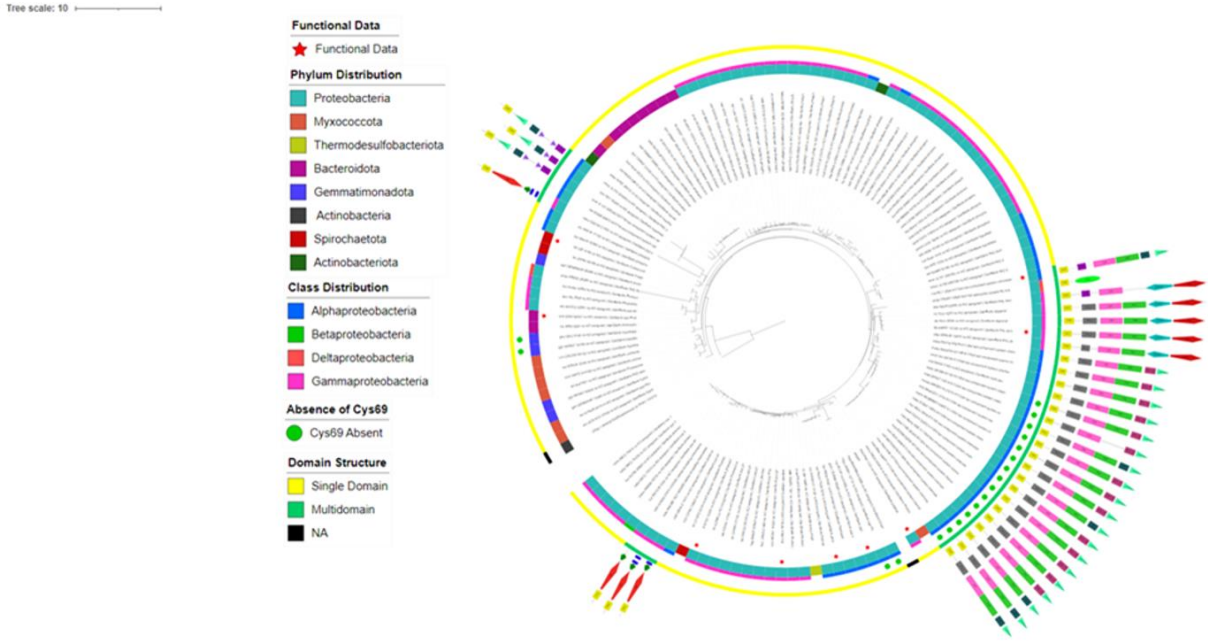


Figure 7 Phylogenetic tree of identified single and multi-domain pyp homologs using the KEGG database. Not all identified pyp homologs from NCBI are identified within the KEGG database. 135 PYP homologs were found in KEGG, using a structurally related PAS domain from *Rhodococcus jostii* RHA1 as an outgroup to root the tree. PYP homologs for which biochemical data has been reported are indicated with an *. Only 9/11 of the PYP with functional data are annotated on the tree. *R. salexigens* and *H. salexigens* were not present in KEGG so they were not annotated. The phylogenetic tree was created by Rosalie Dohmen.

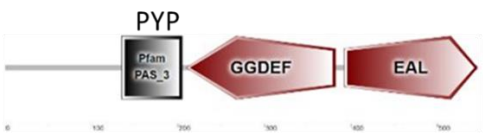
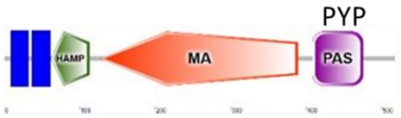
Genes	Pictogram
PYP-GGDEF-EAL	
PYP-PHY-GGDEF-EAL	
MCP-PYP	
PYP-PHY-HK	

Table 11: Pictogram representing the multi-domain structure of flanking genes to the *PYP* denoted as PAS domain. PHY= phytochrome, MCP= methyl-accepting chemotaxis protein, HK= histidine kinase. Domain structures were created using the SMART database. Note: MCPs contain HAMP and MA domains.

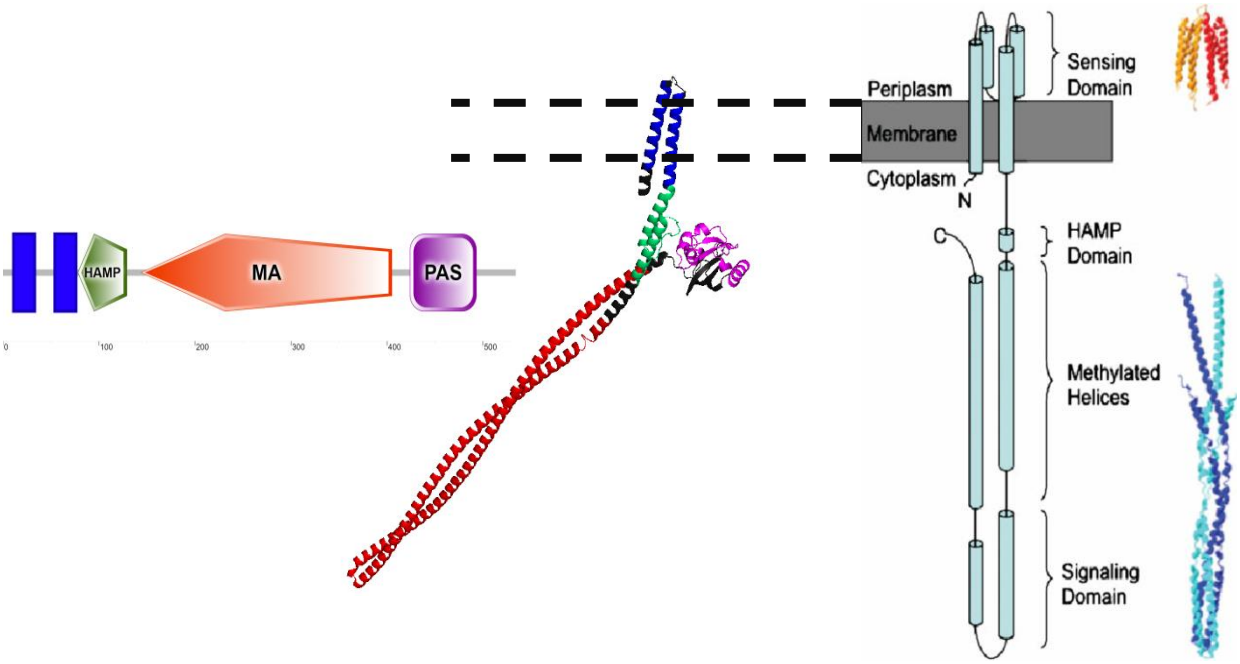


Figure 80 Predicted structure of *N. alkalilacustris.* This prediction was based off of the *E. coli* structure and created by Rosalie Dohmen using the Robetta Fold program.and the SMART database.

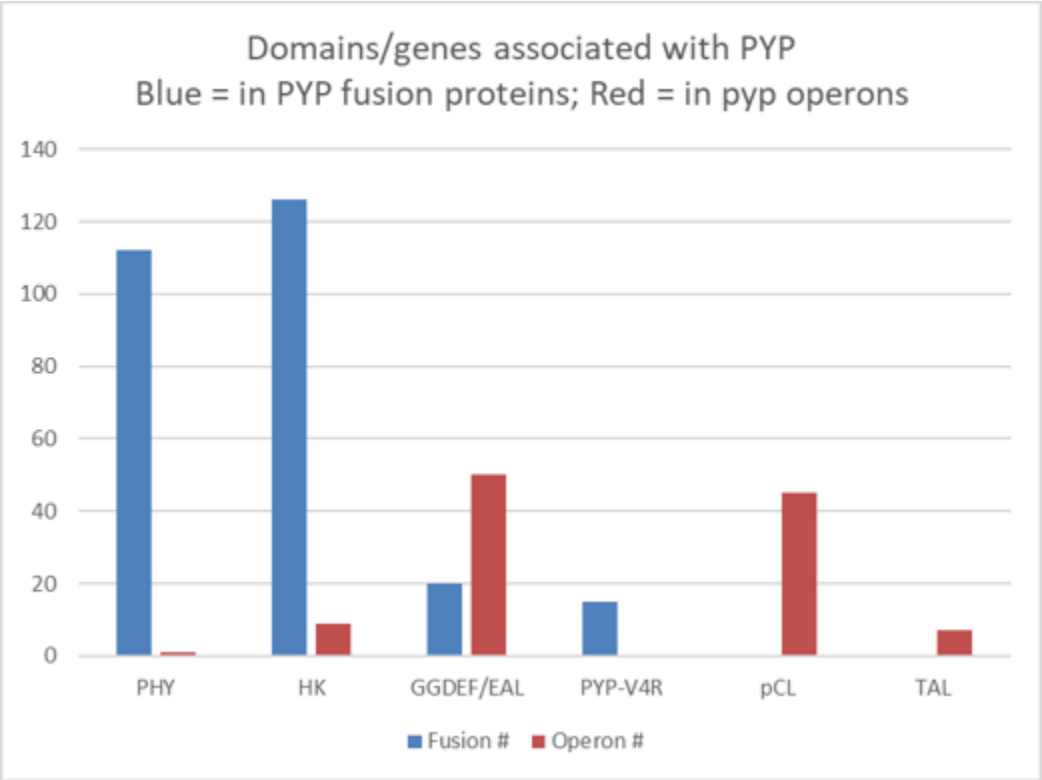


Figure 11 Gene analysis with respect to their relationship to the PYP. Blue = fusion proteins, Red= within the pyp operon. PHY=phytochrome, HK= histidine kinase, pCL= p-coumaroyl CoA ligase, Tal= tyrosine ammonia lyase. This figure summarizes information on selected functional domains associated with *pyp* genes by being present in a PYP multidomain protein (red) or by being part of a predicted *pyp* operon *blue).

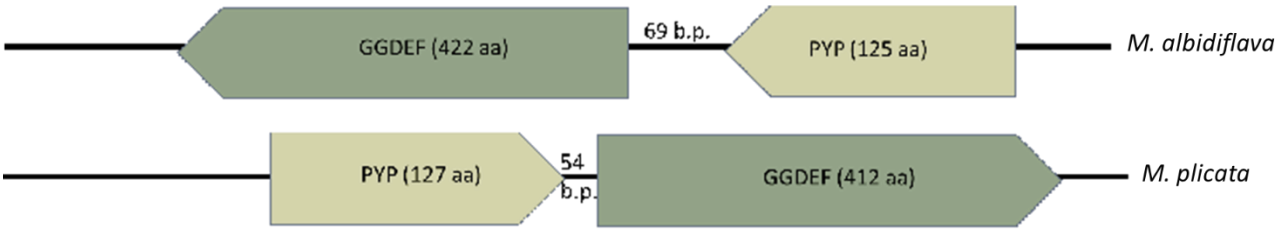


Figure 12 Pictorial representation of the *M. albidiflava* and *M. plicata* species operons indicating the number of amino acids within each gene and the base pair length between each gene indicating they are a part of the operon.

3.2 Bacterial growth and morphology

Nitrincola alkalilacustris is a newly described species discovered in 2017 therefore TEM imaging had yet to be published. Since imaging was limited, we performed transmission electron microscopy of *N. alkalilacustris* conducted using both a pelleted form and broth culture suspension yielded in various cell shape and size, see *Figures 13 & 14* for resultant images. The notable cell shape and size variation is intriguing and warrants further investigation. Of note many of the cells were congregated near apparent salt crystals. This is of interest for further studies for identification of the ions in the crystal component and their relation to cellular growth. Other notable components related to growth upon examination of the TEM images includes the possible presence of a viral component and possible spore production.

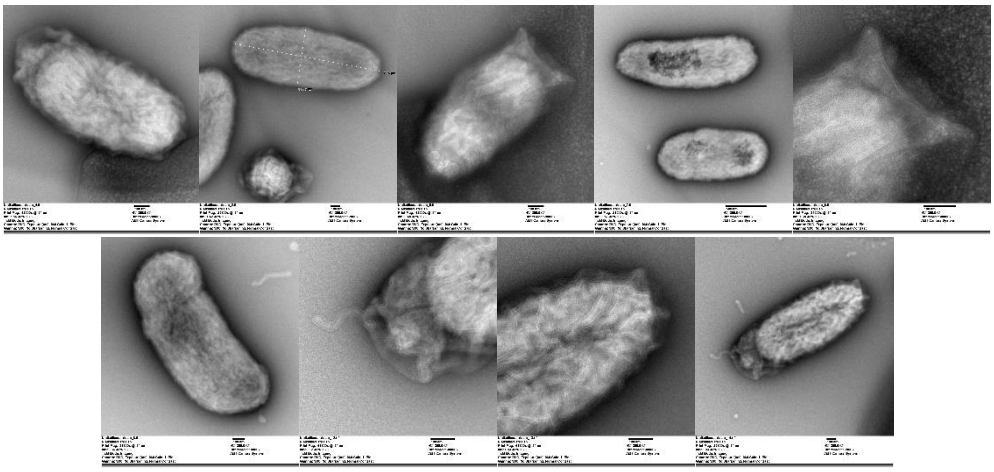


Figure 13 N. alkalilacustris sp transmission electron microscopy images of 48-hour growth in marine broth culture. The culture was then centrifuged to form a pellet and resuspended in the fixative agent. Images were obtained at OSU microscopy laboratory.

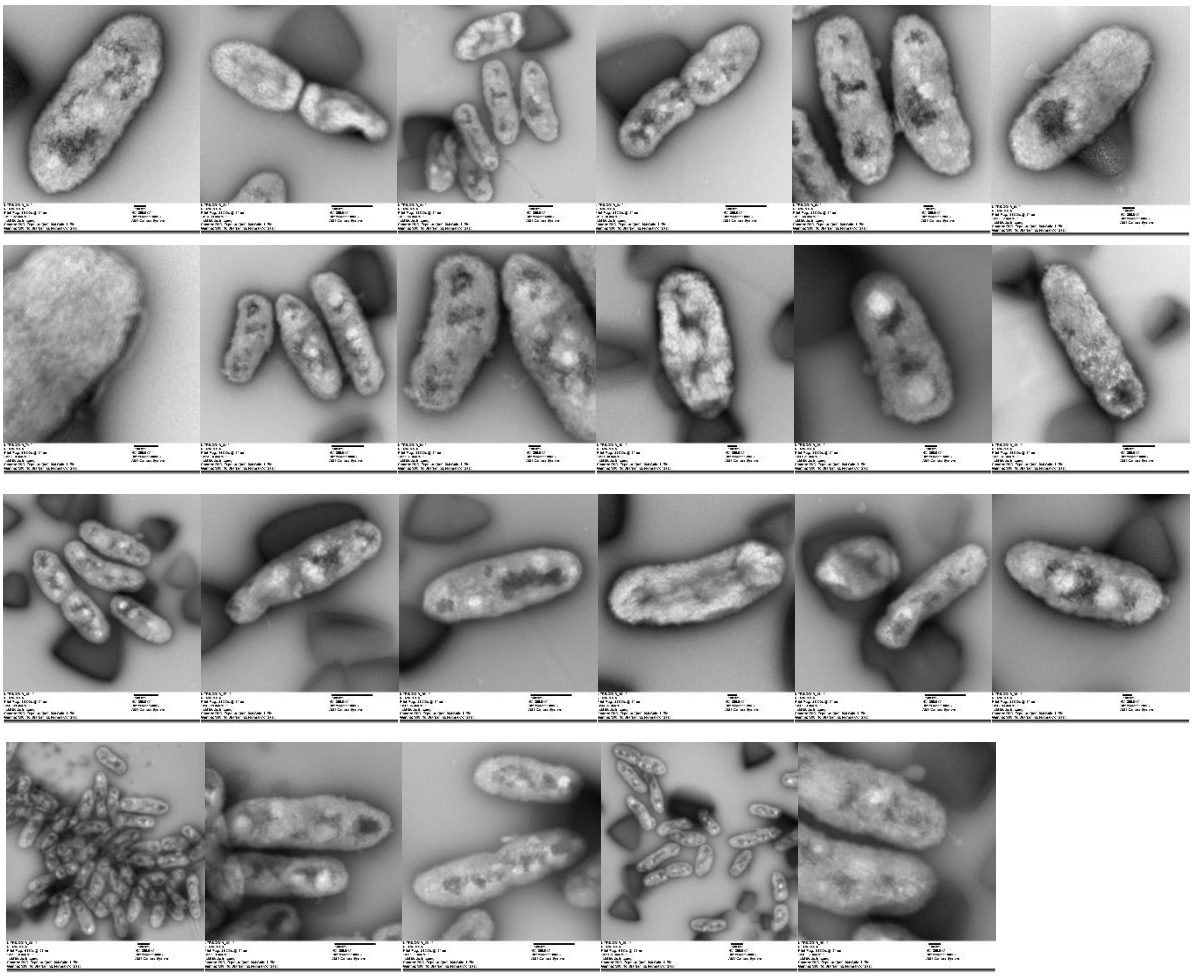


Figure 14 N. alkalilacustris sp transmission electron microscopy images of 48-hour growth in marine broth culture. The culture was then placed into the fixative agent at a 20:100 ratio of culture to fixative. Images obtained by the OSU Microscopy Laboratory.

3.3 gDNA Extraction, PCR, and Gel-Electrophoresis

Before commencing physiological studies, we aimed to use PCR to ascertain that the genes of interest related to PYP were indeed present in the genomes of the strains that we received. *Table 12* depicts the concentrations and purity obtained using the 260/280 and 260/230 purity ratios from the genomic DNA extraction and purity testing for *N. alkalilacustris, I.*

loihiensis, M. albidiflava, and *M. plicata*. For the 260/280 ratio we utilized the approximate cutoff for DNA as ~1.8. Each sample was within the ideal range for pure DNA for the 1.8 260/280 cut off. However, we had relatively low 260/230 ratio purity which should be remedied in future studies. Nonetheless, the gDNA was used for identification of components of interest: PYP, MCP and GGDEF.

Upon completion of PCR and gel-electrophoresis, the *N. alkalilacustris sp pyp* gene was expected to be 427 base pairs (bp) in size, MCP 867bp, and the MCP-PYP fusion 1,294bp. The *I loihiensis pyp gene* was about 409bp in size. The *M. albidiflava pyp* gene was absent and GGDEF motif was 565bp. The *M. plicata pyp* gene was 413bp and GGDEF motif was absent, which we assume is due to poor PCR amplification. See figures *Figure 15, 16, & 17* for simulated gels using the SnapGene database and the experimentally observed gels. Since the PCR product for the *pyp* gene for *M. albidiflava* and the GGDEF motif *M. plicata was not detected,* a second round of DNA extraction, PCR, and gel needs to be performed. However, the predicted gels state the size of the genes are as follows, for *M. albidiflava* and *M. plicata*, the *pyp* gene was 403 bp, and the GGDEF was 443 bp respectively. Of note, for the *M. albidiflava* and *M. plicata* gels there appears to be double banding for which is due to the selected primers binding to other locations within the genome, for further studies one should obtain primers free of other binding sites within the genome. The presence of each respective gene for the given organisms at the relative gene size given by the simulated gels indicated that all needed components for the downstream signaling to occur are contained within the organism.

Organism	Concentration (ng/µL)	260/280	260/230
N. alkalilacustris	124.3	1.88	1.70
M. albidiflava	285.6	1.84	0.95
M. plicata	279.4	1.87	0.97
I. loihiensis	88.1	1.95	0.55

Table 12 gDNA extraction concentrations and purity performed using the 260/280 and 260/230 cut offs.

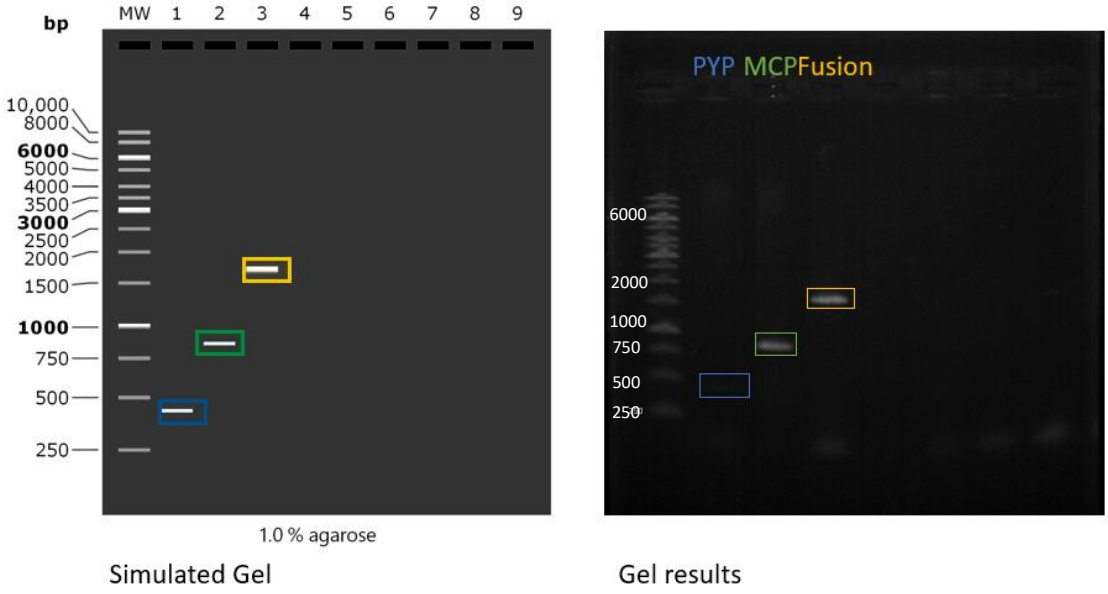


Figure 15 PCR and Gel-electrophoresis performed on *N. alkalilacustris* for confirmation of PYP, MCP, and PYP-MCP fusion protein. Left image is a simulated gel using the SnapGene database for a 1.0% agarose gel. The right image is the experimentally observed pattern of PCR fragments. PYP enclosed in a blue box, MCP enclosed in a green box, and the PYP-MCP fusion enclosed in a yellow box in both images. For reference the simulated gel provides the following information about well components: 1: Nitrincola alkalilacustris genome map genomic1_R + PYP_F 1. 427 bp 2: Nitrincola alkalilacustris genome map 2 mcp-pyp_R + Genomic1_F 1. 867 bp 3: Nitrincola alkalilacustris genome map 3 genomic1_R + Genomic1_F 1. 1679 bp

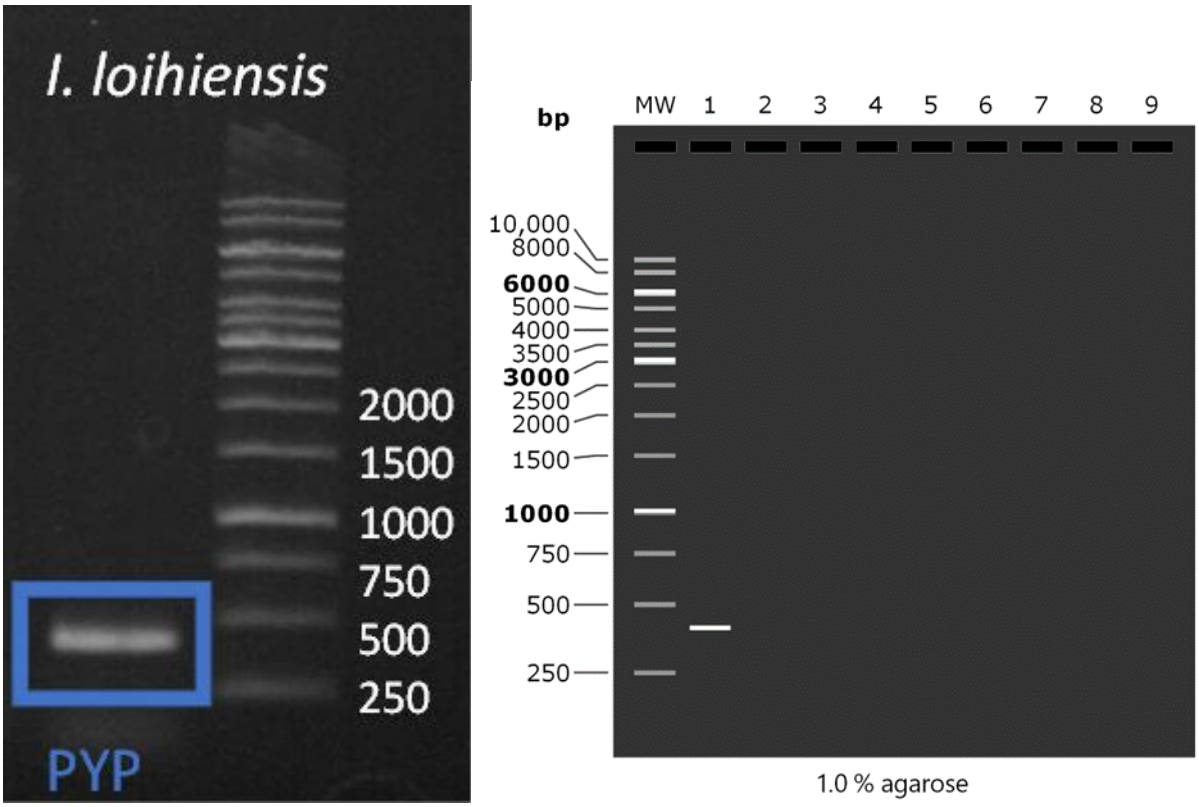


Figure 16 PCR and Gel-electrophoresis performed on *I. loihiensis* for confirmation of the presence of the *pyp* gene. The right image is a simulated gel using the SnapGene database for a 1.0% agarose gel while the left image is the performed gel. PYP enclosed in a blue box, the GGDEF motif was not included due to size limitations of PCR. The simulated gel provides the following information regarding the well components: 1: AE017340 I.l_PYP_F + I.l_PYP_R 1. 409 bp.

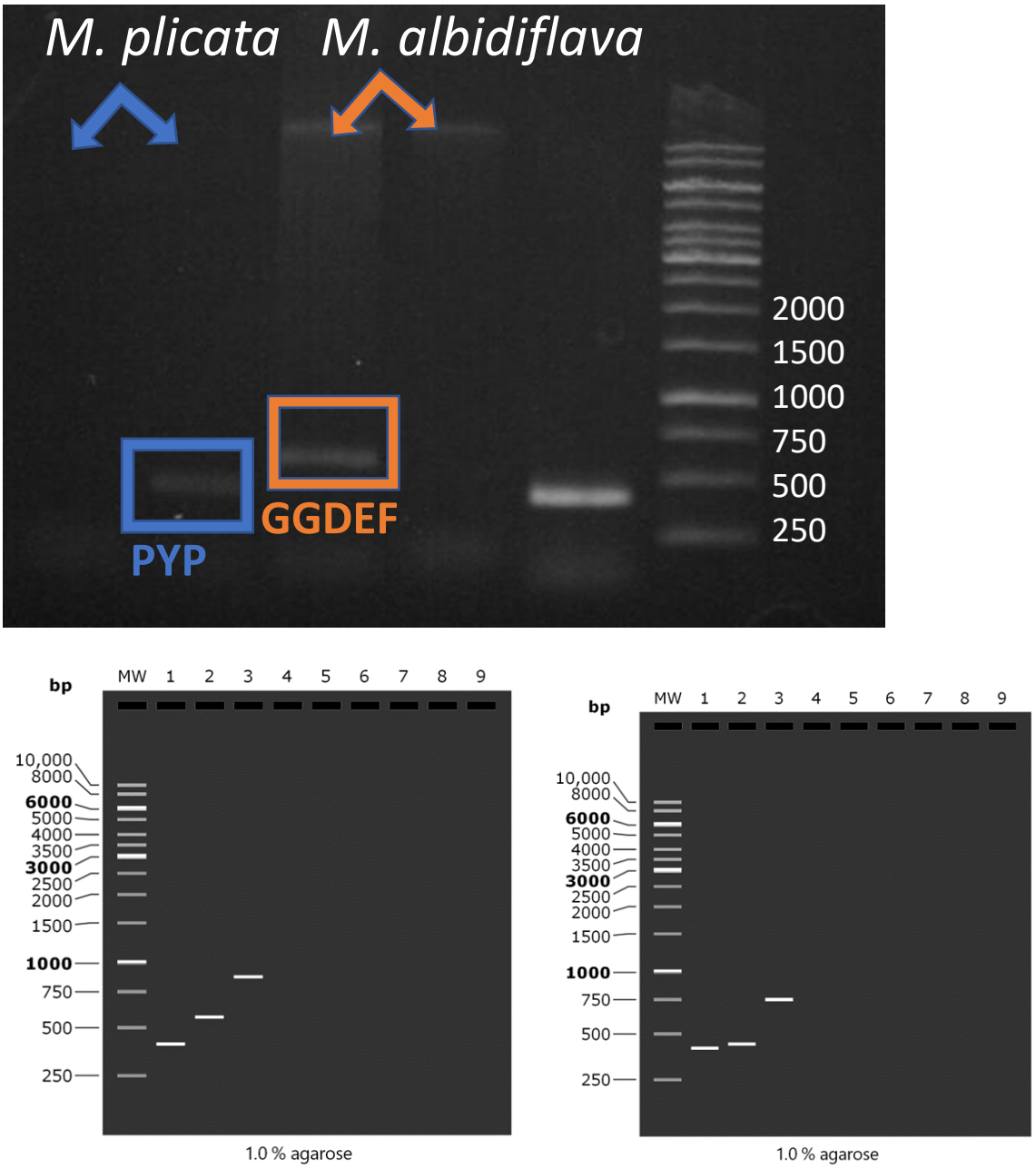


Figure 17 PCR and Gel-electrophoresis performed on *M. albidiflava* and *M. plicata* for confirmation of the PYP and GGDEF motif. The bottom left image is a simulated gel using the SnapGene database for a 1.0% agarose gel for *M. albidiflava*, and the bottom right is a simulated gel for *M. plicata*. PYP for *M. plicata* enclosed in a blue box GGDEF gene enclosed in an orange

box for *M. albidiflava*. The simulated gels provide the well component information as follows; *M. albidiflava*, 1: CP036401 M.a_PYP_F + M.a_PYP_R 1. 403 bp 2: CP036401 M.a_PYP-GGDEF_F + M.a_PYP-GGDEF_R 1. 565 bp 3: CP036401 2 M.a_PYP_F + M.a_PYP-GGDEF_R 1. 868 bp, and *M. plicata*, 1: CP038026 M.p_PYP_F + M.p_PYP_R 1. 413 bp 2: CP038026 M.p_PYP-GGDEF_F + M.p_PYP-GGDEF_R 1. 443 bp 3: CP038026 M.p_PYP_F + M.p_PYP-GGDEF_R 1. 747 bp.

3.4 Motility

N. alkalilacustris is a part of the *Gammaproteobacteria* and family *Oceanospirillaceae*. *N. alkalilacustris* was reported to be a motile bacterium propelled by a singular polar flagellum [38]. Using this information, an analysis of motility was conducted to differentiate between swimming, swarming, twitching, and gliding motility. While the exact definition and determination between the motility types is somewhat debated, for our purposes swimming motility is defined as the movement of cells through a liquid medium by means of a singular flagellum, while swarming motility consists of multi-cell coordinated interactions though liquid films on surfaces, often accompanied by surfactant production and multiple flagella[43]. Twitching is known for movement upon semi-solid to solid surfaces through type IV pili causing an intermittent or uneven movement [46] and gliding is known for a more uniform motion through the use of cellular proteins across semi-solid to solid surfaces [43]. These differentiations in motility can be determined through macroscopic techniques such as motility in altered concentrations of semi-solid agar assays as well as testing for surfactants and performing imaging techniques for observation of flagella.

Initial altered semi-solid agar concentration plate assays were performed, showing that *N. alkalilacustris* is motile under the growth conditions used. For 0.5% concentrated agar there was little motility observed and a 0.2% agar concentration resulted in the greatest motility, indicating the possibility of either swimming or swarming motility based on the analysis conducted from published results [43-49]. However, it is important to note that not all species react in the same way using this analysis method. *Figure 18* shows motility results for all agar concentrations tested under white light illumination and during 30°C incubation.

TEM images observed for the presence of flagella were inconclusive as there were not obvious flagella on all cells, but rather the appearance of a double stranded tail and polar end of few cells leading to inconclusive results, however, since there was a lack of multiple flagella this suggests that our organism is undergoing swimming motility. *Figure 19* shows all TEM images with possible flagellar like projections. Possible errors in preparation could have led to the destruction or disturbance of flagellar identification through TEM, however, with the results from our TEM images there is inconclusive results of flagellum presence which contraindicates the results shown by Borsodi, Korponai [38].

Drop-collapse assays resulted in the conclusion of no surfactant production suggesting that swimming motility is detected as the water drop placed upon the plate reaming in rigid formation, see *Table 13* for images of drop-collapse assay for 0.3% and 0.2% concentrated agar plates. *Figure 20* shows addition images for drop-collapse assay performed on 0.2% concentrated agar plates contained 0.5mM solution of pCA. The addition of pCA was added as *N. alkalilacustris* lacks the TAL and pCL genes which are needed for production of pCA and binding to the apoPYP for confirmational change to the active state.

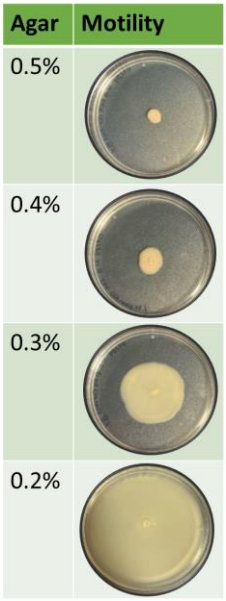


Figure 18 Altered semi-solid agar-concentration for analysis of motility performed under white light emissions. All plates were inoculated concurrently and kept in a controlled environment at 30°C for 48 hours. Each plate was inoculated using a sterile inoculating loop that was inserted half-way through the agar media at approximately the center of the plate to ensure the culture would not move upon moving the plate to the incubation room.

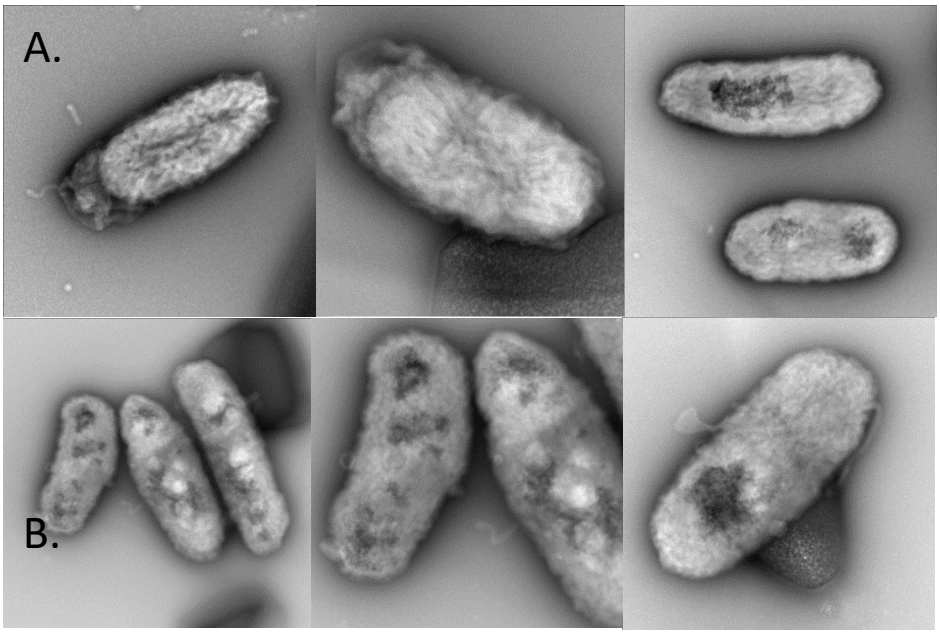


Figure 19 N. alkalilacustris sp transmission electron microscopy images of 48 hour growth in marine broth culture. A) The culture was then centrifuged to form a pellet and resuspended in the fixative agent, B) the culture was then placed into the fixative agent at a 20:100 ratio of culture to fixative. These images are of selected cells which appear to contain a single flagella like appendage.

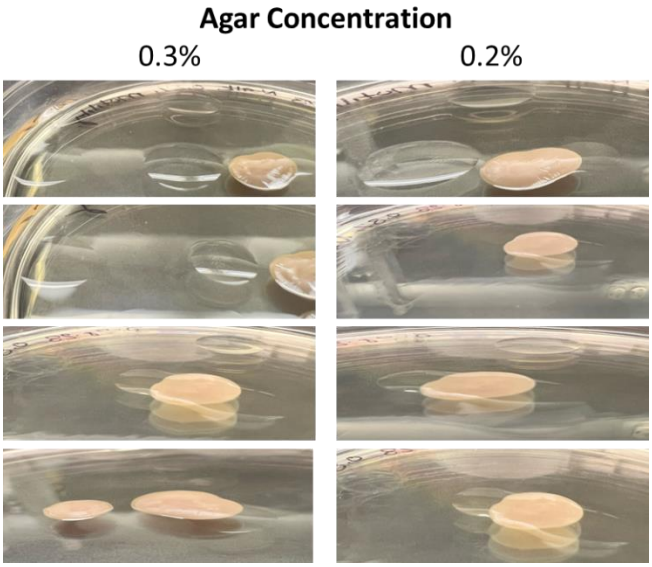


Table 13 Drop-collapse assay to probe surfactant production performed at 0.3% and 0.2% concentrated agar. Each plate the assay involved several drops of water placed on the plate both near the inoculate growth and at the edge of the plate. The resultant images are presented. No clear signs of surfactant production were observed.

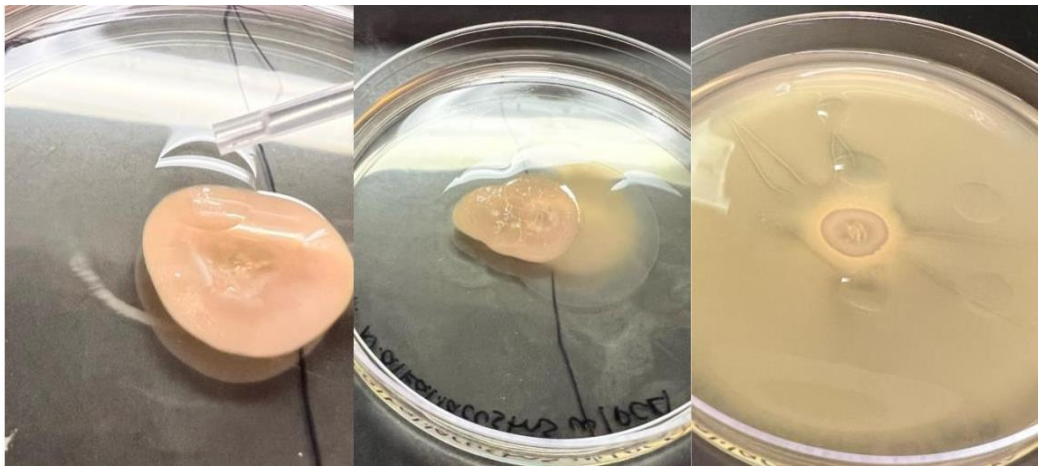


Figure 20 Drop-collapse assay to probe surfactant production performed on 0.2% concentrated agar containing a 0.5mM solution of pCA. Each plate involved several drops of water placed on the plate. No clear signs of surfactant production were observed.

3.5 Phototactic Motility

Initial phototactic assays were performed under white light and were inconclusive, as there was no distinction between the negative control plates, positive control plates, or test plates performed on altered concentration agar plates from 0.5%-0.2% as depicted in *Table 14*. Further testing using the Fenix UC35 V2.0 1000 Lumen Rechargeable Tactical Flashlight with LumenTac Organizer on the 150 lumen setting and the 'Blue Circle" filter (see Figure 2) were performed to test phototactic results under blue light illumination of about 450nm. The results of this experiment were also inconclusive, as there again was no differentiation between the plates of the negative control, positive control, or test plates, as shown in *Table 15* which lists the phototactic results of *N. alkalilacustris* on 0.2% concentrated agar plates, which is the optimized agar concentration for motility. It is of note that there were problems with humidity which resulted in a layer of moisture on the lids of the plates that could have altered the results by adding surface moisture allowing motility to occur by other means.

We next considered the bioinformatics results with the indication that *N. alkalilacustris* does not contain clear homologs for the TAL or *p*CL genes as present in *H. halophila*, implying that this organism may be unable to produce its own *p*CA. We therefore hypothesized that the addition of *p*CA may be needed for PYP-derived photo-responses, as this species would need to take up *p*CA from its environment in order to activate its apoPYP. With this new information, *p*CA was added to our growth media to obtain a 0.5mM solution, and the phototaxis experiments were performed again. In this experiment a negative control, positive control, and a test plate were all inoculated and placed under white and blue light illuminations. We used the same white light and the 'Blue LED" light. See *Table 16* for phototactic results for *N. alkalilacustris* in the

presence of *p*CA. The addition of pCA in *Table 16* appears to restore positive phototactic motility which is to be confirmed with biological replicates in future studies.

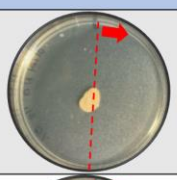
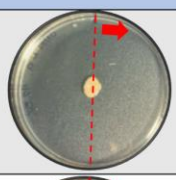
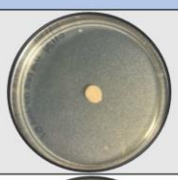
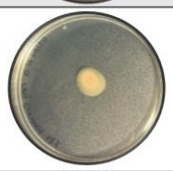
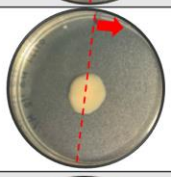
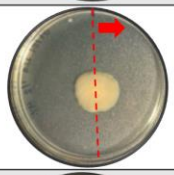
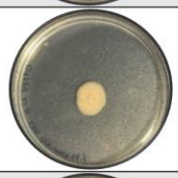
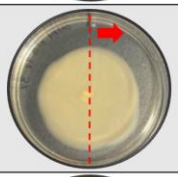
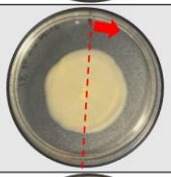
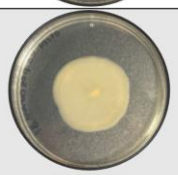
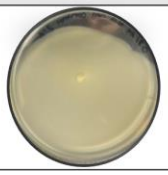
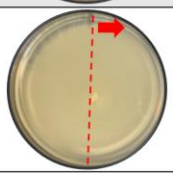
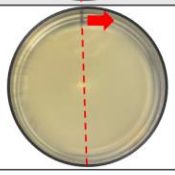
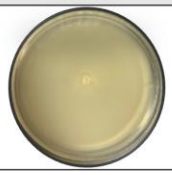
Agar concentration	Negative Control	Test	Positive Control
0.5%		 	
0.4%		 	
0.3%		 	
0.2%		 	

Table 14 Phototactic motility assay performed on *N. alkalilacustris* on varying semi-solid agar concentration plates from 0.5%-0.2% under white light illumination. Each contained a negative control plate completely covered with aluminum foil to eliminate all light exposure, a positive control plate unshielded from light, and two test plates consisting of partial light exposure and partial coverage by aluminum foil. The side of the test plates exposed to white light is indicated with the arrowhead facing the light and the dotted line indicates the placement of the light shield. These plates contain Marine broth and granulated agar to form the desired concentrations.

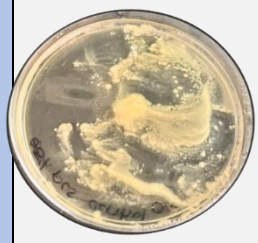
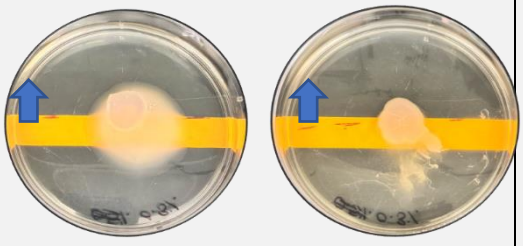
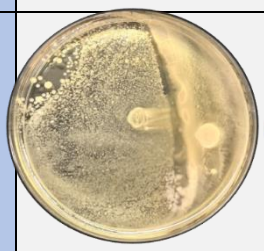
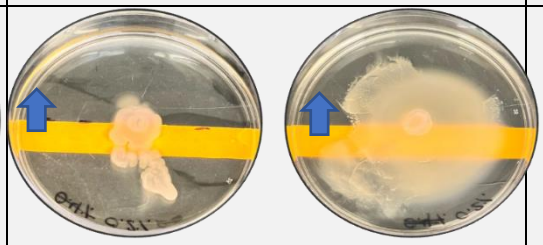
Agar Concentration	Negative control	Test Plates	Positive control
0.3%			
0.2%			

Table 15 Phototactic motility assay performed on *N. alkalilacustris* on varying semi-solid agar concentration plates from 0.3%-0.2% under blue light illumination. Each contained a negative control plate completely covered with aluminum foil to eliminate all light exposure, a positive control plate unshielded from light, and two test plates consisting of partial light exposure and partial coverage by aluminum foil. The side of the test plates exposed to blue light is indicated with the arrowhead facing the light and the orange taped line indicates the placement of the light shield. These plates contain Marine broth and granulated agar to form the desired concentrations.

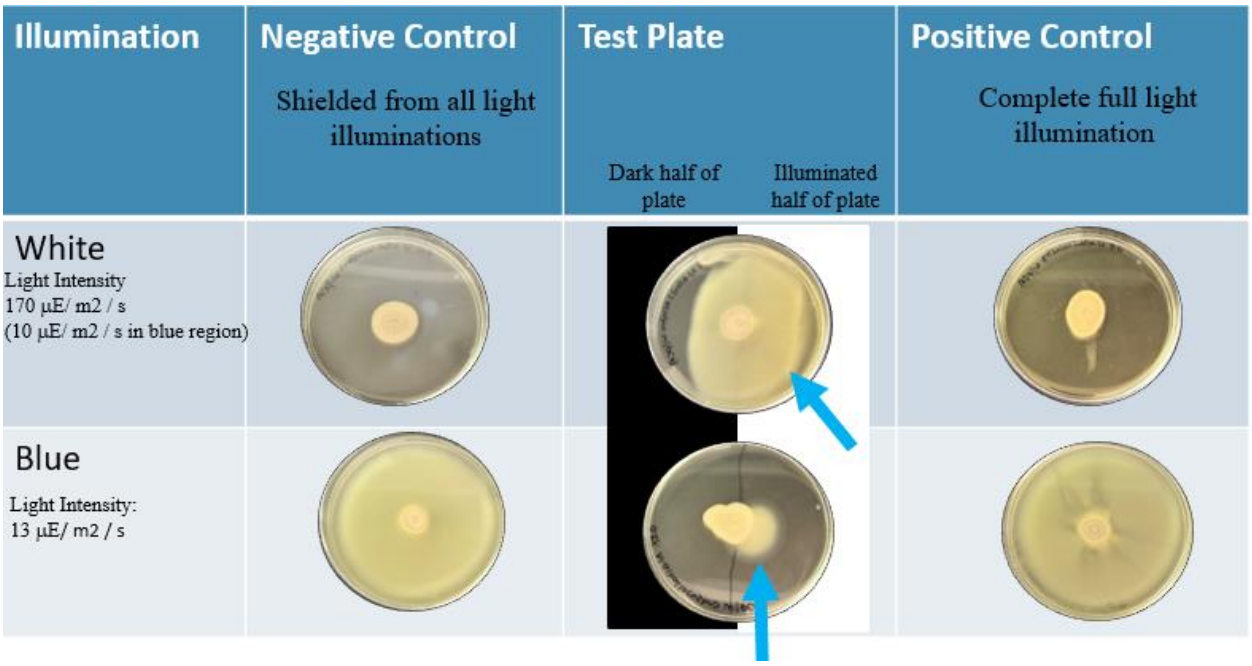


Table 16 Phototactic motility assay performed on *N. alkalilacustris* on 0.2% semi-solid agar concentration plates under blue and white light illumination with the addition of a 0.5mM solution of pCA. Each contained a negative control plate completely covered with aluminum foil to eliminate all light exposure, a positive control plate unshielded from light, and a test plates consisting of partial light exposure and partial coverage by aluminum foil. The side of the test plates exposed to light is indicated with the arrowhead facing the light and the orange taped line indicates the placement of the light shield. These plates contain Marine broth and granulated agar to form the desired concentrations. Each sample shows the intensity of the light upon the plates.

3.6 Biofilm Formation

Both *Massilia* species studied here are a part of the taxonomic class *Betaproteobacteria* and family *Oxalobacteraceae*, and other (not studied in our laboratory) *Massilia* species have been shown to produce biofilms at the surface of broth cultures [56]. Each species studied in our

laboratory contain the necessary genes for biofilm formation including the GGDEF motif as shown in out bioinformatic results and confirmed PCR and gel-electrophoresis. *Idiomarina loihiensis* was used as our positive control as it has been reported to contain both the PYP and GGDEF motif through bioinformatics studies as well as experimental studies indicating that the PYP containing cis-*p*CA resulted in decreased biofilm formation compared to its *trans*-locked state[21]. Using this information, we conducted several laboratory studies to observe the effects of PYP on photoregulation of biofilm formation in *Massilia*.

Initial static growth tube culture assays were performed on all species under full light exposure to observe biofilm formation, *I. loihiensis* contained a biofilm but all other results were inconclusive after 72 hours of growth. This experiment was replicated with the addition of each inoculate containing a one culture sample completely exposed to light and one culture sample shielded for light exposure, and again the results were inconclusive.

The genome of each of these three bacteria was then analyzed using bioinformatics for the presence of genes encoding TAL and *p*CL. This analysis indicated that *I. loihiensis* contained the pCL component and a possible TAL component with altered substrate specificity associated with the production of *p*CA and its covalent attachment to apoPYP, see the supplemental material for further information regarding bioinformatic analysis. In contrast, the *M. albidiflava* and *M. plicata* species however did not contain clear homologs to either component. Based on these results, a next experiment was conducted comparing the *Massilia* samples inoculated in their normal media and media with a 0.5mM concentration of *p*CA. Each of the *Massilia* species were also compared for optimal conditions for oxygen availability by comparing samples in an aerobic condition with samples containing lids which were not tightened and anerobic

conditions with samples containing lids that were tightened. All samples were also tested in full white light illumination and no illumination conditions, see *Figures 21-23* for results. *I. loihiensis* was not tested for oxygen tolerance or using the addition of *p*CA, as the optimal conditions were followed using previously known conditions. Each of the *Massilia* species tested grew optimally in aerobic conditions.

Further analysis was done to quantify biofilm formation under various conditions using 96-well assays. The initial 96-well assay results were compared using standard deviations, yielding inconclusive results. Each set of conditions was compared against their respective dark inoculate as well as across the entire conditions, see *Figures 24-26* for results. These inconclusive results may be related to the observed border effect which occurs when the inoculates close to the border are exposed to more oxygen, more humidity variation, and is exposed to other environmental factors at an increased rate to that of inoculates in the center of the plate, this is accounted for in the second 96-wells assay by having all border wells filled with water. ~~There is also the possibility of inoculation errors in adding the correct sample sizes when inoculating.~~ The Results of the corrected optimal set up were compared using standard deviation as seen in *Figure 27*. This second 96-well assay was read using a plate reader with an optical density of 600nm. When looking at the *M. plicata* species with and without pCA there appears to be no variation leading to the understanding that pCA has no effect. Furthermore, there appears to be a light response which downregulated the production of biofilm formation, however further replicates need to be performed to conclude these results.

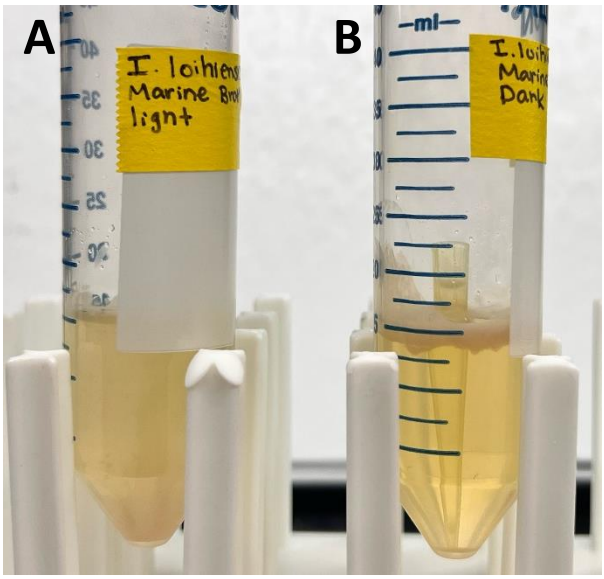


Figure 21 Static tube biofilm formation assay performed on *I. loihiensis* under aerobic conditions comparing the production of biofilm for both (A) full light illumination and (B) no illumination conditions. The cultures were allowed to grow for 72 hours and visually observed for results.

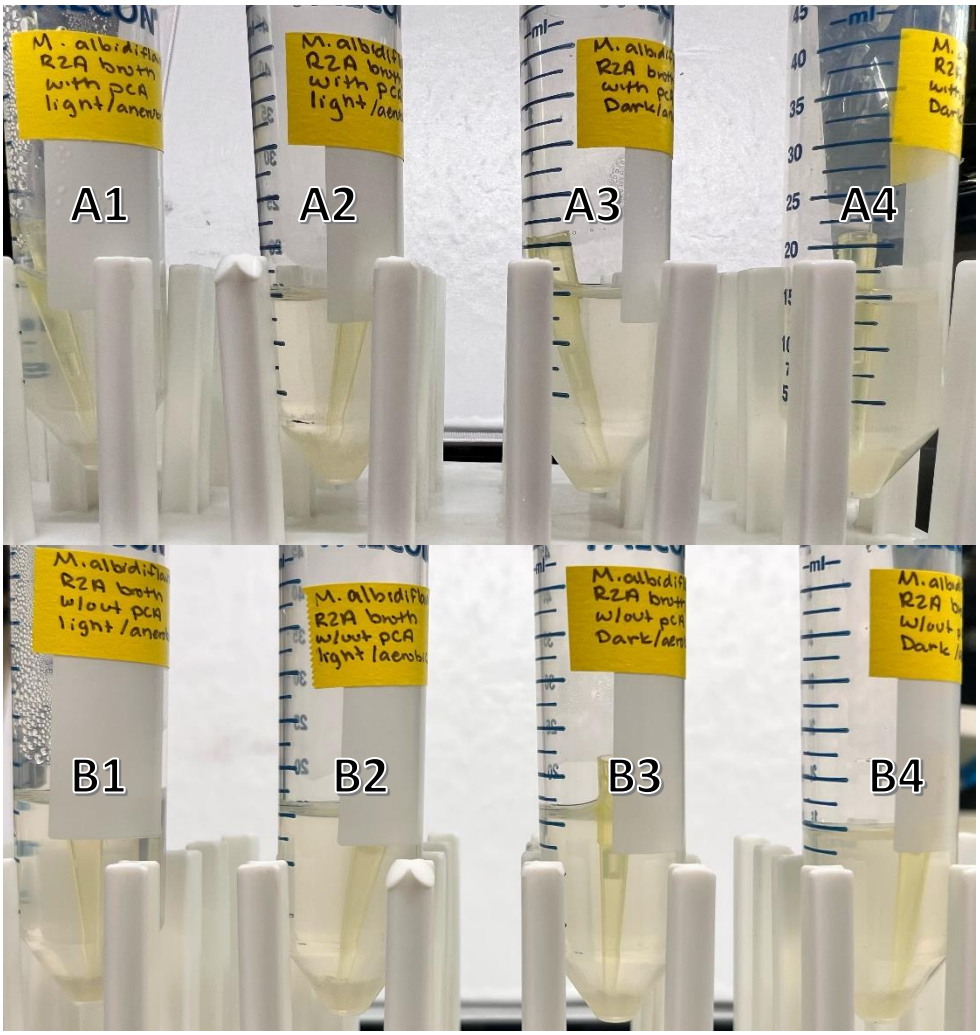


Figure 22 Static tube biofilm formation assay performed on *M. albidiflava.* All samples annotated (A) contain a 0.5mM concentration of pCA within the media. (A1) full light illumination under anerobic conditions, (A2) full light illumination under aerobic conditions, (A3) no illumination under anerobic conditions, and (A4) no illumination under aerobic conditions. All samples annotated (B) do not contain pCA, (B1) full light illumination under anerobic conditions, (B2) full light illumination under aerobic conditions, (B3) no illumination under anerobic conditions, and (B4) no illumination under aerobic conditions. The cultures were allowed to grow for 72 hours and visually observed for results.

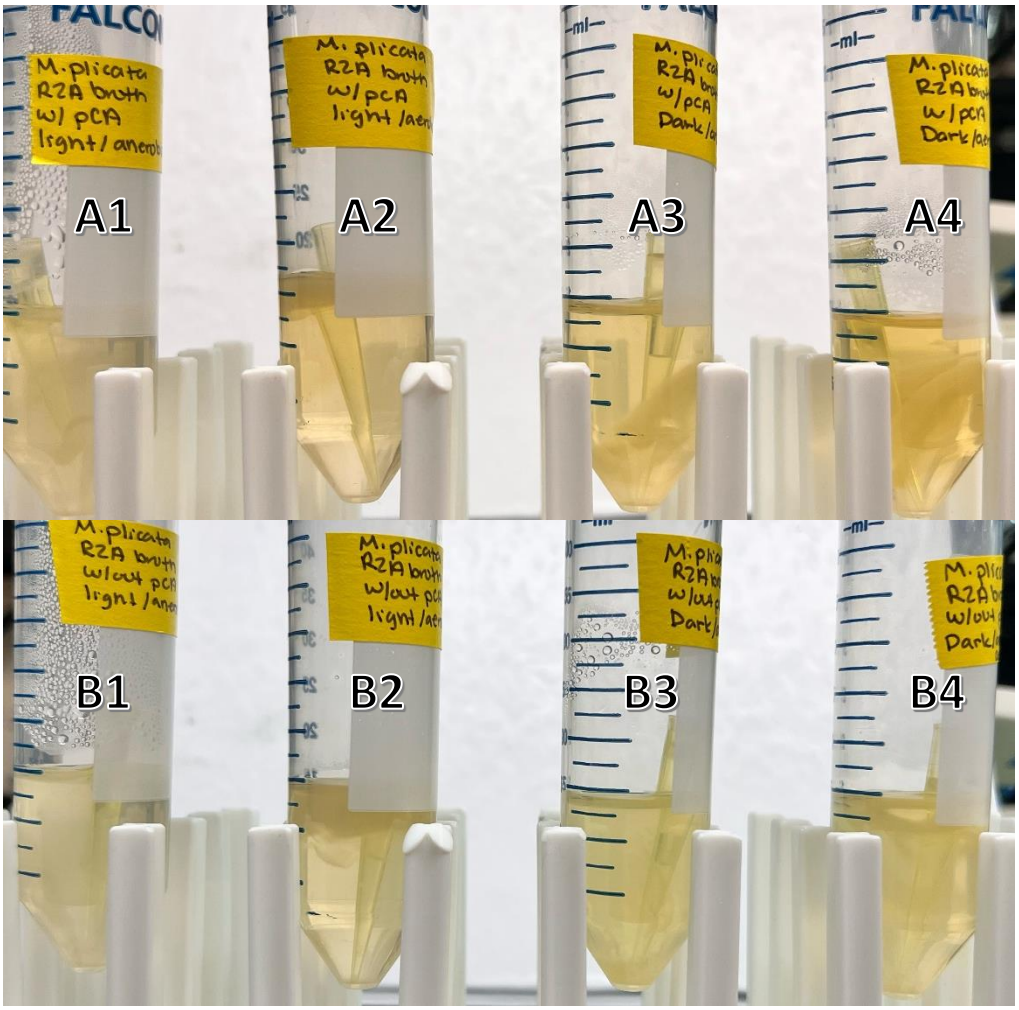


Figure 23 Static tube biofilm formation assay performed on *M. plicata.* All samples annotated (A) contain a 0.5mM concentration of pCA within the media. (A1) full light illumination under anerobic conditions, (A2) full light illumination under aerobic conditions, (A3) no illumination under anerobic conditions, and (A4) no illumination under aerobic conditions. All samples annotated (B) do not contain pCA, (B1) full light illumination under anerobic conditions, (B2) full light illumination under aerobic conditions, (B3) no illumination under anerobic conditions, and (B4) no illumination under aerobic conditions. The cultures were allowed to grow for 72 hours and visually observed for results.

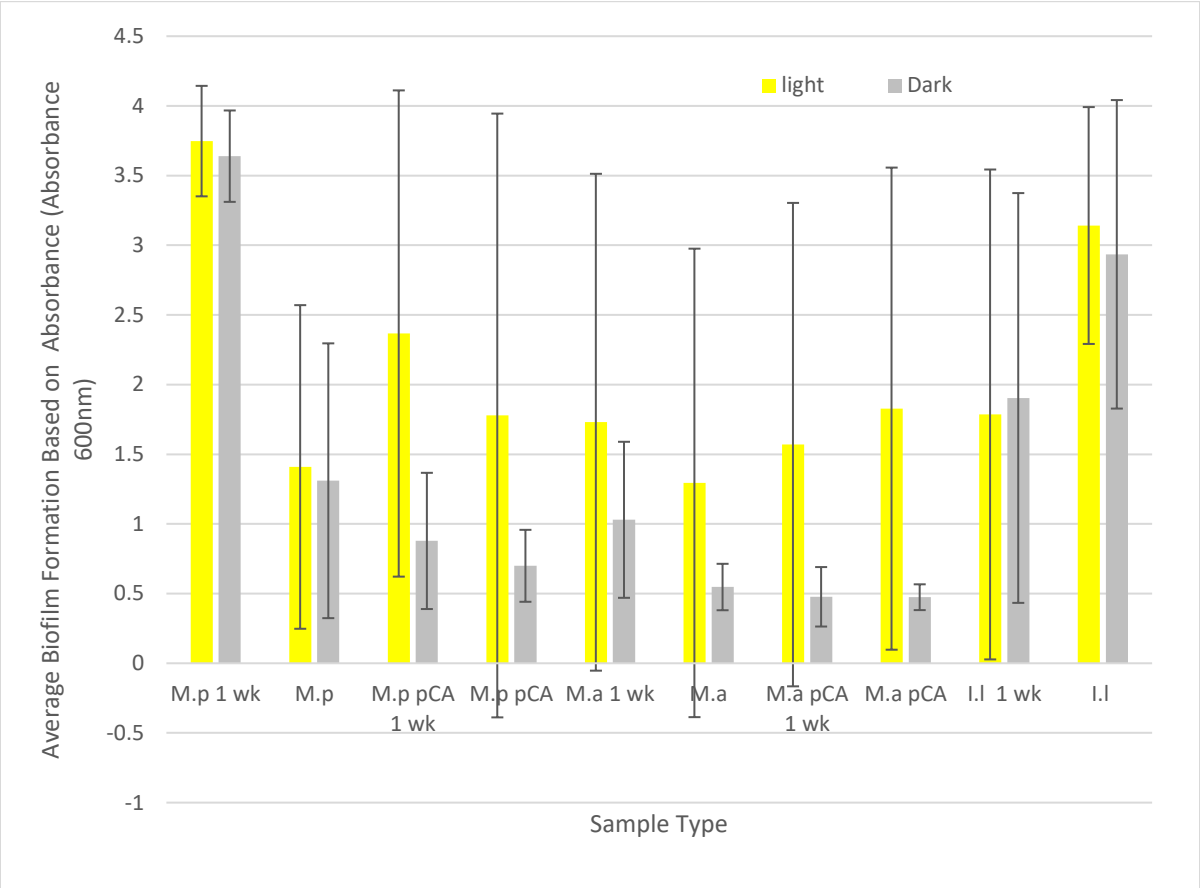


Figure 24 96-well assay comparing white light illumination and dark conditions using standard deviation. Absorbance was measured from 500nm-600nm, and each sample was diluted by a factor of 3. Each resultant absorbance value was then corrected to obtain the natural absorbance value.

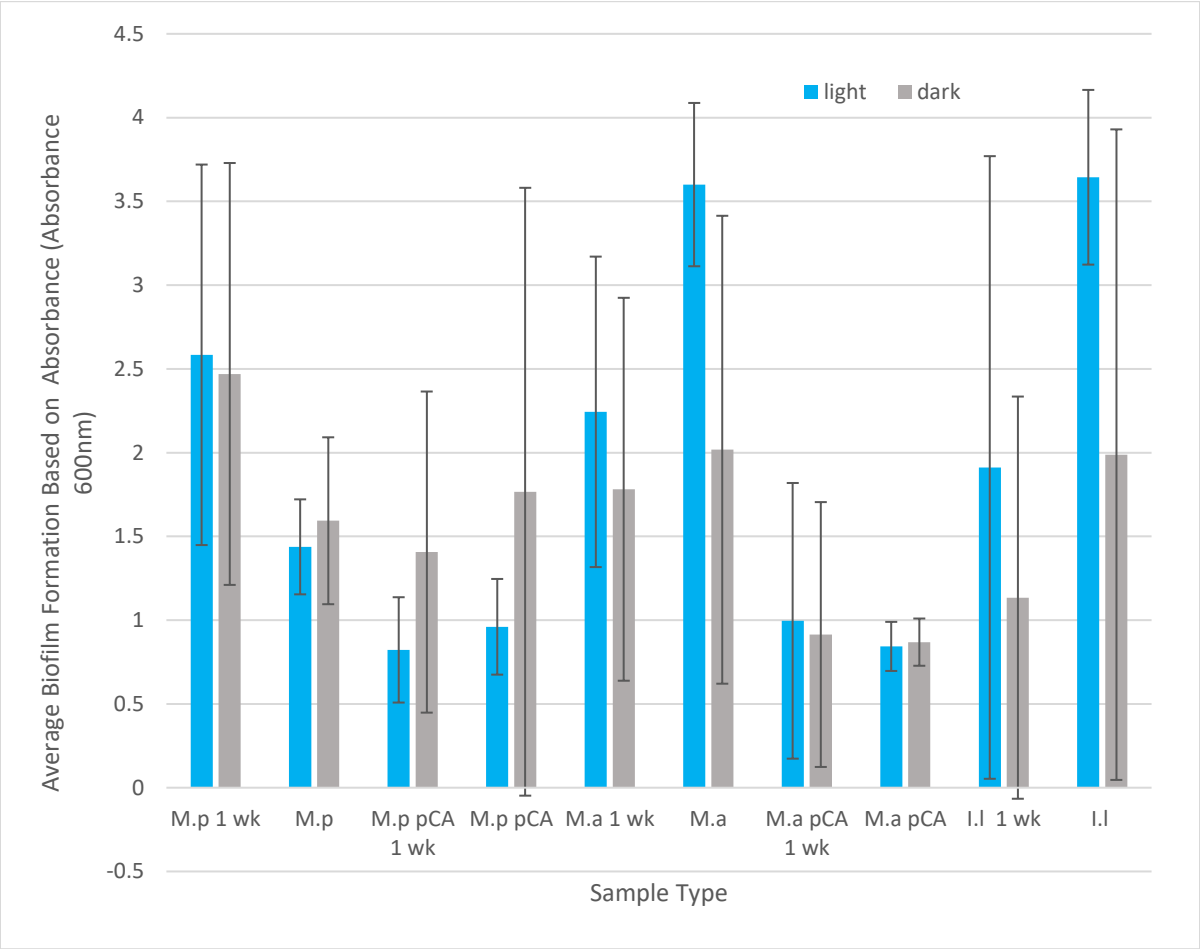


Figure 25 96-well assay comparing blue light illumination and dark conditions using standard deviation. Absorbance was measured from 500nm-600nm, and each sample was diluted by a factor of 3. Each resultant absorbance value was then corrected to obtain the natural absorbance value.

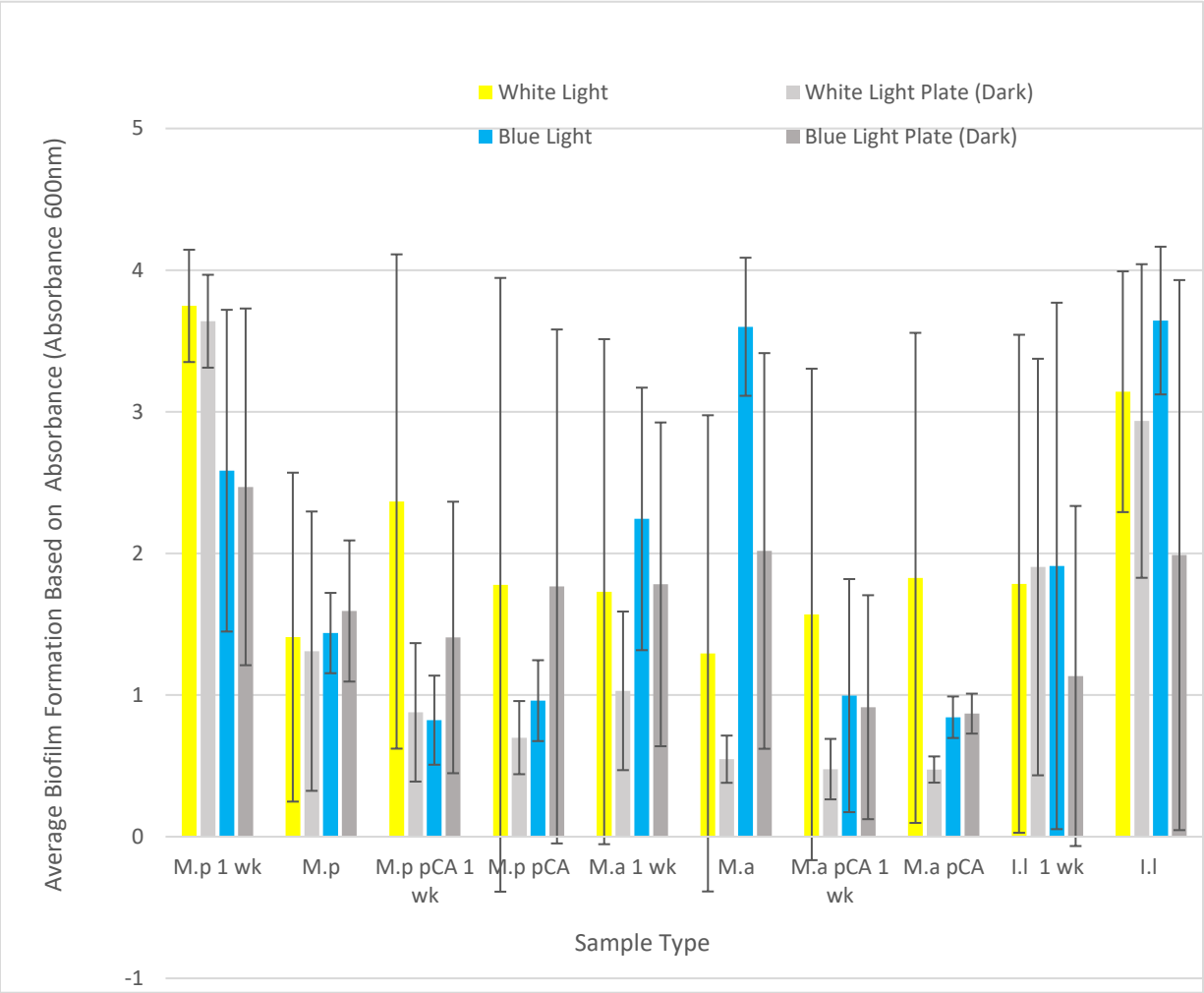


Figure 26 96-well assay comparing white light illumination, blue light illumination and dark conditions for each plate respectively using standard deviation. Absorbance was measured from 500nm-600nm, and each sample was diluted by a factor of 3. Each resultant absorbance value was then corrected to obtain the natural absorbance value.

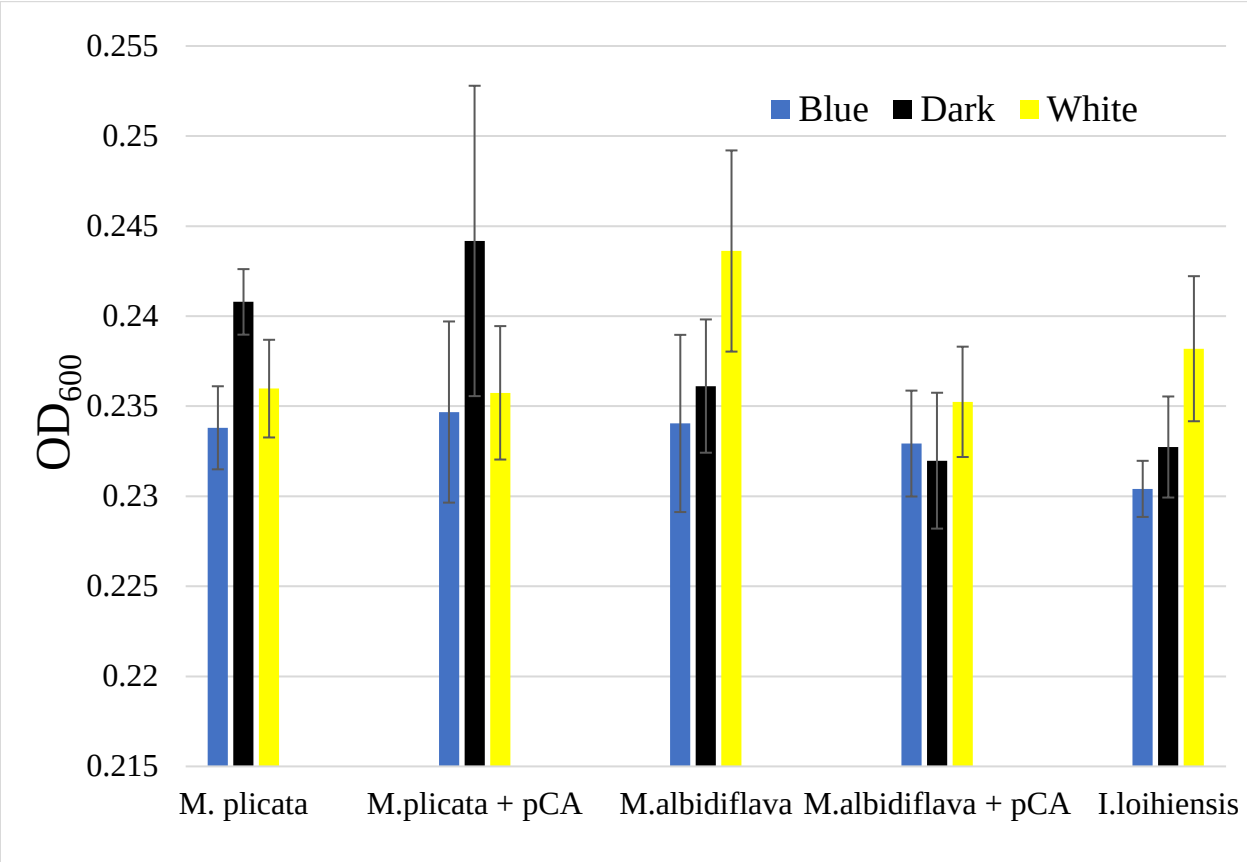


Figure 27 96-well plate assay performed with accounting for the border effect using sterile water in all border wells of the plate. Each plate was inoculated in the same manner and one of each plate was placed in either white light, blue light, or no light conditions. The plates were analyzed using a plate reader with optical density of 600nm.

4. Conclusion

Bioinformatics work yielded >1,100 PYP homologs which were then further inspected for recurring themes. There were 130 *pyp* operons inspected and 250 PYP multi-domain proteins inspected, this resulted in a recurring theme in both multiple *pyp* operons and multiple PYP multi-domain proteins with a genetic association with a GGDEF domain, implying blue light regulation of biofilm formation. Further four different genomes were identified for their MCP-

PYP fusion proteins, implying phototaxis response to blue light illuminations. We analyzed the *N. alkalilacustris sp* for phototaxis assays; and *M. plicata* and *M. albidiflava* for biofilm assays.

Motility assays were performed on *N. alkalilacustris* which yielded results suggesting that it undergoes swimming motility due to its increased motility upon 0.2% concentrated agar media, the presence of only a single flagellum, and the absence of surfactant production. Phototactic assays were performed which yielded inconclusive results until the addition of a 0.5mM solution of pCA was added as these organism's appeared to be lacking the TAL gene needed for the synthesis of the pCA chromophore of PYP, suggesting that this organism takes up pCA from the environment. Upon the addition of pCA *N. alkalilacustris* appears to have restored positive phototactic movement which needs further biological replicates to confirm these results.

Our biofilm results were inconclusive, however *M. plicata* may have reduced biofilm formation upon light illumination. There appears to be difficulty in identifying substrate specificity when performing bioinformatic analysis of the pCL and TAL genes, nonetheless bioinformatics currently suggests that these organisms may not be able to biosynthesize the PYP chromophore pCA. Again, further replicates are needed to further analyze biofilm formation and the signal transduction chain of PYP within these organisms.

5. Future Directions

To further confirm the results of this study additional testing and analysis needs to be performed. Further TEM imaging should be conducted taking care to retain the flagella by filtering cell samples before fixation. This will allow for better imaging as well as understanding of motility mechanisms. Further analysis of PCR and gel-electrophoresis for *M. albidiflava* and

M. plicata should occur to optimize the amplification and results of the product. This will help to identify the needed components within theses organisms. It may be of importance to select different primers to avoid primer dimers or annealing to other unintended locations. Additional biological replicates are needed for both phototaxis plate assays and biofilm formation 96-well plate assays to confirm our results and optimize the protocols including varying humidity, temperature and concentration. It is also of interest to add a *trans*-locked pCA for biofilm formation, as a control for comparable studies, if this blocks the blue light effect, it is likely a valid PYP photo-response. It may also be of importance to perform negative controls using a red-light illumination rather than a dark control to minimize the temperature differences caused by illumination.

References

1. Fraikin, G.Y., et al., *Bacterial photosensory proteins: Regulatory functions and optogenetic applications.* Microbiology, 2015. 84(4): p. 461-472.

2. van der Horst, M.A. and K.J. Hellingwerf, *Photoreceptor Proteins, "Star Actors of Modern Times": A Review of the Functional Dynamics in the Structure of Representative Members of Six Different Photoreceptor Families.* Accounts of Chemical Research, 2004. **37**(1): p. 13-20.

3. Spudich, J.L., et al., *Retinylidene Proteins: Structures and Functions from Archaea to Humans.* Annual review of cell and developmental biology., 2000. **16**(1): p. 365-392.

4. Hoff, W.D., K.-H. Jung, and J.L. Spudich, Molecular Mechanism of Photosignaling by Archaeal Sensory Rhodopsins. Annual review of biophysics and biomolecular structure., 1997. **26**(1): p. 223-258.

5. *The phytochrome family: dissection of functional roles and signalling pathways among family members. Philos. Trans. R. Soc. London B: Biol. Sci.* 1998.

6. *The xanthopsins: a new family of eubacterial blue-light photoreceptors EMBO J.* 1996.

7. Ahmad, M. and A.R. Cashmore, *HY4 gene of A. thaliana encodes a protein with characteristics of a blue-light photoreceptor.* Nature., 1993. **366**(6451): p. 162-166.

8. Huala, E., et al., *Arabidopsis NPH1: A Protein Kinase with a Putative Redox-Sensing Domain.* Science., 1997. **278**(5346): p. 2120-2123.

9. Gomelsky, M. and G. Klug, *BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms.* Trends in biochemical sciences., 2002. **27**(10): p. 497-500.

10. van der Horst, M.A., J. Key, and K.J. Hellingwerf, *Photosensing in chemotrophic, non-phototrophic bacteria: let there be light sensing too.* Trends Microbiol, 2007. **15**(12): p. 554-62.

11. Larsen, D.S., R. Van Grondelle, and K.J. Hellingwerf, *Primary Photochemistry in the Photoactive Yellow Protein: The Prototype Xanthopsin*. 2008, Springer Berlin Heidelberg. p. 165-199.

12. Meyer, T.E., et al., *Properties of a water-soluble, yellow protein isolated from a halophilic phototrophic bacterium that has photochemical activity analogous to sensory rhodopsin.* Biochemistry, 1987. **26**(2): p. 418-423.

13. Meyer, T.E., *Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium Ectothiorhodospira halophila.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1985. **806**(1): p. 175-183.

14. Pellequer, J.-L., et al., *Photoactive yellow protein: A structural prototype for the three-dimensional fold of the PAS domain superfamily.* Proceedings of the National Academy of Sciences, 1998. **95**(11): p. 5884-5890.

15. Cusanovich, M.A. and T.E. Meyer, *Photoactive Yellow Protein: A Prototypic PAS Domain Sensory Protein and Development of a Common Signaling Mechanism.* Biochemistry, 2003. **42**(17): p. 4759-4770.

16. Crosson, S. and K. Moffat, *Structure of a flavin-binding plant photoreceptor domain: Insights into light-mediated signal transduction.* Proceedings of the National Academy of Sciences, 2001. **98**(6): p. 2995-3000.

17. Xie, A., et al., *Formation of a New Buried Charge Drives a Large-Amplitude Protein Quake in Photoreceptor Activation.* Biochemistry, 2001. **40**(6): p. 1510-1517.

18. Xie, A., et al., *Glu46 Donates a Proton to the 4-Hydroxycinnamate Anion Chromophore During the Photocycle of Photoactive Yellow Protein.* Biochemistry, 1996. **35**(47): p. 14671-14678.

19. Sprenger, W.W., et al., *The eubacterium Ectothiorhodospira halophila is negatively phototactic, with a wavelength dependence that fits the absorption spectrum of the photoactive yellow protein.* Journal of Bacteriology, 1993. **175**(10): p. 3096-3104.

20. Jiang, Z., et al., *Bacterial Photoreceptor with Similarity to Photoactive Yellow Protein and Plant Phytochromes.* Science, 1999. **285**(5426): p. 406-409.

21. Horst, M.A.v.d., et al., *Locked Chromophore Analogs Reveal That Photoactive Yellow Protein Regulates Biofilm Formation in the Deep Sea Bacterium Idiomarina loihiensis.* Journal of the American Chemical Society, 2009. **131**(47): p. 17443-17451.

22. Armitage, J.P., *Tactic responses in photosynthetic bacteria.* Canadian Journal of Microbiology, 1988. **34**: p. 475-481.

23. Hazelbauer, G.L., *The bacterial chemosensory system.* Canadian Journal of Microbiology, 1988. **34**(4): p. 466-474.

24. Taylor, B.L. and D.E. Koshland, *Intrinsic and extrinsic light responses of Salmonella typhimurium and Escherichia coli.* Journal of bacteriology, 1975. **123**(2): p. 557-569.

25. Alam, M., et al., *Methyl-accepting taxis proteins in Halobacterium halobium.* The EMBO Journal, 1989. **8**(2): p. 631-639.

26. Spudich, E.N., C.A. Hasselbacher, and J.L. Spudich, *Methyl-accepting protein associated with bacterial sensory rhodopsin I.* Journal of Bacteriology, 1988. **170**(9): p. 4280-4285.

27. Bivin, D.B. and W. Stoeckenius, *Photoactive Retinal Pigments in Haloalkaliphilic Bacteria.* Microbiology, 1986. **132**(8): p. 2167-2177.

28. Hustede, E., M. Liebergesell, and H.G. Schlegel, *The Photophobic Response of Various Sulfur and Nonsulfur Purple Bacteria.* Photochemistry and Photobiology, 1989. **50**.

29. Gomelsky, M. and W.D. Hoff, *Light helps bacteria make important lifestyle decisions.* Trends Microbiol, 2011. **19**(9): p. 441-8.

30. van der Horst, M.A., J. Key, and K.J. Hellingwerf, *Photosensing in chemotrophic, non-phototrophic bacteria: let there be light sensing too.* Trends in Microbiology, 2007. **15**(12): p. 554-562.

31. Hou, S., et al., *Genome sequence of the deep-sea γ-proteobacterium Idiomarina loihiensis reveals amino acid fermentation as a source of carbon and energy.* Proceedings of the National Academy of Sciences, 2004. **101**(52): p. 18036-18041.

32. Hickman, J.W., D.F. Tifrea, and C.S. Harwood, *A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels.* Proceedings of the National Academy of Sciences, 2005. **102**(40): p. 14422-14427.

33. Paul, R., et al., *Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain.* Genes & Development, 2004. **18**(6): p. 715-727.

34. Ryjenkov, D.A., et al., *Cyclic Diguanylate Is a Ubiquitous Signaling Molecule in Bacteria: Insights into Biochemistry of the GGDEF Protein Domain.* Journal of Bacteriology, 2005. **187**(5): p. 1792-1798.

35. Cotter, P.A. and S. Stibitz, *c-di-GMP-mediated regulation of virulence and biofilm formation.* Current Opinion in Microbiology, 2007. **10**(1): p. 17-23.

36. Koblitz, J., et al., *MediaDive: the expert-curated cultivation media database.* Nucleic Acids Research, 2022. **51**(D1): p. D1531-D1538.

37. Reimer, L.C., et al., *BacDive in 2022: the knowledge base for standardized bacterial and archaeal data.* Nucleic Acids Research, 2021. **50**(D1): p. D741-D746.

38. Borsodi, A.K., et al., *Nitrincola alkalilacustris sp. nov. and Nitrincola schmidtii sp. nov., alkaliphilic bacteria isolated from soda pans, and emended description of the genus Nitrincola.* International Journal of Systematic and Evolutionary Microbiology, 2017. **67**(12): p. 5159-5164.

39. Parte, A.C., et al., *List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ.* International Journal of Systematic and Evolutionary Microbiology, 2020. **70**(11): p. 5607-5612.

40. Verslyppe, B., et al., *StrainInfo introduces electronic passports for microorganisms.* Systematic and Applied Microbiology, 2014. **37**(1): p. 42-50.

41. Patrick, J.E. and D.B. Kearns, *Swarming motility and the control of master regulators of flagellar biosynthesis.* Molecular Microbiology, 2012. **83**(1): p. 14-23.

42. Kearns, D.B., *A field guide to bacterial swarming motility.* Nature Reviews Microbiology, 2010. **8**(9): p. 634-644.

43. Palma, V., et al., *Methods to Evaluate Bacterial Motility and Its Role in Bacterial–Host Interactions.* Microorganisms, 2022. **10**(3): p. 563.

44. Partridge, J.D. and R.M. Harshey, *Investigating Flagella-Driven Motility in Escherichia coli by Applying Three Established Techniques in a Series.* Journal of Visualized Experiments, 2020(159).

45. Partridge, J.D. and R.M. Harshey, *Swarming: Flexible Roaming Plans.* Journal of Bacteriology, 2013. **195**(5): p. 909-918.

46. Turnbull, L.W., C.B, *Motility Assay: Twitching Motility.* Pseudomonas Methods and Protocols, 2014: p. 816.

47. Martínez, A., S. Torello, and R. Kolter, *Sliding Motility in Mycobacteria.* Journal of Bacteriology, 1999. **181**(23): p. 7331-7338.

48. Liu, Y., S. Kyle, and P.D. Straight, *Antibiotic Stimulation of a Bacillus subtilis Migratory Response.* mSphere, 2018. **3**(1): p. e00586-17.

49. Tchoufag, J., et al., *Mechanisms for bacterial gliding motility on soft substrates.* Proceedings of the National Academy of Sciences, 2019. **116**(50): p. 25087-25096.

50. Xing, J., V.M. Gumerov, and I.B. Zhulin, *Photoactive Yellow Protein Represents a Distinct, Evolutionarily Novel Family of PAS Domains.* Journal of Bacteriology, 2022. **204**(11): p. e00300-22.

51. Van Beeumen, J.J., et al., *Primary structure of a photoactive yellow protein from the phototrophic bacterium Ectothiorhodospira halophila, with evidence for the mass and the binding site of the chromophore.* Protein Science, 1993. **2**(7): p. 1114-1125.

52. Hoff, W.D., et al., *Thiol ester-linked p-coumaric acid as a new photoactive prosthetic group in a protein with rhodopsin-like photochemistry.* Biochemistry, 1994. **33**(47): p. 13959-13962.

53. Möglich, A., R.A. Ayers, and K. Moffat, *Structure and Signaling Mechanism of Per-ARNT-Sim Domains.* Structure, 2009. **17**(10): p. 1282-1294.

54. Merritt, J.H., D.E. Kadouri, and G.A. O'Toole, *Growing and analyzing static biofilms.* Curr Protoc Microbiol, 2005. **Chapter 1**: p. Unit 1B.1.

55. Thibeaux, R., M. Kainiu, and C. Goarant, *Biofilm Formation and Quantification Using the 96-Microtiter Plate*. 2020, Springer US. p. 207-214.

56. Agematu, H., K. Suzuki, and H. Tsuya, *Massilia sp. BS-1, a Novel Violacein-Producing Bacterium Isolated from Soil.* Bioscience, Biotechnology, and Biochemistry, 2011. **75**(10): p. 2008-2010.

Abbreviations

Abbreviation	Meaning
°C	degrees Celsius
µL	microliter
µM	micromolar
1d	1 day
1wk	1 week
aa	amino acid
apoPYP	inactive unbound state of PYP
ATP	adenosine triphosphate
BLASTP	protein blast
bp	base pairs
c-di-GMP	diguanylate cyclase
CheA	chemotaxis protein A
CheW	chemotaxis protein W
CheY	chemotaxis protein Y
CoA	coenzyme A
Cys69	Cystine 69
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
g	grams
GC	guanine-cytosine content
gDNA	genomic DNA
Glu46	Glucose 46
H. halophila=Hhal	*Halorhodospira halophila*
HK	histidine kinase
hr	hour

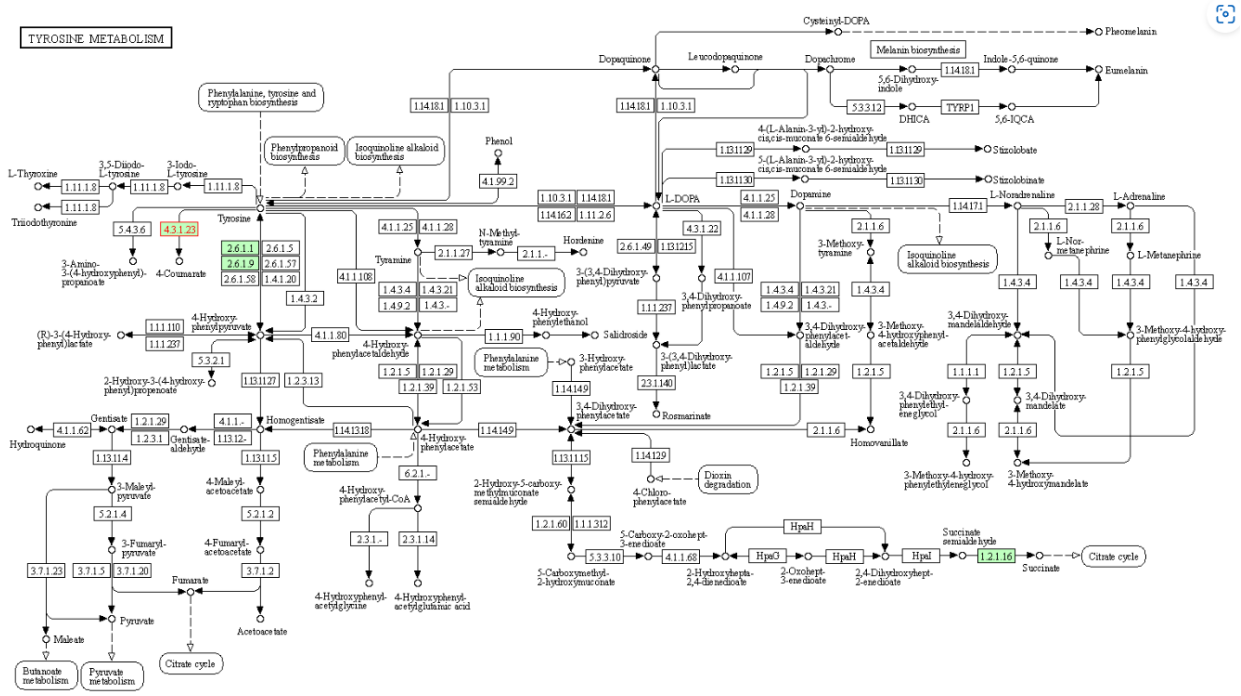
Abbreviation	Meaning
I. loihiensis= I.l=I.I	*Idiomarina loihiensis*
LOV	light, oxygen or voltage
M. albidiflava=M.p=M.P	*Massilia albidiflava*
M. plicata=M.p=M.P	*Massilia plicata*
MCP	methyl-accepting chemotaxis protein
mL	milliliter
mM	millimolar
MSA	multiple sequence alignment
N. alkalilacustris=N.a=N.A	*Nitrincola alkalilacustris species*
ng	nanogram
nm	nanometer
PAS	Per-Arnt-Sim
pCA	p-coumaric acid
pCL	p-coumaroyl CoA ligase
PCR	polymerase chain reaction
PHY	phytochrome
pol	polymerase
PVC	polyvinyl chloride
PYP	photoactive yellow protein
s	seconds
SRI(I)	sensory rhodopsin I(I)
TAL	tyrosine ammonia lyase
TEM	transmission electron microscopy
Tm	melting temperature
V	volts

Supplemental Material

Bioinformatics- TAL gene comparison

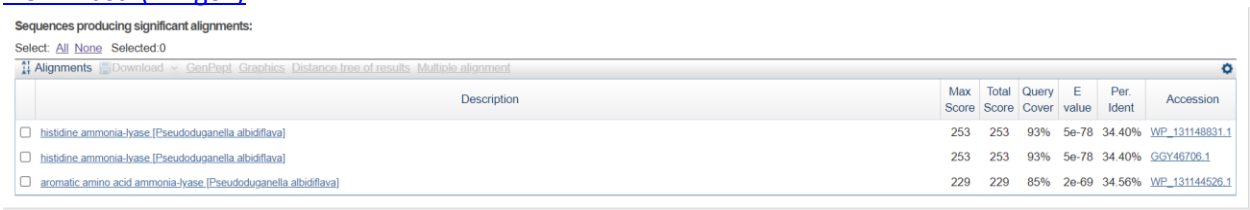
Halorhodospira halophlia sequences used for comparison

Hhal_1818 H.halophlia PYP in KEGG
TAL: Hhal_1820
MAEVDLAGSLSAADIEAIGYGHRTATVSPTGWKRLRSAEAYLQRLVDERRQVYGVTTGYG
PLATSRIDPSASRTLQRNLVYHLCSGVGEPLSRCHTRATLGARIASVTRGHSGVTPAVVE
RLLAWLEHDVVPEVPAIGTVGASGDLTPLAHVARALMGEGRVCINGGEWEPADAAQRRLG
WEPWTLDGKDAIALVNGTSTTAGICAVNGAGAERAAGVCAVLGMVYAELLGGHAEAFQPA
IGAVRPHPGQMRAHAWLTALAEDSQRLQPWTGTPPRLTEGQEAVLPDQPLPQDPYSIRCL
PQALGAVLDSITFHNQTVASELDAASDNPLLFPDEGRVLHGGNFFGQHLAFAADALNNAV
VQLALHSERRISRITDSTRSGFPAFMQPRQTGLHSGFMGAQVTASALVAEMRTGAHPASI
QSIPTNADNQDIVPMSTRAARQAATNLDHLQRILAIEALVLAQGLELADGVGFSSSARRT
LGWVRELAPPLEDDRPLAEEIARVAAALATPYQAHRLVAGLPGAPPGPAS



NCBI Protein blast for M. albidiflava DSM17472 -TAL gene

NCBI Blast: (nih.gov)



Sequences producing significant alignments:

Select: All None Selected:0

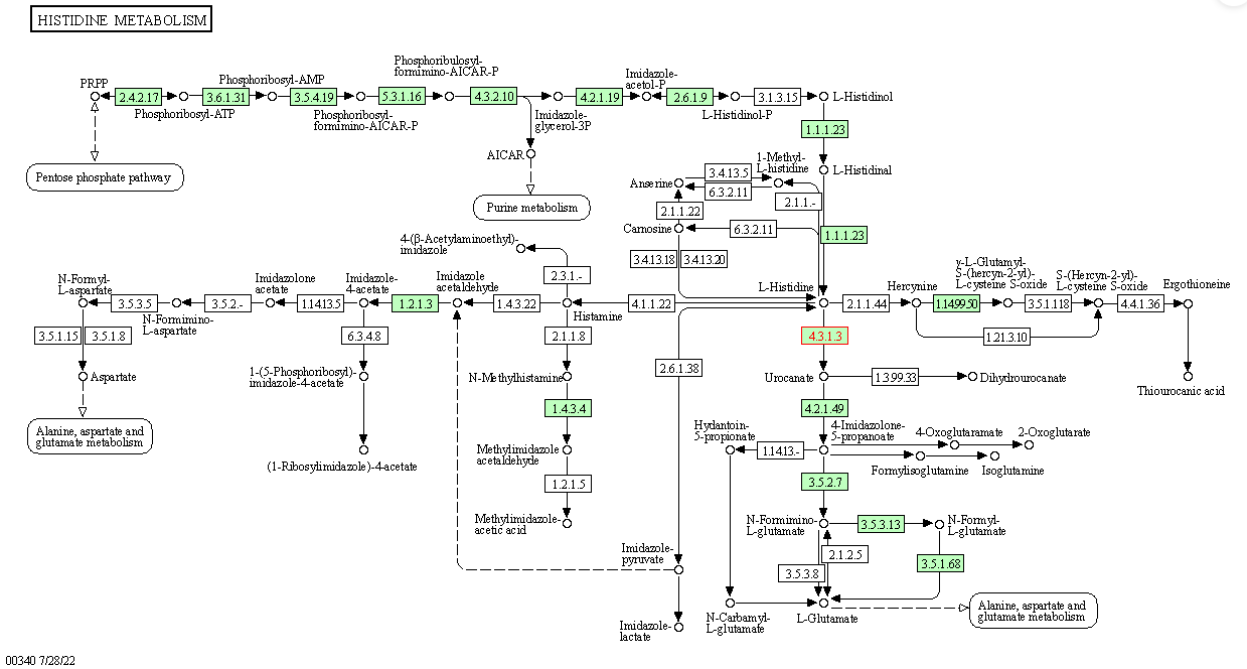
Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
histidine ammonia-lyase [Pseudoduganella albidiflava]	253	253	93%	5e-78	34.40%	WP_131148831.1
histidine ammonia-lyase [Pseudoduganella albidiflava]	253	253	93%	5e-78	34.40%	GGY46706.1
aromatic amino acid ammonia-lyase [Pseudoduganella albidiflava]	229	229	85%	2e-69	34.56%	WP_131144526.1

Each following hit is then analyzed within the KEGG database by coping the FASTA sequence into the KEGG BLAST below are the populated results

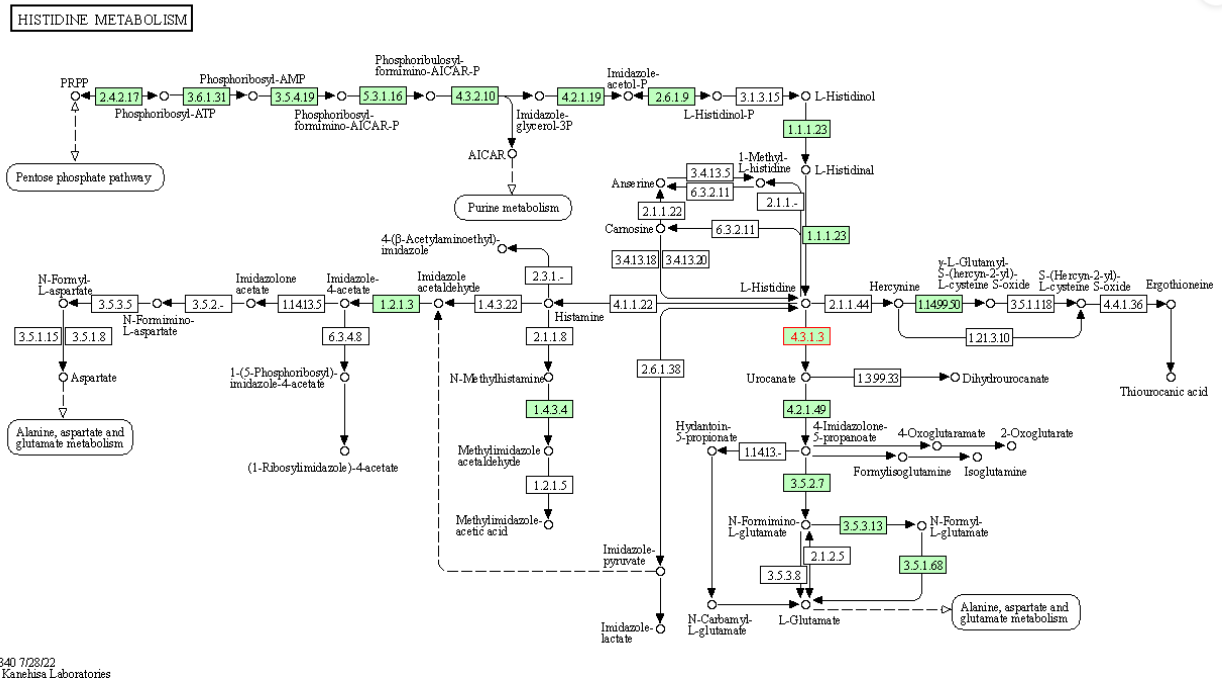
1.

>WP_131148831.1 histidine ammonia-lyase [Pseudoduganella albidiflava]
MTPTSTSTSTPISTSSLILIPGRLALADLRSAWMAHGPIALAPEAYPQIDAAARAVHAIVQKGDAAYGIN
TGFGLLAKTRIPDDQLEQLQRNLILSHSVGTGELMADHVVRLLMLTKIGSLARGFSGIRASVIDTLVALY
NAGIMPAIPVKGSVGASGDLAPLAHMTLALLGVGPVRVKGELMEAHQALAAAGIAPVVLAPKEGLALING
TQASTALALHGLFMAERLLEAGIVTGSLSLDAAKGSDAPFDARVHEVRGQPGQIAAAAIYRQLVAGSAIR
ASHLLGDERVQDPYCLRCQPQVMGACMDLIGNATRTLLIEANAVTDNPLLFREGGDTGDTVTIVSGGNFH
AEPVAFAADTLALAIAEIGSIAERRIALLIDATLSGLPPFLVREPGVNSGFMIAHVTAAALASENKSLAH
PASVDSLPTSANQEDHVSMATFAARRLDEMAHNTAAIVGIELLAAAQGIDFHRPLKTSPHLEHVHAQLRR
QVPFYDADRYFAPDIEAAKRLVLEGALSADCRQLFTPLHP



2.

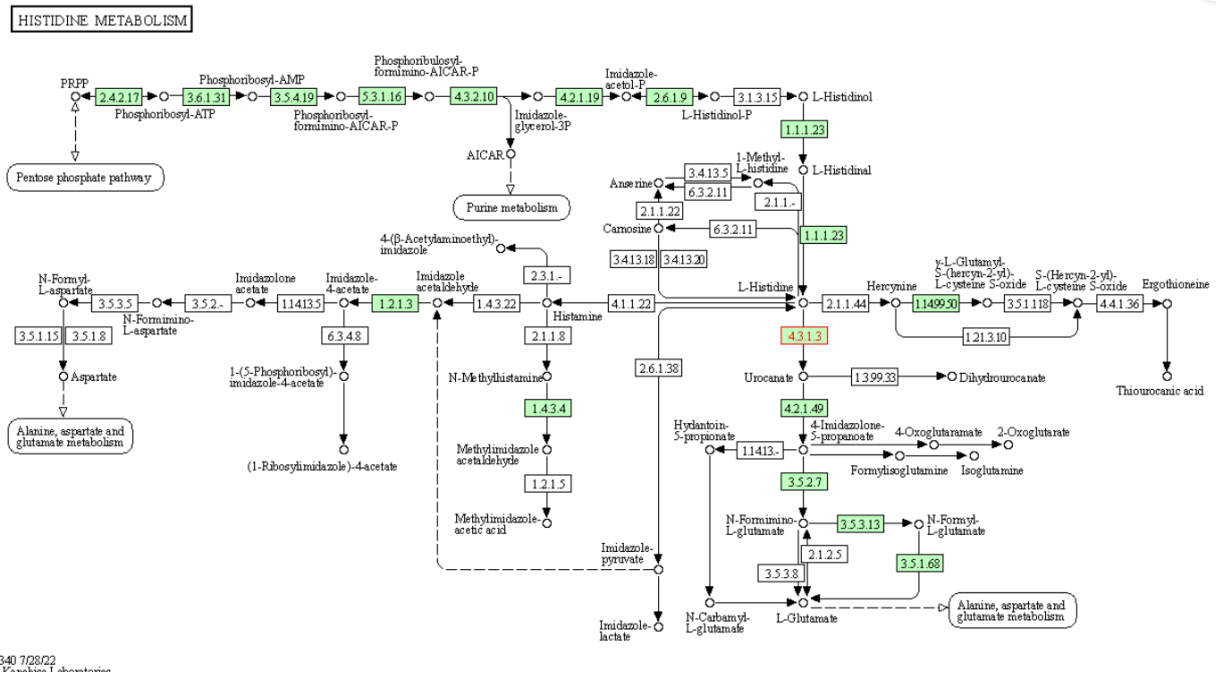
>GGY46706.1 histidine ammonia-lyase [Pseudoduganella albidiflava]
MTVEKMTPTSTSTSTPISTSSLILIPGRLALADLRSAWMAHGPIALAPEAYPQIDAAARAVHAIVQKGDA
AYGINTGFGLLAKTRIPDDQLEQLQRNLILSHSVGTGELMADHVVRLLMLTKIGSLARGFSGIRASVIDT
LVALYNAGIMPAIPVKGSVGASGDLAPLAHMTLALLGVGPVRVKGELMEAHQALAAAGIAPVVLAPKEGL
ALINGTQASTALALHGLFMAERLLEAGIVTGSLSLDAAKGSDAPFDARVHEVRGQPGQIAAAAIYRQLVA
GSAIRASHLLGDERVQDPYCLRCQPQVMGACMDLIGNATRTLLIEANAVTDNPLLFREGGDTGDTVTIVS
GGNFHAEPVAFAADTLALAIAEIGSIAERRIALLIDATLSGLPPFLVREPGVNSGFMIAHVTAAALASEN
KSLAHPASVDSLPTSANQEDHVSMATFAARRLDEMAHNTAAIVGIELLAAAQGIDFHRPLKTSPHLEHVH
AQLRRQVPFYDADRYFAPDIEAAKRLVLEGALSADCRQLFTPLHP



3.

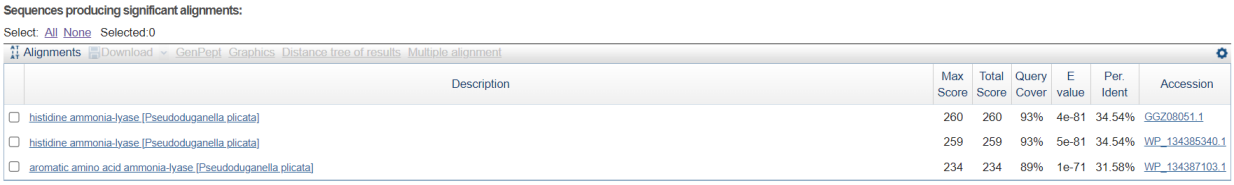
>WP_131144526.1 aromatic amino acid ammonia-lyase [Pseudoduganella albidiflava]

MQLAEHFAVLHAAPRTVRFDRDRLTIEDIVDIAQGRANAELSPDPAFRAAIARGADFLDRLLREDGNVYG
VTTGYGDSCTVAVPPELVAELPHHLYTYHGCGLGEFFDPAQTRAILAARLASLAKGYSGVSVELLLQITR
LLDAGLLPQIPSEGSVGASGDLTPLSYLAAVMCGEREVWRNGQVVCAAVALREAGITPLRLRPKEGLAIM
NGTAVMTALACLAYDRADYLVKLATRITALASYTLDGNAHHFDEVLFAVKPHPGMGRVAAMLREDLPTTV
WERNGKRLQDRYSIRCAPHVIGVLADAMPLFRSNIENELNSANDNPLIDGDNERVLHGGHFYGGHIAFAM
DGMKNAVANVADLLDRQMALLVDARYNNGLPANLSGSTGPRAAINHGLKALQISASAWTAEALKLTMPAS
VFSRSTECHNQDKVSMGTIAARDCLRVLELTEQVVAALLITVRQGAWLRGRASPELALQPHLAAMVEALD
IAPIDEDRRLDSDLHSLLASIRAKAWELYA



NCBI Protein blast for M. plicata DSM17505- TAL gene

NCBI Blast:TAL protein blast for M plicata (nih.gov)



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

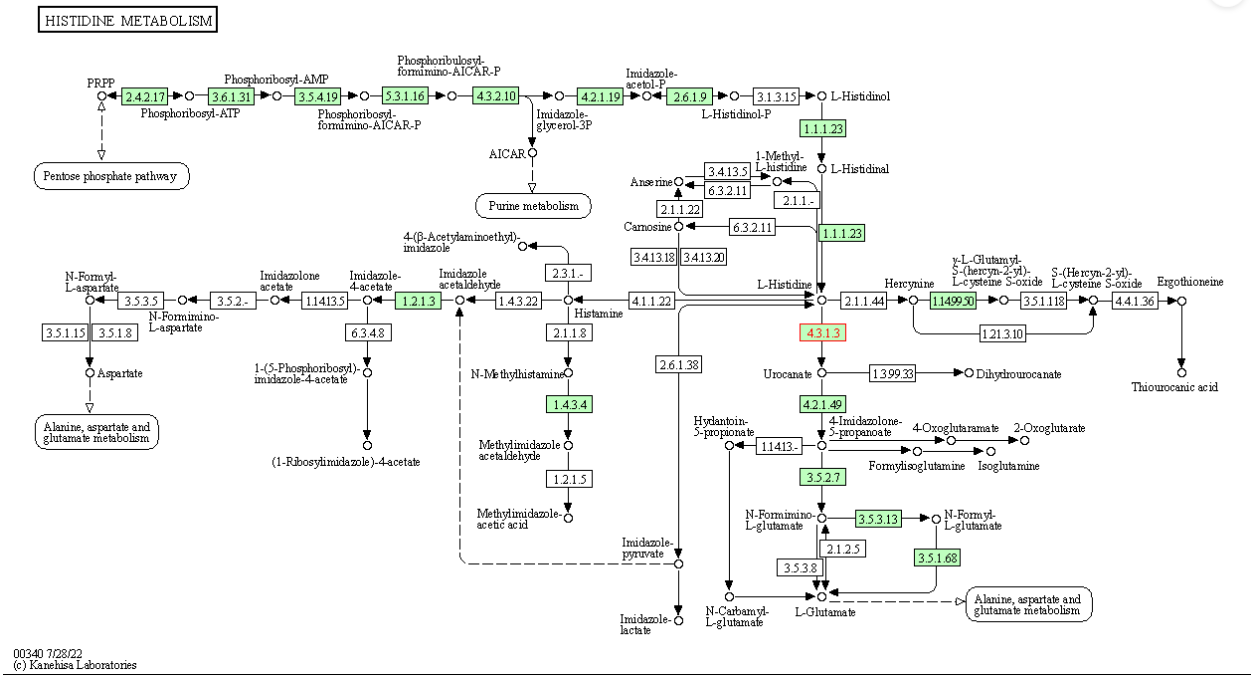
Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
histidine ammonia-lyase [Pseudoduganella plicata]	260	260	93%	4e-81	34.54%	GGZ08051.1
histidine ammonia-lyase [Pseudoduganella plicata]	259	259	93%	5e-81	34.54%	WP_134385340.1
aromatic amino acid ammonia-lyase [Pseudoduganella plicata]	234	234	89%	1e-71	31.58%	WP_134387103.1

Each following hit is then analyzed within the KEGG database by coping the FASTA sequence into the KEGG BLAST below are the populated results

1.

>GGZ08051.1 histidine ammonia-lyase [Pseudoduganella plicata]
MKMTQPNHTLTLNPGKLTLAELRAAWTAHGPIALAAEAYPVIDASAQAVQAIVQKGEAAYGINTGFGLLA
KTRIPDDKLEQLQRNLILSHSVGTGDLMPDHVVRLLMLMKIGSLARGYSGIRAVVIDTLIALYNAGIMPA
IPVKGSVGASGDLAPLSHMTLAILGVGEVRVKGELMQAADALTSAGIAPVVLAAKEGLALINGTQASAAL
ALHGLFMAERLLEAGMVTGSLSLDAAKGSDSPFDARVHEVRGQPGQIAAAAIYRQLVTGSAIRASHLVGD
ERVQDPYCLRCQPQVMGACMDLIANVTRTLLIEANAVTDNPLLFRDGDEVNIVSGGNFHAEPVAFAADTL
ALAIAEIGSIAERRIALLIDATLSGLPPFLVREPGVNSGFMIAHVTAAALASENKSLAHPASVDTIPTSA
NQEDHVSMATFAARRLDEMAHNTAAIIGIELLAAAQGVDFHRPLRTSPHLEHVHAQLRQKVPFFDADRYF

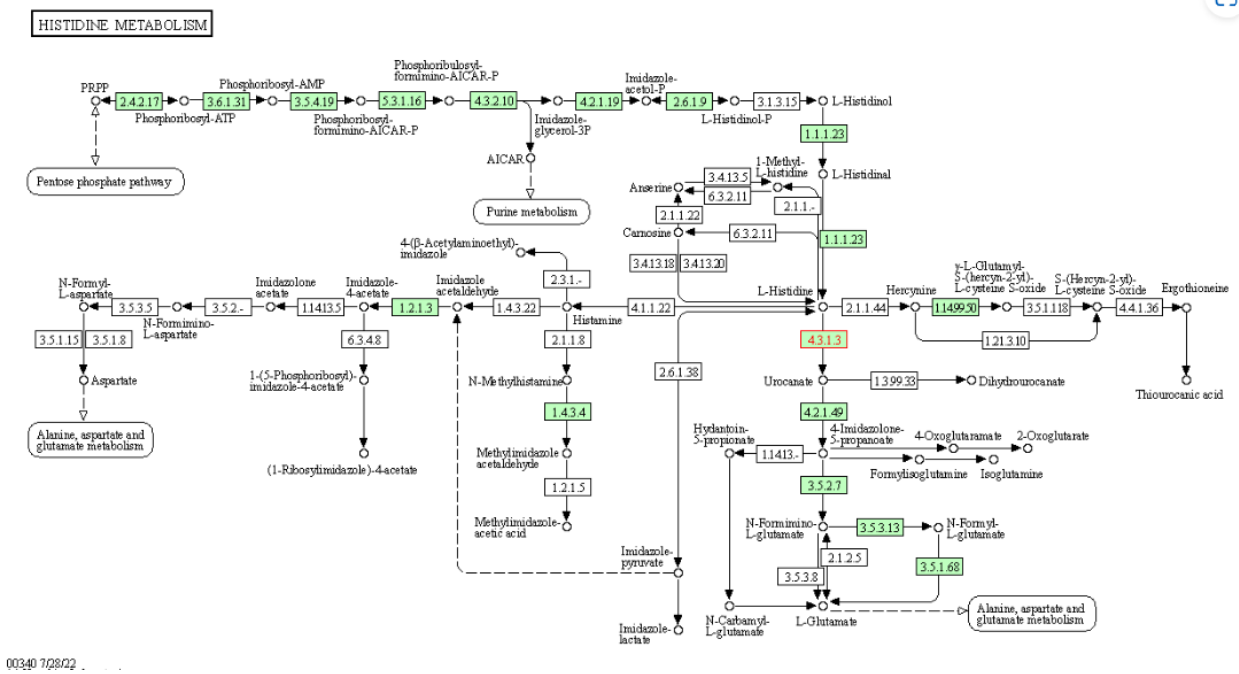
APDIEAAKQMVLKGELSATCRNLFTPLHP



2.

>WP_134385340.1 histidine ammonia-lyase [Pseudoduganella plicata]
MTQPNHTLTLNPGKLTLAELRAAWTAHGPIALAAEAYPVIDASAQAVQAIVQKGEAAYGINTGFGLLAKT
RIPDDKLEQLQRNLILSHSVGTGDLMPDHVVRLLMLMKIGSLARGYSGIRAVVIDTLIALYNAGIMPAIP
VKGSVGASGDLAPLSHMTLAILGVGEVRVKGELMQAADALTSAGIAPVVLAAKEGLALINGTQASAALAL
HGLFMAERLLEAGMVTGSLSLDAAKGSDSPFDARVHEVRGQPGQIAAAAIYRQLVTGSAIRASHLVGDER
VQDPYCLRCQPQVMGACMDLIANVTRTLLIEANAVTDNPLLFRDGDEVNIVSGGNFHAEPVAFAADTLAL
AIAEIGSIAERRIALLIDATLSGLPPFLVREPGVNSGFMIAHVTAAALASENKSLAHPASVDTIPTSANQ
EDHVSMATFAARRLDEMAHNTAAIIGIELLAAAQGVDFHRPLRTSPHLEHVHAQLRQKVPFFDADRYFAP

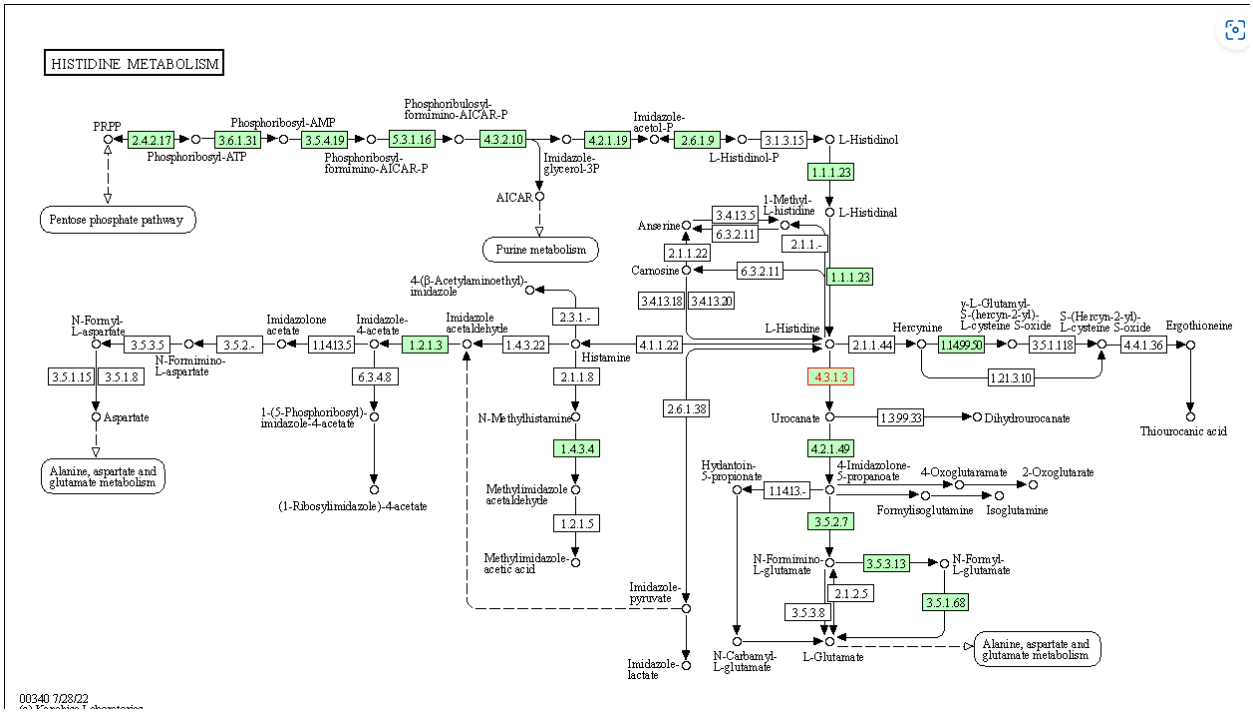
DIEAAKQMVLKGELSATCRNLFTPLHP



3.

>WP_134387103.1 aromatic amino acid ammonia-lyase [Pseudoduganella plicata]
MKRTLTISQQRLTIEDIVDLSQQRVHVALSTDAAFQQRIKKGADFLDRLLETEGEIYGVTTGYGDSVTRQ
VSRELIEQLPVNLLRFHGCGLGEHFDLETSRAIMAARLVSLSRGYSGVTPALLQQLASMLEHDIIPLIPQ
EGSVGASGDLTPLSYVAAAMIGEREVRYRDQVRPSAEVLAECGLTPLRLRPKEGLAMMNGTAVMTALACL
AFKRAEYLSRLCTRTTALAAIALKSNLHHFDEDLFKAKPHPGLQRIAALIRADLMPQHDPAERLQDRYSI
RCAPHVIGVLEDSMAHFRSVIENELNSANDNPLLDPEKEKLLHGGHFYGGHIAHVMDTLKTQIANLADLM
DRQLALLVDTRYNNGMTANLSLPDPSRASINHGFKAVQIGVSAWTAEALQNCMPASVFSRSTECHNQDKV

SMGTIAARSCIRVIELVEQVQAAMVLAATQAIAVRRRVEPGASLPIGLDAHGFIETVEAEFALVGEDRPL
EADLRRFIAYIREQRWTFPA



NCBI Protein blast for N. alkalilacustris DSM29817 – TAL gene

NCBI Blast:TAL blast for N alkalilacustris (nih.gov)

Sequences producing significant alignments:

Select: All None Selected:0

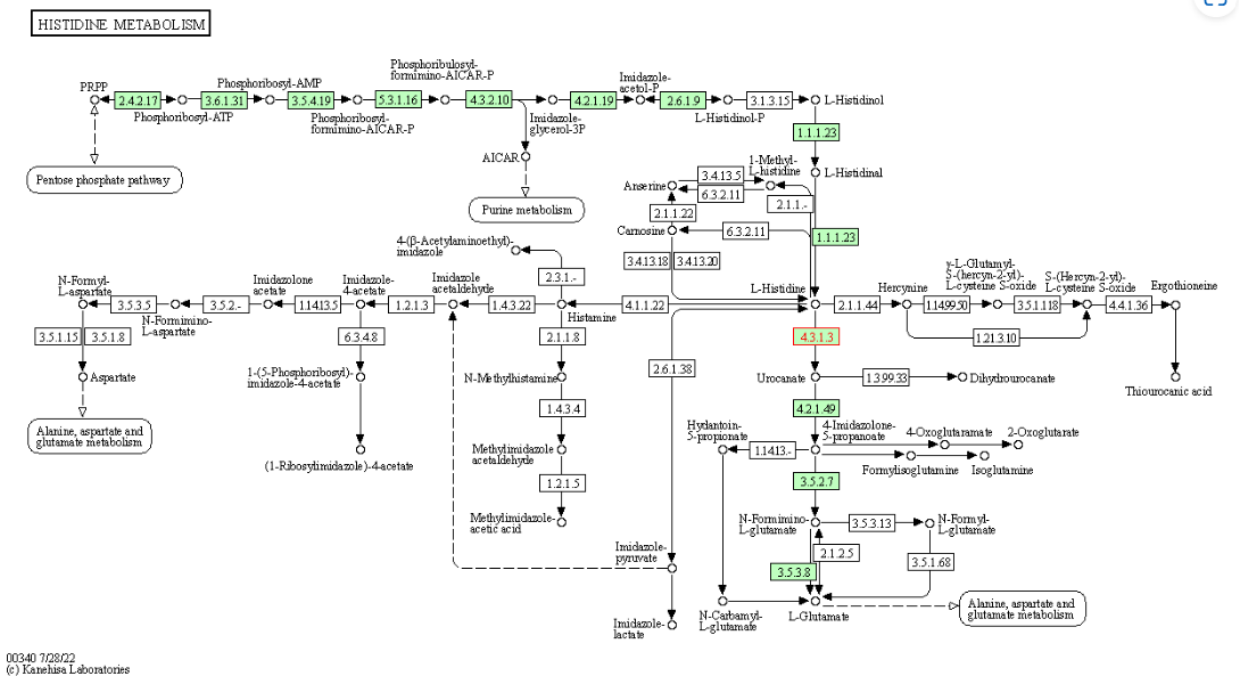
Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
histidine ammonia-lyase [Nitrincola alkalilacustris]	257	257	86%	2e-80	34.78%	WP_151704889.1
aromatic amino acid ammonia-lyase [Nitrincola alkalilacustris]	227	227	94%	1e-68	31.32%	WP_240776273.1

Each following hit is then analyzed within the KEGG database by coping the FASTA sequence into the KEGG BLAST below are the populated results

1.

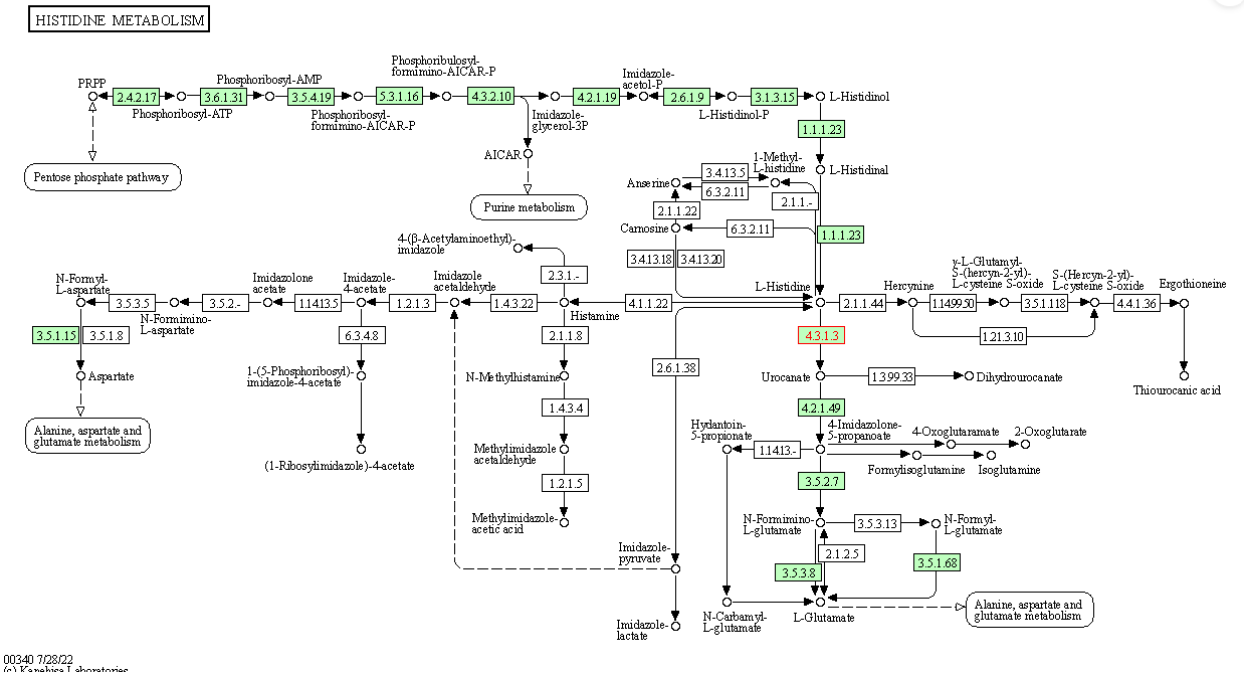
>WP_151704889.1 histidine ammonia-lyase [Nitrincola alkalilacustris]
MSKLVIQPGQMTLAQLRQVYSSPVSIEIPTSANTAIEASVSCVNQVVAEDRTVYGINTGFGLLAQTRIKH
EDLESLQRSLVLSHATGIGAAISDDLVRLIMVLKVNSLARGFSGIRRKVIDALITLINAEIYPCIPVKGS
VGASGDLAPLAHMSAVLLGEGKARYRGEWLSASEALLVAGLEPLALAPKEGLALLNGTQVSTAYALRGLF
EAEDLYAAATVCGSLTVEATLSSRSPFDARIHAIRGQQGQIDAAAGYRHLLQERSEISDSHSHCDKVQDP
YSLRCQPQVMGACLTQIRQAAEVLLVESNAVSDNPLVFAEEGDIISGGNFHAEPVAMAADNLALAIAEIG
SLTERRISLMMDKHMSQLPPFLVENGGVNSGFMIAQVTAAALASENKALAHPHSVDSLPTSANQEDHVSM
APAAGKRLWEMADNTRAVLAIEWLAACQGLDFRNGMKTTAPLEQARQWLREKVDYYDKDRFFAPDIQAAI
ELLESRRLSELVQKPLLPSINS



2.

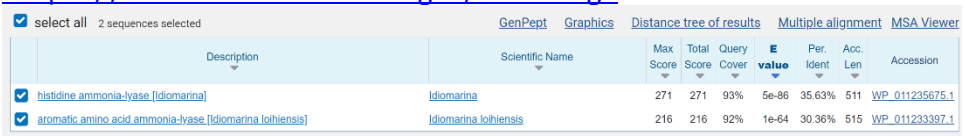
>WP_240776273.1 aromatic amino acid ammonia-lyase [Nitrincola alkalilacustris]

MTSVNDLPSPQPSLESSDNSVDKICFGEGRITIEQINLIARRRLTVSLSSSAEFAHKIDSAVAFLDKLLE
EDGVIYGVTTGYGDSCTVKVPLSQVYELPIHLTRFHGCGLGDNFSPQEGRAILATRLCSLTQGYSGVSWD
LLRQLTLFLNQDVVPVIPQEGSVGASGDLTPLSYLAASLIGEREVYYRGRIRPAAEVLAELNTAPIRLRP
KEGLAIMNGTAVMTALACIAYERAEYLTRLSSRLTALASLALQGNGHHFDEILFSVKPHPGQNEVASWIR
DDLNHHEHPRNSDRLQDRYSIRCAPHITGVLRDALPWFRQMIENELNSANDNPIIDGIGEHVLHGGHFYG
GHIAMVMDSMKNAVANLADLHDRQMALLMDVKFNNGLPANLSAAEAERQFINHGFKAVQIGCSAWTAEAL
KLTMPASVFSRSTECHNQDKVSMGTIAARDCLRVLQLTEQVMAAMLMAVQQGVQLRIRQGQLKQESLSPP
IQSMLAELQESCPLLVEDRRLDSELRQLVSGIQEQRWLLYQEAVRP



NCBI Protein Blast for Idiomarina loihiensis L2TR (taxid:283942)- TAL gene

https://blast.ncbi.nlm.nih.gov/Blast.cgi



1.

>WP_011235675.1 MULTISPECIES: histidine ammonia-lyase [Idiomarina]

MSHFELQPGQLQLSELRDWFYQHQTLKLSDEAKDNIATSAKTVADVLEQGRVVYGINTGFGLLANTRIPP
ERLTDLQRRIVLSHAAGTGDLMEDSVVRLMLLLKINSLSRGFSGVRQVLVDALIKLLNAEVYPCIPEKGS
VGASGDLAPLAHMVLPLVGEGTVRHNGKVLNAEEGLKIAGIEPFELAPKEGLALLNGTQASTALALAGLF
RIERNFHAAIVVGATSVEAAMGSRAPFDERVHAVRGQPGQIKAAEMLRHVLTDSSEIAKDHENCEKVQDP
YSLRCQPQVMGAVLDQIEHASGILVREANGVTDNPLVFSEEQDIISGGNFHAEPVAMAADILAIAASEIG
ALSERRSALLIDSHLSKLPAFLVNDGGVNSGFMLAQVTAAALASENKTLAHPASVDSLPTSANQEDHVSM
ATFAARRLTDIAKNVSDIIAIEWLEAAQGLDFRRPLKGAAAVETAFNCLREQVAYYAEDRFFAPDIKAAS
DLIQNGELAAVVKLPHILSEV

Idiomarina loihiensis GSL 199: K734_12330

Help

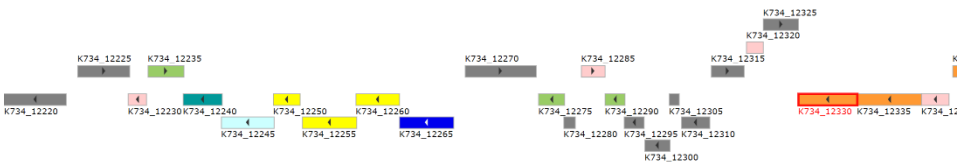
Entry	K734_12330 CDS T02681
Name	(GenBank) histidine ammonia-lyase
KO	K01745 histidine ammonia-lyase [EC:4.3.1.3]
Organism	ili Idiomarina loihiensis GSL 199
Pathway	ili00340 Histidine metabolism ili01100 Metabolic pathways
Module	ili_M00045 Histidine degradation, histidine => N-formiminoglutamate => glutamate
Brite	KEGG Orthology (KO) [BR:ili00001] 09100 Metabolism 09105 Amino acid metabolism 00340 Histidine metabolism K734_12330 Enzymes [BR:ili01000] 4. Lyases 4.3 Carbon-nitrogen lyases 4.3.1 Ammonia-lyases 4.3.1.3 histidine ammonia-lyase K734_12330 BRITE hierarchy BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: Lyase_aromatic DUF348 Motif
Other DBs	NCBI-ProteinID: AGM37325
Position	complement(2621189..2622724) Genome browser
AA seq	511 aa AA seq DB search MSHFELQPGQLQLSELRDWFYQHQTLKLSDEAKDNIATSAKTVADVLEQGRVVYGINTGF GLLANTRIPPERLTDLQRRIVLSHAAGTGDLMEDSVVRLMLLLKINSLSRGFSGVRQVLV DALIKLLNAEVYPCIPEKGSVGASGDLAPLAHMVLPLVGEGTVRHNGKVLNAEEGLKIAG IEPFELAPKEGLALLNGTQASTALALAGLFRIERNFHAAIVVGATSVEAAMGSRAPFDER VHAVRGQPGQIKAAEMLRHVLTDSSEIAKDHENCEKVQDPYSLRCQPQVMGAVLDQIEHA SGILVREANGVTDNPLVFSEEQDIISGGNFHAEPVAMAADILAIAASEIGALSERRSALL IDSHLSKLPAFLVNDGGVNSGFMLAQVTAAALASENKTLAHPASVDSLPTSANQEDHVSM ATFAARRLTDIAKNVSDIIAIEWLEAAQGLDFRRPLKGAAAVETAFNCLREQVAYYAEDR FFAPDIKAASDLIQNGELAAVVKLPHILSEV

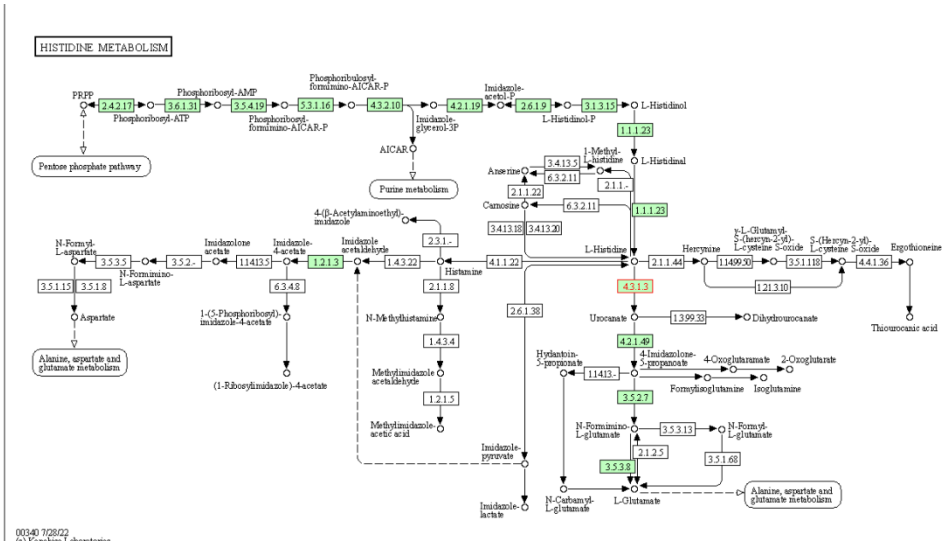
KEGG Genome Browser - Idiomarina loihiensis GSL 199

[Copy URL | Image file | Help]

ili; Idiomarina loihiensis GSL 199 Length : 2,839,759 (circular)

Position : 2600417 - 2643497 KID Select mode





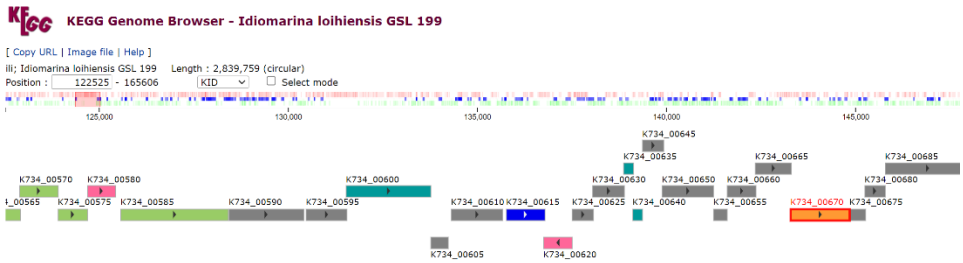
2.

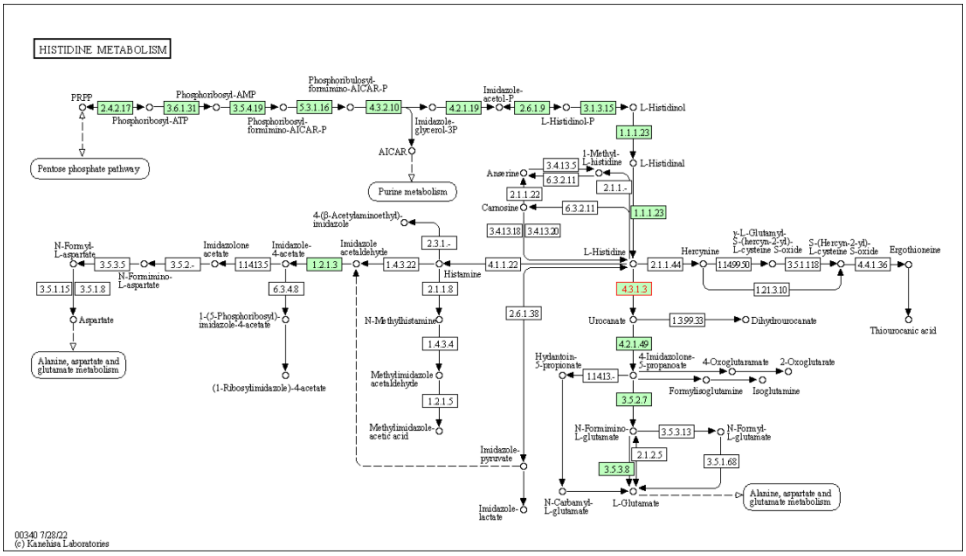
```
>WP_011233397.1 aromatic amino acid ammonia-lyase [Idiomarina loihiensis]
MTTSIIAFGGRPLTIKDVVEIANGKAQIALSDAPEFVEKVDAGVRFLDELLEQDGVIYGVTTGYGDSCTV
SVPENLVNELPIHLTRFHGCGLGDTFDEQETRAILATRLSSLAQGYSGVSWELLERLVVMLNENMLPLIP
KEGSVGASGDLTPLSYIAGALIGERDVRFRNKVMNSAEAFELLGMSVHKLRPKEGLAIMNGTAVMTALAC
LAWDRAEYLTRLSSRITSLASIALEGNSNHFDDLLFAVKPHKGQQQVASWIQQDLNHVEHPRNSSRLQDR
YSIRCAPHVIGVLKDSLPWFKETIENELNSANDNPIIDGLGEHVLHGGHFYGGHIAMVMDSMKTAVANLA
DLHDRQMALLMDTKMNHGLPSNLSAAEGERKSINHGFKAVQIGCSAWTAEALKLTMPASVFSRSTECHNQ
DKVSMGTIAARDCIRILELTEQVVVATLLAAYQGVELRKRITDKPVNPSAGVLEMLRSLADQFELLTEDR
QLEPELRYWLTQLRLKVWNLYQEPA
```

Idiomarina loihiensis GSL 199: K734_00670

Help

Entry	K734_00670 CDS T02681
Name	(GenBank) histidine ammonia-lyase
KO	K01745 histidine ammonia-lyase [EC:4.3.1.3]
Organism	ili Idiomarina loihiensis GSL 199
Pathway	ili00340 Histidine metabolism ili01100 Metabolic pathways
Module	ili_M00045 Histidine degradation, histidine => N-formiminoglutamate => glutamate
Brite	KEGG Orthology (KO) [BR:ili00001] 09100 Metabolism 09105 Amino acid metabolism 00340 Histidine metabolism K734_00670 Enzymes [BR:ili01000] 4. Lyases 4.3 Carbon-nitrogen lyases 4.3.1 Ammonia-lyases 4.3.1.3 histidine ammonia-lyase K734_00670 BRITE hierarchy BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: Lyase_aromatic TAFH Motif
Other DBs	NCBI-ProteinID: AGM35001
Position	143291..144838 Genome browser
AA seq	515 aa AA seq DB search MTTSIIAFGGRPLTIKDVVEIANGKAQIALSDAPEFVEKVDAGVRFLDELLEQDGVIYGV TTGYGDSCTVSVPENLVNELPIHLTRFHGCGLGDTFDEQETRAILATRLSSLAQGYSGVS WELLERLVVMLNENMLPLIPKEGSVGASGDLTPLSYIAGALIGERDVRFRNKVMNSAEAF ELLGMSVHKLRPKEGLAIMNGTAVMTALACLAWDRAEYLTRLSSRITSLASIALEGNSNH FDDLLFAVKPHKGQQQVASWIQQDLNHVEHPRNSSRLQDRYSIRCAPHVIGVLKDSLPWF KETIENELNSANDNPIIDGLGEHVLHGGHFYGGHIAMVMDSMKTAVANLADLHDRQMALL MDTKMNHGLPSNLSAAEGERKSINHGFKAVQIGCSAWTAEALKLTMPASVFSRSTECHNQ DKVSMGTIAARDCIRILELTEQVVVATLLAAYQGVELRKRITDKPVNPSAGVLEMLRSLA DOFELLTEDROLEPELRYWLTOLRLKVWNLYOEPA





Bioinformatics- pCL gene comparison

***Halorhodospira halophlia* sequences used for comparison**

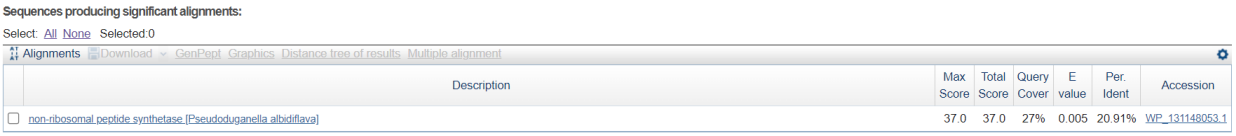
Hhal_1818 H.halophlia PYP in KEGG

pCL Hhal_1819

MQGLNADEVLRLLRSLIPGELAEGRGHRNDPPEGTDICADTRLDHTPIRADSLDRLHLAS
ALNRLFCLHETGVEDRLLTVRRIGDIAELIAEGSQHTSGLSFSTSGSTGTPQSHHHSWSA
LTQEAEALAAALGHHRRVIAWLPLHHLYGFVFGVALPRTLGSTVVESHEAPAALFRNPAP
DDLIASVPARWRYLLDSDHRFPGGTGVSSTAPLEAACRHGLPRAGLDALVEVYGATETGG
IGLRWAPAEDYRLLPYWQCNADGNLRRALPEGSAVTITPLDRLEWLDERVFRPRGRIDDI
IQIGGVNVSPQHVARRFESHEAVAACAIRSHGEGSQRRLKAFIVPAHPETDPEELRQALE
TWAWEHLPAVERPTDLRIGPELPRNAMGKLQDWD

NCBI Protein blast for M. albidiflava DSM17472- pCL gene

NCBI Blast:pCL protein blast for M albidiflava (nih.gov)



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
non-ribosomal peptide synthetase [Pseudoduganella albidiflava]	37.0	37.0	27%	0.005	20.91%	WP_131148053.1

Each following hit is then analyzed within the KEGG database by coping the FASTA sequence into the KEGG BLAST below are the populated results

1.

KEGG

Pseudoduganella albidiflava: EYF70_26465

Help

Entry	EYF70_26465 CDS T05903
Name	(GenBank) amino acid adenylation domain-containing protein
Organism	mali Pseudoduganella albidiflava
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: AMP-binding Condensation NAD_binding_4 PP-binding AMP-binding_C Epimerase GDP_Man_Dehyd 3Beta_HSD NAD_binding_10 Motif
Other DBs	NCBI-ProteinID: QBI03963 UniProt: A0A411X500
Position	complement(6374866..6382407) Genome browser
AA seq	2513 aa AA seq DB search MNKQDIQEVHRLTPLQQGMLFHTLEAPGSGVYVEQFACPAGGRIDGARWREAWNLVLGAS PVLRAAIAWEGLEHPLLVVTKSATVDVIEIDATGDDDSAFAARLDALRREEAARGFDLRR PPLLRQALLHREDRSVVFWTYHHVLLDGWSAFVVLGSLLSAYAALADGQAWQPRPAPGWG NYLGWLRGRDKAAAESYWRGRMAGFHAPTPLGLVKRPAPDSGADAGATATAELGEALGSR LRDTARTLRVTPGTLLQAAWAHLLALYSGEDDVAFGATVAGRPPELAGADAMVGLFINTV PVRVRIDPAETVRGFAARLGAALVAQREHEHASLTDIAGWSEVPRGRALFDSMLAIESFP TTDGFSLPDVSVWQHTNFPLALVIDPVGRMRIKALYDARQFDAGAIAGLLSHYRELLSRM VGDVDLPVGTLSIAPGTDLALQDAWNHRPAVHAASDLGELLARQVAERPDAPAIVSGDGH ATYAELDLRARRLGARLLAAGVRAGDTVAFAFEPGVPMIAALVAITRLGCAYAPLDARLP APRLAQMMTDLRIRHVVTDGAHAALFDIAGVTVLLAGEDEKDVPLKDWPETDPARILYVI HTSGSTGKPKAAGVFHDSFTRFLGWWNREFAFAPGERCLLVNKITFDLAQKCVWGALATG GILHLAPTRHFDPVHAREQVAAHGIGWINCTPSMAYAMVEADSCSHDALASLRLLFVGGE PVDKRRLAPWLLSPGCTTELVNTYGPTECTDLCTTHRFTRDEFVDLAAPVTVGSVLPGLA VHVLDRFGNRLPAGVTGEVAIAGGSVGAGYLNNAQMSAQKFLPDPGMRGARLYLTGDLGY FRADGTLVVQGRVDFQVKLRGYRIELDAIGHELRGHPAVLDAVAAVTPDGQRLVAYVVPE GGAAWTPALQDECRAWLAARLPDYMVPARYVPLAALPLNANGKLDRGALPSPELGIGLGN SDVERIAPRDDIEAKLAGIWAAVLGTDEFGVTDDFFDLGGHSLSITQAYARLQKTFGVRI PLSVLFEQPTIAGQAAALREAGADGADGANGNQAAARQPAIATGQRPERLPLSFSQSRLW FLHQYDPASMAYHVPNALPLAGTVDRAALQRALDWLHARHESLRTRYPEADGMPWQEVTP PTSVALEYDDLRGAADAAERLSAIATAEAAQPFDLRAGPVARYRLVQGDDRAILLVSLHH IATDGWSMDVMMRELSAAYAAFSAGGAPQAAPLPIQYADYALWQRADLAGDKRGRLVDYW RTELDGSQPQITLPYDFPRPAARSSAGGLHVSHIPADVCAALHGAASAAAGSGSTAFMTW LALYELLLYRWSGQPDFNVGSPIANRNLEETEGVIGFFVNTLVLRARMEGGQTFAGLLDD VRRTARAAYDHAELPFELLVDELNPPRSANTLPFFQVGFALQRAYEDTSLIDSTEWIARF DLQLVLHESADGGLRAHWEFARDLFLPDTVARLADSFALLAROVAQAPAAPLRDLALVDG

All links

Genome (1)
KEGG GENOME (1)
Gene (1)
NCBI-PROTEINID (1)
Protein sequence (1)
UniProt (1)
Protein domain (9)
Pfam (9)
All databases (12)

Download RDF

```
>WP_131148053.1 non-ribosomal peptide synthetase [Pseudoduganella
albidiflava]
MNKQDIQEVHRLTPLQQGMLFHTLEAPGSGVYVEQFACPAGGRIDGARWREAWNLVLGASPVLRAAIAWE
GLEHPLLVVTKSATVDVIEIDATGDDDSAFAARLDALRREEAARGFDLRRPPLLRQALLHREDRSVVFWT
YHHVLLDGWSAFVVLGSLLSAYAALADGQAWQPRPAPGWGNYLGWLRGRDKAAAESYWRGRMAGFHAPTP
LGLVKRPAPDSGADAGATATAELGEALGSRLRDTARTLRVTPGTLLQAAWAHLLALYSGEDDVAFGATVA
GRPPELAGADAMVGLFINTVPVRVRIDPAETVRGFAARLGAALVAQREHEHASLTDIAGWSEVPRGRALF
DSMLAIESFPTTDGFSLPDVSVWQHTNFPLALVIDPVGRMRIKALYDARQFDAGAIAGLLSHYRELLSRM
VGDVDLPVGTLSIAPGTDLALQDAWNHRPAVHAASDLGELLARQVAERPDAPAIVSGDGHATYAELDLRA
RRLGARLLAAGVRAGDTVAFAFEPGVPMIAALVAITRLGCAYAPLDARLPAPRLAQMMTDLRIRHVVTDG
AHAALFDIAGVTVLLAGEDEKDVPLKDWPETDPARILYVIHTSGSTGKPKAAGVFHDSFTRFLGWWNREF
AFAPGERCLLVNKITFDLAQKCVWGALATGGILHLAPTRHFDPVHAREQVAAHGIGWINCTPSMAYAMVE
ADSCSHDALASLRLLFVGGEPVDKRRLAPWLLSPGCTTELVNTYGPTECTDLCTTHRFTRDEFVDLAAPV
TVGSVLPGLAVHVLDRFGNRLPAGVTGEVAIAGGSVGAGYLNNAQMSAQKFLPDPGMRGARLYLTGDLGY
FRADGTLVVQGRVDFQVKLRGYRIELDAIGHELRGHPAVLDAVAAVTPDGQRLVAYVVPEGGAAWTPALQ
DECRAWLAARLPDYMVPARYVPLAALPLNANGKLDRGALPSPELGIGLGNSDVERIAPRDDIEAKLAGIW
AAVLGTDEFGVTDDFFDLGGHSLSITQAYARLQKTFGVRIPLSVLFEQPTIAGQAAALREAGADGADGAN
GNQAAARQPAIATGQRPERLPLSFSQSRLWFLHQYDPASMAYHVPNALPLAGTVDRAALQRALDWLHARH
ESLRTRYPEADGMPWQEVTPPTSVALEYDDLRGAADAAERLSAIATAEAAQPFDLRAGPVARYRLVQGDD
RAILLVSLHHIATDGWSMDVMMRELSAAYAAFSAGGAPQAAPLPIQYADYALWQRADLAGDKRGRLVDYW
```

RTELDGSQPQITLPYDFPRPAARSSAGGLHVSHIPADVCAALHGAASAAAGSGSTAFMTWLALYELLLYR
WSGQPDFNVGSPIANRNLEETEGVIGFFVNTLVLRARMEGGQTFAGLLDDVRRTARAAYDHAELPFELLV
DELNPPRSANTLPFFQVGFALQRAYEDTSLIDSTEWIARFDLQLVLHESADGGLRAHWEFARDLFLPDTV
ARLADSFALLARQVAQAPAAPLRDLALVDGAGRARMLALARGAKPALPIATVHELFAQQALLAPAKVALV
QGELRLGYDELNRRANRLAHHLLGEGVAPGGIVGVALPRGPQLIVALLAVLKAGCAYLPLDTEYPRERTA
YMLANAGAGAVLTDVHGAALLPAFDGVVLDVDAPQVAALRDTNPDVPNDPRALAYVMYTSGTTGKPKGVM
IEHRSIVRLVREAGYADFSADGTYLQYAPVGFDASTFEIWGSLLNGARLVQAPAGVVGLDQLAGLVAGER
IDTVFVTAALFNQLVDQHPESLAGVRQLLTGGEAMSSPHAARAIEAMRAGSKGELIHVYGPTECTTFATA
GKVSMADALAGTVPIGRPIDHTDTYVLDEALEPVPVGVPGELYIGGAGVARGYLNAPQQSAERFVADPFG
ADSAAAGSRLYRTGDKVRWNAGGTLEFLGRMDDQVKIRGYRIEPAEIEAQLAACPGVNGALVLALGDSAD
KRLVAYVTCTEGTGVASPTPALLRDYLAQRLPAFMIPGAYVLLDRFPVTANGKVDRRALPLPDADALGAG
DYVAPETALEESIATVWGEMLGLERVSVTADFFALGGNSLTATQLLARVRASLGHTVTLPEFFSEPTVRA
MAFRIEHADTARAMDAAESRLDGEAESALVLPEPLPPLASSLEHVLLTGATGFVGAYVAADMLERWPRVT
LHCHVRAAHPATGLRRLRDNLDQYGLWKEGYAARIRVLTGDLADPRLGLDDEHYATLARDVDLVVHNASR
LNHVLPYQALRRDNVEPTRRLLELAATAKRKGFVLVSTAGVLQGESGGTFDEDVRVEEIGQSAREGYNAS
KWVSELMVRRAARAGIPAQIMRLGRVVVDSRSGAGRMDDFVALFVRTCLQVGAWPDRPFVEQFVPVDHVA
RAITALAADYTSTGVHHLVGDDRRDWSRLLPDFVDCADAGLKRMPIRDWVDAVKERSAVEPLPFAPYLFW
WDTDAAAPEEKRLKVKQAKTVRRLAGHGLREPRIDSEAWQRYVGAIFAAEGMSAKPKKRGIFG

NCBI Protein blast for M. plicata DSM17505 -pCL gene
NCBI Blast:Protein Sequence (nih.gov)

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
class I adenylate-forming enzyme family protein [Pseudoduganella plicata]	60.1	60.1	72%	1e-10	27.74%	WP_134387101.1
AMP-binding protein [Pseudoduganella plicata]	37.7	37.7	28%	0.002	25.44%	WP_189568707.1
non-ribosomal peptide synthetase [Pseudoduganella plicata]	33.9	33.9	27%	0.034	31.36%	WP_134382918.1

Each following hit is then analyzed within the KEGG database by coping the FASTA sequence into the KEGG BLAST below are the populated results

1.

KEGG **Pseudoduganella plicata: E1742_21150**

Help

Entry	E1742_21150 CDS T06563
Name	(GenBank) long-chain fatty acid--CoA ligase
Organism	mpli Pseudoduganella plicata
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: AMP-binding Motif
Other DBs	NCBI-ProteinID: QBQ38400 UniProt: A0A4P7BJ12
Position	4764322..4765620 Genome browser
AA seq	432 aa AA seq DB search MATPNQAVVARFDDGRTVHREQAGRDIAGLMAALAGAPGAILLDCQCSYRFYVALLACLA AGRQVVLLPNTLPATCARYRDQYDSVIDDAGGVLRLRHGTSAVALVESGGTPWDGQLDPA AMVLLYTSGSTGEPKKVVKTVAGLMAEVSDLQALFLGPPALLIAATVPHYHLYGLLFRIL LPFQSGYPFLAGTIRSPAELAELADYVLVSSPAFLRRLTEQEPIGGAALVFSSGGALMPD EELLVQQVFGAPCFEIYGSSETGGIAYRRAEAGQHLWPLPGTEVKMGTDGRLWLWSRYCS EAGWIERDDLIQLHADGSFRVLGRADDIVKIEEKRISLAHLASLIRGTSPAIRTARLMVY DEPMRKRLAAIVVVDAAVPESQFAQIEQAAKAELQGHVERMFVPRKWKVVHEAVLDGMGK LSRSSMLELMHE
NT seq	1299 nt NT seq +upstream 0 nt +downstream 0 nt gtggcaacgccgaaccaggcggtggtggcgcgattcgacgatgggcgcacggtccatcgc gaacaggccggtcgggacattgccggcctcatggcggccctggccggggcgcctggcgcc atcctgctcgactgccagtgcagttaccgcttctatgtcgccctgctcgcctgcctggcg gcgggccgccaggtggtgctgctgcccaatacgctgccggccacctgcgcgcgctatcgg gaccagtatgacagcgtgatcgacgatgccggcggggtgctgcgcctgcgtcacggcacc agcgcggtcgccctggtggagtcaggcggcacgccgtgggatggccagctcgatccggcg gcgatggtgttgctgtatacctccggatcgaccggtgaaccgaaaaaagtcgtcaagaca gtggccggcctgatggcggaagtgagcgacctgcaagccctgttcctggggccgccagcg ctgctgatcgccgctaccgtgccgcactaccacctgtatggcctgctgtttcggattctg ttgccgttccagtcgggctacccgttcctggccggcacgatccgctcgccggctgagctg gcggagctggccgactatgtgctggtcagcagcccggccttcctgcgccggctgacggag caggagccgatcggcggcgctgccctggtgttctcgtcgggcggtgccttgatgcctgac gaggaactgctggtgcaacaggtgttcggcgcgccgtgcttcgagatctacggcagcagt gaaacgggcggtatcgcctaccgccgcgccgaggcggggcagcatttgtggccgcttccc ggtaccgaggtcaagatgggtaccgatggccggctgtggctctggtcgcgctattgcagt gaagcgggctggatcgaacgcgacgacctgatccagctgcatgccgacggcagcttccgc gtgctgggccgccgccgacgatatcgtcaagatcgaagagaaacgcatttcgctggcccat

>WP_134387101.1 class I adenylate-forming enzyme family protein [Pseudoduganella plicata]

MATPNQAVVARFDDGRTVHREQAGRDIAGLMAALAGAPGAILLDCQCSYRFYVALLACLAAGRQVVLLPN
TLPATCARYRDQYDSVIDDAGGVLRLRHGTSAVALVESGGTPWDGQLDPAAMVLLYTSGSTGEPKKVVKT
VAGLMAEVSDLQALFLGPPALLIAATVPHYHLYGLLFRILLPFQSGYPFLAGTIRSPAELAELADYVLVS
SPAFLRRLTEQEPIGGAALVFSSGGALMPDEELLVQQVFGAPCFEIYGSSETGGIAYRRAEAGQHLWPLP
GTEVKMGTDGRLWLWSRYCSEAGWIERDDLIQLHADGSFRVLGRADDIVKIEEKRISLAHLASLIRGTSP
AIRTARLMVYDEPMRKRLAAIVVVDAAVPESQFAQIEQAAKAELQGHVERMFVPRKWKVVHEAVLDGMGK

LSRSSMLELMHE

2.

KEGG

Telluria chitinolytica: PX653_13400

Help

Entry	PX653_13400 CDS T08857
Name	(GenBank) AMP-binding protein
KO	K08295 2-aminobenzoate-CoA ligase [EC:6.2.1.32]
Organism	tct Telluria chitinolytica
Pathway	tct00627 Aminobenzoate degradation tct01110 Biosynthesis of secondary metabolites tct01120 Microbial metabolism in diverse environments
Brite	KEGG Orthology (KO) [BR:tct00001] 09100 Metabolism 09111 Xenobiotics biodegradation and metabolism 00627 Aminobenzoate degradation PX653_13400 Enzymes [BR:tct01000] 6. Ligases 6.2 Forming carbon-sulfur bonds 6.2.1 Acid-thiol ligases 6.2.1.32 anthranilate---CoA ligase PX653_13400 BRITE hierarchy
Motif	Pfam: AMP-binding AMP-binding_C Motif
Other DBs	NCBI-ProteinID: WEF35696
Position	complement(3029787..3031403) Genome browser
AA seq	538 aa AA seq DB search MSTDARSASGHVDPFARAHLPPREEWPELLFELPELHYPDRLNCTAALLDDTARAHGERV ALHGVHGSWTYAQLLERVDRIAHVLRGPLGLVPGNRVLLRGANTPMMAAALLAVFKAGCI AVPTMPLLRARELATIVAKAEVNAVLCAADLCAELEQLPACPPAMHFGGTDTQLEALMAA QPATFAACDTAADDVCLIGFTSGTTGVPKGTMHFHRDLLAICDTFARHTLRAQPADLFIG TPPLAFTFGLGGLLLFPLRIGAASVLLEKLTPETLLQAIERHRATVCFTAPTFYRQMAAL ASRYDLACLAKTVSAGEALPIATRQAWQQATGLAMIDGIGATEMLHIFISAAGDEIRPGA TGKPVPGYRACVLDEAGRPLGPGIVGRLAVKGPTGCRYLADERQRDYVRDGWNLTGDAYE MDADGYFIYRSRTDDMIISAGYNIAGPEVEDALLRHPAVAECGVIGRPDPERGQLVEAHV VLRPGHAPSAQLTAQLQDFVKNQIAPYKYPRSIVYAEALPRTETGKLQRFRLRAKETA
NT seq	1617 nt NT seq +upstream 0 nt +downstream 0 nt atgagcacagatgcaagaagcgccagcggccatgtcgatccgttcgctagggcgcacctg

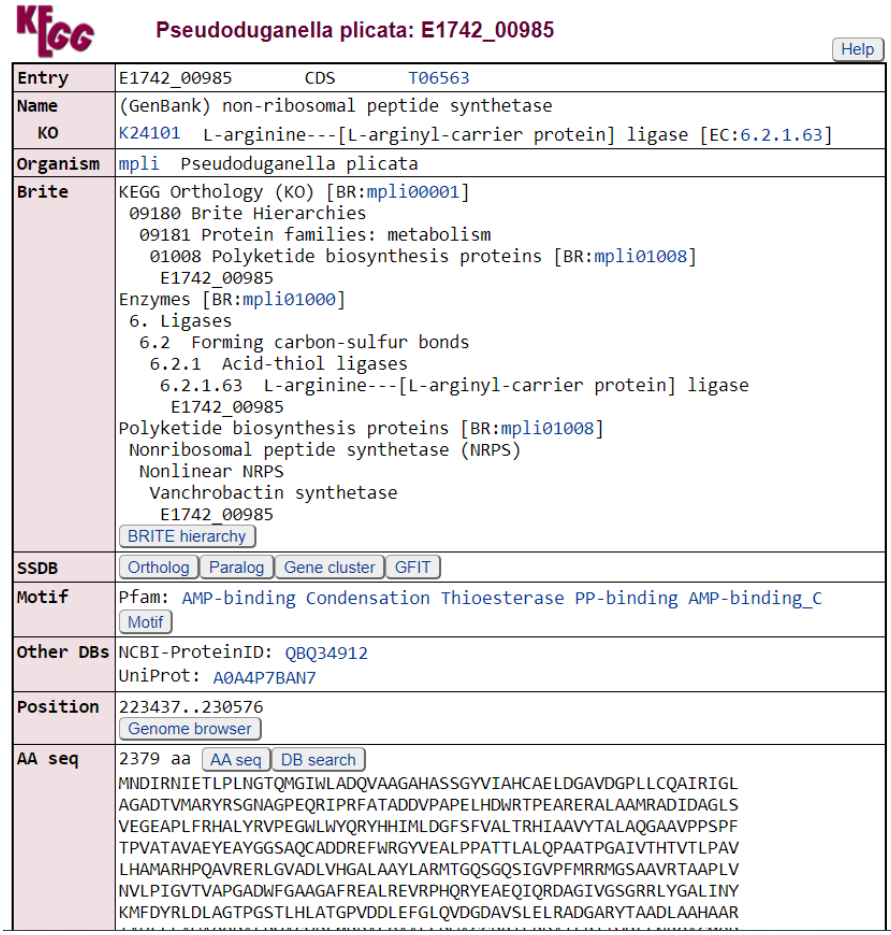
All links

Ontology (2)
KEGG BRITE (2)
Pathway (3)
KEGG PATHWAY (3)
Chemical substance (6)
KEGG COMPOUND (6)
Chemical reaction (2)
KEGG ENZYME (1)
KEGG REACTION (1)
Genome (1)
KEGG GENOME (1)
Gene (2)
KEGG ORTHOLOGY (1)
NCBI-PROTEINID (1)
Protein domain (2)
Pfam (2)
All databases (18)

Download RDF

>WP_189568707.1 AMP-binding protein [Pseudoduganella plicata]
MIPDARSASGHVDTFARQHLPPREQWPELLFELPELRYPDRLNCAAELLDHANGARPALYGVRGTWTYAE
LREKVDRIAHVLRGPLGLVPGNRVLLRGANTPMMAAAILAVFKAGCIAVPTMPLLRARELGTIMARAQVD
AVLCAGDLCAELDLLPDCPRTVRFGGDGTQGLEALMAAQPAAFSPCDTAADDVCLIGFTSGTTGVPKGTM
HFHRDILAICDCFPKHTLRAQPDDLFIGTPPLAFTFGLGGLLLFPLRIGAAGVLLEKLTPETLLGAIERH
RATVCFTAPTFYRQMATLAPRYDLSSLARTVSAGEALPLATRTAWQQATGLRMIDGIGATEMLHIFISAA
GDDIRPGATGKPVPGYRACILDDAGRPAGPGTIGRLAVKGPTGCRYLADERQKDYVIDGWNVTGDAYEMD
ADGYFIYRSRVDDMIISAGYNIAGPEVEDALLRHPAVAECGVIGRPDVERGQIVEAHVVLRPGHEPSPAF
TAQLQEFVKREIAPYKYPRSIVYAQALPRTETGKLQRFRLRTQAAKDGA

3.



KEGG

Pseudoduganella plicata: E1742_00985

Help

Entry	E1742_00985 CDS T06563
Name	(GenBank) non-ribosomal peptide synthetase
KO	K24101 L-arginine---[L-arginyl-carrier protein] ligase [EC:6.2.1.63]
Organism	mpli Pseudoduganella plicata
Brite	KEGG Orthology (KO) [BR:mpli00001] 09180 Brite Hierarchies 09181 Protein families: metabolism 01008 Polyketide biosynthesis proteins [BR:mpli01008] E1742_00985 Enzymes [BR:mpli01000] 6. Ligases 6.2 Forming carbon-sulfur bonds 6.2.1 Acid-thiol ligases 6.2.1.63 L-arginine---[L-arginyl-carrier protein] ligase E1742_00985 Polyketide biosynthesis proteins [BR:mpli01008] Nonribosomal peptide synthetase (NRPS) Nonlinear NRPS Vanchrobactin synthetase E1742_00985 BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: AMP-binding Condensation Thioesterase PP-binding AMP-binding_C Motif
Other DBs	NCBI-ProteinID: QBQ34912 UniProt: A0A4P7BAN7
Position	223437..230576 Genome browser
AA seq	2379 aa AA seq DB search MNDIRNIETLPLNGTQMGIWLADQVAAGAHASSGYVIAHCAELDGAVDGPLLCQAIRIGL AGADTVMARYRSGNAGPEQRIPRFATADDVPAPELHDWRTPEARERALAAMRADIDAGLS VEGEAPLFRHALYRVPEGWLWYQRYHHIMLDGFSFVALTRHIAAVYTALAQGAAVPPSPF TPVATAVAEYEAYGGSAQCADDREFWRGYVEALPPATTLALQPAATPGAIVTHTVTLPAV LHAMARHPQAVRERLGVADLVHGALAAYLARMTGQSGQSIGVPFMRRMGSAAVRTAAPLV NVLPIGVTVAPGADWFGAAGAFREALREVRPHQRYEAEQIQRDAGIVGSGRRLYGALINY KMFDYRLDLAGTPGSTLHLATGPVDDLEFGLQVDGDAVSLELRADGARYTAADLAAHAAR

All links

Ontology (3)
KEGG BRITE (3)
Chemical reaction (1)
KEGG ENZYME (1)
Genome (1)
KEGG GENOME (1)
Gene (2)
KEGG ORTHOLOGY (1)
NCBI-PROTEINID (1)
Protein sequence (1)
UniProt (1)
Protein domain (5)
Pfam (5)
All databases (13)

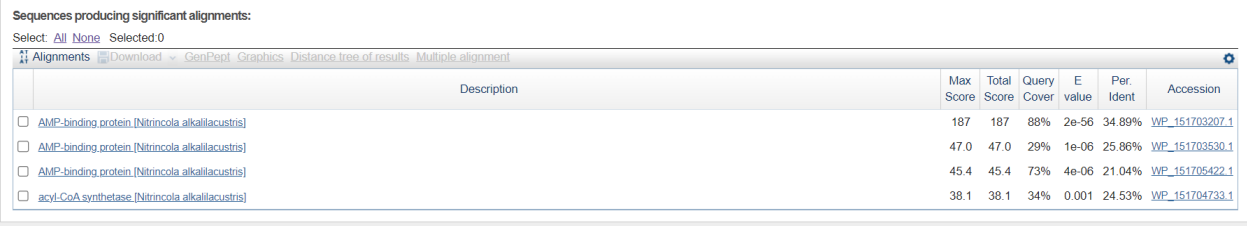
Download RDF

```
>WP_134382918.1 non-ribosomal peptide synthetase [Pseudoduganella plicata]
MNDIRNIETLPLNGTQMGIWLADQVAAGAHASSGYVIAHCAELDGAVDGPLLCQAIRIGLAGADTVMARY
RSGNAGPEQRIPRFATADDVPAPELHDWRTPEARERALAAMRADIDAGLSVEGEAPLFRHALYRVPEGWL
WYQRYHHIMLDGFSFVALTRHIAAVYTALAQGAAVPPSPFTPVATAVAEYEAYGGSAQCADDREFWRGYV
EALPPATTLALQPAATPGAIVTHTVTLPAVLHAMARHPQAVRERLGVADLVHGALAAYLARMTGQSGQSI
GVPFMRRMGSAAVRTAAPLVNVLPIGVTVAPGADWFGAAGAFREALREVRPHQRYEAEQIQRDAGIVGSG
RRLYGALINYKMFDYRLDLAGTPGSTLHLATGPVDDLEFGLQVDGDAVSLELRADGARYTAADLAAHAAR
IAHLLEAWAQQPALPVASVPLMPPAEQAALEDWAGGPRIERDATLRTIVDLLNRQAGMQPEMTALVCGEE
RFTFAEVAGKVAQLARLLRSCGAGQGRVVAVALPRCADAVIAMLAVLESGATFLPLDLDYPLERIAMMCE
DTQPVLALCSERAAVPLPAGLPCLRIDAAGVRADVAGMPAHPLNDNERGTLAIDAIGYVIFTSGSTGRPK
GVMNTHTALLNLIGVHRDTIYEPARAAVRANFSGRALRAAHTHSFSFDSSWLQLFWMLLGEELHVIDDEM
RRDAWALARHVDSVGIDAMDLPPSFLAQLLNSGLMAGGTHAPTVLLIGGEAAPEALWRQLRAYPALQAHN
LYGPTEYTVDTLRAPILASARPVVGRPIGNTRVYVLDARLQPVPPGALGELYVSGAGLALGYLARPDLSA
ARFVADPFGQDGARMYRTGDLVRWNLSGQLEFIGRCDDQVKVRGYRVELGEVENALSLLPGVESVIVLAQ
PVNGSHRLVAYCAVPGIAGEARAQTCRELLAQLAQVLPDYMVPALLLVLDAFPRNVSGKIDRKRLPAPES
IVMTTGGGAPAAGTDAALLCDAMARTLKLAGVGADDDFFALGGDSIAAIVLCGELRRAGRQLRPSAVFAL
RTPRAMAASLTAIDTPAAGWSLPPAQHAALAARHGAFAAAAPVLPLQQGMLFHTGMGGNYSAFTRLRLEG
AVDPSRLRRALDALVLRYPQLGGLFDTETVEQPVFLLPAVGTAGWFWREDDVADLAPHARAERIAEVEAD
LLAGSCHTDRFGGMLAAGLTRTGAREYTLLLAIHHLVIDGWSTPLLLRDLLAAYRDDAALLPLPVGYEKV
VGALAVRDQTASIAQWRAALEGVQPTVLFGEPAGTVTESLLALTVQESADLTNRLRAAGVTLNVMMQAVW
GIVIGALAGRDDVVFGTPVAGRSAPVDGIGDQVGLFLNTVPVRVRLDAHAPLWAQLPALQARHAALQEHD
GPGLAEIQRAAGGATLFDTLLVVENYPDNSYLAQQLAGADGAPLRAGDVQNRGYSHYPLALLVLPGDEIA
FLLENRSAVADAAALAERVVGLLRTALATPDLPLARYPLVTPAETAAIGAVNATAHPLPPATLRSALAIQ
```

AERTPRAPALCDADHALDYGQLRQQVRVLAGRLAAAGVGPGDIVAVALPRSVRLTIALLAVIEAGAAYLP
LELGYPDERLQYMLEDAVPRLLIAQESDRARFTAAAVLTFDRLAPPGECGAELEPVLTPDHPAYLIYTSG
TTGRPKGALVSHRAIVNRIAWMQHQYALGPLDVVLQKTPCGFDVSVWEFFWPLMTGASLAMAAPDAHKDP
AALLATVEAYGATCMHFVPSMLATFTAHLATLDERRAASLRLVFCSGEALTKAQAAQFARYSAARLHNLY
GPTEAAVDVTYMPADEVTATGGSGVPIGRPVWNTQLRVLDRWLRPVPPGAVGELYLCGVQLALGYLGKAA
LTAGRFVADPFGDGGRMYRTGDVVRWLDDGTVEYLGRADDQVKIRGQRIELGEIESVLTALPEVAQAAVN
AVALGQAGGAVDNRQLVAWLVPVAGATLPEDAALRQLLATRLPAHMVPVAFVPLDRLPLSANGKLDRKSL
PLPVPDDSGRRIPARGLESRLAEVFAAVLGVPQVYADDDFFAIGGHSLLAMRLASEIRRVLERPVSVGQI
MTAPTVARLAARLNEDGMRNDFGGDGFEPVIHLRGGTGTPLFCFYPGSGFAWQYSVLARYLEEEVPIIGL
QSPRPDGLIATSPDMETLVTRQLAIVRGVQPDGPYRLLGYSLGGTVAYALAARLRAAGEEVSFLGLLDTY
PAEVHDWTDPQGAEAAMGAEREQQQLLADAYDGDDGAAAAALAAEREALLAQIFANYGDAVRLLARTRTP
GYDGDLTLFVAEQSLPAYIEPERDWRRHAASLRVHRLAQASHENIMSPQSLETLGPLLRAALAASEGTP

NCBI Protein blast for N. alkalilacustris DSM29817 -pCL gene

NCBI Blast:Protein Sequence (nih.gov)



Sequences producing significant alignments:

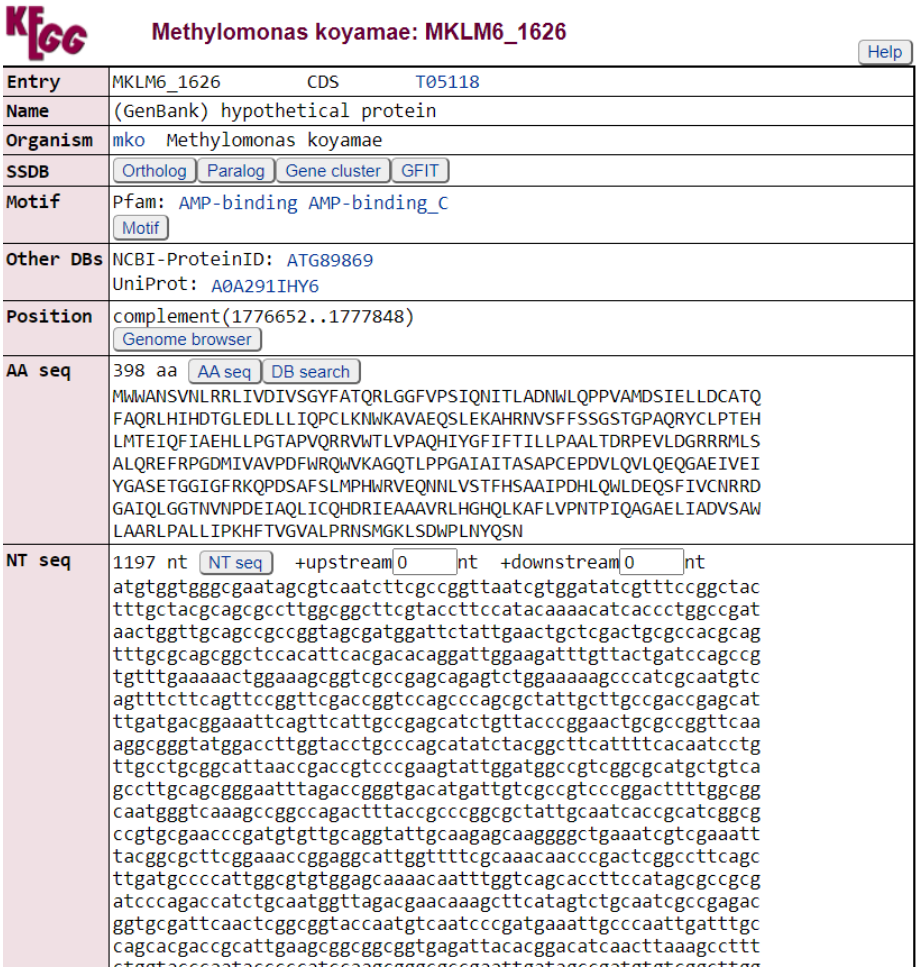
Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
AMP-binding protein [Nitrincola alkalilacustris]	187	187	88%	2e-56	34.89%	WP_151703207.1
AMP-binding protein [Nitrincola alkalilacustris]	47.0	47.0	29%	1e-06	25.86%	WP_151703530.1
AMP-binding protein [Nitrincola alkalilacustris]	45.4	45.4	73%	4e-06	21.04%	WP_151705422.1
acyl-CoA synthetase [Nitrincola alkalilacustris]	38.1	38.1	34%	0.001	24.53%	WP_151704733.1

Each following hit is then analyzed within the KEGG database by coping the FASTA sequence into the KEGG BLAST below are the populated results

1.



KEGG

Methylomonas koyamae: MKLM6_1626

Help

Entry	MKLM6_1626 CDS T05118
Name	(GenBank) hypothetical protein
Organism	mko Methylomonas koyamae
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: AMP-binding AMP-binding_C Motif
Other DBs	NCBI-ProteinID: ATG89869 UniProt: A0A291IHY6
Position	complement(1776652..1777848) Genome browser
AA seq	398 aa AA seq DB search MWWANSVNLRRLIVDIVSGYFATQRLGGFVPSIQNITLADNWLQPPVAMDSIELLDCATQ FAQRLHIHDTGLEDLLLIQPCLKNWKAVAEQSLEKAHRNVSFFSSGSTGPAQRYCLPTEH LMTEIQFIAEHLLPGTAPVQRRVWTLVPAQHIYGFIFTILLPAALTDRPEVLDGRRRMLS ALQREFRPGDMIVAVPDFWRQWVKAGQTLPPGAIAITASAPCEPDVLQVLQEQGAEIVEI YGASETGGIGFRKQPDSAFSLMPHWRVEQNNLVSTFHSAAIPDHLQWLDEQSFIVCNRRD GAIQLGGTNVNPDEIAQLICQHDRIEAAAVRLHGHQLKAFLVPNTPIQAGAELIADVSAW LAARLPALLIPKHFTVGVALPRNSMGKLSDWPLNYQSN
NT seq	1197 nt NT seq +upstream 0 nt +downstream 0 nt atgtggtgggcgaatagcgtcaatcttcgccggttaatcgtggatatcgtttccggctac tttgctacgcagcgccttggcggcttcgtaccttccatacaaaacatcaccctggccgat aactggttgcagccgccggtagcgatggattctattgaactgctcgactgcgccacgcag tttgcgcagcggctccacattcacgacacaggattggaagatttgttactgatccagccg tgtttgaaaaactggaaagcggtcgccgagcagagtctggaaaaagcccatcgcaatgtc agtttcttcagttccggttcgaccggtccagcccagcgctattgcttgccgaccgagcat ttgatgacggaaattcagttcattgccgagcatctgttacccggaactgcgccggttcaa aggcgggtatggaccttggtacctgcccagcatatctacggcttcattttcacaatcctg ttgcctgcggcattaaccgaccgtcccgaagtattggatggccgtcggcgcatgctgtca gccttgcagcgggaatttagaccgggtgacatgattgtcgccgtcccggacttttggcgg caatgggtcaaagccggccagactttaccgcccggcgctattgcaatcaccgcatcggcg ccgtgcgaacccgatgtgttgcaggtattgcaagagcaaggggctgaaatcgtcgaaatt tacggcgcttcggaaaccggaggcattggttttcgcaaacaacccgactcggccttcagc ttgatgccccattggcgtgtggagcaaaacaatttggtcagcaccttccatagcgccgcg atcccagaccatctgcaatggttagacgaacaaagcttcatagtctgcaatcgccgagac ggtgcgattcaactcggcggtaccaatgtcaatcccgatgaaattgcccaattgatttgc cagcacgaccgcattgaagcggcggcggtgagattacacggacatcaacttaaagccttt

All links

Genome (1)
KEGG GENOME (1)
Gene (2)
NCBI-PROTEINID (1)
OC (1)
Protein sequence (1)
UniProt (1)
Protein domain (2)
Pfam (2)
All databases (6)

Download RDF

>WP_151703207.1 AMP-binding protein [Nitrincola alkalilacustris]
MWWRDRKQLRRFTLDLVMHRLTIQRPGITLPSPANALLAQRWRDAPLNLDSLELVESATLFAQALHITET
GLEDLLLAQPSLNGWVDIACDSLEQYHDHISFFSSGSTGEPARHKQPLTHLLREASFLRECLQRQTAPDQ
QARVWSLVPAHHIYGFIFTVLLPDLLGPDTEVIDARTRLLTSIERNLKPGDILVAVPEFWQRWVSTGMRL
PPGCKAVTAAGPSNSDTLRQLINQGAELLEIYGSSETAGLGYRRSPESPLTLMPHWTIEGDTAVSDCLRA
DLPDRLEWSDARHFVVAGRKDKAVQIAGVNVYPDRIAGLISTHEAVAEAAVRLQGQRLKAFVVPEKDLPD
SDFASLEHSLRRWLSTQLPAGQIPGHFAFGDSLPRNSMNKLSDWLIE

2.

Azoarcus olearius BH72: azo1996

Help

Entry	azo1996 CDS T00441
Symbol	acS1
Name	(GenBank) putative acetyl-coenzyme a synthetase
KO	K08295 2-aminobenzoate-CoA ligase [EC:6.2.1.32]
Organism	azo Azoarcus olearius BH72
Pathway	azo00627 Aminobenzoate degradation azo01110 Biosynthesis of secondary metabolites azo01120 Microbial metabolism in diverse environments
Brite	KEGG Orthology (KO) [BR:azo00001] 09100 Metabolism 09111 Xenobiotics biodegradation and metabolism 00627 Aminobenzoate degradation azo1996 (acS1) Enzymes [BR:azo01000] 6. Ligases 6.2 Forming carbon-sulfur bonds 6.2.1 Acid-thiol ligases 6.2.1.32 anthranilate---CoA ligase azo1996 (acS1) BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: AMP-binding AMP-binding_C Motif
Other DBs	NCBI-ProteinID: CAL94613 RhizoBase: azo1996 UniProt: A1K708
Position	2197112..2198719 Genome browser
AA seq	535 aa AA seq DB search MNSSAHLDRFVIDNLPPAEQWPELIFERPELQFPEQLNATAELLDRALAEGHAERPAIIG KNIHWTYRQLHQQVNRLAHVLVEDMGLVPGNRVLLRGANTPWLAACWLAVWKAGGVAVGT MPLLRAKELKEILRLAEISHALCDASLAEELEYARAESPTLRHVMLFGEGGELAARMDGK PDQFDTVATAATDPALIGFTSGTTGIPKGTIHFHRDVMAMCEVFPRHCLKPTQDDVFIGT PPLAFTFGLGGLLCFPLWARASTVLLEKLAPEPLMQAIQDHQATVCFTSPTAYRQMAGLV GKYDIASLKKCVSAGEALPTDTRDKWREASGIQIHDGIGGTEMIHIYIASGPDDYRPGAL

All links

Ontology (2)
KEGG BRITE (2)
Pathway (3)
KEGG PATHWAY (3)
Chemical substance (6)
KEGG COMPOUND (6)
Chemical reaction (2)
KEGG ENZYME (1)
KEGG REACTION (1)
Genome (1)
KEGG GENOME (1)
Gene (9)
KEGG ORTHOLOGY (1)
RefGene (5)
NCBI-PROTEINID (1)
OC (1)
RHIZOBASE (1)
Protein sequence (1)
UniProt (1)
Protein domain (2)
Pfam (2)
All databases (26)

Download RDF

>WP_151703530.1 AMP-binding protein [Nitrincola alkalilacustris]
MAHTAHIDTFVRDHLPPAEQWPELIFNRPELQFPEYLNAVDQLLDRAVDEGDGERQAIIGKDVSWSYAEL
QRQVNRIANVLVQDLQLQTGNRVLLRGMNSPMLAACWLAVLKAGGVAVGTMPLLRAKELKDIAQAARITH
AICDESLAPELQLAAESCETLTQVIHYGGTGELEKLAADKTDEFAAVNTAAEDPALIAFTSGTTGIPKGT
VHFHRDIMAMCEVFPRHCLKPTRDDVFIGTPPLAFTFGLGGLLCFPLWARASTVLLEKLAPEPLLKAIEE
HGATICFTSPTAYRHMTPLVPEYNISSLQKCVSAGEALPTDTRDKWREATGIEIHDGIGGTEMIHIYLGS
SPEDYRAGALGRMLPGYIGMIVDEQMQPLPTGEVGKLAVKGPTGCRYLADERQTDYVVNGWNLPGDAFYM
DDDGYFYYQARVDDMIVTSGYNVASPEVESALLAHPAVAECGVVGVPDPHRGQVIKAYVVLNVDYVGDDE
MTAELQNFVKNTVAPYKYPRQIKYIDALPRTETGKLQRFKLKQI

3.

KEGG

Nitrincola iocasae: F5I99_07830

Help

Entry	F5I99_07830 CDS T06212
Name	(GenBank) AMP-binding protein
KO	K00666 fatty-acyl-CoA synthase [EC:6.2.1.-]
Organism	nik Nitrincola iocasae
Brite	KEGG Orthology (KO) [BR:nik00001] 09180 Brite Hierarchies 09181 Protein families: metabolism 01004 Lipid biosynthesis proteins [BR:nik01004] F5I99_07830 Enzymes [BR:nik01000] 6. Ligases 6.2 Forming carbon-sulfur bonds 6.2.1 Acid-thiol ligases 6.2.1.- F5I99_07830 Lipid biosynthesis proteins [BR:nik01004] Acyl-CoA synthetase Unknown F5I99_07830 BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: AMP-binding AMP-binding_C Motif
Other DBs	NCBI-ProteinID: QEW06423 UniProt: A0A5J6LD73
Position	complement(1692286..1693974) Genome browser
AA seq	562 aa AA seq DB search MPTLPSYTSGTSEKPLLGMTIGDMFDATVNRYPDNDALIVLHQNIRWTYKELHSQVNLCA RAFLAAGVRRGDRVAIWSPNRYEWTLTQFATAKIGAILVNINPAYRTHELEYALNQSESR YLVTADSFKASDYRGMLYELAPELKMCAEGQLDSQKLPLLKGVINLDDHKHPGMWRWQNF MALAESMPQEILDEAQAELQFDDAINIQYTSGTTGFPKGATLSHHNILNNGYFVAESMGF TEQDRLVIPVPLYHCFGMVMGNLGCVTHGSTMIYPDEGFEPGSVLKAVHQERATALYGVP TMFIAELGHPDFASMDLSSLRTGIMAGSLCPTEVMKQVINKMHMKEVQIAYGMTETSPVS MQTGAMDSLEKRVTTVGRTQPHLESKIIDPGNGSILPRGEVGELCTRGYSVMLKYWNNDQ GTADAIDEADWMHTGDLASMDDEGYIQIVGRIKDMVIRGGENIYPKEVEEFLYTHPSIAE

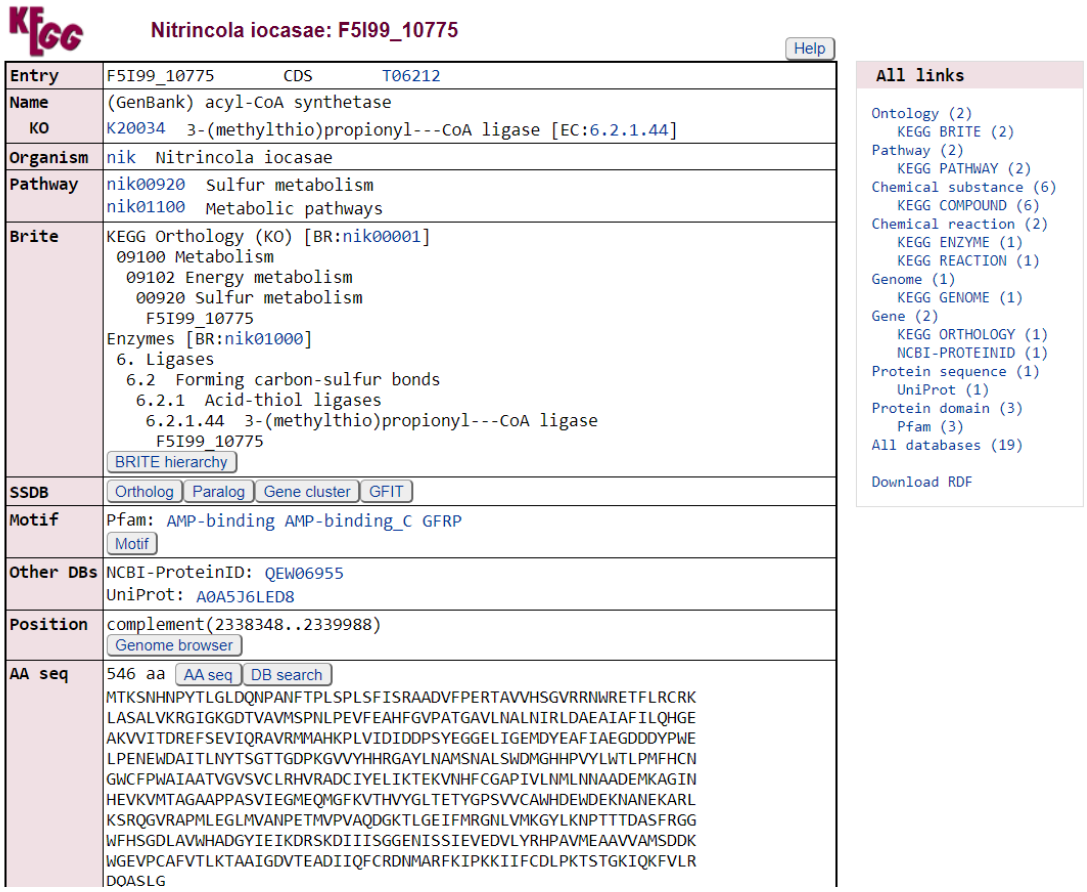
All links

Ontology (3)
KEGG BRITE (3)
Genome (1)
KEGG GENOME (1)
Gene (2)
KEGG ORTHOLOGY (1)
NCBI-PROTEINID (1)
Protein sequence (1)
UniProt (1)
Protein domain (2)
Pfam (2)
All databases (9)

Download RDF

>WP_151705422.1 AMP-binding protein [Nitrincola alkalilacustris]
MSKTLPSYTSGTAEQPLIGMTIGDMFDATANRYPDNDALIVLHQNIRWSYRELQAEVNRCARALLSIGVK
PGDRIGVWSPNRYEWTVSQFATAKVGAILVNINPAYRLHELEYALNQSETRFLISADSFKSSDYRHMIYE
LAPELKMCGEGQLKCAKLPHLQGVINLDSNKHPGMWSWNEFMLGSEATTQEAVDEIQAQLQFDDAINIQY
TSGTTGFPKGATLSHHNILNNGFFVAESMRFTDQDRLVIPVPLYHCFGMVMGNLGCVTHGATMIYPEEGF
EPGAVLRAVHQERATALYGVPTMFIAELEHPEFNDMDLSSLRTGIMAGSICPTEVMKQVIEKMHMSEVQI
AYGMTETSPVSTQTGADDPIAKRVSTVGRTQPHLESKVIDPATGRIMPRGEIGELCTRGYSVMLKYWNND
KATAETIDEAGWMHTGDLAIMDDEGYIQIVGRIKDMVIRGGENVYPKEVEEYLYTHPAISDVQVTGVPDK
KYGEELIAWVKLHPGAEDVTAEALREFCKGKITHFKIPRYFKFVDQFPMTVTGKIQKFKMREISIEELGL
KDD

4.



KEGG

Nitrincola iocasae: F5I99_10775

Help

Entry	F5I99_10775 CDS T06212
Name	(GenBank) acyl-CoA synthetase
KO	K20034 3-(methylthio)propionyl---CoA ligase [EC:6.2.1.44]
Organism	nik Nitrincola iocasae
Pathway	nik00920 Sulfur metabolism nik01100 Metabolic pathways
Brite	KEGG Orthology (KO) [BR:nik00001] 09100 Metabolism 09102 Energy metabolism 00920 Sulfur metabolism F5I99_10775 Enzymes [BR:nik01000] 6. Ligases 6.2 Forming carbon-sulfur bonds 6.2.1 Acid-thiol ligases 6.2.1.44 3-(methylthio)propionyl---CoA ligase F5I99_10775 BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: AMP-binding AMP-binding_C GFRP Motif
Other DBs	NCBI-ProteinID: QEW06955 UniProt: A0A5J6LED8
Position	complement(2338348..2339988) Genome browser
AA seq	546 aa AA seq DB search MTKSNHNPYTLGLDQNPANFTPLSPLSFISRAADVFPERTAVVHSGVRRNWRETFLRCRK LASALVKRGIGKGDTVAVMSPNLPEVFEAHFGVPATGAVLNALNIRLDAEAIAFILQHGE AKVVITDREFSEVIQRAVRMMAHKPLVIDIDDPSYEGGELIGEMDYEAFIAEGDDDYPWE LPENEWDAITLNYTSGTTGDPKGVVYHHRGAYLNAMSNALSWDMGHHPVYLWTLPMFHCN GWCFPWAIAATVGVSVCLRHVRADCIYELIKTEKVNHFCGAPIVLNMLNNAADEMKAGIN HEVKVMTAGAAPPASVIEGMEQMGFKVTHVYGLTETYGPSVVCAWHDEWDEKNANEKARL KSRQGVRAPMLEGLMVANPETMVPVAQDGKTLGEIFMRGNLVMKGYLKNPTTTDASFRGG WFHSGDLAVWHADGYIEIKDRSKDIIISGGENISSIEVEDVLYRHPAVMEAAVVAMSDDK WGEVPCAFVTLKTAAIGDVTEADIIQFCRDNMARFKIPKKIIFCDLPKTSTGKIQKFVLR DQASLG

All links

Ontology (2)
KEGG BRITE (2)
Pathway (2)
KEGG PATHWAY (2)
Chemical substance (6)
KEGG COMPOUND (6)
Chemical reaction (2)
KEGG ENZYME (1)
KEGG REACTION (1)
Genome (1)
KEGG GENOME (1)
Gene (2)
KEGG ORTHOLOGY (1)
NCBI-PROTEINID (1)
Protein sequence (1)
UniProt (1)
Protein domain (3)
Pfam (3)
All databases (19)

Download RDF

>WP_151704733.1 acyl-CoA synthetase [Nitrincola alkalilacustris]
MNKHNLYALGLEQNQANYTPLTPLSFITRAADVYPNRTAVVHQSLRRTWAETYRRCCQLASALHKRGIGK
GDTVAVMAPNLPETFEAHFAVPMAGAVLNALNIRLDAEAISFILQHGEAKVVIVDREFSEVIQRAVRMMA
NKPLVIDIDDPFYEGGELIGEKDYEALLSEGDPDFAWQMPEDEWEAITLNYTSGTTGDPKGVVYHHRGAY
LNAMSNIVSWDMGHHPVYLWTLPMFHCNGWCFPWAIAASAGVSVCLRHVRADHIYDAIKLEKVNHFCGAP
IVLNMLNNAPDEMKTGIQHEVKVMTAGAAPPASVIEGMEQMGFRVTHVYGLTETYGPSVVCAWHDEWDEK
TATEKARLKSRQGVRAPMLEGLMVANPETLEPVPQDGKTIGEIFMRGNLVMKGYLKNPGTTDKSFAGGWF
HSGDLAVWHADGYVEIKDRSKDIIISGGENISSIEVEDVLYRHPAVLEAAVVAMHDEKWGEVPCAFIALK
PSAGNVSEETIINFCRDNMARFKAPKKVIFCDLPKTSTGKIQKFVLRDQANSKS

NCBI Protein Blast for Idiomarina loihiensis L2TR (taxid:283942)- pCL gene
NCBI Blast:Protein Sequence (nih.gov)

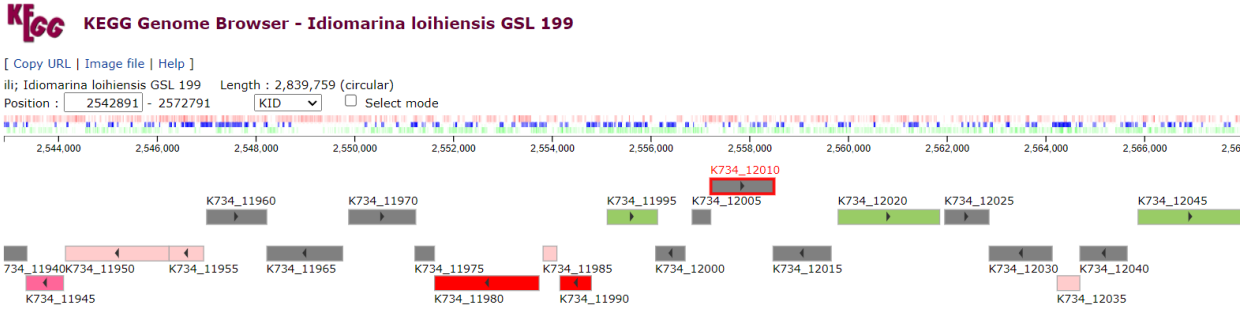
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☐	AMP-binding protein [Idiomarina]	Idiomarina	189	189	87%	3e-57	34.08%	422	WP_011235612.1
☐	AMP-binding protein [Idiomarina loihiensis]	Idiomarina loihiensis	64.3	64.3	75%	2e-12	24.85%	449	WP_011233393.1
☐	acetate--CoA ligase [Idiomarina]	Idiomarina	37.4	37.4	24%	0.001	22.55%	646	WP_011235235.1
☐	acetate--CoA ligase [Idiomarina loihiensis]	Idiomarina loihiensis	34.7	34.7	26%	0.008	27.27%	623	WP_011234314.1

1.

>WP_011235612.1 MULTISPECIES: AMP-binding protein [Idiomarina]
MTDSLQQHLIVTVIGDLIADELARMRPAESEYWKRRQWHEDDTLVAKNKSTKDNGEDDVVVDSLERLALA
GRVVQFFHMGDSGVEDYLLRRNSLAEWAEVVLKSRQVHTQNLTVTTSGSTGQPKACEHSWSALVEEVREF
VRIFDNDYELSPVRIVALVPSHHIYGFLFTVLLPHLVDAPVLRGFKAYSHVRNGGLRAGDAVVGFPELLT
QLSSEMPPLPPGVLFISSAGPCPASTVHQLYAIGAARAVEIYGSSETAGMAYRSKPENNYRLLSRWRKNT
ENHQQLIDRQTKVIYEIPDNTQWHTEDEFQITGRVDKAVSIRGINVFPAHIAKCLRQHPAVADATVRPMR
SDEGYGLKAFIVLQENISETVTEQSVQTWLSDNLCAAEIPERISFGEQLPINSMGKAQDWSIDNSPTGKP
LN

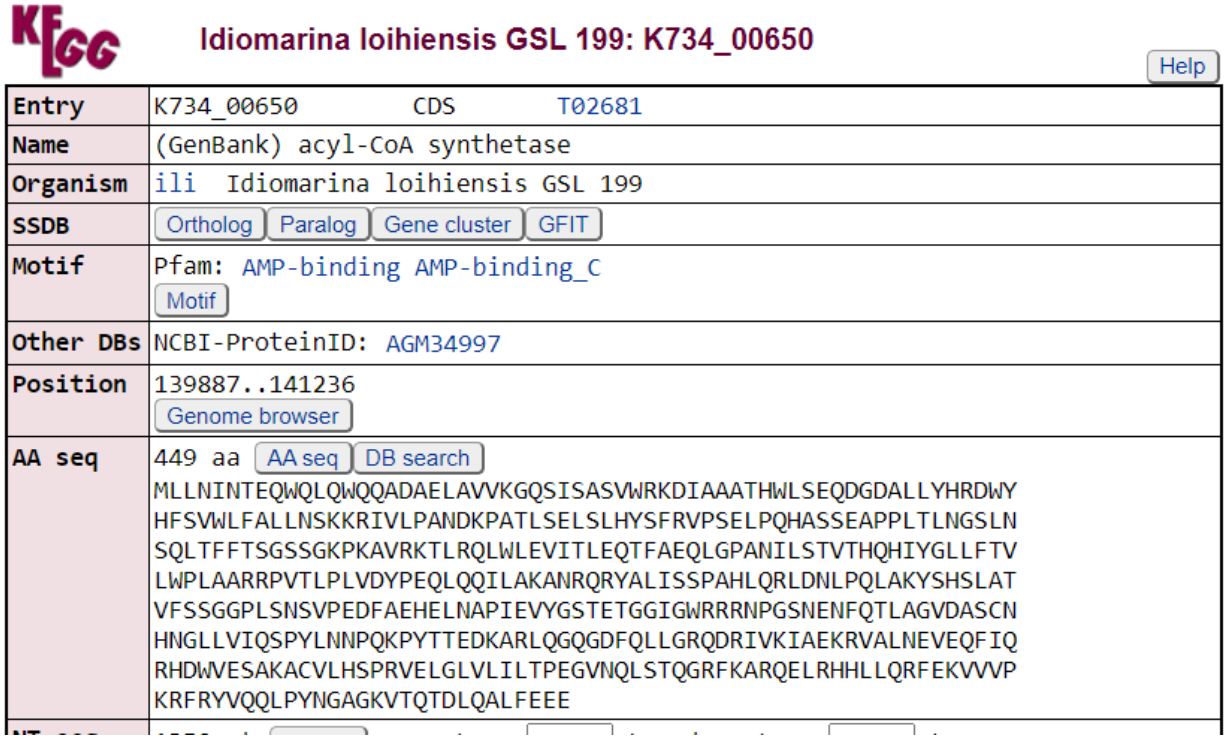
KEGG **Idiomarina loihiensis GSL 199: K734_12010** Help

Entry	K734_12010 CDS T02681
Name	(GenBank) 4-coumarate--CoA ligase
Organism	ili Idiomarina loihiensis GSL 199
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: AMP-binding AMP-binding_C Motif
Other DBs	NCBI-ProteinID: AGM37261
Position	2557206..2558474 Genome browser
AA seq	422 aa AA seq DB search MTDSLQQHLIVTVIGDLIADELARMRPAESEYWKRRQWHEDDTLVAKNKSTKDNGEDDVV VDSLERLALAGRVVQFFHMGDSGVEDYLLRRNSLAEWAEVVLKSRQVHTQNLTVTTSGST GQPKACEHSWSALVEEVREFVRIFDNDYELSPVRIVALVPSHHIYGFLFTVLLPHLVDAP VLRGFKAYSHVRNGGLRAGDAVVGFPELLTQLSSEMPPLPPGVLFISSAGPCPASTVHQL YAIGAARAVEIYGSSETAGMAYRSKPENNYRLLSRWRKNTENHQQLIDRQTKVIYEIPDN TQWHTEDEFQITGRVDKAVSIRGINVFPAHIAKCLRQHPAVADATVRPMRSDEGYGLKAF IVLQENISETVTEQSVQTWLSDNLCAAEIPERISFGEQLPINSMGKAQDWSIDNSPTGKP LN



2.

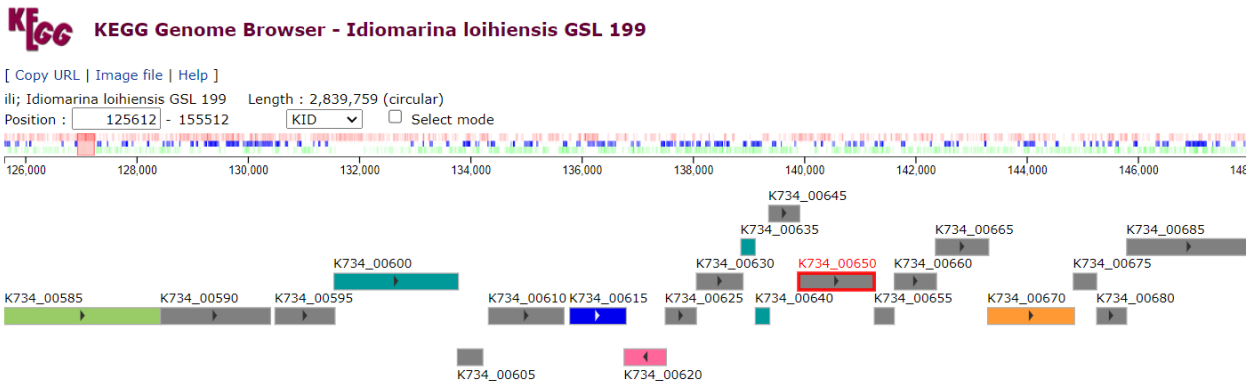
>WP_011233393.1 AMP-binding protein [Idiomarina loihiensis]
MLLNINTEQWQLQWQQADAELAVVKGQSISASVWRKDIAAATHWLSEQDGDALLYHRDWYHFSVWLFALL
NSKKRIVLPANDKPATLSELSLHYSFRVPSELPQHASSEAPPLTLNGSLNSQLTFFTSGSSGKPKAVRKT
LRQLWLEVITLEQTFAEQLGPANILSTVTHQHIYGLLFTVLWPLAARRPVTLPLVDYPEQLQQILAKANR
QRYALISSPAHLQRLDNLPQLAKYSHSLATVFSSGGPLSNSVPEDFAEHELNAPIEVYGSTETGGIGWRR
RNPGSNENFQTLAGVDASCNHNGLLVIQSPYLNNPQKPYTTEDKARLQGQGDFQLLGRQDRIVKIAEKRV
ALNEVEQFIQRHDWVESAKACVLHSPRVELGLVLILTPEGVNQLSTQGRFKARQELRHHLLQRFEKVVVP
KRFRYVQQLPYNGAGKVTQTDLQALFEEE



Idiomarina loihiensis GSL 199: K734_00650

Help

Entry	K734_00650 CDS T02681
Name	(GenBank) acyl-CoA synthetase
Organism	ili Idiomarina loihiensis GSL 199
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: AMP-binding AMP-binding_C Motif
Other DBs	NCBI-ProteinID: AGM34997
Position	139887..141236 Genome browser
AA seq	449 aa AA seq DB search MLLNINTEQWQLQWQQADAELAVVKGQSISASVWRKDIAAATHWLSEQDGDALLYHRDWY HFSVWLFALLNSKKRIVLPANDKPATLSELSLHYSFRVPSELPQHASSEAPPLTLNGSLN SQLTFFTSGSSGKPKAVRKTLRQLWLEVITLEQTFAEQLGPANILSTVTHQHIYGLLFTV LWPLAARRPVTLPLVDYPEQLQQILAKANRQRYALISSPAHLQRLDNLPQLAKYSHSLAT VFSSGGPLSNSVPEDFAEHELNAPIEVYGSTETGGIGWRRRNPGSNENFQTLAGVDASCN HNGLLVIQSPYLNNPQKPYTTEDKARLQGQGDFQLLGRQDRIVKIAEKRVALNEVEQFIQ RHDWVESAKACVLHSPRVELGLVLILTPEGVNQLSTQGRFKARQELRHHLLQRFEKVVVP KRFRYVQQLPYNGAGKVTQTDLQALFEEE



3.

>WP_011235235.1 MULTISPECIES: acetate--CoA ligase [Idiomarina]
MSQLYSVPEHIKAKAKIDNDGYKKLYQQSVDDPEGFWSEHGQRITWFTPYTKVKNTSFEPGKVSVKWYED
GTLNACYNCVDRHLADKADKTAIIWEGDDPSVDKHITYRELHKEVSRFANGLKKLGISKGDRVAIYMPMV
PEAAYAMLACARIGAVHSVIFGGFSPNAIADRINDCQAKAIITADEGRRGGGTIGLKANVDKAIADDACP
TLEHSIVCRVTEGDVDWTEGRDVWWHELVGNVSDECQAEVMNAEDPLFILYTSGSTGKPKGVVHTTGGYM
VYASMTHEYVFDYHEDEVYWCAADVGWITGHSYIVYGPLANGATTLMFEGVPTYPGVGRIGEIVDKHKVN
ILYTAPTAIRALMAKGDEAAKTSTRESLRILGSVGEPINPEAWEWYHRAIGNGKCPIMDTWWQTENGGIL
ITPLPGATDLKPGSATRPFFGIQPALVDSEGNIQEGEAEGGLVIKDSWPGQMRTLWGDHERFEQTYFSTF
KGMYFSGDGARRDADGYYWITGRMDDVLNVSGHRLGTAEIESALVSHKAVAEAAVVGYPHDLKGQGIYVY
VTPVEGVEVTDELTKEVRNWVRSELSPIATPDLIHWTTGLPKTRSGKIMRRILRKIAANEYENLGDTSTL
ADPSVVDSLIENRMNK

Idiomarina loihiensis GSL 199: K734_10105

Help

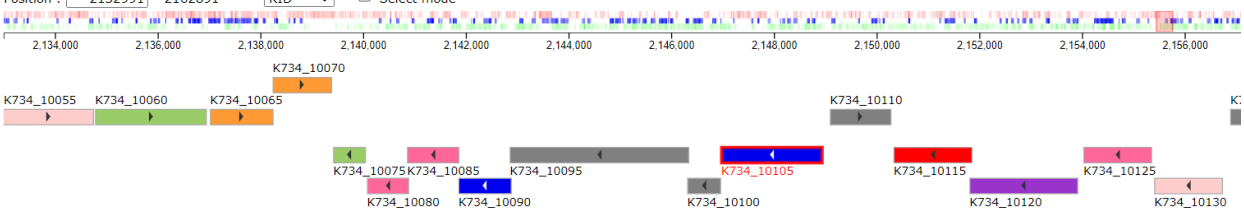
Entry	K734_10105 CDS T02681
Name	(GenBank) acetyl-CoA synthetase
KO	K01895 acetyl-CoA synthetase [EC:6.2.1.1]
Organism	ili Idiomarina loihiensis GSL 199
Pathway	ili00010 Glycolysis / Gluconeogenesis ili00620 Pyruvate metabolism ili00630 Glyoxylate and dicarboxylate metabolism ili00640 Propanoate metabolism ili00680 Methane metabolism ili01100 Metabolic pathways ili01110 Biosynthesis of secondary metabolites ili01120 Microbial metabolism in diverse environments ili01200 Carbon metabolism
Brite	KEGG Orthology (KO) [BR:ili00001] 09100 Metabolism 09101 Carbohydrate metabolism 00010 Glycolysis / Gluconeogenesis K734_10105 00620 Pyruvate metabolism K734_10105 00630 Glyoxylate and dicarboxylate metabolism K734_10105 00640 Propanoate metabolism K734_10105 09102 Energy metabolism 00680 Methane metabolism K734_10105 09180 Brite Hierarchies 09181 Protein families: metabolism 01004 Lipid biosynthesis proteins [BR:ili01004] K734_10105 Enzymes [BR:ili01000] 6. Ligases 6.2 Forming carbon-sulfur bonds 6.2.1 Acid-thiol ligases 6.2.1.1 acetate---CoA ligase K734_10105

KEGG Genome Browser - Idiomarina loihiensis GSL 199

[Copy URL | Image file | Help]

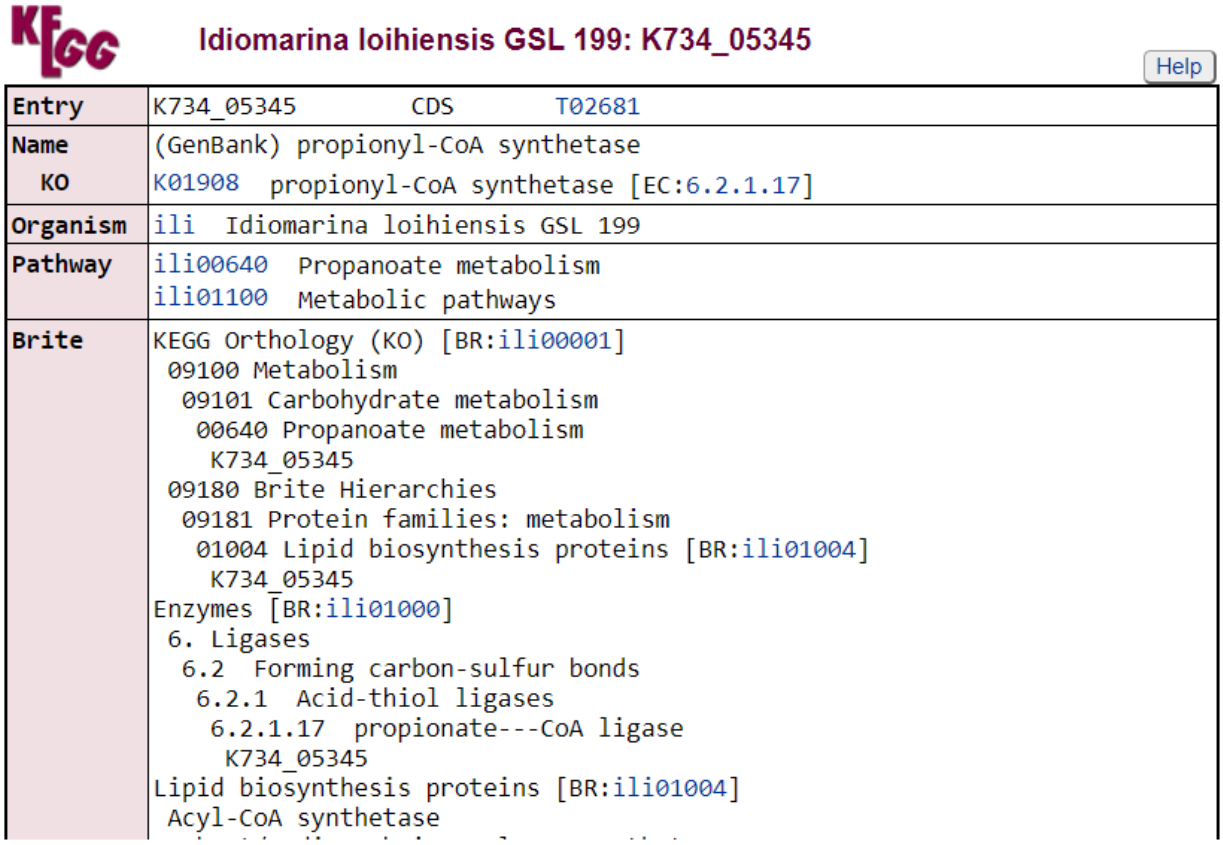
ili; Idiomarina loihiensis GSL 199 Length : 2,839,759 (circular)

Position : 2132991 - 2162891 KID Select mode



4.

```
>WP_011234314.1 acetate--CoA ligase [Idiomarina loihiensis]
MQAYKNVYSEAQQDPEGFWLEQARRLNWFKTPKVAAQVKKNGLADWFPDGEVNISYLALDAQIEAGRGEQ
VALYYDSPVTQSKANYTFNQLREQVACFAHVLKQQGVEKGDRVIIYLPMIPQAAIAMLACARIGAIHSVV
FGGFAAHELAVRIDDAAPKLIITASCGIEGTKVLPYKPIVDKALAEATYKPRTIVYQREQCQAEMQEGQD
TDWEQALTTASPLEAIPLPATHPLYILYTSGTTGKPKGVVRDHGGYAVALNFSMNYVYGLEPGEVFFTAS
DVGWVVGHSYIVYGPLLFGCSSVLYEGKPVNTPDAGAFWRIVQDYNVSAIFSAPTAFRAIKKEDPEGEFI
QRYDLSSLKRIYMAGERLDPATYEWTTELTDLPVYDHWWQTESGWPMCANPVGIGAVKVKPGSSTFPVPG
YEISVLDPLGKGLGSKEEGAICVRLPLPPGCLTTVWGDEQRLHSSYLNAFPGYYCSGDGGYIDEQGYVFI
MGRTDDVINVAGHRLSTGEMEEILAAHKDVAECAVVAQPDELKGELPVGFVILKNSSTIAHEQLQQELIQ
SIRAEIGAIACLKTLHIVSRLPKTRSGKILRRLIRQVLQGDDVAVPSTIDDPSSIDAIKETMR
```

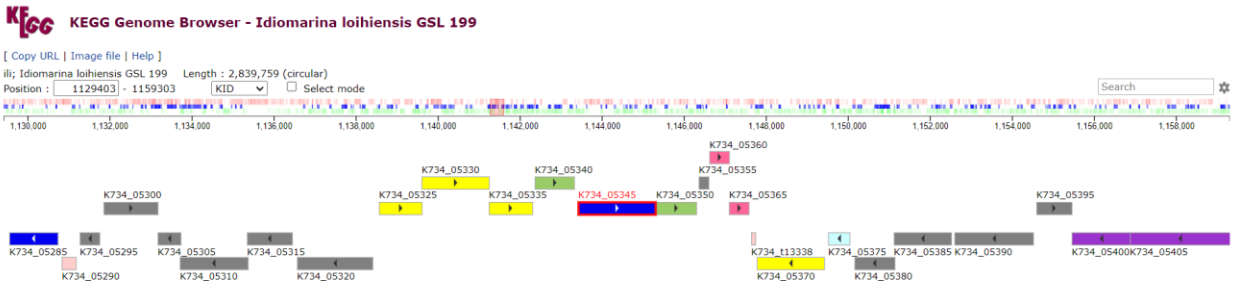


KEGG

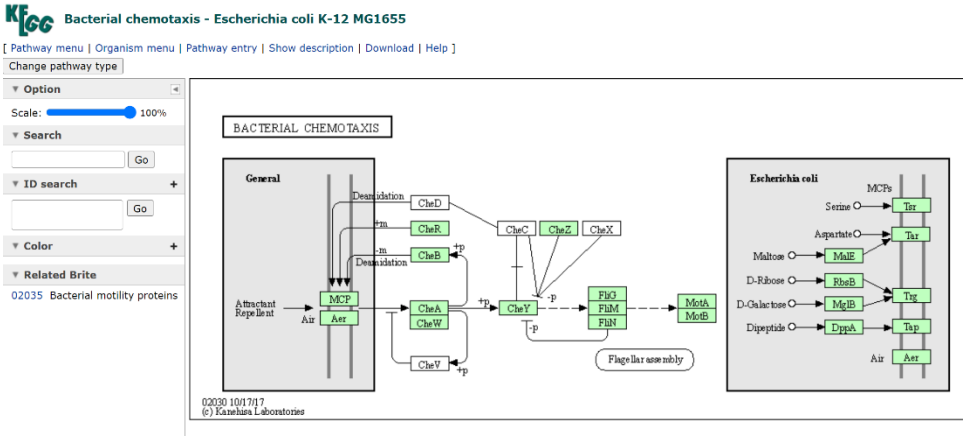
Idiomarina loihiensis GSL 199: K734_05345

Help

Entry	K734_05345 CDS T02681
Name	(GenBank) propionyl-CoA synthetase
KO	K01908 propionyl-CoA synthetase [EC:6.2.1.17]
Organism	ili Idiomarina loihiensis GSL 199
Pathway	ili00640 Propanoate metabolism ili01100 Metabolic pathways
Brite	KEGG Orthology (KO) [BR:ili00001] 09100 Metabolism 09101 Carbohydrate metabolism 00640 Propanoate metabolism K734_05345 09180 Brite Hierarchies 09181 Protein families: metabolism 01004 Lipid biosynthesis proteins [BR:ili01004] K734_05345 Enzymes [BR:ili01000] 6. Ligases 6.2 Forming carbon-sulfur bonds 6.2.1 Acid-thiol ligases 6.2.1.17 propionate---CoA ligase K734_05345 Lipid biosynthesis proteins [BR:ili01004] Acyl-CoA synthetase

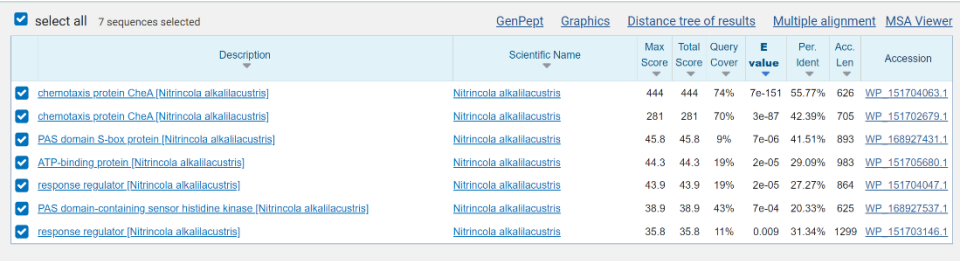


Bioinformatics analysis of *N. alkalilacustirs sp* for Chemotaxis proteins using *E. coli*



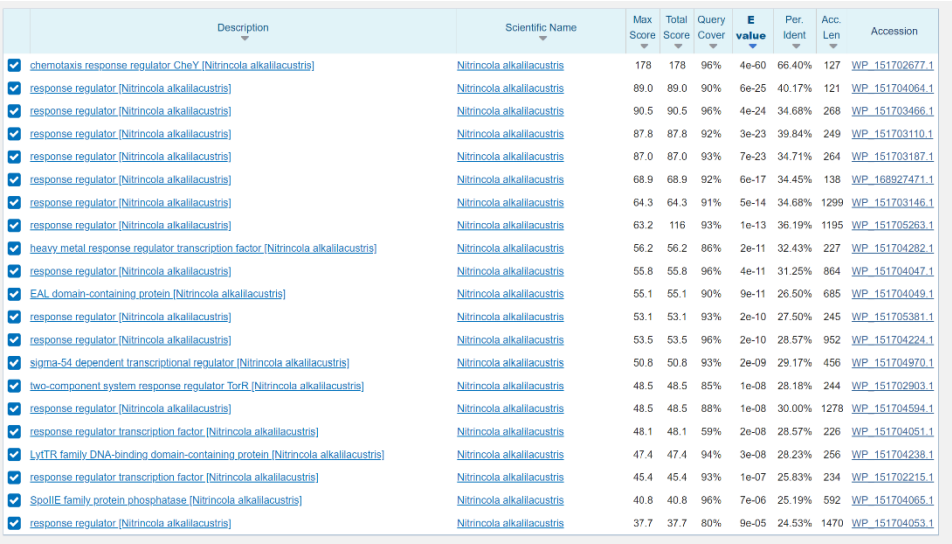
CheA

MQEQLDAYKQSQEPDAASFDYICQALRQLALEAKGETPSAVTRLSVVAKSEPQDEQSRSQ
SPRRIILSRLKAGEVDLLEEELGHLTTLTDVVKGADSLSAILPGDIAEDDITAVLCFVIE
ADQITFETVEVSPKISTPPVLKLAAEQAPTGRVEREKTTRSNESTSIRVAVEKVDQLINL
VGELVITQSMLAQRSSELDPVNHGDLITSMGQLQRNARDLQESVMSIRMMPMEYVFSRYP
RLVRDLAGKLGKQVELTLVGSSTELDKSLIERIIDPLTHLVRNSLDHGIELPEKRLAAGK
NSVGNLILSAEHQGGNICIEVTDDGAGLNRERILAKAASQGLTVSENMSDDEVAMLIFAP
GFSTAEQVTDVSGRGVGMDVVKRNIQKMGGHVEIQSKQGTGTTIRILLPLTLAILDGMSV
RVADEVFILPLNAVMESLQPREADLHPLAGGERVLEVRGEYLPIVELWKVFNVAGAKTEA
TQGIVVILQSGGRRYALLVDQLIGQHQVVVKNLESNYRKVPGISAATILGDGSVALIVDV
SALQAINREQRMANTAA



CheY

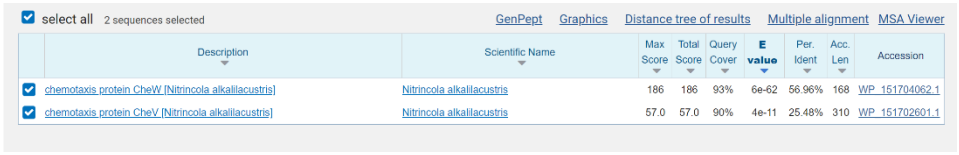
MADKELKFLVVDDFSTMRRIVRNLLKELGFNNVEEAEDGVDALNKLQAGGYGFVISDWNM
PNMDGLELLKTIRADGAMSALPVLMVTAEAKKENIIAAAQAGASGYVVKPFTAATLEEKL
NKIFEKLGM



Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
chemotaxis response regulator CheY [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	178	178	96%	4e-60	66.40%	127	WP_151702677.1
response regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	89.0	89.0	90%	6e-25	40.17%	121	WP_151704064.1
response regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	90.5	90.5	96%	4e-24	34.68%	268	WP_151703466.1
response regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	87.8	87.8	92%	3e-23	39.84%	249	WP_151703110.1
response regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	87.0	87.0	93%	7e-23	34.71%	264	WP_151703187.1
response regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	68.9	68.9	92%	6e-17	34.45%	138	WP_168927471.1
response regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	64.3	64.3	91%	5e-14	34.68%	1299	WP_151703146.1
response regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	63.2	116	93%	1e-13	36.19%	1195	WP_151705263.1
heavy metal response regulator transcription factor [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	56.2	56.2	86%	2e-11	32.43%	227	WP_151704282.1
response regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	55.8	55.8	96%	4e-11	31.25%	864	WP_151704047.1
EAL domain-containing protein [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	55.1	55.1	90%	9e-11	26.50%	685	WP_151704049.1
response regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	53.1	53.1	93%	2e-10	27.50%	245	WP_151705381.1
response regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	53.5	53.5	96%	2e-10	28.57%	952	WP_151704224.1
sigma-54 dependent transcriptional regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	50.8	50.8	93%	2e-09	29.17%	456	WP_151704970.1
two-component system response regulator TorR [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	48.5	48.5	85%	1e-08	28.18%	244	WP_151702903.1
response regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	48.5	48.5	88%	1e-08	30.00%	1278	WP_151704594.1
response regulator transcription factor [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	48.1	48.1	59%	2e-08	28.57%	226	WP_151704051.1
LytTR family DNA-binding domain-containing protein [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	47.4	47.4	94%	3e-08	28.23%	256	WP_151704238.1
response regulator transcription factor [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	45.4	45.4	93%	1e-07	25.83%	234	WP_151702215.1
SpoIIE family protein phosphatase [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	40.8	40.8	96%	7e-06	25.19%	592	WP_151704065.1
response regulator [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	37.7	37.7	80%	9e-05	24.53%	1470	WP_151704053.1

CheW

MTGMTNVTKLASEPSGQEFLVFTLGDEEYGIDILKVQEIRGYDQVTRIANTPAFIKGVTN
LRGVIVPIVDLRIKFSQVDVDYNDNTVVIVLNLGQRVVGIVVDGVSDVLSLTAEQIRPAP
EFAVTLSTEYLTGLGALGDRMLILVNIEKLLNSEEMALLDSAASEVA



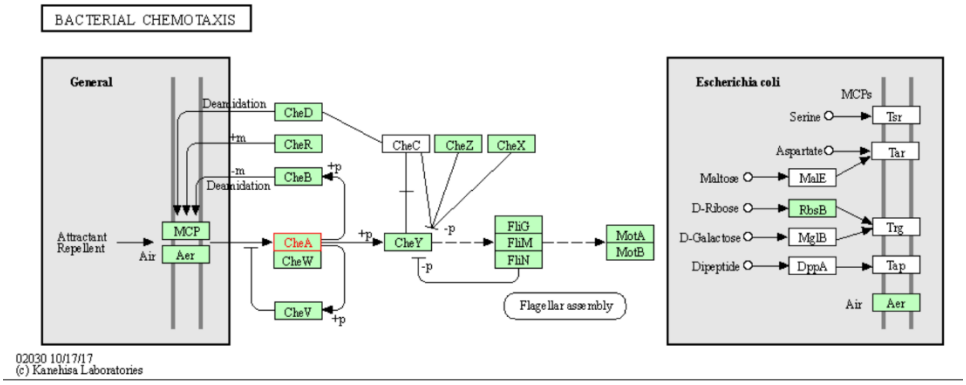
select all 2 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
chemotaxis protein CheW [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	186	186	93%	6e-62	56.96%	168	WP_151704062.1
chemotaxis protein CheV [Nitrincola alkalilacustris]	Nitrincola alkalilacustris	57.0	57.0	90%	4e-11	25.48%	310	WP_151702601.1

Using the top hits for each, performed a KEGG BLAST

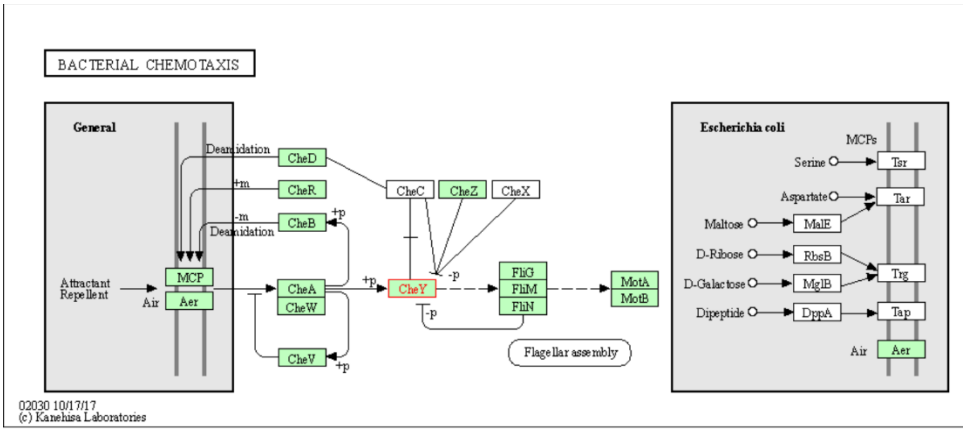
CheA- top hit amino acid sequence

>WP_151704063.1 chemotaxis protein CheA [Nitrincola alkalilacustris]
MSIDMSQFHATFFEEGHEHLDEMERVLMAIDLSAPQAEDLDTIFRAAHSIKGGSGIFGFDALTTVTHVME
NILEQVRSGSMSLTQPLVDCLLGTSDHLRAILESYRNQSEINWTEIGDITHSLEVWLTAEGPAPQVEQGY
GFFEPLQQPGRDDSYGFFDQDESEQAGSDDGFGFFEPLPATAELPEVPVAGAAPTDVAATSSPAQPSNTT
ETAPAKPVVRPDSSQPAQRSEREASSIRVDIVKVDQLVNMVGELVITQSMLNLIGSEVQGPETERLHAAL
VELERNTRELQESIMSIRMLPISFVFNRFPRVVRDLSSKLHKRVELVVEGGQTEIDKGLIERLADPLTHL
VRNSLDHGIEQPEERLRKGKPECGRLTLRAEQKGGNILISVIDDGAGLNRERILQKAAEKSIELPEHPTD
SQVWQLIMAPGFSTAAEVTDVSGRGVGMDVVKRNIEHMGGRIEIESRTGEGAKFQIRLPLTLAILDGMSV
AVGEQHFIIPLVNIIESVQPSPEQIKQIKKQEMLELRDAYWPIVPLYQVMQVEGARTAISEAILVLIETS
KTRFALMVDDLIGQQQVVIKSLEQHYRRVPGIAGATIMGDGSVALILDAESLAQRVDESLMMKETV



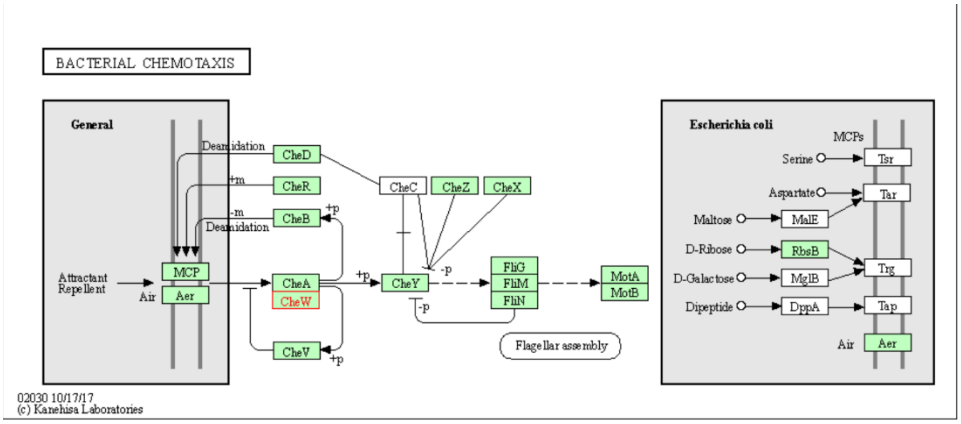
CheY- top hit amino acid sequence

>WP_151702677.1 chemotaxis response regulator CheY [Nitrincola alkalilacustris]
MKKDIKILVVDDFSTMRRIIKNLLRDLGFTNVDEADDGKTALPILQQGRIDFLITDWNMPGMTGIDLLKA
VRADPNLSHIPVLMVTAEAKKEQIIAAAQAGVNGYVIKPFTAVVLKEKIEKIFERIG



CheW- top hit amino acid sequence

>WP_151704062.1 chemotaxis protein CheW [Nitrincola alkalilacustris]
MSDNAGDTIKPESSTCEFLTFTLGDEHYGLDIMQVREIRGYEPVTKIANAPPFIKGVLNLRGDVVPIVDL
RIKFDVGRAEYNEFTIVIMLHIGDRIVGIVVDAVSDVINFQKQDIKPPPEFGVAFDSQYLFGLAPINNNM
IILLNIEKLISSQELGLFDASSAAQETQ


BACTERIAL CHEMOTAXIS
General
Deamidation
CheD
+m
CheR
-m
Deamidation
CheB
+p
CheC
CheZ
CheX
Attractant
Repellent
MCP
Air
Aer
CheA
CheW
+p
CheY
-p
-p
FliG
FliM
FliN
MotA
MotB
Flagellar assembly
CheV
+p
Escherichia coli
MCPs
Serine
Tsr
Aspartate
Tar
Maltose
MalE
D-Ribose
RbsB
D-Galactose
MglB
Trg
Dipeptide
DppA
Tap
Air
Aer
02030 10/17/17
(c) Kanehisa Laboratories