Annotation of the Moraxella catarrhalis Genome at Locus Tags MCR_RS00815 and MCR_RS00820

Sadnan Nawz, Elizabeth Pearce, Dillon Sullivan, Fernando Vega, Kira Mioducki & Dawn Weihrich Research Laboratory High School, Buffalo Public Schools (BPS#366) and the Western New York Genetics in Research and Health Care Partnership

Abstract

Two consecutive genes within the genome of Moraxella Catarrhalis were annotated and compared to the GenBank proposed gene product name. They were then assessed in terms of the general genomic information, amino acid sequence-based similarity data and structure-based evidence from the amino acid sequence. The GenBank proposed gene product name did not differ significantly from the proposed gene annotation for each of the genes in the group and as such, the genes appear to be correctly annotated by the database.

Gene Browser (reset)	
----------------------	--

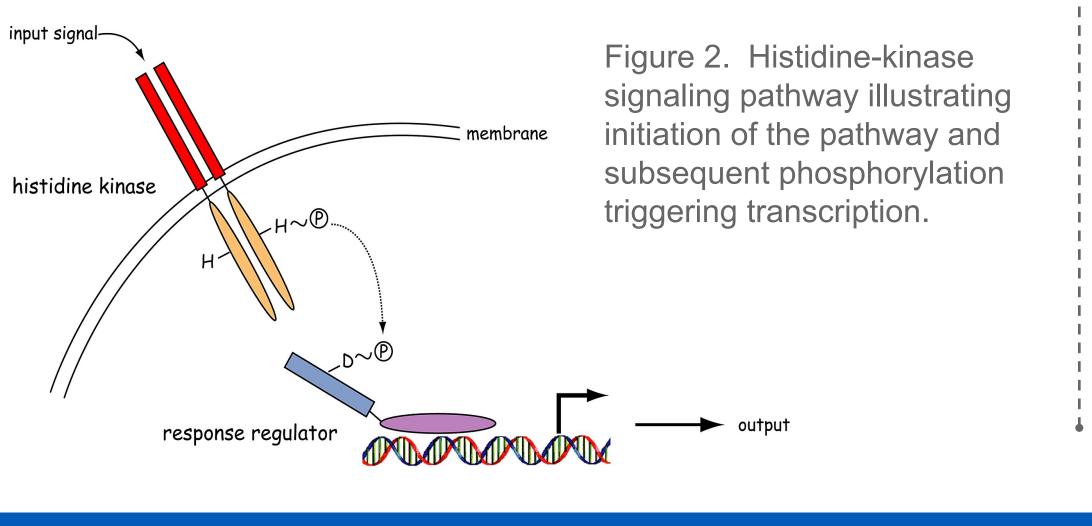
Gene browse	(reset)					
189k	190k	191k	192k	193k	194k	
MCR_RS00805	MCR_R500810	MCR_R500815		MCR_R500820	R_R500825	

Figure 1. The locus tags and relative position of the genes under investigation including MCR RS00815 and MCR RS00820.

Introduction

Moraxella catarrhalis is a round bacterium that typically occurs in the form of two joined cells (Brophy, 2017). It is a common culprit in pediatric cases of respiratory infections including bronchitis as well as otitis media (infections of the inner ear). Moraxella catarrhalis can be diagnosed based on a sample of fluid from the ear, the nasopharynx, or the roof of the mouth.

The group of proteins investigated under the proposed annotation make up a two-component regulatory system, receiving signals from outside of the cell and transmitting signals to regulate functions within the cell via transcription. This system is situated across the membrane and with specific protein domains. An external environmental signal serves as a stimulus that initiates the pathway. This stimulus causes a change in the formation of the histidine protein kinase, which in turn causes the transfer of phosphoryl groups from ATP to a conserved histidine residue. This phosopho-group is then moved to an aspartate residue on the response regulator which is the second component of the system/pathway. This will interact with the response regulator to allow for binding of the DNA in order to regulate the transcription of its target genes.



Methods

Modules of the GENI-ACT (<u>http://www.geni-act.org</u>) were used to complete *Moraxella catarrhalis* genome annotation. The modules are described below:

Modules	Activities	Questions Investigated
Basic Information	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of the gene and protein? Where is it located in the genome?
Sequence-Based Similarity	Blast, CDD, T-Coffee, WebLogo	How similar is the protein under investigation to other proteins in GenBank?
Structure-Based Similarity	TIGRfam, Pfam, PDB	What functional domains are present in the protein under investigation?
Final Annotation	Evaluate data from all modules	Has the gene been correctly called by the pipeline annotation?

Results

MCR_RS00815:

The proposed initial gene product was a two-component sensor histidine kinase. This hypothesis was supported by both h sequence and structure-based evidence. The gene annotation is between 191194 and 192759 coordinates with a length of 521 amino acids and 1156 nucleotides. Significant BLAST hits revealed alignment to proteins in *Pseudomonas mendocina* and Psychrobacter sp. (see Figure 3). The top two BLAST hits revealed gene products with the names: sensor histidine kinase TmoS and HAMP domain-containing histidine kinase. The scores of the BLAST were 77.8 bits and 572 bits with e-values \ of 8e-14 and 0.0 respectively. Significant COG hits revealed signal transduction histidine kinase. Pfam revealed three specific domains, all aligned to the proposed annotation, including the HATPase_c (PF02518), HisKA (PF00512), and a HAMP domain (PF00672), which are necessary components of a histidine kinase regulatory pathway. T-coffee and WebLogo results revealed poor conservation throughout the midsection, but well conserved portions between residues 193 to 383. A Protein data bank (PDB) search further revealed alignment to the proposed gene annotation matching 4CTI: a histidine kinase revealing a HAMP domain as evidenced by Figure 4.

HAMP domain-containing histidine kinase [Ps	ychrobacter sp. 1501(2011)]
Sequence ID: WP_007395864.1 Length: 584 Numb	per of Matches: 1
▶ <u>See 1 more title(s)</u>	

tange 1: 6 to 572 GenPept Graphics (dentities Positives 572 bits(1473) 0.0 Compositional matrix adjust. 304/569(53%) KFGFRSISARLFISVFLALLTFTIAFVLLSQFAHNNSDYTRNRAIASQIMNQIEPFLAEA +F F S+ ++LF SV L L+ F A V L HNNS R +A QI++QI+PF+ E RFAFTSVFSKLFTSVVLTLVLFAAAMVTLMNLVHNNSADARWEVLAKQIVSQIDPFVDEL HTLSAKNNLLQARFSLVVIKKSFDIFDESLNAKIGLYDSQGRLMLQTENTDLPETLVPNP 127 + N LLQARF L VIKKSFD+FDESL AK+GLYD G L++QT+++DLP+TL +P DNATTMNELLQARFMLAVIKKSFDVFDESLQAKMGLYDRNGHLLIQTDDSDLPKTLPDDP 125 PWISQI----FAPSPTHIVVNSTLGYSLWYENRILSPERPLMAWFNLFSGTVLLFTIMSA 1 +IS++ F+ ++ S GY+L YE R L A N+F+GT+ L IMS SFISRMLPAFFSTDQVNVQAMSDRGYTLLYEPRNPPKRSELWAVLNIFTGTIALIAIMSF 1 VLWWISHSITWRINQLSRQMSRLGEGDFSVRVSEEGNDEIAVLAHGFNQSAQKIEQLIDA VLWWI+HS+TWRI+Q+S+QM++LGEGDFSVRV +G DEIA LA GFNQ+AQKIEQLIDA VLWWIAHSMTWRIDQMSKQMAKLGEGDFSVRVPAKGTDEIASLARGFNQAAQKIEQLIDA SLLLAHASHEFRTPITRIRLQIEMMDMLTSQLDETTKAKFDKRAAAVNRDLTGLNDLVE 303 +LLLAHASHE RTPITRIRLQIEMM+ML S++ E T+AKF KRA AVNRDLTGLNDLVE NLLLAHASHELRTPITRIRLQIEMMEMLASEMSEETQAKFHKRAQAVNRDLTGLNDLVE 305 LLVSRLDAGHALQATEQVDLYELIRSEVQHYPEATLIGESVSMMAQPKLLTHLVRNLL 363 LLVSRLDAGHA Q E+ D YEL+ E QHYPEATL G+S+ + QPKLL HL+RNLL LLVSRLDAGHAAQNNERFDFYELVNQECQHYPEATLYGKSLVLEGQPKLLIHLIRNLL 365

 NNAMIHGIPPVQVYLYHATNIQEATSIPQYLL
 395

 NNA+IHG PP++V+LY
 +EA
 P+
 LL

 NNALIHGEPPIEVHLYGVVAQEEALYPPKQLLDQIEAAKIEATRELNDSLEDQSTATKGL
 425

-----DSFCNI----GDGDILHSKSDEANAL-----DNKTPTD------ 424 DS ++ G+ ++ S +E ++L +NK D DSDTEADSLSSLEDQNGNTELATSNGEEQSSLTAGQVASKEQDKGNNKAKADGSKLNEIS 485 FLKRLTRPKDKEQKADQPNFAVLAFIDQGLGIPKDKRKDIFSPFVRLKQEKKGSGLGLSL F KRL RP K + +PNFAVLA IDQG GIP DKR++IFSPFVRL+Q+KKGSGLGLSL FTKRLRRP--KAEPIIEPNFAVLAVIDQGKGIPLDKREEIFSPFVRLQQQKKGSGLGLSL 485 VSQIVEAHHGTIITDTWLGHTRFLVILPL 513 V+QIVEAH+G I TDT+ G TRF+V+LP+ 544 VAQIVEAHNGRIRTDTYNGRTRFIVLLPV 572

Figure 3 (left). BLAST results from nr database revealed alignment to HAMP domains from *Psychrobacter sp.*

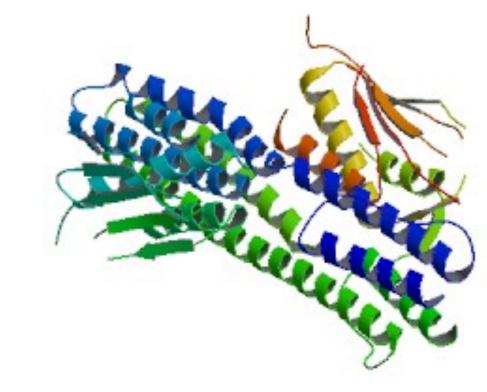
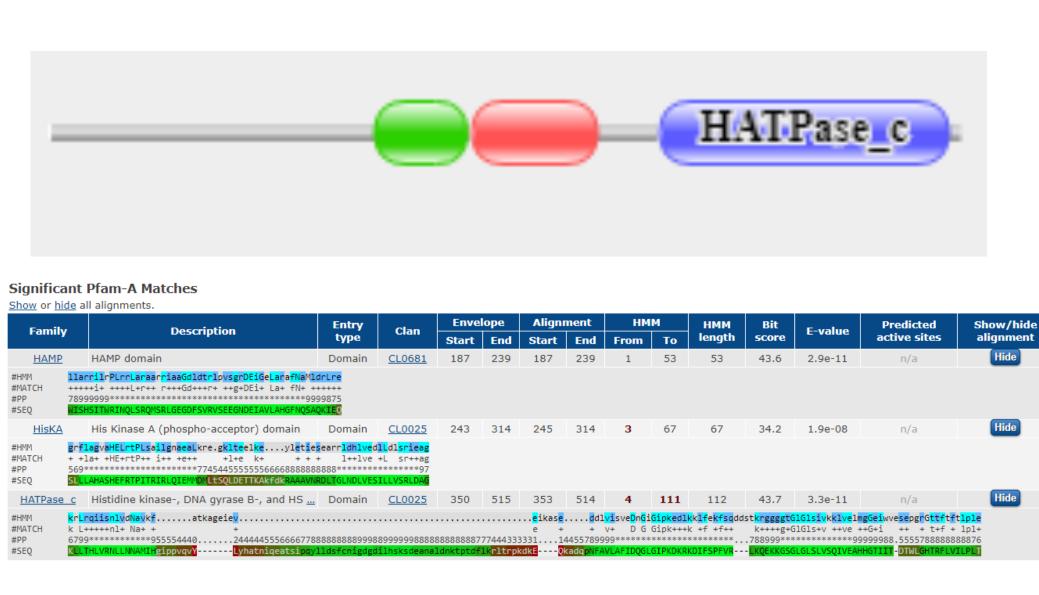


Figure 4 (above). MCR RS00820 PDB results aligned to 4CTI: Entity 1 containing Chain A, B, C, D Escherichia coli EnvZ histidine kinase catalytic part fused to Archaeoglobus fulgidus Af1503 HAMP domain.

Acknowledgments



Thank you to Dr. Stephen Koury, Dr. Rama Dey-Rao, Dr. Sandra Small & Jonathan E. Bard for your invaluable support and expertise.



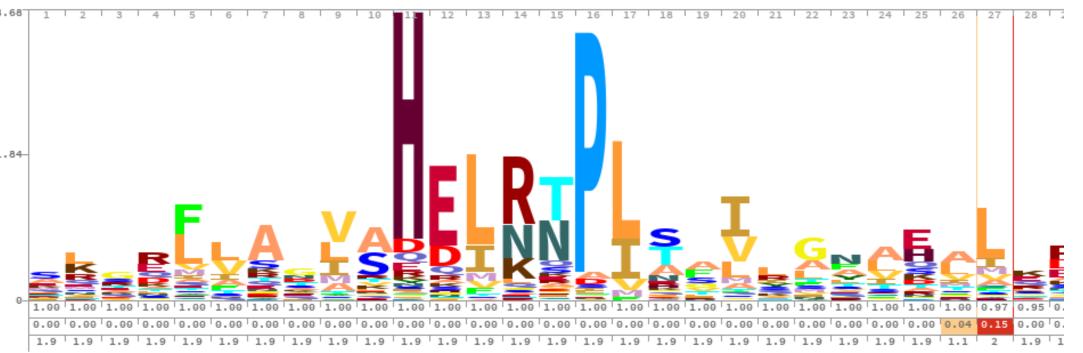


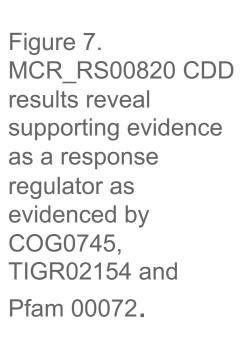
Figure 6. MCR_RS00815 HMM logo shows key residues His:11 and P:16. Histidine is integral in the function of proposed gene product, two-component sensor histidine kinase.

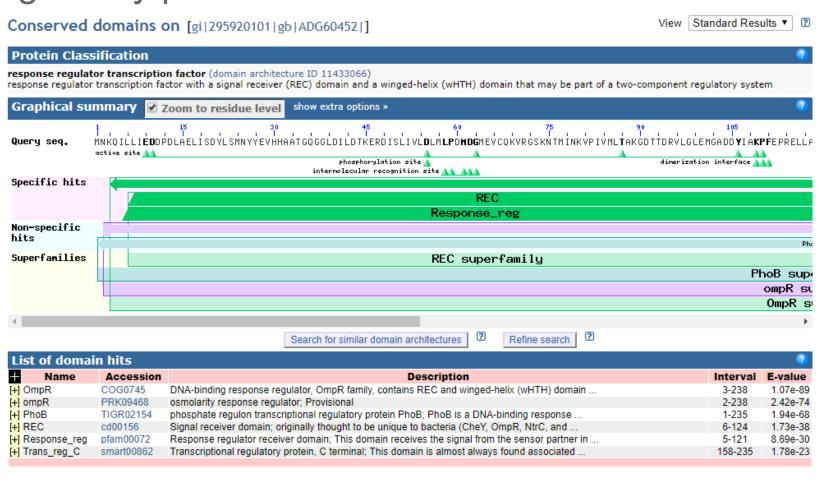
The computer pipeline proposed product of this gene was a DNAbinding response regulator. The gene annotation is between the second se 192,895 and 193,617 coordinates in the genome. It is 723 ¹ nucleotides in length and 240 amino acids. Significant BLAST hits were transcriptional regulatory protein Wa1R and response ¦ regulator transcription factor. The scores of the BLASTs were 194 bits and 458 bits, with e-values of 7e-61 and 2e-162 respectively. The CDD results showed COG0745, DNA binding response regulator DNA binding response regulators, OmpR would include completing GENIACT Module 4: Cellular family and contains REC and winged-helix. T-coffee and the second WebLogo results revealed that there are numerous areas of high conservation as shown by the wide large letters. Significant Pfam ; results also showed response regulator receiver domains and the second s transcriptional regulatory proteins.

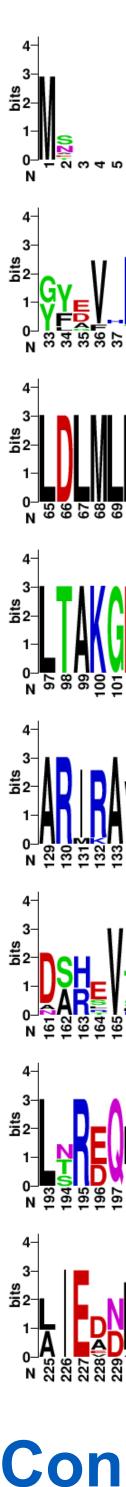
Figure 7

Figure 5. MCR RS00815 Pfam results reveal three domains aligned to function to the proposed gene annotation, a two component sensor histidine kinase.

MCR RS00820:







_ocalization Data to confirm transmembrane domains and

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the group and as such, the genes appear to be correctly annotated by the computer database. Suggestions for gathering further evidence to support proper annotation signal peptides to help determine location of receptor proteins within the cell and the two-component sensor histidine kinase within the cellular membrane.

Locus Ta MCR_RS MCR_RS



Wolanin, P. M., Thomason, P. A., & Stock, J. B. (2002). Histidine protein kinases: key signal transducers outside the animal kingdom. Genome biology, 3(10), REVIEWS3013. doi:10.1186/gb-2002-3-10-reviews3013

buffalo.edu



MPNDSSEN RADINGRA

Figure 8 (below). MCR RS00820 Protein data bank results aligned to 20QR: Entity 1 containing Chain A. The structure of the response regulator RegX3 from Mycobacterium tuberculosis.

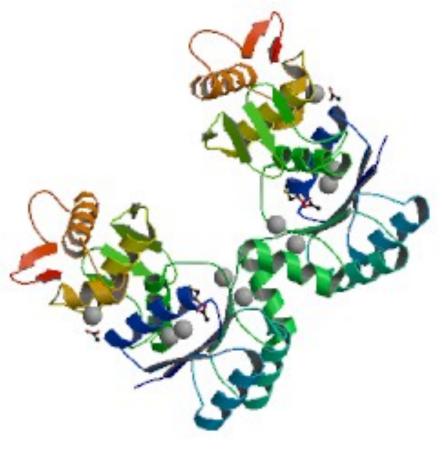


Figure 9 (left). MCR RS00820 VebLogo highly shows conserved residues throughout the aminoon terminal revealing poorer conservation. Well conserved histidine residues are present at 218, and 38.

Conclusion

g	Pipeline Annotation Product Name		Changes Proposed?
00815	Two-component sensor histidine kinase	Two-component sensor histidine kinase	No
00820	DNA-binding response regulator	DNA-binding response regulator	No

References

Supported by an NIH Science Education Partnership (SEPA) Award - R25GM129209

Jacobs School of Medicine and Biomedical Sciences