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

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
Supported by the National Institutes of Health

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.

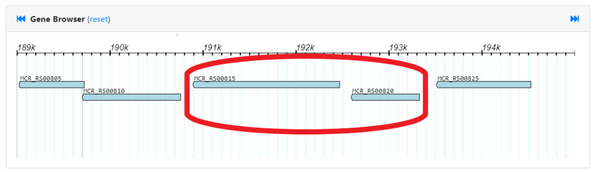

Figure 1. The locus tags and relative position of the genes under investigation including MCR_RS00815 and MCR_RS00820.

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 phospho-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.

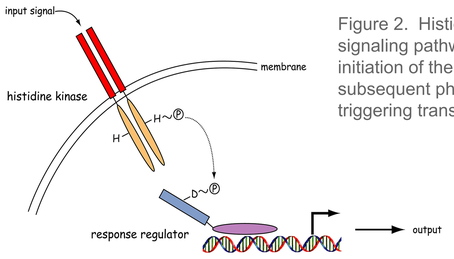

Figure 2. Histidine-kinase signaling pathway illustrating initiation of the pathway and subsequent phosphorylation triggering transcription.

Methods

Modules of the GENI-ACT (http://www.geni-act.org) 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 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.

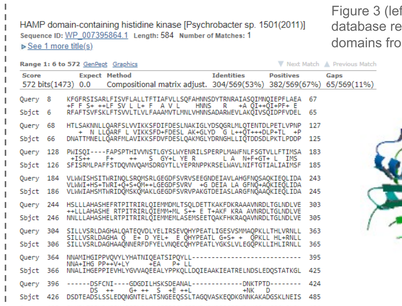

Figure 3 (left). BLAST results from nr database revealed alignment to HAMP domains from *Psychrobacter sp.*

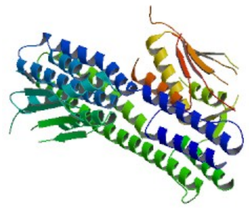

Figure 4 (above). MCR_RS00820 PDB results aligned to 4CTI: Entity 1 containing Chain A, B, C, D *Escherichia coli* EnvZ histidine kinase catalytic part fused to *Archaeoglobus fulgidus* Af1503 HAMP domain.

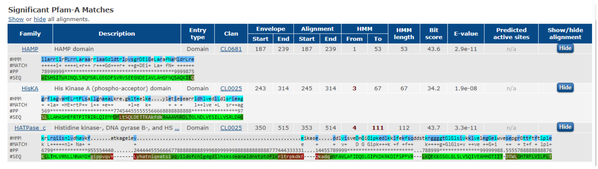

Figure 5. MCR_RS00815 Pfam results reveal three domains aligned to function to the proposed gene annotation, a two component sensor histidine kinase.

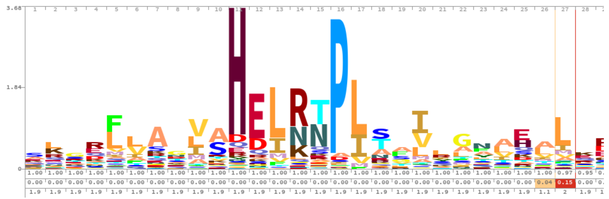

Figure 6. MCR_RS00815 HMM logo shows key residues His:11 and P:16. Histidine is integral in the function of proposed gene product, two-component sensor histidine kinase.

MCR_RS00820:

The computer pipeline proposed product of this gene was a DNA-binding response regulator. The gene annotation is between 192,895 and 193,617 coordinates in the genome. It is 723 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 family and contains REC and winged-helix. T-coffee and 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 transcriptional regulatory proteins.

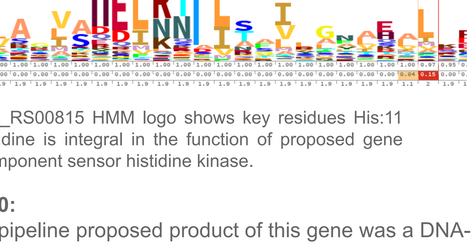

Figure 7. MCR_RS00820 CDD results reveal supporting evidence as a response regulator as evidenced by COG0745, TIGR02154 and Pfam 00072.

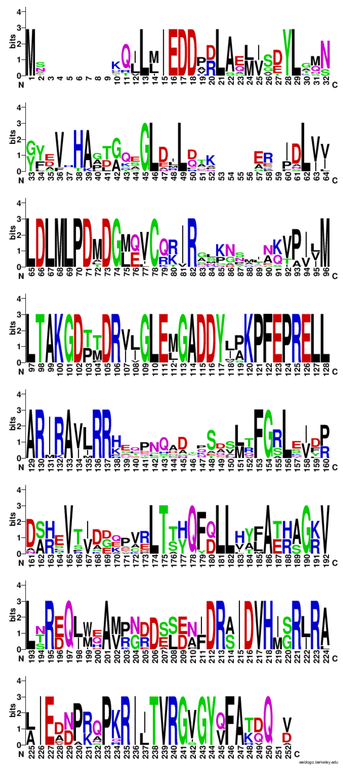

Figure 8 (below). MCR_RS00820 Protein data bank results aligned to 2OQR: Entity 1 containing Chain A. The structure of the response regulator RegX3 from *Mycobacterium tuberculosis.*

Figure 9 (left). MCR_RS00820 WebLogo shows highly conserved residues throughout with areas on the amino-terminal revealing poorer conservation. Well conserved histidine residues are present at 218, and 38.

Conclusion

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 would include completing GENIACT Module 4: Cellular Localization Data to confirm transmembrane domains and signal peptides to help determine location of receptor proteins within the cell and the two-component sensor histidine kinase within the cellular membrane.

Locus Tag	Pipeline Annotation Product Name	Proposed Annotation	Changes Proposed?
MCR_RS00815	Two-component sensor histidine kinase	Two-component sensor histidine kinase	No
MCR_RS00820	DNA-binding response regulator	DNA-binding response regulator	No

References

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

University at Buffalo The State University of New York

Acknowledgments

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

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

Jacobs School of Medicine and Biomedical Sciences

buffalo.edu