Annotation of the *Moraxella catarrhalis* Genome at Locus Tags MCR_RS00120 to MCR_RS00140

Fatma Abdulrahman, Muhammad Bari, Sabina Bayrahtarova, Mohaddeseh Golzar, Suzan Habtezion, Frida Ininahazwe, Bien Aimee Ishimwe, Sadnan Nawz, Bre'jon Sanders 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

A group of 5 consecutive genes from the microorganism *Moraxella catarrhalis* (MCR_RS00120 to MCR_RS00140) were annotated using the collaborative genome annotation website GENI-ACT (see Figure 1). The GenBank proposed gene product name for each gene was 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.

Introduction

Moraxella catarrhalis is a non-motile bacterium that spreads rapidly on its host. Colonies of *M. catarrhalis* may have a rough surface, are pinkish-brown in color, and opaque. This bacterium causes otitis media, a middle ear infection. It can also cause bronchitis, which inflames the lining of bronchial tubes that carry air to and from the lungs, making breathing difficult (Racaniello, 2019).

This bacterium can be transmitted through breathing and coughing from person to person in a hospital setting, and they occur more in people with the complications of the respiratory system such as COPD or infants with underdeveloped immune systems (Urgent Care, 2015). This bacteria adheres and colonizes the respiratory tract by infecting mucous membranes and epithelial cells.

The collection of our gene products indicated that they are all CRISPR-associated proteins. CRISPR, also known as Clustered Regularly Interspaced Short Palindromic Repeats, is an adaptive immune system in many types of bacteria. Bacteria use many CRISPR-associated genes (Cas) that bind foreign DNA from invading viruses and incorporate that DNA into the bacterial genome (Cooper et. al, 2018). This DNA then will allow the bacteria to transcribe it and make a guide RNA that will detect the foreign DNA in the future and degrade it before invading the bacteria (Hille et. al, 2016) (See Figure 2).

Figure 1. The locus tags and relative position of the genes under investigation in this research.

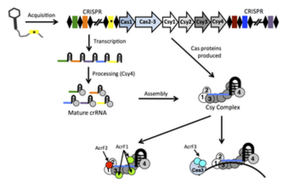


Figure 2. Image above shows how CRISPR operates in bacteria and depicts the proteins and their function within the cell.

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_RS00120:

The computer pipeline proposed product of this gene was a type I-F Crispr-associated helicase Cas 3. The gene location is between 32260 and 35673. It is 3414 nucleotides and 1137 amino acids in length. The BLAST results support the hypothesis and showed a gene product of Crispr-associated nuclease/helicase Cas3 subtype I-F/Y-Pest. Significant COG hits indicate alignment with a Crispr/Cas system-associated endonuclease/helicase Cas, which is a component in the defense mechanism of bacteria that unwinds the DNA of the virus before cleaving. WebLogo results show high conservation at the N-terminal and up to 150 amino acid residues. There is little to no conservation in the remaining protein.

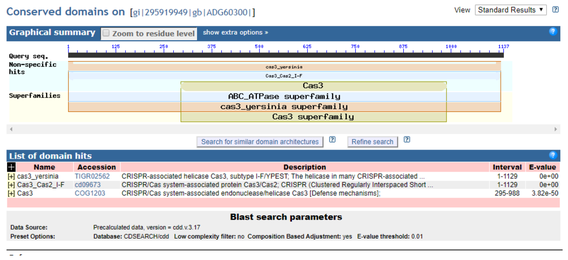


Figure 3. MCR_RS00120 CDD results revealing COG and TIGRFAM hits that support the product of a CRISPR-associated nuclease/helicase Cas 3 protein.

MCR_RS00125:

The proposed initial gene product was type I-F CRISPR-associated protein Csy1. This hypothesis was supported by the top two BLAST hits, TIGRFAM, PFAM and PDB results. The NR database shows good alignment with *Psychrobacter* with a score of 287 bits and a product of the same name as the proposed product. WebLogo reveals lack of conservation at the amino terminal with periods of good alignment but also sections of no conservation. The HMM logo indicates 10 larger letters that represent more conserved amino acid residues. The protein data bank result is consistent with this protein being CRISPR-associated. PDB results indicate a 5B7I: Entity 1 containing Chain A Cas3-AcrF3 complex.

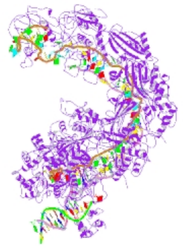


Figure 4. MCR_00125 PDB results revealing a 5B7I: Entity 1 containing Chain A Cas3-AcrF3 complex which supports the hypothesis of this proteins function as being a part of the Csy complex in CRISPR.

MCR_RS00130:

The computer pipeline proposed product indicated a type I-F CRISPR-associated protein Csy2. The gene annotation of MCR_RS000130 is between 37222 and 38157 coordinates in the genome. Both significant BLAST hits showed that the gene product was correctly indicated by the proposed product. T-COFFEE and WebLogo results revealed lack of conservation at the N-terminal and C-terminal as well as no conservation between 231 and 250. TIGRFAM revealed Cas-Csy2: CRISPR type I-F/Y-Pest-associated protein Csy2. This protein works with other proteins in the CRISPR complex to protect bacteria against viruses.

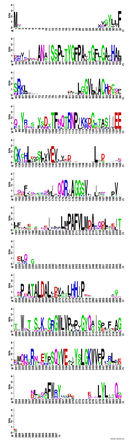


Figure 5. MCR_RS00130 WebLogo shows regions of alignment but incomplete alignment at the N-terminal, C-terminal and gaps throughout.

MCR_RS00140:

The initial proposed product of this gene was type 1-F CRISPR-associated endoribonuclease Cas6/Csy4. The gene length is 196 amino acids. BLAST hits show that the gene product name is type 1-F CRISPR-associated endoribonuclease Cas6/Csy4 and is well aligned with a *Acinetobacter species* with an E-value of 7e-71. Significant COG hits showed alignment to a siroheme synthase (precorrin-2 oxidase/ferro chelatase domain) [coenzyme transport and metabolism]. T-Coffee and WebLogo results revealed small gaps of lack of conservation with an overall highly conserved protein. PFAM results with a score of 128 revealed a protein family with the name of Cas_Csy4 CRISPR-associated protein.

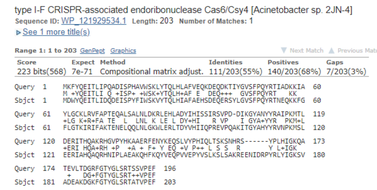
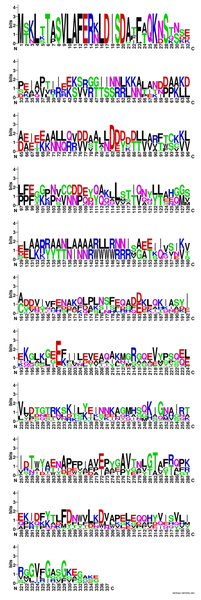


Figure 7. MCR_RS00140 reveals a well aligned BLAST hit with *Acinetobacter sp.*



MCR_RS00135:

The proposed initial gene product was a CRISPR-associated Csy3. This gene annotation is between 38192 and 39199 coordinates on the genome. It is 1008 nucleotides and 335 amino acids in length. The two top BLAST hits both revealed the gene product of CRISPR-associated Protein Csy3. The sequence aligned with the organisms *Pseudomonas aeruginosa* and *Acinetobacter marinus*. WebLogo revealed areas of higher conservation at the N-terminal with good alignment of remaining residues. Both TIGRFAM and PFAM hits also support the proposed product showing Cas_Csy3 domains.

Figure 6 . MCR_RS00135 WebLogo reveals good alignment throughout with higher conservation at the N-terminal.

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.

Locus Tag	Pipeline Annotation Product Name	Proposed Annotation	Changes Proposed?
MCR_RS00120	type I-F CRISPR-associated helicase Cas 3	type I-F CRISPR-associated helicase Cas 3	No
MCR_RS00125	type I-F CRISPR-associated protein Csy1	type I-F CRISPR-associated protein Csy1	No
MCR_RS00130	type I-F CRISPR-associated protein Csy2	type I-F CRISPR-associated protein Csy2	No
MCR_RS00135	CRISPR-associated protein Csy3	CRISPR-associated protein Csy3	No
MCR_RS00140	type I-F CRISPR-associated endoribonuclease Cas6/Csy4	type I-F CRISPR-associated endoribonuclease Cas6/Csy4	No

References

First Med Urgent Care OKC & Edmond. (2015, Sept). Urgent Care Reports: Signs of Ear Infections for Children and Adults. Retrieved from http://firstmedok.com/2015/02/25/signs-of-ear-infections-for-children-and-adults/

Hille, F. and Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. *Phil. Trans R. Soc. B.* 371:20150496. Retrieved from http://dx.doi.org/10.1098/rstb.2015.0496

Racaniello, V. (2009, May). Viruses and the Respiratory Tract. [Blogpost]. Retrieved from http://www.virology.ws/2009/05/21/viruses-and-the-respiratory-tract/

Shabbir, M. A., Hao, H., Shabbir, M. Z., Wu, Q., Sattar, A., & Yuan, Z. (1970, January 01). Bacteria vs. Bacteriophages: Parallel Evolution of Immune Arsenals - Semantic Scholar. Retrieved from https://www.semanticscholar.org/paper/Bacteria-vs.-Bacteriophages:-Parallel-Evolution-of-Shabbir-Hao/eb1ca8e14f461c1c31159caab5ec4358dfb1c383

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