Annotation of the Moraxella catarrhalis Genome within DNA Coordinates 32260 to 39199 (Locus Tags MCR_RS00120 to MCR_RS00135) Fatma Abdulrahman, Muhammad Bari, Sabina Bayrahtarova, Frida Ininahazwe, Suzan Habtezion, Bien Aimee Ishimwe,

Kira Mioducki, and Dawn Weihrich

Abstract

consecutive genes (locus tags: MCR_RS00120, A group of MCR_RS00125, MCR_RS00130, MCR_RS00135) from the bacteria Moraxella catarhalis were annotated using the collaborative genome annotation website GENI-ACT. This bacteria is a nonmotile pathogen that causes ear infections and lung infections. 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. All of our gene products were CRISPR associated proteins. CRISPR is an adaptive immune system in bacteria.

Introduction

Moraxella catarrhalis is a non-motile type of bacteria that spreads rapidly on its host. Moraxella has a rod-shaped form and is a diplococcus, which is a round bacterium that typically occurs in the form of two joined cells. (Brophy, 2017). 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).

M. catarrhalis can be diagnosed based on a sample of fluid from the ear or from the nasopharynx, at the base of your skull, and above the roof of your mouth. Its mechanism of infection is unknown. It adheres and colonizes the respiratory tract by infecting mucous membranes of respiratory tract and epithelial cells (See Figure 1).

The collection of our gene products indicated that they are all CRISPRassociated 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-Moraxella catarhalis http://microbe-canvas.com/Bacteria.php?p=1195



Figure 2 Explanation of how CRISPR-Cas immunity works. (Ledford, 2017)

Research Lab Program for Bioinformatics and Life Sciences, Buffalo, NY and the WNY Genetics in Research And Health Care Partnership

Methods

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

Modules	Activities	Questions Investigated
Module 1- Basic Information Module	DNA Coordinates and sequence, protein sequence	What is the sequence of my gene and protein? Where is it located in the genome?
Module 2- Sequence-Based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Is my sequence similar to other sequences in Genbank?
Module 5- Structure-Based Evidence	Pfam, PDB	Are there functional domains in my protein?

Results

MCR_RS00120:

The initial proposed product of the gene with the locus tag MCR_RS00120 gene by GENI-ACT was I-F CRISPER-associated helicase Cas3. The BLAST database provides multiple hits that indicate the protein is involved with CRISPER which is a family of DNA sequences in bacteria and archaea. Protein Data Bank was used to identify the protein by name, as Cas3-AcrF3 complex. The cellular location of the protein is within the cytoplasm. As a result, the proposed annotation is a helicase. (See Figure 5).

MCR RS00125:

The initial proposed product of this gene by GENI-ACT was a CRISPRassociated protein Csy1. This was supported by the top BLAST hit. Web logo analysis revealed good alignment except for gaps in sequence at 1-11, 85-88, 118-121, 159-160, 162-164, 226-229, 264-278, 323-327, 345-346. The protein was determined to be part of Cas_Csy1 family as supported by Pfam. The PDB results indicated a 3-D structure with the name: Cryo-EM structure of Type I-F CRISPR crRNA-guided Csy surveillance complex with bound target dsDNA (see figure 6).



The initial proposed product was CRISPR-associated protein Csy 3. The top BLAST hit indicated alignment with type I-F CRISPR-associated protein Csy3. It was 72 percent aligned. Web logo showed very good alignment, but 6 small gaps which indicates similarities to other sequences in Genbank. (see Figure 8). The top PFAM hit identified a protein family as Cas_Csy 3 which indicates that the protein is part of the CRISPR array which is a type of adaptive immunity in some prokaryotes.



MCR RS00130:

The initial proposed product was a CRISPR associated protein, Csy2. This gene product proposal was supported by the top BLAST for the amino acid sequence and was aligned to, specifically, *Psychrobacter* (see Figure 7) The gene product name was CRISPR-associated protein Csy2. The second NR BLAST hit showed a 75% positive alignment with Pectobacterium astrosep. Results from the CDD search indicated alignment with the COG CRISPR/Cas system-associated RAMP superfamily protein Csy2.

MCR RS00135:



Figure 5-Crystal Structure of Locus Tag MCR_RS00120



Crystal Figure Structure of Locus Tag MCR_RS00125

A- 3- 3- 1- 0- 9- 9- 9- 9- 9- 9- 9- 9- 9- 9
A B B C C C C C C C C C C C C C
4- 3- 92 1- N 82 82 82 82 82 82 82 82 82 82 82 82 82
⁴ ³ ⁴ ³ ⁴ ⁴ ⁴ ³ ⁴ ⁴ ⁵ ⁴ ⁴ ⁵ ⁴ ⁴ ⁵ ⁴ ⁴ ⁵ ⁴ ⁴ ⁴ ⁵ ⁴ ⁴ ⁵ ⁵ ⁵ ⁵ ⁵ ⁵ ⁵ ⁵ ⁵ ⁵
$4 - 3 - \frac{3}{2}$ $\frac{3}{2} - 1 - 1 - 0 - 3$

Figure 8 – Weblogo of Locus Tag MCR RS00135

		0.00	Der <u>oenre</u>							11124	A Photochi	-	-
S(core 84 hite	(087	Expect	Metho	osition	al matrix	v adjuct	Iden 176	itities /314(56%)	Positives	(74%)	Gaps 6/314(1%)	
		(907	/ 46-151	comp	osicion	armach	x aujust	. 170	(JU-10)	234/314	(/ + /0)	0/514(170)	-
Qu	ery 2		NDKVRVGCV ND +G +	F K+V	IYDANT I ANT	ISSPLTY ISSPLTY	GLPALTG G PA+TG	FTGAFH F G+FH	HALSRTLMADI HA+SR + DI	ERFADLALG E+ ++++LG	G 61 G		
Sb	jct 8		NDIGVIGYL	FKKVV	INGANT	ISSPLTY	GFPAITG	FLGSF	HAMSRKMADD	EQLSNVSLG	G 67		
Qu	ery 6	2	VLLACYECT	PKTYRA	GRYHDH	TENOTRN	PLKKDGK	TAPIVE	EEGKCNLVMS	FAVEVLASD	L 121		
Sb	jct 6	8	VLLACHDCQ	PQMYRF	NSYNNY	TENQTRN	PIKKDGK	TASII	EEGKCHLVMS	FVVEVLADT	Q 127		
Qu	ery 1	22	LSEKRSQEL	rqflsq	AIFRQR	IAGGSVM	NIAKTHP	VRYFG	FESMHQIVPL		A 181		
Sb	jct 1	28	L+ ++ + LTTEQQASV	Q SQ LQKTSQ) I +QR WIQQQR	+AGGSV MAGGSVR	+A+ P GLARFEP	V+YF VQYFDS	+ ++P SSDISAVIPQ	L+PAFVLMD LLPAFVLMD	A A 187		
Qu	ery 1	82	SSEFSELID	AQSTO	PKATPL	DVLLDVC	TLHHLPK	LDLDG	VIKWHIKSIK	TGHGWIVPM	P 241		
Sb	ict 1	88	EF+++I+ ODEFAOTIN	Q +	P A PL	D L+DVC	LHH+P+ ALHHTPE	+G+ AOKNGI	+TKW T S K- DTKWLTTSRK	+GHGW+VPM SGHGWLVPM	P P 247		
	erv 2	42		FAGEN		EVESOVV	EVANGLO	KPUVDE	KDI		 н 206		
Qu		42	+GYQ IS++	++AG M	NVRNP	EYPSQYV	E +Y LG	KW+YP		+I AFWR	H 290		
50	jct 2	48	TGYQGISDR	rdagvi	QNVRNP	EYPSQYV	EATYSLG	KWIYPO	2KENKIGGEH	DIVNAFWRY	H 307		
Qu	ery 2	97	YD-ETGLYLI YD + LYL·	TONE	309								
	fict 2	08	VDADDSI VI V	/ΤἶΟΟ	321								

Conclusion

The GENI-ACT proposed gene products 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.

Each of these locus tags codes for a Cas protein. These specific Cas proteins (Csy1, Csy2, and Csy3) in combination with CRISPR RNA, form a surveillance Csy complex that identifies the target sequence of foreign DNA (Hille et. al, 2016). The Cas3 protein then unwinds and degrades the foreign DNA (Jackson, 2014).

Gene	Locu
MCR_	_RS0(
MCR_	_RS0(
MCR_	RS0
MCR_	RS0



Special thanks to Dr. Rama Dey-Rao, Dr. Stephen Koury, and Dr. Sandra Small for their assistance with this project.

This work was supported by the National Science Foundation ITEST Strategies Award Number 1311902

s tag	Geni-Act proposed gene product	Proposed Annotation
0120	Type I-F CRISPR-associated helicase Cas3	I-F CRISPR-associated helicase Cas3
0125	Type I-F CRISPR-associated protein Csy1	CRISPR-associated protein Csy1
0130	Type I-F CRISPR-associated protein Csy2	CRISPR-associated protein Csy2
0135	CRISPR-associated protein Csy 3	CRISPR-associated protein Csy3

References

Brophy Martinez, K. (2017, July). MLAB 2434: Microbiology. [Powerpoint Slides]. Retrieved from http://slideplayer.com/slide/5794094/

Cooper LA, Stringer AM, Wade JT. 2018. Determining the specificity of Cascade binding, interference, and primed adaptation in vivo in the Escherichia coli type I-E CRISPR-Cas system. mBio 9:e02100 17. https://doi.org/10.1128/mBio.02100-17

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-forchildren-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

Jackson, R. N., Lavin, M., Carter, J., & Wiedenheft, B. (2014). Fitting CRISPR-associated Cas3 into the Helicase Family Tree. Current Opinion in Structural Biology, 0, 106–114.

http://doi.org/10.1016/j.sbi.2014.01.001 Ledford, H. (2017, Jan). Five big mysteries about CRISPR's origins. Retrieved from https://www.nature.com/news/five-big-mysteries-about-crispr-s-origins-1.21294#auth-1 Racaniello, V. (2009, May). Viruses and the Respiratory Tract. [Blogpost]. Retrieved from http://www.virology.ws/2009/05/21/viruses-and-the-respiratory-tract/

Acknowledgments

