

Annotation of the *Listeria monocytogenes* EDG-e Genome from Locus Tags Imo_0154 to Imo_0156)

Jarrett Evans, David Palmer, Nicholas Cacciotti, Melanie Roth

Erie 1 BOCES at Maryvale Work Readiness and Western New York Genetics in Research Partnership



Abstract

A group of 3 consecutive genes from the microorganism *Listeria monocytogenes* EDG-e (Imo_0154 – Imo_0157) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name for each gene was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, cellular localization data, and enzymatic function. The Genbank proposed gene product name did differ with one proposed gene annotation but did not differ significantly from the other two genes in the group and as such, the genes appear to be correctly annotated by the computer database.

Introduction

Listeria monocytogenes EDG-e can stay alive in different environments, including a living or non-living host (Chatterjee et al., 2006). This bacterium will induce uptake into its host cell and then enter the host cell cytosol by getting into the vacuolar membrane (Chatterjee et al., 2006).

According to Chatterjee et al. (2006), *Listeria monocytogenes* EDG-e is a microorganism of interest for several reasons. *Listeria monocytogenes* EDG-e is a biological pathogen that causes a disease called listeriosis. Listeriosis is a disease found on food with a high mortality rate. Any of the genes that can make you sick in *L. monocytogenes* EDG-e are controlled by positive regulator factor A (Chatterjee et al., 2006). The bacterium moves into the host cell and disrupts the phagosomal membrane of the host by expressing a pore-forming toxin (Chatterjee et al., 2006).

Methods

Modules of the GENI-ACT (<http://www.geni-act.org/>) were used to complete *Listeria monocytogenes* EDG-e 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 3- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 5- Structure-Based Evidence	TIGRFam, Pfam, PDB	Are there functional domains in my protein?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?

Results

Imo_0154		
Basic Information	DNA Coordinates	159663..161378 forward
	DNA Length	705 bp
	Amino Acid Sequence Length	234 aa
Sequence Based Similarity	COGs	COG1121 Zinc transport system ATP-binding protein
	TCOFFEE	highly conserved throughout
	WEBLOGO	highly conserved especially residues 0-202
Cellular Localization	Transmembrane Helices	None
	Signal Peptide	No
	PSORTB Final Prediction	Cytoplasmic membrane
Structure Based Evidence	TIGRFAMs	TIGR03771
	PFAMs	PF00005
	HMM Logo Key Residues	None
Enzymatic Function	PDB	2NQ2
	KEGG	zinc ABC transporter ATP-binding protein

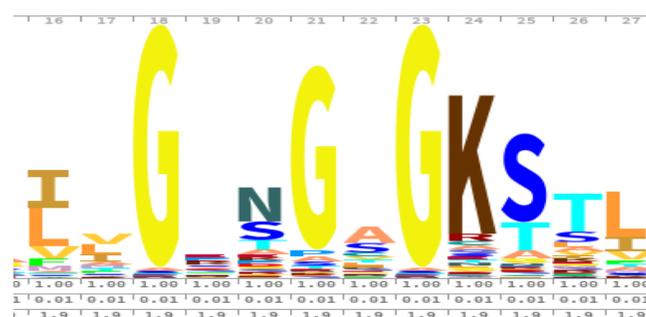


Figure 2. HMM logo of Imo_0154. G18, G21, G23, and K24 show high conservation.

Imo_0154 Summary:

The initial proposed product of this gene by GENI-ACT was a zinc ABC transporter ATP-binding. This gene product proposal is supported by TIGRFAM, Weblogo and Pfam. PSORTb predicted this gene to reside in the cytoplasmic membrane. The proposed annotation is zinc ABC transporter ATP-binding.

Imo_0155		
Basic Information	DNA Coordinates	151839..152645
	DNA Length	807
	Amino Acid Sequence Length	207
Sequence Based Similarity	COGs	transport system permease component
	TCOFFEE	Highly conserved throughout
	WEBLOGO	Highly conserved throughout
Cellular Localization	Transmembrane Helices	7
	Signal Peptide	No
	PSORTB Final Prediction	Cytoplasmic Membrane
Structure Based Evidence	TIGRFAMs	None
	PFAMs	PF 00950 Transport Family
	HMM Logo Key Residues	H40, G188, P199, G228
Enzymatic Function	PDB	None
	KEGG	zinc transport system permease protein

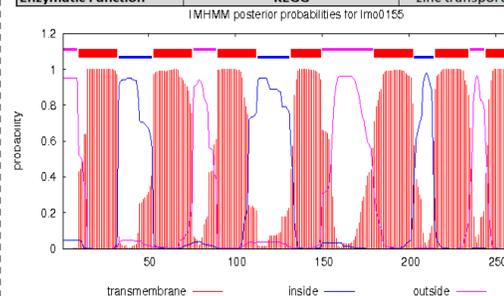


Figure 3 – TMHMM of Imo_0155 predicted 7 transmembrane helices.

Imo_0155 Summary:

The initial proposed product of this gene by GENI-ACT was zinc ABC transporter permease. This gene product proposal was supported by the top two BLAST hits for the amino acid sequence with low e-values. TMHMM results predicted seven transmembrane helices. PSORTb showed that this gene most likely resides in the cytoplasmic membrane. As such, the proposed annotation is a zinc ABC transporter permease.

Imo_0156		
Basic Information	DNA Coordinates	159663..161378
	DNA Length	681
	Amino Acid Sequence Length	226
Sequence Based Similarity	COGs	COG0664 cAMP-binding domain of CRP
	TCOFFEE	None
	WEBLOGO	None
Cellular Localization	Transmembrane Helices	Unknown
	Signal Peptide	None
	PSORTB Final Prediction	None
Structure Based Evidence	TIGRFAMs	None
	PFAMs	None
	HMM Logo Key Residues	None
Enzymatic Function	PDB	None
	KEGG	hypothetical protein

Crp/Fnr family transcriptional regulator [Listeria innocua]
Sequence ID: WP_003772539.1 Length: 226 Number of Matches: 1
See 1 more title(s)

Score	Expect	Method	Identities	Positives	Gaps
358 bits (919)	1e-123	Compositional matrix adjust.	168/224 (75%)	196/224 (87%)	0/224 (0%)
Query 1	MPKLNLYQEFIRLSHEGKIAYEQIDVPRNASLLTEKIDIDNHYLVVDGVIALLINDGS		60		
Sbjct 1	MEK+LINT EET LSH+G+I +E+I+YFGRA LL D+DQ+PE+DGYAL LINDG		60		
Query 61	QNSKIYSIQGKGFILNYFILLDSNNHFNFKLSGCSLYKYSWADMEVFLSMFFENFGFQ		120		
Sbjct 61	SKVYSIQG+G FLNYFILLD S N F +GILS C LY+YK D+EVFLSMFFENFGFQ		120		
Query 121	FFIMQNTIHHVYFKSLMASSPASEKLTITFSMALLHGVLSDDGTVILPQAIKTHLLSY		180		
Sbjct 121	FFIMQNTIHHVYFKSLMASSPASEKLTITFSMALLHGVLSDDGTVILPQAIKTHLLSY		180		
Query 181	SNLSKSCFYKDLQVLTITNQIEKKEKWIHDQALYTMINGQT 224				
Sbjct 181	SNLSKSCFYKDLQVLTITNQIEK+ WIH+ L M+QNG+I 224				

Figure 4 – BLAST of Imo_0156. A low e value of 1e-123 makes it very likely that Imo_0156 is a transcriptional regulator.

Imo_0156 Summary:

The initial proposed product of this gene by GENI-ACT was a hypothetical protein. The BLAST search showed that this gene is most likely a transcriptional regulator. It has a low e-value, high positive match and no gaps in the sequence with a transcriptional regulator. The proposed annotation for this gene is a transcriptional regulator.

Predicted ABC transporter of Zn2+

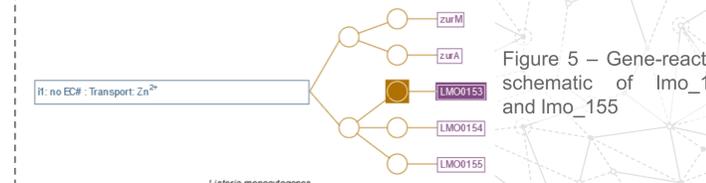
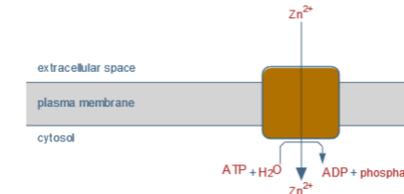


Figure 5 – Gene-reaction schematic of Imo_154 and Imo_155

Figure 6 – Predicted ABC transporter of Zn2+ enzymatic activity of Imo_154 and Imo_155



Conclusion

The GENI-ACT proposed gene product did differ from the proposed gene annotation for one of the genes but did not differ significantly on the other two genes in the group and as such, three genes appear to be correctly annotated by the computer database and one was not.

Gene Locus	GENI-ACT Gene Products	Proposed Annotation
0154	zinc ABC transporter ATP-binding protein	zinc ABC transporter ATP-dependent protein
0155	zinc ABC transporter permease	zinc ABC transporter permease
0156	hypothetical protein	transcriptional regulator

References

Chatterjee, S. S., Hossain, H., Otten, S., Kuenne, C., Kuchmina, K., Machata, S., Domann, E., Chakraborty, T., Hain, T. (2006). Intracellular gene expression profile of *Listeria monocytogenes*. *Infection and Immunity*, 74(2), 1323-1338.

Acknowledgments

Thank you to Dr. Stephen Koury for his guidance, support and dedication to this project. Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536

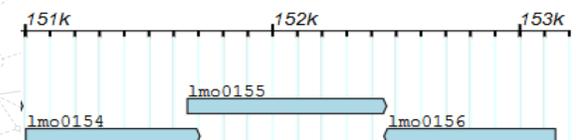
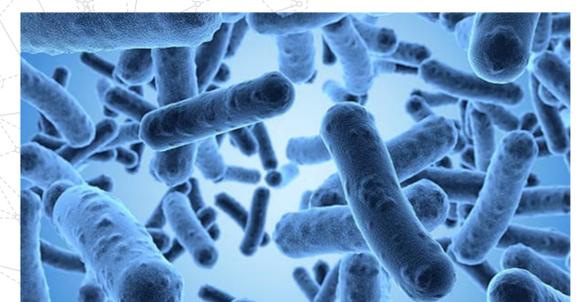


Figure 1. The locus tags and relative position of the genes under investigation in this research



Listeria monocytogenes EDG-e
Photo Credit: Guido Vrola