# Annotation of the Listeria monocytogenes Genome from Locus Tags Imo\_0157 to Imo\_0160

Lilith Adelman\*, Lori Krajewski\*, Matthew Ressing\*, Dana Whalen\* and Melanie Roth Erie 1 BOCES at Maryvale Work Readiness and Western New York Genetics in Research and Health Care Partnership \*indicates the same effort

Abstract

A group of 4 consecutive genes from the microorganism Listeria monocytogenes EDG-e (Imo\_0157 – Imo\_0160) 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 of the other three 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 vaculolar membrane (Chatterjee et al., 2006).

According Chattterjee et (2006),Listeria to monocytogenes EDG-e is a microorganism of interest for reasons. Listeria monocytogenes EDG-e is a several 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).



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

The manual genome annotation of *Listeria monocytogenes* EDG-e was carried out using GENI-ACT (Genomics Education National Initiative Annotation Collaboration Toolkit).



Listeria monocytogenes EDG-e www.foodpoisonjournal.com

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

lmo_0157			
	DNA Coordinates	151839152645	
<b>Basic Information</b>	DNA Length	2540	
	Amino Acid Sequence Length	779	
Soguoneo Pacod	COGs	COG-1199 CRad3-related DNA helicase	
Similarity	TCOFFEE	Moderately conserved throughout	
	WEBLOGO	Well conserved between nucleotides 194-224	
	Transmembrane Helices	No	
<b>Cellular Localization</b>	Signal Peptide	No	
	PSORTB Final Prediction	No peptide detected	
	TIGRFAMs	None	
Structure Based	PFAMs	DEAD_2 PF06733	
Evidence	HMM Logo Key Residues	P109, Y109, P128, Y154, E174	
	PDB	Crystal structure of an XPD DNA complex	
Enzymatic Function	KEGG	DNA Replication	



Figure 2 – Imo 0157 is involved in DNA replication. DNA Pol III holoenzyme has a proofreading capabilities that correct replication mistakes by means of exonuclease activity.

## Lmo\_0157 Summary:

The initial proposed product of this gene by GENI-ACT was a ATP-dependent helicase. This gene product proposal was supported by the top BLAST hit (ATP-dependent helicase) for the amino acid sequence. The amino acid was also located by running the CDD search and came up as a Rad3-related DNA Helicase. T-Coffee also showed results of an ATP-Dependent DNA helicase.

lmo_0158			
	DNA Coordinates	155993156805 forward	
<b>Basic Information</b>	DNA Length	813bp	
	Amino Acid Sequence Length	270aa	
Soguence Peeed	COGs	0561 (hyproxymethylpyrimidine pyrophosphatse)	
Sequence based	TCOFFEE	highly conserve throught out	
Similarity	WEBLOGO	highly conserve throught out	
	Transmembrane Helices	0	
Cellular Localization	Signal Peptide	no	
	PSORTB Final Prediction	cytoplasmic	
	TIGRFAMs	Cof-subfamily: Cof-like hydrolase	
Structure Based	PFAMs Hydrolase3 haloacid dehalogenase		
Evidence	HMM Logo Key Residues	D6 G63 K188 G210 D215	
	PDB	Crystal structure of HAD-like phosphatase yidA from E. col	
nzymatic Function	KEGG	thiamine metabolism	





## Lmo 0158 Summary:

The initial proposed product of this gene by GENI-ACT was hypothetical protein. This proposed gene annotation is a phosphatase or hydrolase. Phosphatase is supported by Tcoffee due to the highly conserved region at the amino terminus, and BLAST resulted in a low e value score of 8e-74. Hydrolase is supported by TIGRFAMs and PFAM.

lmo_0159			
	DNA Coordinates	157089159470	
<b>Basic Information</b>	DNA Length	2382	
	Amino Acid Sequence Length	793	
Sequence Based Similarity	COGs	COG4932	
		Uncharac Surface Anchored Protein	
	TCOFFEE	Not well conserved throughout	
	WEBLOGO	Well conserved throughout	
Cellular Localization	Transmembrane Helices	1	
	Signal Peptide	Yes. Cleavege between 31 and 32.	
	PSORTB Final Prediction	cell wall	
Structure Based Evidence	TIGRFAMs	TIGR01167	
	PFAMs	PF05737	
	HMM Logo Key Residues	G8, W20, D38	
	000	3RKP: Entity 1 containing Chain A, B	
	PUB	Crystal structure of BcpA*(D312A)	
nzymatic Function	KEGG	None	



Figure 3 – This graph represents the Phobius output for Imo 0159. The graph confirms transmembrane helices found on the TMHMM site.

## Lmo\_0159 Summary:

The initial proposed protein predicted from GENI-ACT was peptidoglycan binding protein. This gene product proposal was supported by the top BLAST hit because wall associated proteins are found in the cell wall just like peptidoglycan binding proteins. Furthermore, CDD also supports the predicted protein because Uncharacterized surface anchored proteins are also binding proteins that are located in the cell wall.

Enzym



Lmo 0The initial proposed product of this gene by GENI-ACT was peptidoglycan binding protein. This gene proposal was supported by Pfam due to the similarity in the protein sequence alignment. PDB determined that the amino acid sequence is involved in cell adhesion. PSORT-B predicted the amino acid sequence to be in the cell wall.

# 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 three 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
0157	ATP-dependent helicase	ATP-dependent helicase
0158	hypothetical protein	phosphotase or hydrolase
0159	peptidoglycan binding protein	peptidoglycan binding protein
0160	peptidoglycan binding protein	peptidoglycan binding protein

## 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 - R250D010536

	Supported by the Natio	SCIENCE EDUC PARTNERSHIP / nal Institutes of	ATION AWARD Health
es	Unive at Bu	<b>B</b> <b>rsity</b> <b>ffalo</b>	
	DNA Coordinates	15	59663161378
asic Information	DNA Length		1716bp
	Amino Acid Sequence Length	COC4022 page	571aa
nce Based Similarity	TCOFFFF	COG4932 need	wly conserved
ince bused similarity	WEBLOGO	highly conser	ved especially at 177-200
	Transmembrane Helices		2
llular Localization	Signal Peptide	yes cleva	ge between 36 and 37
	PSORTB Final Prediction		Cellwall
	TIGRFAMs		none
ture Based Evidence	PFAMs	PF05737 Co	llagen_binding protein
ture based Evidence	HMM Logo Key Residues	W20 , G8 , N25	
	PDB	3RKP Entity	1 containing Chain A, B
natic Function	KEGG		none
ТМНМ	A posterior probabilities for Imo0160		
			Figure 4 –
			localization
			from TMHMM
			database for
			shows the
			presence of two
			transmomhrana
I	ı		
100 2	00 300 400	500	helices.
trop and such uses -		nutaida	
transmemprane ——	inside ———	outside ——	
160 Summa	rv·		
	u y .		

## www.buffalo.edu