University at Buffalo The State University of New York | REACHING OTHERS

KENDALL CENTRAL SCHOOL

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

Annotation of the *Listeria monocytogenes* Genome from DNA Coordinates 1450..2513 and 5886..7337

Alycia A. Davenport*, Bridget K. Miller*, and Kathryn S. Hoppe

Kendall High School and the Western New York Genetics in Research and Health Care Partnership

Abstract

Listeria monocytogenes is sprin cipally found in pathogenic cases of food poisoning. This organism which causes Listeriosis, is one of the leading causes of death of foodborne pathogens. The bacterium can spread from the site of infection in the intestines to the nervous system to the placenta in pregnant individuals. The purpose of this study was to identify and explain different coding regions of the genome of *Listeria monocytogenes*. The two gene loci studied were *Listeria monocytogene* 5923_0009 and 5923_0004. Methods for studying Listeria monocytogenes were the GENI-ACT online modules one through seven. The gene annotation of locus 5923_0009 showed it was an enzyme and locus 5923_0004 was a transmembrane protein. This study exhibited that the computer process of gene annotation was correctly identified, even though accuracy was previously questioned.

Introduction

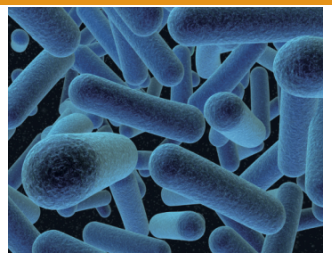

Figure 1: *Listeria monocytogenes*

Listeria monocytogenes is a facultative anaerobic, motile, encapsulated, and non-spore forming gram positive rod-shaped bacterium, found predominantly in a biofilm formation. This bacterium is catalase-positive and oxidase-negative. It was first isolated in 1924 by E.D.G. Murray, but originally titled *Bacterium monocytogenes*. It was reverted back to the known name of *Listeria monocytogenes* by Harvey Pirie in 1940. It was related to neonatal infection, sepsis, and meningitis in East Germany in 1952. *Listeria monocytogenes* was announced as a foodborne illness 30 years later in Nova Scotia. *Listeria monocytogenes* grows abundantly at temperatures of four to thirty degrees celsius. Four additional vitamins including riboflavin, thiamine, biotin, and thioctic acid and several other amino acids are needed for *Listeria monocytogenes* to grow.

Listeria monocytogenes is relevant in multiple cases across the scientific community. Initially found in cases of adolescent rabbits, it is most recently found in food poisoning cases, and studies with pancreatic cancer with radioactive *Listeria monocytogenes*. *Listeria monocytogenes* is abundantly found in dairy products and vegetables. One of the most recent cases of *Listeria monocytogenes*, was a multistate outbreak in raw milk in Miller's Organic Farm in Pennsylvania, as of March 2016. It is recommended that people consume pasteurized dairy products for this reason, the utmost safety for our people. *Listeria monocytogenes* has been investigated in using Nontoxic Radioactive Listeria as a highly effective therapy against pancreatic cancer. *Listeria monocytogenes* was used in presenting to tumor associated antigens, and was successful in efficiently clearing the immune system of pathogens.

Methods and Materials

Modules of the GENI-ACT (http://www.geni-actorg/) were used to complete *Listeria monocytogenes* 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- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 4- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 5- Alternative Open Reading Frame	IMG Sequence Viewer For Alternate ORF Search	Has the amino acid sequence of my protein been called correctly by the computer?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?
Module 7- Gene Duplication/Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?

Results

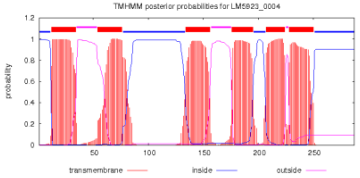

Figure 2: A TMHMM result for *Listeria monocytogenes 5923_0004* predicts six transmembrane helixes.

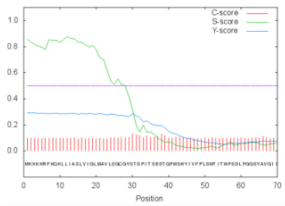

Figure 3: A SignalP output for *Listeria monocytogene s 5923_0004* shows the presence of a signal peptide.

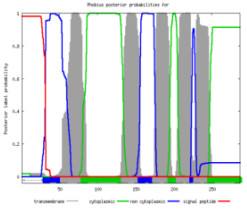

Figure 4: A phobius output of *Listeria monocytogenes 5923_0004* showing the presence of five transmembran e helixes and a signal peptide.

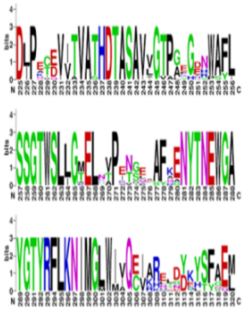

Figure 5: WebLogo of *Listeria monocytogenes* 5923_0009: Throughout the whole alignment of the amino acids there is higher conservation. There is a greater predominance of polar amino acids throughout the whole sequence.

Family	Description	Entry type	Clan
FGGY_N	FGGY family of carbohydrate kinases, N-t ...	Domain	CL0108
FGGY_C	FGGY family of carbohydrate kinases, C-t ...	Domain	CL0108

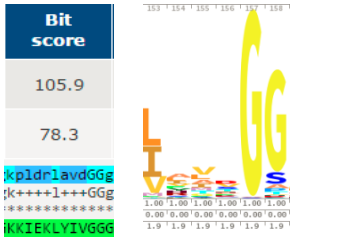

Figure 6: PFAM results of *Listeria monocytogenes* 5923_0009 (Top and bottom left). Key structural residues in the HMM logo for FGGY_C include G6, T7, A126, G157, A170, G189, and A194 Bottom right)..

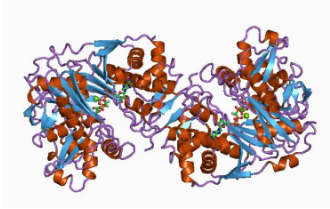

Figure 7: Example FGGY family of Carbohydrate Kinases, N-terminal domain. Specifically, *Enterococcus casseliflavus* glycerol kinase complexed with glycerol

LM5923_0009:

The initial proposed product of this gene by GENI-ACT was hypothetical protein. The top BLAST hits for this gene were, however, Rhamnulokinase. The BLAST hits had significant and scores and E-values. The correct annotation being Rhamnulokinase is also supported by TIGRFAM and PFAM hits for this gene. It is thus proposed that LM5923_0009 should be annotated as Rhamnulokinase. Rhamnulokinase (EC 2.7.1.5) is an enzyme that catalyzes the chemical reaction ATP + L-rhamnulose $\rightleftharpoons$ ADP + L-rhamnulose 1-phosphate. *Listeria monocytogenes* as a whole has roughly 881 genes predicted to be enzymes. *Listeria monocytogene* s 08-5923 has an estimated 3039 genes, of which 2966 are predicted to be proteins. Although both participants in the GENI-ACT had genes from the same genome, each had a different locus.

Listeria monocytogenes 5923_0004:

The initial proposed product of this gene by GENI-ACT was also a hypothetical protein. However, the top BLAST, CDD, and TIGRFAM hits suggested this gene was membrane protein insertase YidC. YidC is required for the insertion and/or proper folding and/or complex formation of integral inner membrane proteins. The cellualr localization module predicted LM5923_0004 to be a transmembrane protein. All together, the existence of well put together functional domains, transmembrane topography, and the cell location within the amino acid sequence was confirmed by the participants data. The annotation of this gene should therefore be membrane protein insertase YidC.

Conclusions

Throughout this study the GENI-ACT guided the participants through the gene annotation. Evidence was found that both genes, which were originally annotated as hypothetical proteins on their respective GENI-ACT gene information pages, had better had better calls based on our research. The gene annotation provided significant data for the *Listeria monocytogenes* including the unique shape and size of the gene between the two participants different models. The two genes annotated were *LM5923_0009* and *LM monocytogenes 5923_0004*. We propose that *LM5923_0009* should be annotated as Rhamnulokinase and that LM5923_0004 should be annotated as membrane protein insertase YidC.

References

Sims et al. (2009). Complete genome sequence of *Listeria monocytogenes*. *Standards in Genomic Sciences*, 12 - 20.

Listeria. (n.d.). Retrieved May 24, 2016, from http://www.foodsafety.gov/poisoning/causes/bacteriavirusesli steria/

Paterson, & Maciaqi. (2005, October 7). Listeria-based vaccines for cancer treatment.

U.S. Food and Drug Administration. (n.d.). Retrieved May 24, 2016, from http://www.fda.gov/Food/RecallsOutbreaksEmergencies/Outbreaks/ucm359588.htm

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award -5R25OD0105

Dr.Stephen Koury (University of Buffalo)

Tige E. Noni (teacher at KCS)