Annotation of the *Wolbachia pipientis* Genome at Locus Tags DNA WP0029, WP0037, WP0045 and WPOO47

Hannah LoVullo, Sarah Seguin, Sophie Serpas, Maddie Tavernier, and Pamela Patterson
Holland High School, Holland, NY 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 4 genes from the microorganism *Wolbachia pipientis* (WP0029, WP0037, WP0045 and WPOO47) 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, potential alternative open reading frames, enzymatic function, presence or absence of gene duplication and degradation, the possibility of horizontal gene transfer, and the production of an RNA product. 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 in the database.

Introduction

Wolbachia pipientis is a maternally inherited gram-negative bacterium that falls under the genus *Rickettsia* (González G, 2009). This bacterium is present in over 70% of insect species (originally discovered in the mosquito genus *Culex)* as well as some arthropods and nematodes. This endosymbiont bacterium is a horizontally transferable disease that performs best in 25°C with a range of +/- 4°C; typically active in warmer climates. It is best known for manipulating its host's reproductive system by means of its four different phenotypes; parthenogenesis, male-killing, feminization, and cytoplasmic incompatibility (unidirectional or bidirectional) to eliminate male offspring in the larvae stage (WerrenLab, 2011). This occurrence aids in the destruction of diseases carried by insects such as Zika, Yellow Fever, and Dengue Fever. Further comprehension of this genome could prove to be very beneficial to the whole world.

Holland's UB Genome group took a deeper look into *Wolbachia's* roots as an attempt to help the scientific community further launch itself into research for disease ending bacterium. Students worked quickly and diligently to get as far as possible in their research as well as practicing gathering data and using practical websites like BLAST, WebLogo, T-Coffee, and many others.

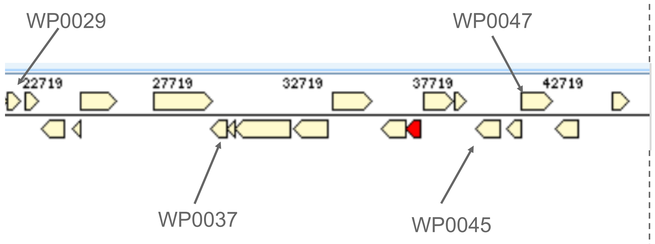

Figure 1:The locus tags and relative position of the genes under investigation in this research.

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Wolbachia pipientis* 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?
Cellular Localization	Gram Stain, TMHMM, SignalP, LipoP, Psortb , Phobius	Is the protein under investigation located in the cytoplasm, secreted, in the periplasm or embedded in the cell membrane or cell wall?
Final Annotation	Evaluate data from all modules	Has the gene been correctly called by the pipeline annotation?

Results

WP0037 *Wolbachia* Translation Initiation Factor IF-1 is one of the factors required for *Wolbachia* growth and viability. The precise function of Translation Initiation Factor IF-1 remains unknown. Interestingly, there appears to be extra amino acids added to two regions of WP0037 as compared to other species. This causes two gaps in the T-Coffee graph (Figure 2: Residues 16-19, 77-78). It is unknown why these extra amino acids are present.

Figure 2: WP0037 T-Coffee results showing gaps at residues 16-19 and 77-78

WP0029 is a transposon, meaning its function is to cut sections of DNA and move them to other areas of the genome. WP0029 causes a lot of shuffling of DNA in the two types of cells. *Wolbachia* bacteria appears to play a key role in sterilizing mosquitoes. Understanding gene transfer from host to infection are likely to prove useful in eliminating diseases harmful to humans that are carried by mosquitoes. According to literature residues D6, D86,and E123 of the HMM logo (Figure 3) are in similar relative locations to known conserved residues in DDE motif transposases. (http://www.jbc.org/content/278/3/1904)

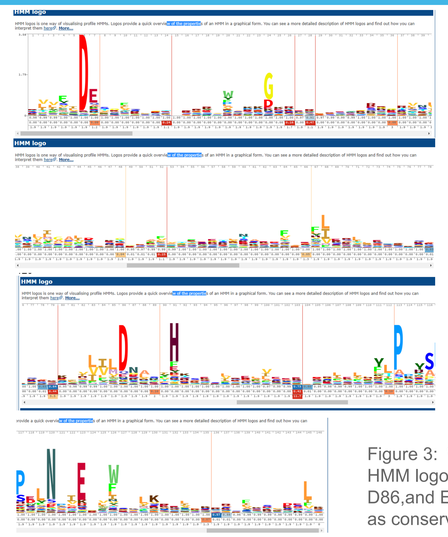

Figure 3: WP0029 HMM logo showing D6, D86,and E123 residues as conserved

WP0045

A permease is a membrane transport protein, a class of multi-pass transmembrane proteins that allow the diffusion of a specific molecule in or out of the cell, controlling what goes in and out of the cell. An interesting fact is that commonly permease usually has high levels of homology, but in contrary within research it can be revealed that permease is not as conserved as one would think. Can be seen in Figure 4.

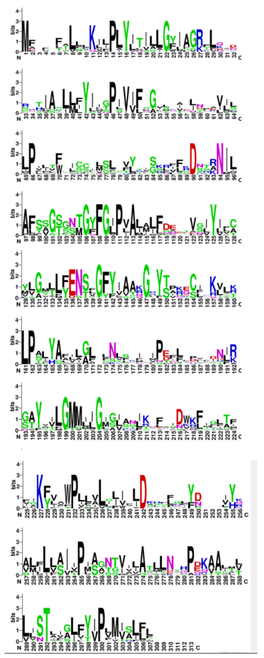

Figure 4: WP0045 WebLogo showing less than predicted conservation.

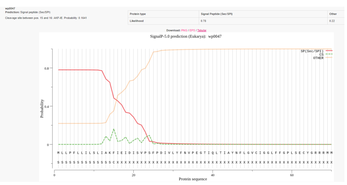

Figure 5: WP0047 Signal P showing predicted a cleavage point between 13 and 14

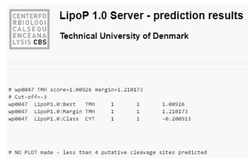

Figure 6: WP0047 LipoP showing no predicted cleavage point

Wp0047: Major Facilitator Superfamily (MFS) Transporter. transports a single class of compounds. Compounds transported by MFS permeases include simple sugars, oligosaccharides, drugs, amino acids, nucleosides, and a large variety of organic and inorganic anions and cations. This is very important to further drug resistant gene research and to identify what makes bacteria resistant the drugs being used. There are eleven predicted transmembrane helices (Figure 7) and a signal peptide at the starting point of the transmembrane topology graph. Signal P predicted a cleavage point between 13 and 14 but LipoP P did not show any (Figure 5 and 6). It seems to be located in the cytoplasmic membrane.

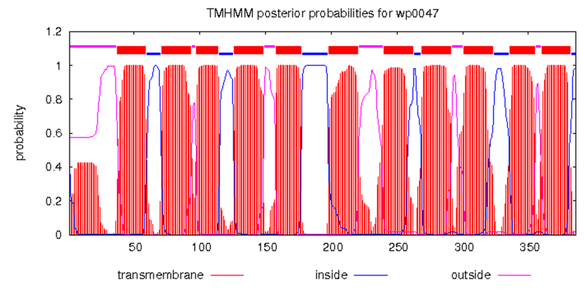

Figure 7: WP0047 TMHMM with 11+ predicted transmembrane helices

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?
WP0029	Putative transposase	Transposase	No
WP0037	Bacterial translation initiation factor (IF-1)	Translation initiation factor IF-1	No
WP0045	Transmembrane helix (Auxin efflux carrier)	Permease	No
WP0047	Drug resistance transporter	Major Facilitator Superfamily (MFS) Transporter	No

References

(1)González1, G., & Brazil, M. F. (2009, January 01). C. I. Espino. Retrieved May 01, 2018, from http://aem.asm.org/content/75/2/547.full#
(2)Werren Lab, University of Rochester, (2011, Jan 31), "Wolbachia", http://www.sas.rochester.edu/bio/labs/WerrenLab/WerrenLab-WolbachiaBiology.html

University at Buffalo The State University of New York
Jacobs School of Medicine and Biomedical Sciences
buffalo.edu

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536
-Mr. Guidotti, Mrs. Gregory, Mrs. Fabiatos, Dr. Koury