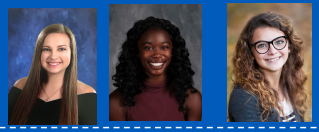
Annotation of the *Wolbachia endosymbiont of Culex quinquefasciatus* Genome from Locus Tags WP0002,WP0005, WP0006, WP0008 and WP0015)

Sydney Aldrich, Emily Gronquist, Lindsay Lotter, Deidra Osula, Andrea Smith, Jackson Booth, Dominique Riforgiato*, Selin Arnavut*, Aurora Merwin* and Mrs. Lee *control group
Fredonia High School and the Western New York Genetics in Research and Health Care Partnership



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

University at Buffalo

Abstract

Our group used the GENI-ACT computer programs from four modules to annotate five genes from the bacterium *Wolbachia*. Our goal was to annotate each gene and compare the information to the identity of these genes determined by GENI-ACT. We used the BLAST, TIGRFAM, Pfam and CDD programs to look at amino acid sequence similarity to other orthologs, We used T-Coffee and WebLogo to compare our gene product with orthologs from other organisms to determine the amount of conservation of our gene product across other organisms. The TMHMM, SignalP, PSORT-b, and Phobius programs were used to determine if the gene product was embedded within a membrane, in the cytoplasm of the cell or secreted outside of the cell. All of this information was taken together to decide if it made sense based on the function of the gene product determined by GENI-ACT. We concluded that all five of our genes were correctly called by GENI-ACT.

Introduction

Wolbachia are a group of gram negative bacteria that live within the cells of many insects. The bacteria are transmitted to the next generation of the host through the cytoplasm of the eggs. *Wolbachia* has evolved to manipulate the host's reproduction in order to increase its own transmission.

Wolbachia is able to induce feminization, male killing, parthenogenesis and cytoplasmic incompatibility in its host, all of which promote an increase in the number of females produced. This is important to *Wolbachia* since it is only transmitted by females. *Wolbachia* is not infectious to humans although many of the host insects that they infect also carry human diseases including West Nile virus, dengue virus, lyme disease and Zika. There has been some evidence that when *Wolbachia* is present in the mosquito, *Aedes aegypti*, these viruses are unable to survive and infect humans. *Wolbachia* is being studied for use as a biological control. In the GENI-ACT program we have annotated five genes with differing functions. We have determined that each of these genes were correctly called by the GENI-ACT program. (Werren,JH., 1997).

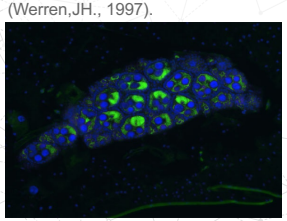


Figure 1. *Wolbachia* is shown in green in the ovaries of an *Aedes aegypti* mosquito.

http://www.eliminatedengue.com/our-research/wolbachia

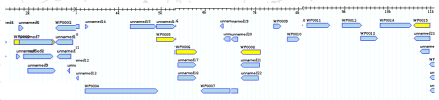


Figure 2 – Gene map showing the *Wolbachia* gene locus tags, indicated in yellow, under investigation in this study.

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Wolbachia* 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 4- 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 5- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?

Results

WP0002:

The initial proposed name of this gene by GENI-ACT was protein-export membrane protein SecF. The BLAST search matched other orthologs of a protein translocase with subunit SecF with high scores and low E values. The TIGRFAM and PFAM tests also supported the name of the gene as a protein-export membrane protein SecF. The T-coffee and WebLogo showed high conservation throughout the protein. The TMHMM, SignalP, PSORT-B, and Phobius tests all support a transmembrane protein with 6 transmembrane helices. This is consistent with a function of transporting a protein across the membrane.

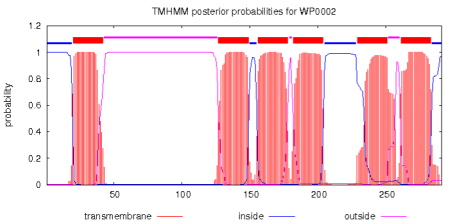


Figure 3 – Results of the TMHMM program showing six transmembrane helices for the WP0002 gene product..

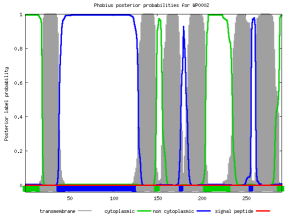


Figure 4 – Results of the Phobius program corroborating the results of the TMHMM showing six transmembrane helices for the WP0002 gene product .

WP0005:

The initial proposed name of this gene by GENI-ACT was 6,7-dimethyl-8-ribityllumazine synthase. This initial proposal was supported by multiple hits on BLAST that had high scores and low E-values. The TIGRFAM and PFAM tests also supported the name of the gene as 6,7-dimethyl-8-ribityllumazine synthase. Based on these results, it can be assumed that the correct name of this gene is 6,7-dimethyl-8-ribityllumazine synthase. 6,7-dimethyl-8-ribityllumazine synthase is an enzyme that acts as a catalyst in the biosynthesis of riboflavin. The synthesis of one riboflavin molecule requires one molecule of GTP, which is an energy-rich nucleotide that can be compared to ATP, and two molecules of ribulose 5-phosphate as substrates. 6,7-dimethyl-8-ribityllumazine synthase is used during the final step of the biosynthesis of the riboflavin. The final step involves the dismutation, or simultaneous oxidation and reduction, of the 6,7-dimethyl-8-ribityllumazine synthase catalyzed by the riboflavin synthase.

WP0006

The initial proposed name of this gene by GENI-ACT was nitrogen utilization protein B. The hits on BLAST and CDD supported the initial proposition of protein B with high scores and low E-values, showing the similarity to the gene's DNA structure. The proposed name of the gene, nitrogen utilization protein B, was also supported by the TIGRFAM and PFAM tests. The T-Coffee and WebLogo tests supported the findings of high conservation to other orthologs. The Phobius, TMHMM, PSORT-b, and SignalP tests also confirm the initial call of nitrogen utilization protein B. Protein B initiates a utilization process that starts when levels of nitrogen are detected and ends when nitrogen is traced into the organism's metabolism or the cell. These various tests support the initial proposal of the gene and its function.

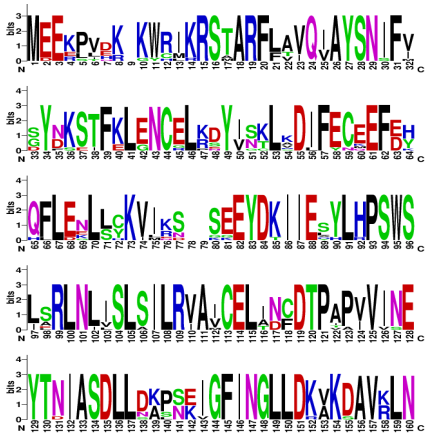


Figure 5 – WebLogo of WP0006 showing high conservation throughout the gene product.

WP0008

The initial proposed name of this gene by GENI-ACT was deoxyuridine 5'-triphosphatenucleotidohydrolase. The BLAST search found two matches with high E values and a high score. Through the WebLogo and T-Coffee tests, I found that this gene was conserved to other orthologs. To test the transmembrane helices I used TMHMM, SignalP, and PSORT-B, I then found that there were none. This gene product is an enzyme and is involved in nucleotide metabolism.

WP0015

The initial proposed name of this gene by GENI-ACT was holo-(acyl-carrier-protein) synthase. The BLAST, TIGRFAM and Pfam all match a phosphopantetheine--protein transferase domain. The CDD matched a phosphopantetheinyl transferase for lipid metabolism, which backs up the results of the other programs. The BLAST, T-Coffee and WebLogo all show high conservation to other orthologs except at the end of the protein. The TMHMM, SignalP, PSORT-B, and Phobius tests all support a cytoplasmic protein. Since the enzyme functions in fatty acid biosynthesis, one would expect to find it in the cytoplasm.

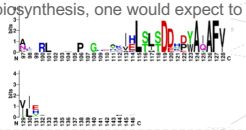


Figure 6 – WebLogo of WP0015 showing low conservation of amino acids at the end of the gene product.

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 Numbers	Proposed Annotation	Most Likely Annotation
WP0002	protein-export membrane protein SecF	protein-export membrane protein SecF
WP0005	6,7-dimethyl-8-ribityllumazine synthase	6,7-dimethyl-8-ribityllumazine synthase
WP0006	N utilization substance protein B	N utilization substance protein B
WP0008	deoxyuridine 5'-triphosphatenucleotidohydrolase	deoxyuridine 5'-triphosphatenucleotidohydrolase
WP0015	holo-(acyl-carrier-protein) synthase	holo-(acyl-carrier-protein) synthase

References

Werren, J. H. (1997). Biology of *Wolbachia*. Annual Review of Entomology. Vol. 42:587 - 609.

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536

www.buffalo.edu