Annotation of the Wolbachia pipientis Genome at Locus Tags DNA WP0033, WP0039, WP0043, WP0046 and WP0050

Andrew Hibit, Jared Montgomery, Lauren Oleksy, Joshua Ostrowski, Renee Wright, and Pamela Patterson Holland Jr./Sr. High School, Holland, NY and The Western New York Genetics in Research and Health Care Partnership

Abstract

A group of 5 genes from the microorganism Wolbachia *pipientis* were re-annotated using the genome annotation website celled GEN-ACT. Once a FASTA sequence was entered, students were given basic information about their gene such as genetic location, horizontal transfers, enzymatic activity, how well they appear to be conserved, and presence of proteins and helices. After weeks of research, students found no significant difference between the GENI-ACT result and the BLAST result. Overall, according to the databases, all genes have been correctly annotated and perceived as true.

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 (Werren Lab, 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 using practical websites like gathering data and BLAST, WebLogo, T-Coffee, and many others.



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

WP0033: Sodium Alanine Symporter: This gene is responsible for the transportation of alanine and sodium through the membrane of Wolbachia. Specifically, this protein is only found in bacteria and archaea. As seen in the diagram (Figure 2), Wolbachia possess approximately 11 transmembrane helices. The literature 8 transmembrane helices reports (https://www.ebi.ac.uk/interpro/entry/IPR001463). By controlling this specific gene, scientists can limit, or change what is allowed into the cell; largely affecting the organism's ability to maintain homeostasis, and or changing its properties.



Figure 2: WP0033 Phobius results showing 11 transmembrane helices



Figure 4: WP0046 X-ray crystallography PDB image showing 4 α helices

WP0043: Protein chain release factor 1 ends the translation of mRNA to a protein. This protein has an important and ubiquitous function in all prokaryotes and thus should be highly conserved. The web logo shows a high degree of conservation between residues 106-260 in Figure 5.

WP0039:

- Adenylosuccinate synthase
- catalyzes the first committed step in the synthesis of adenosine, part of DNA. (http://www.biochemj.org/cont ent/341/3/537.abstract).
- Adenylosuccinate synthase is a protein that randomly floats through the cytoplasm of a cell acting on the first step to crafting adenosine, one of the crucial building blocks for DNA. It's structure is likely highly conserved, and WebLogo results support this claim as shown in Figure 3.
- **WP0046:** GNAT acetyltransferases catalyze the transfer of an acyl group from acyl coenzyme A (acyl-CoA) to a diverse group of They are widely substrates. distributed among all life forms, with diverse roles such as stress regulation, transcription control, anti-oxidants, and many more.
- Abu Ud-Din et al. found these families of protein to contain 6-B strands with four α helices (https://www.ncbi.nlm.nih.gov/pmc/a rticles/PMC4964394/).
- The α helices can be seen in the Protein Database match of the query sequence in Figure 4.

ECKCK | VDwLSE ADxvVRFCCCCHNCHT | V | DE XYKLSLLPSEVER SKLSU GNGVXLDPH **WKLL**ONGTYP EVTSSNTVAGOA - GSG-G-A YLGUAKA TTRVGGPFPTEQ.NE.G. LG. GEGTV TCR RCCIEDAVLVROAV TSGISGIALTKL ET GARS APPAK *<u>1</u>222330

Figure 3: WebLogo for WP0039 showing conserved residues

Figure 5: WP0043 WebLogo showing high degree of conservation from residues 106-260

WP0050: BLAST and CDD produce two different names for the gene product, Endolytic peptidoglycan transglycosylase and peptidoglycan hydrolase respectively. These products perform unique roles that may not have been recognized by the computer. Endolytic peptidoglycan transglycosylase is used in the separation of daughter cells and the maintenance of a rod shape, while peptidoglycan hydrolase regulates cell wall growth and can enlarge the pores in the peptidoglycan for the assembly of large trans-envelope complexes, such as flagella. See Figure 6 for CDD results.

Query seq. Specific hits Non-specific Superfamilies

List of domain hits

User Options:

Figure 6: WP0050 CDD results supporting a match with peptidoglycan hydrolase

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 T Wp_ Wp_

Wp_

(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/WerrenL ab-WolbachiaBiology.html

Acknowledgments Supported by an NIH Science Education Partnership (SEPA) Award - R250D010536 -Mr. Guidotti, Mrs. Gregory, Mrs. Fabiatos, Dr. Koury





atabase: CDSEARCH/cdd v3.17 Low complexity filter: no Composition Based Adjustment: yes E-value threshold: 0.01 Maximum number of hits: 500

g	Pipeline Annotation Product Name	Proposed Annotation	Changes Proposed?
)33	Sodium alanine transporter	Sodium alanine symporter	no
)39	adenyl-succinate synthetase (purine conversions)	Adenylosuccinate synthase	no
)43	Bacterial peptide chain release factor	Protein chain release factor 1	no
)46	Amino glucoside 6'N-acetyl transferase (antibiotic resistance)	GNAT acetyltransferases	no
)50	Putative rare lipoprotein A	Endolytic peptidoglycan transglycosylase/ peptidoglycan hydrolase	no

References