Annotation of the *Wolbachia pipientis* Genome from DNA Coordinates 10607 to 21548

Leesa Applebee, Mallory Drews, Hannah LoVullo, Jacob Ostrowski, Joshua Ratajczak, Maddy Scull, Renee Wright
Holland Jr./Sr. 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

University at Buffalo

Abstract

A group of 8 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 pipientits 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 gathering data and using practical websites like BLAST, WebLogo, T-Coffee, and many others.

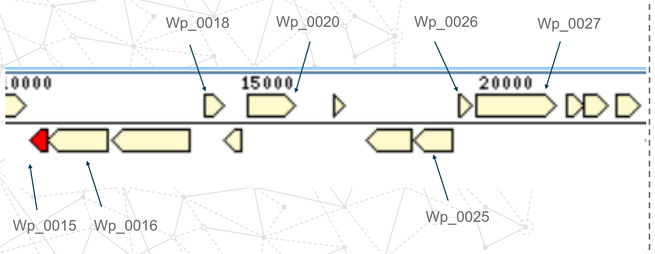

Figure 1. The locus tags and relative position of the genes under investigation in this research.

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
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?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?

Results:

Wp0015: This gene codes for a holo-(acyl-carrier-protein) synthase. From the HMM logo (Figure 1), it was discovered that the FASTA sequence has a non-conserved domain sequence. This particular matched the literature, showing the vast amount of variety. This enzyme belongs to the family *transferases.* That specifically transfer non-stranded phosphate groups. This is likely involved in ATP which occurs in every living being, and it's essential.

Wp_0016: This gene codes for a Proline--tRNA ligase that makes tRNA molecules. A literature search revealed that this protein is conserved at the C-terminus, but the Weblogo (Figure 3) for this protein shows conservation at the N-terminus for the prokaryotes examined.

Wp_0019: This is a subunit of a protein used in the outer membrane of gram negative bacteria. Based on the scores in Signal P (Figure 2), it is likely that there is a signal peptide for this molecule. Signal peptides are what tell the protein when to activate, and BamE is an outer membrane protein that is made in the cytoplasm. Thus a signal peptide would be needed to tell it to activate once it reaches the outer membrane. BamE is of interest, specifically in *Wolbachia*, since they are responsible for the efflux of small molecules like antibiotics.

Wp0020: The initial product of this gene by GEN-ACT was a probable ribonuclease-diphosphate. This gene holds an enzyme, ribonucleotide reductase, which is the responsible factor in the conversion of ribonucleotides to deoxyribonucleotides for building DNA. This enzyme helps maintain the stability of the genome as well as protecting against mutations. A quick search reveals that the ribonucleotide reductase is a normally based in the cytoplasm and is used for synthesizing DNA. From the first glance of the data there is an interesting anomaly where the protein identified is said to have a membrane component in Phobius and TMHMM (Figure 5-6) yet the ribonucleotide reductase normally does not exist there.

Wp_0025: The Wp_0025 gene product is a putative transferase folate-binding protein. According to T-coffee, it was not very conserved in prokaryotes. Swiss Prot revealed matches in eukaryotes such as *Taeniopygia guttata* (zebra finch), *Gia japonicus* (salamander), *Apis mellifera* (honey bee) (Figure 7). This gene is found in the mitochondria of eukaryotes, supporting the theory that mitochondria originated from free-living prokaryotes.

Wp_0026 and Wp_0027:
In the protein database (PDB- Figure 4), these genes code for chaperonin cpn60 and cpn10 with high alignment lengths. Cpn60 and cpn10 also have low e-values and high scores, meaning that these genes are good matches for *Wolbachia*. This shows that these genes will probably fold in the same manner as the given gene. The folds will probably look alike, as well as share similar properties. This would lead one to believe that the functions of these genes may also be similar.
The initial proposed product of these genes by GEN-ACT was an uncharacterized anaerobic dehydrogenase. This gene product proposal was supported by the top BLAST hits for the amino acid sequences, the presence of well-curated functional domains within the amino acid sequences, the cellular location of the amino acid sequences, and the enzymatic function of the amino acid sequences.

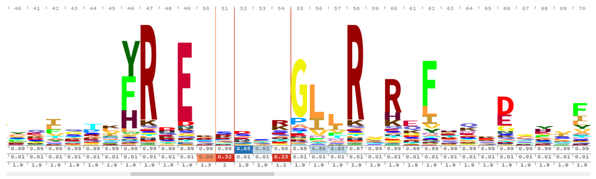

Figure 1- wp_0015 appears to be slightly conserved in middle regions when reviewed by Pfam.

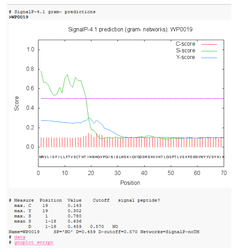

Figure 2- Wp_0019 signal P shows to have no signal peptides at any of the cutoffs.

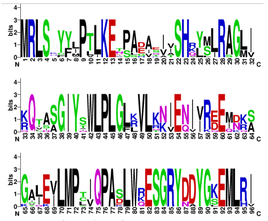

Figure 3- wp_0016 defies the literature (3) by presenting with a conserved end terminus

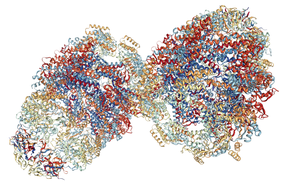

Figure 4 – Wp_0026 shown as a 3D model when inserted into the PBD program.

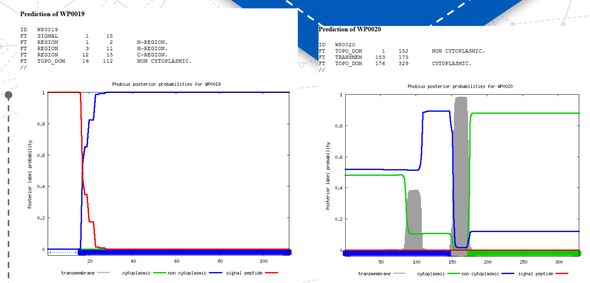

Figure 5 and 6– Wp_0019 (left) and Wp_0020 (right) show predicted proteins and their corresponding region in the sequence as well as the differentiation between cytoplasmic and non- cytoplasmic.

Iba57 IBA57 homolog, iron-sulfur cluster assembly [*Mus musculus* (house mouse)]

Gene ID: 216792, updated on 11-Mar-2018

RecName: Full=Putative transferase CAF17 homolog, mitochondrial; AltName: Full=Iron-sulfur cluster assembly factor homolog; Flags: Precursor
Sequence ID: Q8CAK1.1 Length: 358 Number of Matches: 1

Range 1: 53 to 326 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
119 bits(299)	1e-30	Compositional matrix adjust.	77/274(28%)	131/274(47%)	31/274(11%)

```
Query  6    LNWRSLISLVGPDTRDFLQQVITNDIN--------KLSSQQAIYSLLLNPDGKKYLYDFFL  57
            L  R+L+ + GPD   FL G+ TN++                S +A Y+   LN QG+ LVD  L
Sbjct  53   LDGRALVRVRGPDAAPFLLGLSTNELPLSGPFTGAAQPSARAAYAHFLNVQGRTLYDVIL  112

Query  58   ------IEHDKYIYLECNAHLQQIIEKLDLLKTYLRVRKIKDISSLVKGVLF------  104
                  E    LEC+++ L  + L + K   +V ++      L+   VL
Sbjct  113  YGLPECTEGAPSFLECDSSVLGALQKHLSMVYKIRRKVTVEPSPELHVWAVLPCVPQISE  172

Query  105  DAKLAWCSSKSQVIFQDPRHKLLGWRIIHEDE-----IKEPVGDFIQYEKVRIHNLVPDG  159
             A L +    + ++ +DPR  +G R++ +D+         + +GD   Y K R +  +P+G
Sbjct  173  TAPLEERVEGTMLIRDFPIARMGWRLLTQDDGPALVPRQGLDDLQDYHKIRYQQGIPEG  232

Query  160  AKDWVQNSSFPLQYLIDKINGISFNKGCYIQQEVVNRKMSRQEKFRRKLYLVEGKDALPNI  219
             D+      + FL+  + +HD+SF KGCYIQQE+  R      R++L+ V+ +  LP
Sbjct  233  VCDLEPGMALPLESNLVFMDGVSFTKGCYIQQELIARTHHIGVIRKRLFPVKLEGPLPAS  292

Query  220  GTK-----VISEHNEEVGELRSSVDNIGLALLNI  248
             G      ++ +  G+ R+   ++GLALL +
Sbjct  293  GVSPGAIVTVATGQAAGKFRAQQDRVGLALLRS  326
```

Figure 7 – Wp_00 BLAST results revealing identities, positives, and gaps in the gene sequence

Conclusion

Gene Locus	GENI-ACT Gene Products	Proposed Annotation
Wp_0015	Holo- carrier protein snythase	Holo- carrier protein snythase
Wp_0016	t RNA ligase	t RNA ligase
Wp_0019	Outer membrane protein assembly factor	Outer membrane protein assembly factor
Wp_0020	Ribonuclease-diphosphate	Ribonuclease-diphosphate
Wp_0025	Folate- binding protein	Folate- binding protein
Wp_0026	10 KDa chaperonin	10 KDa chaperonin
Wp_0027	60 KDa chaperonin	60 KDa chaperonin

The GENI-ACT proposed gene products appeared to be the same as the proposed annotation. Significant differences were not noted in any genes so the re-annotation proved successful.

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

Acknowledgments

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

www.buffalo.edu