Annotation of the *Vibrio cholera* O1 El Tor Genome from Locus Tags VC2765, VC2766, VC2767, and VC2768

Maxin McKenzie, Patrick Richardson, Ethan Riggs, Lucianna Russo, and Michael Chiulli
Le Roy Junior-Senior High School, Le Roy, 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 consecutive 4 genes from the microorganism *Vibrio cholera* O1 El Tor (VC2765 – VC2768) 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, and cellular localization data. 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 by the in the database.

Introduction

Vibrio cholera O1 El Tor strain, is a Gram-negative, facultative anaerobe, highly motile, non-encapsulated, and non-endospore forming curved rod-shaped bacterium, found predominantly in brackish water or saltwater. This organism is classified as a chemoheterotroph. Originally discovered as the pathogen that causes cholera in 1854 by Filippo Pacini, it was rediscovered and more widely publicized by Robert Koch in 1884.

Vibrio cholerae is a microorganism of interest for several reasons. This bacterium has been implemented as the etiological agent in cholera, which according to the World Health Organization, infects 1.3 to 4 million people annually, killing 21,000 to 143,000 deaths. (Ali, et al., 2015).

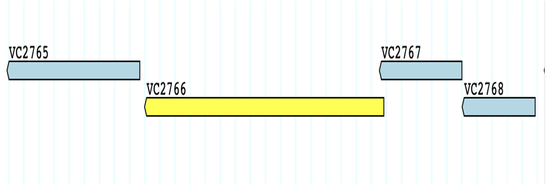

Figure I – Gene neighborhood of the four Vibrio cholerae O1 El Tor genes, VC2765, VC2766, VC2767, and VC2768

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Vibrio cholera* 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

Prediction of VC2768

```
ID   VC2768
FT   TOPO_DOM      1      5       NON CYTOPLASMIC.
FT   TRANSMEM      6     26
FT   TOPO_DOM     27    156       CYTOPLASMIC.
//
```

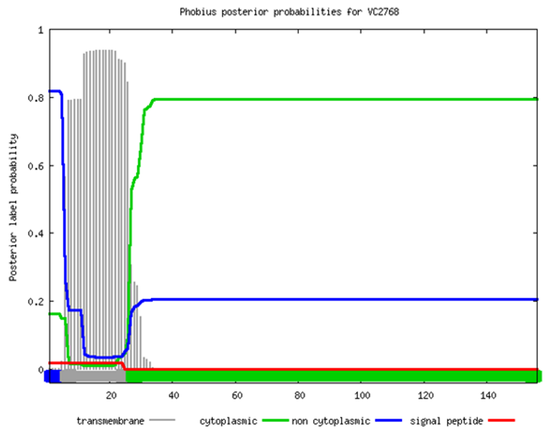

Figure 2 - This picture indicates that 1 transmembrane helix is present in the gene VC2768. This portion of the graph indicated that the protein is embedded in the membrane.

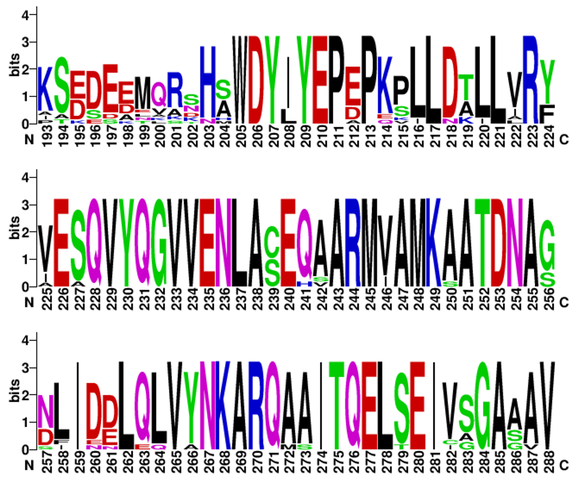

Figure 3 - This is a WebLogo of VC2765 showing that this part of the gene is very important to the construction and function of the protein because it is highly conserved indicated by the large letters.

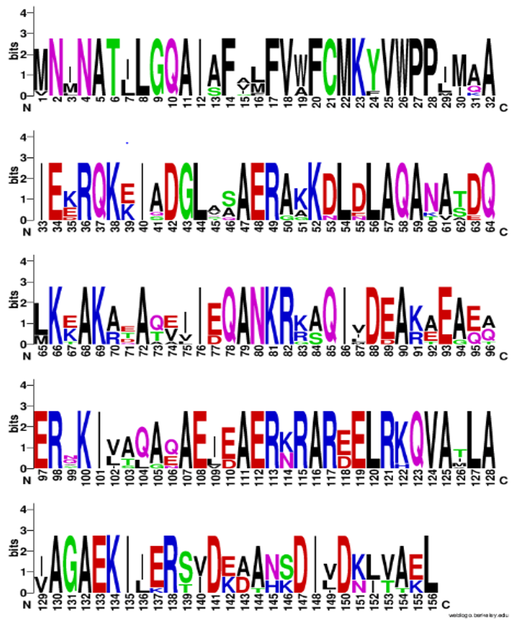

Figure 4 - The large letters mean that the amino acid in the sequence has been conserved. Small letters indicate that the sequence has changed over time and from our results, our amino acid sequence has been mostly conserved throughout time.

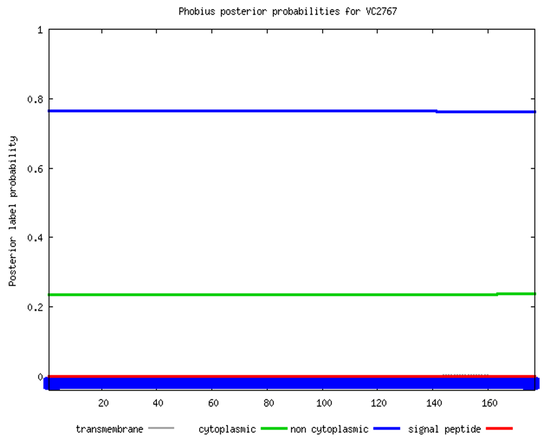

Figure 5 - This Phobius graph demonstrates that the gene VC2767 does not code for any transmembrane helices and therefore falls within the F1 section of ATP synthase.

VC2765:
The initial proposed product of this gene by GENI-ACT was ATP Synthase Gamma Chain. This was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence.

VC2766:
The initial proposed product of this gene was ATP Synthase F0F1 subunit alpha. This was supported by the top BLAST hit for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence.

VC2767:
The initial proposed product of this gene was ATP Synthase Subunit Delta. This was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence.

VC2768:
The initial proposed product of this gene by GENI-ACT was ATP synthase. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence.

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?
VC2765	ATP synthase F0F1 subunit gamma	ATP synthase F0F1 subunit gamma	No
VC2766	ATP synthase F0F1 subunit alpha	ATP synthase F0F1 subunit alpha	No
VC2767	ATP synthase F0F1 subunit delta	ATP synthase F0F1 subunit delta	No
VC2768	ATP synthase F0F1 subunit B	ATP synthase F0F1 subunit B	No

References

Ali et al. (2015). Updated Global Burden of Cholera in Endemic Countries. PLoS Neglected Tropical Diseases 9(6):e0003832. https://doi.org/10.1371/journal.pntd.0003832

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 - R25GM129209