



Annotation of the *Vibrio cholerae* O1 El Tor Genome from Locus Tags VC2769, VC2770, and VC2771

Calvin Herrick, McKenna Cone, Marshall Schultz, and Michael Chiulli
Le Roy Jr./ Sr. High School, Le Roy, NY and The Western New York Genetics in Research and Health Care Partnership

Abstract

A group of three consecutive genes from the microorganism *Vibrio cholerae* O1 El Tor (VC2769 – VC2771) 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 in the database.

Introduction

Vibrio cholerae O1 El Tor strain, is a Gram-negative, highly motile, non-encapsulated, facultative anaerobe, 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 because it has been identified as the etiologic 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).

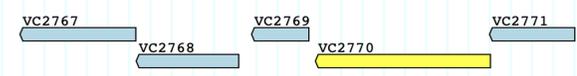


Figure 1 – Gene neighborhood of the three *Vibrio cholerae* genes: VC2769, VC2770, and VC2771

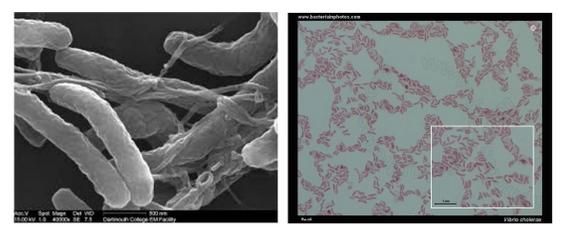


Figure 2 – Microscopic images of *Vibrio cholerae* (<http://www.bacteriainphotos.com>)

Methods

Modules of the GENI-ACT (<http://www.geni-act.org/>) were used to complete *Vibrio cholerae* 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, Psorb, 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

VC2769:

The computer proposed product of this gene was an ATP synthase subunit C. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, information found through TIGRFAM, Pfam, and PDB. As such, the proposed annotation remains ATP synthase. ATP synthase is an enzyme that generates ATP during cellular respiration. The enzyme is found in all living organisms, and is used to power all cellular activities in organisms.

VC2770:

The computer proposed product of this gene was an ATP synthase subunit A. This was supported by the top BLAST hits for the amino acid sequence and by using information found through TIGRFAM, Pfam, and PDB. ATP synthase subunit A is responsible for allowing protons to flow through the membrane at any of its six transmembrane helices to produce ATP.

VC2771:

The computer proposed product of each of these genes was an ATP synthase subunit I. This proposal was further supported by information gathered by using TIGRFAM, Pfam, and PDB. This helps the ATP synthase produce the required ATP.

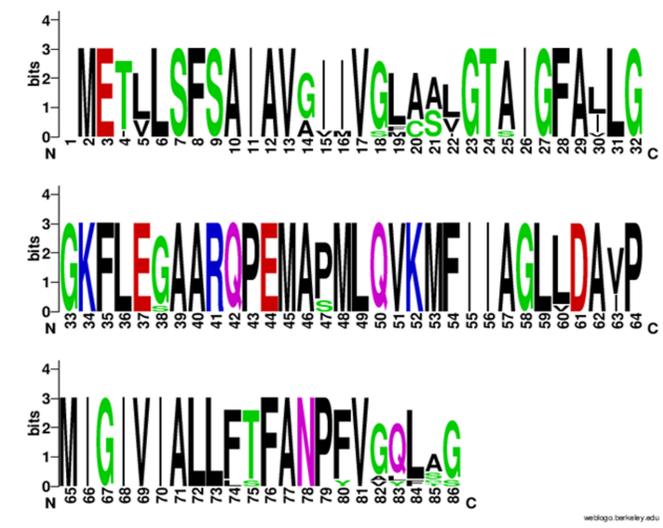


Figure 3 –WebLogo results for VC2769.. The image shows predominantly highly conserved amino acid regions in the VC2769 genome. There are small amounts of amino acid variability in the genome, especially from amino acids 18 to 22.

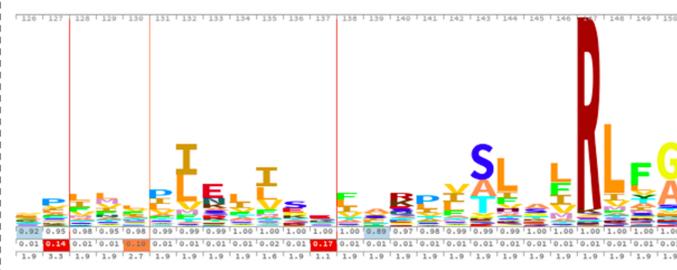


Figure 4 – A portion of the HMM logo for VC2770. On the HMM logo, the larger letters represent more conserved genes. Colors correspond to different types of amino acids such as neutral and acidic. Letters are sorted in descending order depending on their probability of occurring at a given position in a sequence that contains the domain.

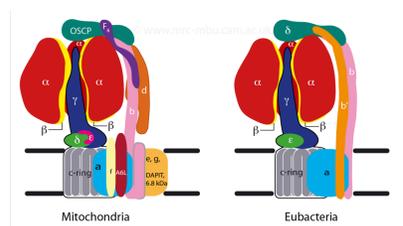


Figure 5 – Diagram of ATP synthase, which is composed of many subunits. Subunits A and C consist of transmembrane helices which allow for the transportation of protons to form ATP.

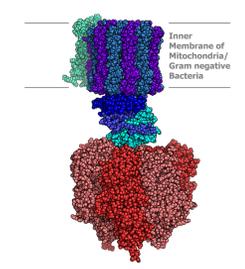


Figure 6 – Diagram of ATP synthase, which is composed of two parts (F0 and F1). Both parts generate ATP through motors. VC2769's proposed product was ATP synthase subunit C, which makes up part of the F0 portion of ATP synthase.

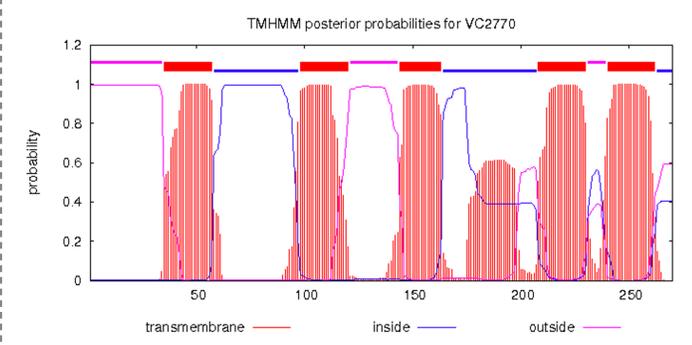


Figure 7 – TMHMM results for VC2770. The six red spikes on the diagram indicate the portions of the Cholera protein that are embedded in the membrane.

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? |
|-----------|----------------------------------|-----------------------------|-------------------|
| VC2769 | ATP synthase F0F1 subunit C | ATP synthase F0F1 subunit C | No |
| VC2770 | ATP synthase F0F1 subunit A | ATP synthase F0F1 subunit A | No |
| VC2771 | ATP synthase F0F1 subunit I | ATP synthase F0F1 subunit I | 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>