# Annotation of the *Vibrio cholerae* Genome at Locus Tag VC395_1566

Ziona Cotton and Karen McCann
School Without Walls Rochester, New York 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 gene from the microorganism *VC395_1566* was annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name for the 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 differ from the proposed gene annotation for the gene and as such, the gene does not appear to be correctly annotated by in the database.

## Introduction

*Vibrio cholerae* is a gram negative, non-spore forming, curved rod. It is oxidase positive which means when an oxidase test is performed it shows bacteria produces cytochrome. It is very motile, has a single flagellum,and is a facultative anaerobe. *Vibrio cholerae* is a part of the Vibronaceae family. Serogroups O1 and O139 are mostly responsible for cholera breakouts. When this bacteria enters the body some of the bacteria are killed by stomach acid. However, some are still able to travel to the small intestine, causing a loss of large amounts of salt and water in the form of diarrhea. This results in dehydration and when severe can cause death. *V. cholerae* is a natural inhabitant of enclosed coastal bodies of brackish water. These organisms mostly spend their entire life cycle in estuarine environments and when a human host is encountered by the organism the disease cholera starts it's development. Cholera outbreaks normally happen due to feces contaminating the waters.

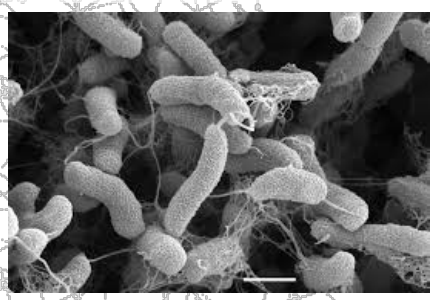
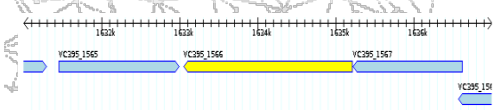


Figure 1. The locus tag and relative position of the gene under investigation in this research.




## 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 |
|---|---|---|
| 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? |
| Module 7- Gene Duplication/ Gene Degradation | Paralog, Pseudogene | Are there other forms of my gene in the bacterium? Is my gene functional? |
| Module 8- Evidence for Horizontal Gene Transfer | Phylogenetic Tree, | Has my gene co-evolved with other genes in the genome? |
| Module 9- RNA | RFAM | Does my gene encode a functional RNA? |

## Results

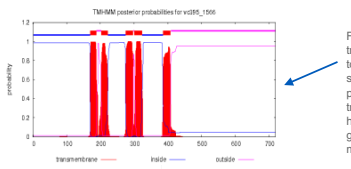
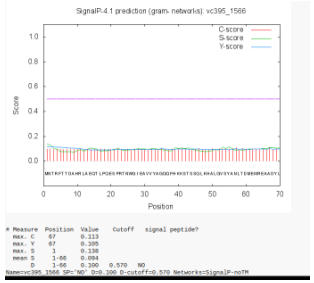



Figure 2. This is a transmembrane topology graph that shows the number of predicted transmembrane helices in the genomes cell membrane

Figure 3. The SignalP graph indicates if there is a signal peptide present and will suggest that the protein is secreted in my bacterium.

This graph shows no signal peptide which indicated that the bacterium in NOT secreted.



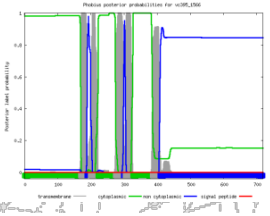




Figure 4. The phobius combines the TMHMM and signalp in one and acts as a confirmation test for the transmembrane helixes and signal peptides.

TMHMM indicates that there were 5 transmembrane helices, SignalP reveals low probability for signal peptide, pSORTB predicted VC395_1566 to be a cytoplasmic protein. Due to my results I believe that VC395_1566 is a cytoplasmic membrane protein

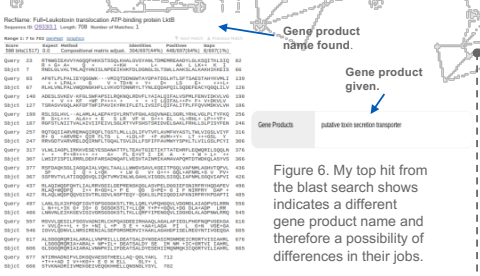



Figure 6. My top hit from the blast search shows indicates a different gene product name and therefore a possibility of differences in their jobs.

COG Hit- When doing my COG search the results came back with only one COG named ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal double-glycine peptidase domain which was significant because I received a e-value of 0e+00.

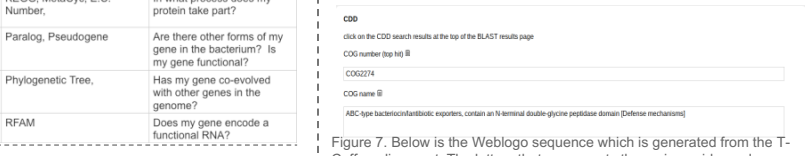



Figure 7. Below is the Weblogo sequence which is generated from the T-Coffee alignment. The letters that represents the amino acids are large even though they are stemming from different genus. The gaps throughout the sequence is caused by one protein having extra amino terminals

From 35-135aa in the weblogo sequence the amino acid alignment was not as well conserved and fluctuated a lot. From 136-145aa there was a gap left, and from 154-740aa the amino acid alignment began to have high conservation. The one area with the most conservation showed that, that part of the amino acid sequence is most likely important.

VC395_1566

- The initial proposed product of this gene by GENI-ACT was a putative toxin secretion transporter. This gene product was not supported but was more of a hypothetical product.
- This was not supported by the top BLAST hits for the amino acid sequence which gave the function to be a leukotoxin translocation ATP- binding protein. Combining this information with the evidence collected from the cellular localization data module it was found to be a cytoplasmic membrane protein.
- The proposed annotation for this gene product is part of an ABC toxin transporter that has multiple parts to translocate a product across a membrane This was supported by the BLAST hits for the amino acid sequence as well as the COG hit.

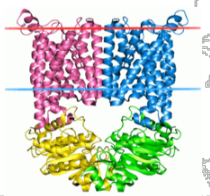


Figure 8. This is an example of an ABC transporter although not my specific protein

## Conclusion

The GENI-ACT proposed gene product did differ from the proposed gene annotation for the gene and as such, the gene appear to be incorrectly annotated by the computer's database. The proposed annotation is a ABC toxin transporter protein.

| Gene Locus | GENI-ACT Gene Product | Proposed Annotation |
|---|---|---|
| VC395_1566 | Putative toxin secretion transporter | Part of an ABC toxin transporter |

## References

Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (541T). Standards Genomic Sciences,12 - 20.

## Acknowledgments

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

www.buffalo.edu