

Annotation of Locus Tags CJE0004, CJE0005 and CJE0012 from the *Campylobacter jejuni* Genome

Christopher Murphy, June Fortner, Nafim Rahman and Dana Pryor-Moncrieffe

Frederick Law Olmsted Science Club and the Western New York Genetics in Research and Health Care Partnership

Abstract

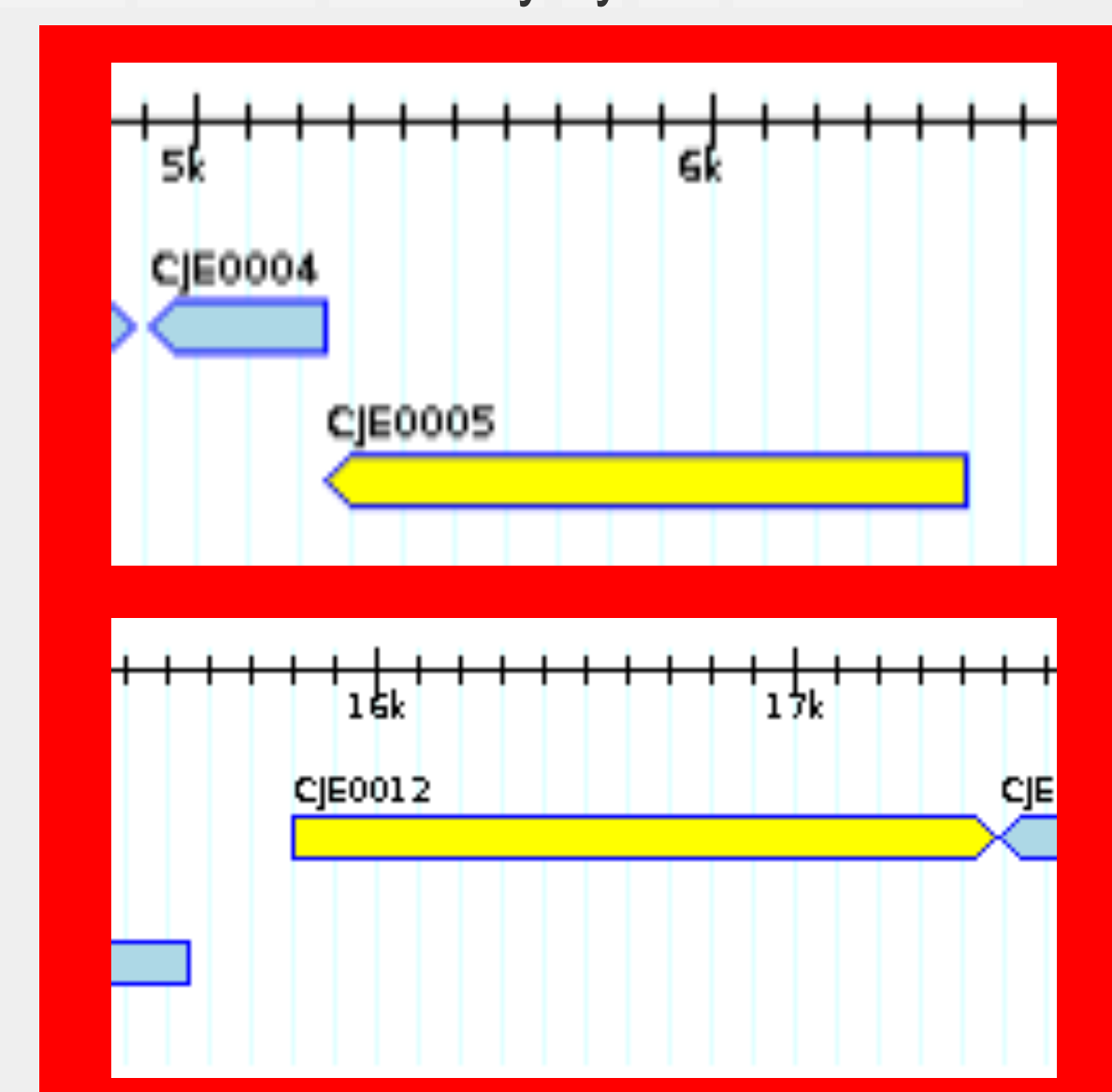
A group of 3 genes from the microorganism *Campylobacter jejuni* (CJE0004, CJE0005, and CJE0012) were annotated using the collaborative genome annotation website GENI-ACT (Genomics Education National Initiative - Annotation Collaboration Toolkit). The Genbank proposed gene product name. Each gene was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence, cellular localization data, and potential alternative open reading frames. While the Genbank proposed that our genes were very similar, they also had some variations in certain aspects of the research, such as in the Structure Based Evidence module. The gene products are a hypothetical protein, a molybdopterin oxidoreductase family protein, and dihydroxy-acid dehydratase. The genes are well discovered but we still have much to learn about them. While we did not finish all of the GENI-ACT modules, we did learn a lot about these genes.

Introduction

Campylobacter jejuni is a gram negative, helical shaped bacterium most commonly related to food poisoning. The *Campylobacter* genome is most commonly obtained when poultry is eaten raw, but it can be obtained through almost any other raw meats. The genome is most commonly related to bacterial enteritis in the United States. *Campylobacter jejuni* was first identified as a bacterial pathogen in 1973. The pathogen is not just known to react in a human's immune system and can lead to illness in cats, dogs, and other animals (Sean F. Altekruze, Norman J. Stern, Patricia I. Fields, and David L. Swerdlow, 2010).

Campylobacter jejuni was identified as bacterial caused infection and was first examined in 1886 by Escherich in a chicken's diarrhea. The genome was further examined and was first researched in humans in 1957 where we were informed of the bacteria which causes the illness. Guillain-Barre Syndrome is often associated with an infection by the *Campylobacter jejuni*, and there are a lot of unknown questions which could be answered with more information about the genome. For example, we might be able to answer the unknown question of why this syndrome hits some but not others.

Our team annotated three different gene sequences from the *Campylobacter jejuni* genome. After analyzing the sequences, we have determined the gene products: a hypothetical protein, a molybdopterin oxidoreductase family protein, and dihydroxy-acid dehydratase to have been named correctly by the GENI-ACT.



The image to the left shows the DNA coordinates of our gene sequences. CJE0004 starts at about 4900 and ends at around 5200. CJE0005 starts at around 5200 and ends at around 6500. CJE0012 starts at around 15800 and ends around 17500.

Methods and Materials

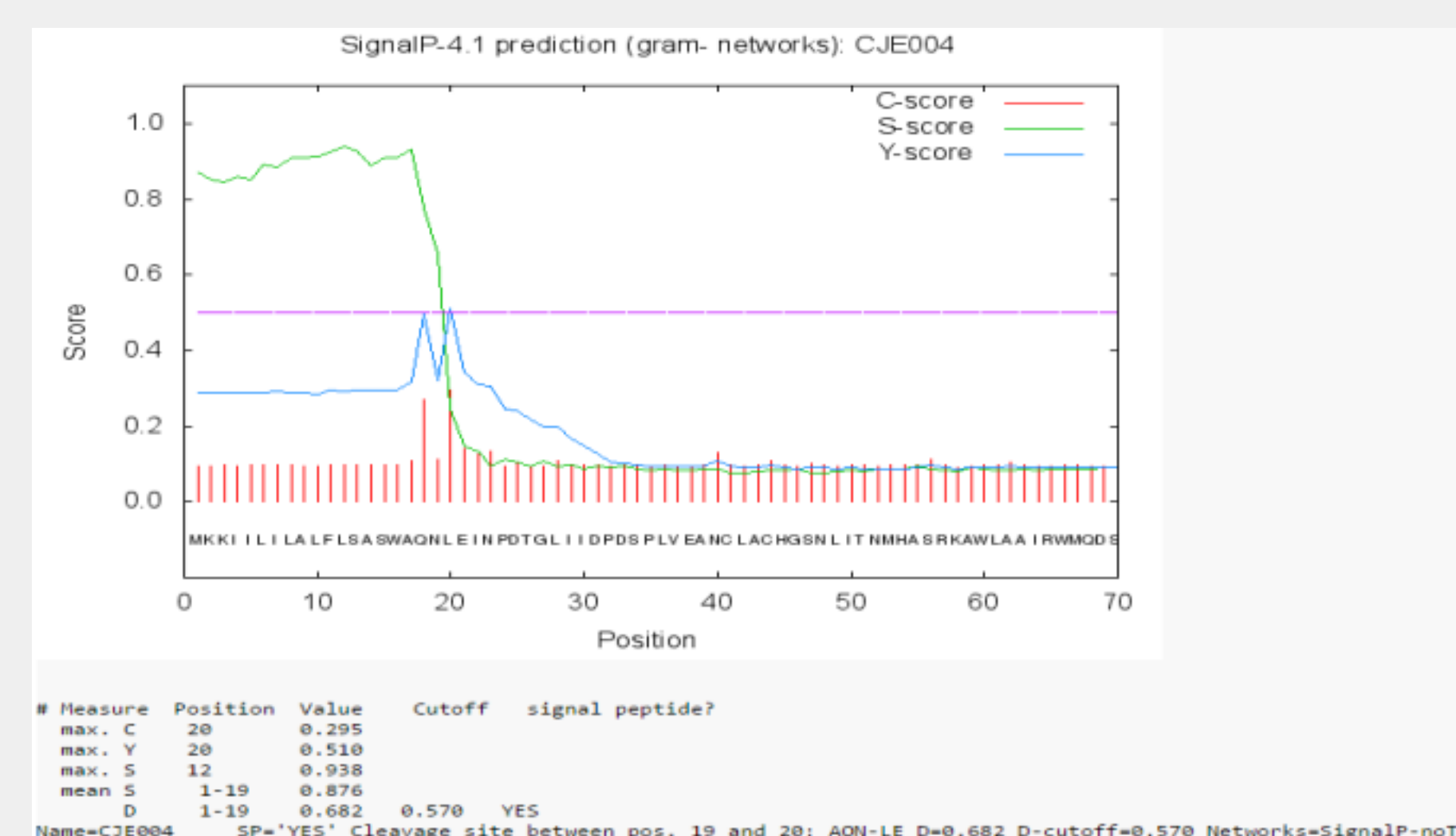
We used GENI-ACT (<http://www.geni-act.org/>) modules to complete our annotation of the *Campylobacter jejuni* genome. The modules we completed are described below.

Modules	Activities	Questions
1: Basic Information	DNA Coordinates DNA Sequence Protein Sequence	What are the protein sequence and DNA sequence of the gene? What is the gene's DNA coordinates?
2: Sequence-Based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Does my gene share any similarities with other genes in Genbank?
3: Structure Based Evidence	TIGRFam, Pfam, PDB	Are any of the domains in my gene's protein functioning?
4: Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	What is the Gram stain of the microbe? Where is the protein located?
5: Alternative Open Reading Frame	IMG Sequence Viewer for Alternate ORF Search	Was the protein sequence given by the computer correct?

Results

Campylobacter jejuni 0004:

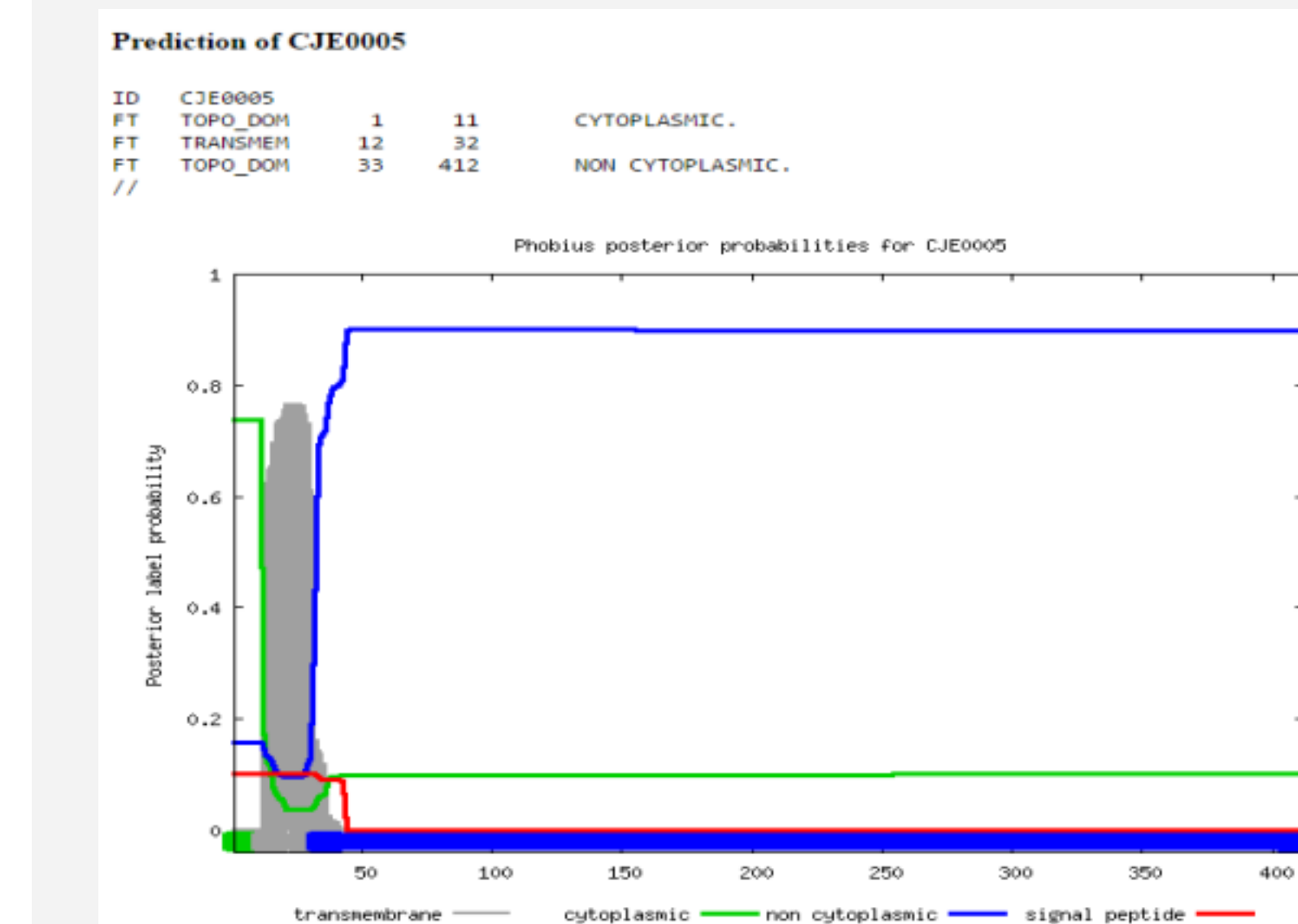
The initial proposed product of this gene by GENI-ACT was a conserved hypothetical protein. This gene product proposal was supported by the first BLAST hit using the amino acid sequence of the gene and the presence of well-curated functional domains within the amino acid sequence. Based on all the Cellular Localization data it was determined that the gene is strongly predicted to be a signal peptide that is very close to the cytoplasmic membrane based on the results of the data we received. Given these results the gene is most likely a hypothetical protein.



This image shows the signalP probability of a signal peptide present in the membrane of the gene. The graph shows a signal peptide cleavage point between amino acids 19 and 20. LipopP and Phobius also predicted the presence of a signal peptide, but with cleavage between amino acids 17 and 18 in the sequence. The presence of a signal peptide indicates the protein encoded by CJE0004 is likely secreted to the periplasm or the exterior.

Campylobacter jejuni 0005:

The initial proposed product of this gene by GENI-ACT was a molybdopterin oxidoreductase family protein. This was supported by the first blast hit for the amino acid sequence in the non-redundant protein sequences database. This gene product proposal was also supported by the TIGRFAM HMM search. The location of the protein in the cell was difficult to find due to the fact that its PSORT-B scores are all the same and it has no identifiable cleavages. According to the Phobius prediction, however, the protein may be found in the membrane since a single transmembrane helix is predicted.



Phobius results for CJE0005 showing the presence of a single transmembrane helix

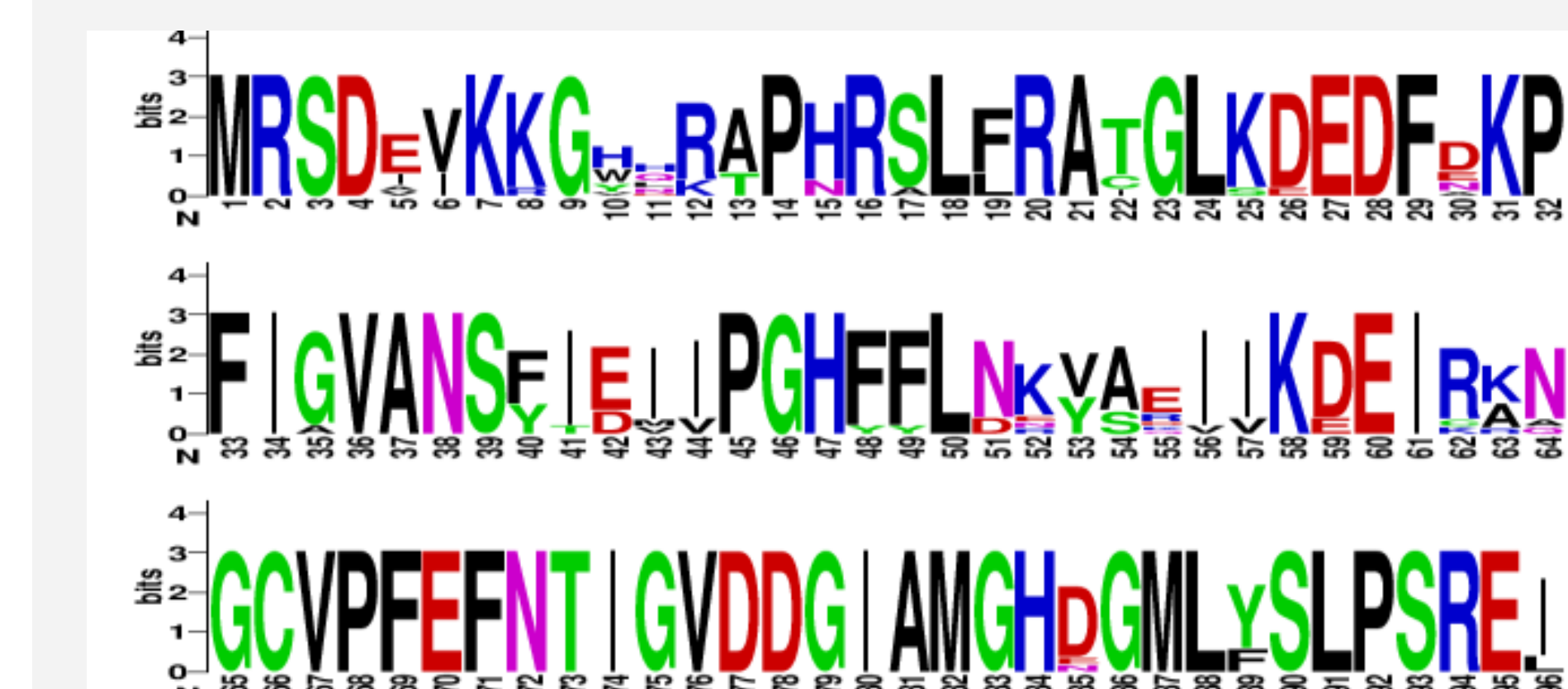
PSORTb Results

SeqID: CJE0005 amino acid sequence
Analysis Report:
 CNSV- Unknown [No details]
 CytoSVM- Unknown [No details]
 ECVS- Unknown [No details]
 ModHMM- Unknown [No internal helices found]
 Motif- Unknown [No motifs found]
 OMPMotif- Unknown [No motifs found]
 OSMV- Unknown [No details]
 PPSVM- Unknown [No details]
 Profile- Unknown [No matches to profiles found]
 SCL-BLAST- Unknown [No matches against database]
 SCL-BLAST- Unknown [No matches against database]
 Signal- Unknown [No signal peptide detected]
 Localization Scores:
 Cytoplasmic: 2.00
 CytoplasmicMembrane: 2.00
 Periplasmic: 2.00
 OuterMembrane: 2.00
 Extracellular: 2.00
 Final Prediction: Unknown

PSORTb results for CJE0005 showing the tool was unable to the protein's location..

Campylobacter jejuni 0012:

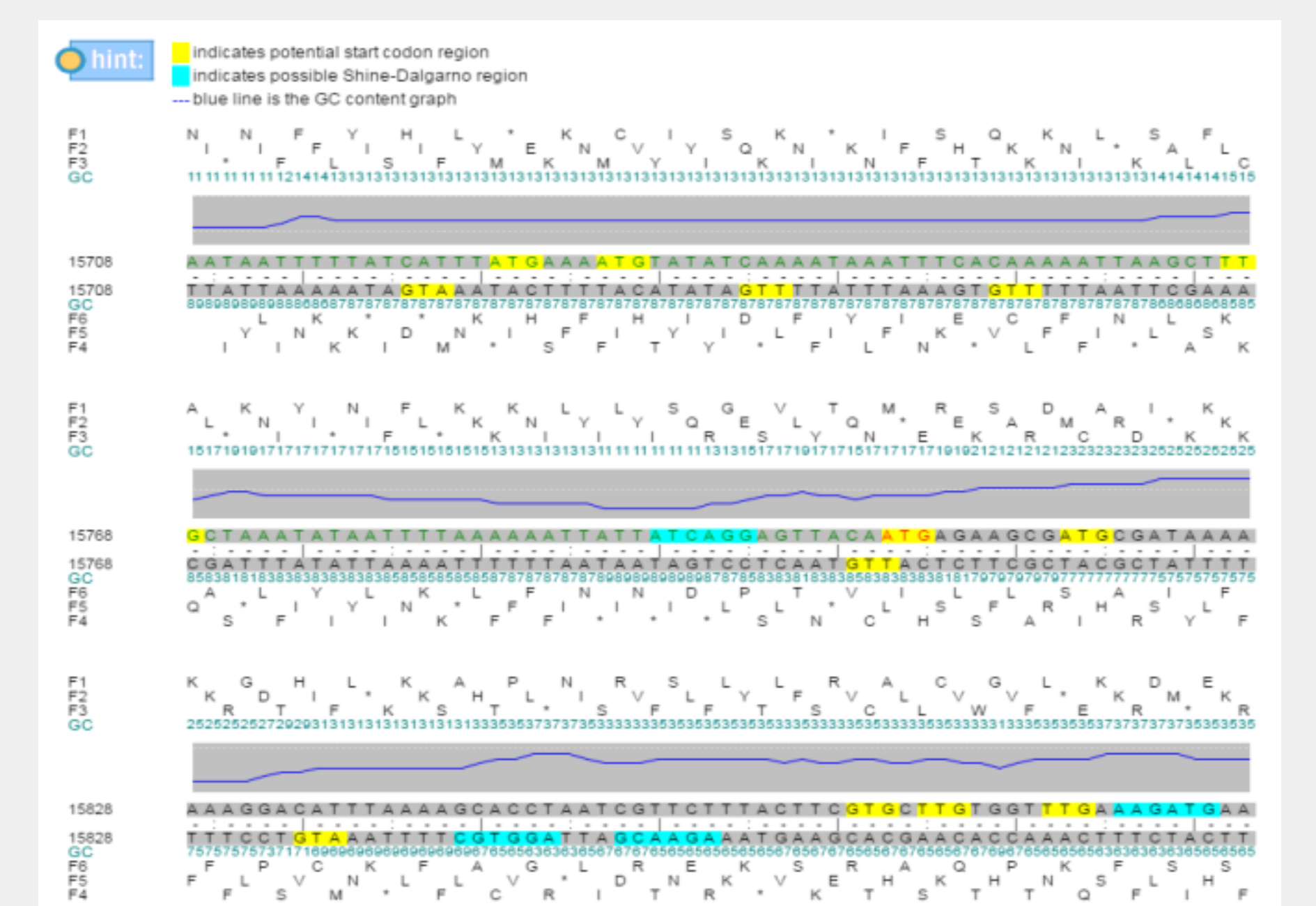
The initial proposed product of these genes by GENI-ACT was a dihydroxy-acid dehydratase. This gene product proposal was supported by the top BLAST hits for the amino acid sequences from curated Swiss-Prot database, the cytoplasmic location of the amino acid sequences, and structural evidence. This proposal is also supported by the IMG sequence viewer, which determined the protein's start codon correctly called. As such the proposed annotation is a dihydroxy-acid dehydratase.



The above image is from the WebLogo representation of the sequences found among different organisms. This demonstrates that the gene is well conserved across different organisms.

The chart to the right shows the PSORTb results, which predict CJE0012 to be cytoplasmic.

Localization Scores:	
Cytoplasmic	9.97
CytoplasmicMembrane	0.01
Periplasmic	0.01
OuterMembrane	0.00
Extracellular	0.00
Final Prediction:	
Cytoplasmic	9.97



IMG Sequence Viewer for alternative open reading frame search for CJE0012. A Shine-Dalgarno region (aqua) is upstream from the predicted start codon (in red), suggesting the protein sequence was most likely called correctly

Conclusion

The GENI-ACT proposed gene product did not differ at all 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 and GENI-ACT.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
CJE0004	Hypothetical Protein	Hypothetical Protein
CJE0005	molybdopterin oxidoreductase family protein	molybdopterin oxidoreductase family protein
CJE0012	Dihydroxy-acid Dehydratase	Dihydroxy-acid Dehydratase

References

Altekruze SF, Stern NJ, Fields PI, Swerdlow DL. *Campylobacter jejuni*—An Emerging Foodborne Pathogen. *Emerg Infect Dis.* 1999, Feb, [cited May 9 2016].
 Acheson, David, and Ban M. Allos. *Campylobacter jejuni Infections: Update on Emerging Issues and Trends.* Oxford: Oxford University Press, 2001. Web. 9 May 2016.

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25OD010536
 Dr. Stephen Koury and Dr. Rama Dey-Rao