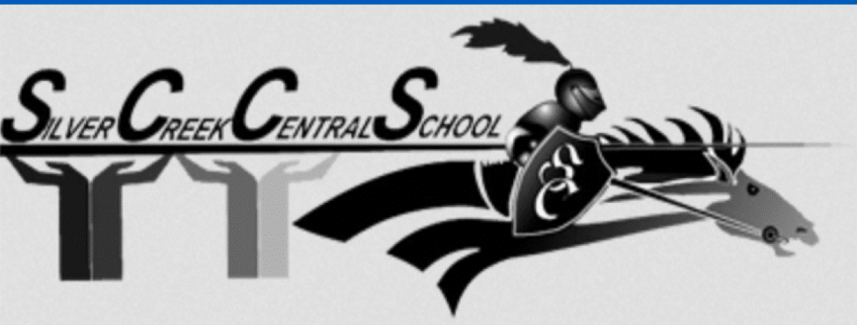


# Annotation of the *Camphylobacter jejuni* RM1221 Genome at Locus Tag Cj\_0281c to Cj\_0285c

Kristen Wielgasz\*, Olivia Watroba\*, Joseph Bowers\*, Amanda Roth and Cheryl A. Smith

Silver Creek High School and the Western New York Genetics in Research and Health Care Partnership



## Abstract

The purpose of the Geni-Act bioinformatics research was to prove the computer programs used correctly analyzed genes from *Camphylobacter jejuni* RM1221. The genes researched were at locus tags Cj\_0281c through Cj\_0285c. Geni-Act was connected to different modules designed to analyze the genes using the basic amino acid sequences, sequence similarities, structure-based evidence, cellular localization, open reading frame analysis and enzymatic function. Results for the gene products of these genes were generally as expected from the Genbank.

## Introduction

*Camphylobacter jejuni* RM1221 is a gram negative bacteria. The name *Camphylobacter* ,meaning “curved rod”, was first proposed in 1961 by Sebald and Veron(1) due to the spiral shaped structure of the bacteria. *Camphylobacter* is found in fecal matter of animals and estimated to affect 1.3 million people every year(2) by causing food poisoning and diarrhea. Most people get infections from undercooked meat, raw milk or contaminated water.

*Camphylobacter jejuni* is highly susceptible to fluctuations in the environment including changes in temperature and pH which is why proper food preparation typically kills the bacteria. Infections are typically treated with antibiotics such as azithromycin and ciprofloxacin.

The purpose of this study was to manually annotate the genome of the bacterium and to describe the function of the gene products of *C. jejuni* RM1221.



Figure 1: *Camphylobacter jejuni* (CDC, 2017)

## Methods

Modules of the GENI-ACT (<http://www.geni-act.org/>) were used to complete *Camphylobacter jejuni* RM1221 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

**Cj\_0281c:**  
Cj\_0281c consists of 325 amino acids which make the protein transaldolase. This enzyme is located in the cytoplasm, does not have a signal peptide and does not have any transmembrane helices according to results from TMHMM and Signal P. The location of the gene may be incorrect due to a lack of a Shine-Delgarno sequence prior to the start codon. The proposed new coordinate for the start would be 259077.

**Cj\_0282c:**  
The proposed gene product for the gene Cj\_0282c from the GenBank was phosphoserine phosphatase SER B. In comparison to the orthologs, the gene shows a high degree of alignment and conservation. This was evidenced by BLAST results from Swiss-Prot and Non-Redundant databases as well as Weblogo.

**Cj\_0284c:**  
The gene product for locus tag Cj\_0284c is a Histidine Protein Kinase. It has a well conserved amino acid sequence as shown through WebLogo; prominent with red acidic amino acids and black hydrophobic amino acids. Cj\_0284c annotation results from SignalP and PSORT-B conclude that it is a cytoplasmic protein. The function of this protein seems to be involved in DNA replication with evidence from MetaCyc and KEGG.

**Cj\_0285c:**  
The proposed gene product for Cj\_0285c is chemotaxis protein. The coordinates are consistent with the predicted values. This gene is well conserved as evidenced by Weblogo. There is no evidence to show the protein is an enzyme and it is most likely the product of an enzyme (reaction). Cj0285c has a paralog, but it is not a pseudogene. Cj0285c is cytoplasmic according to TMHMM, SignalP, and LipoP.

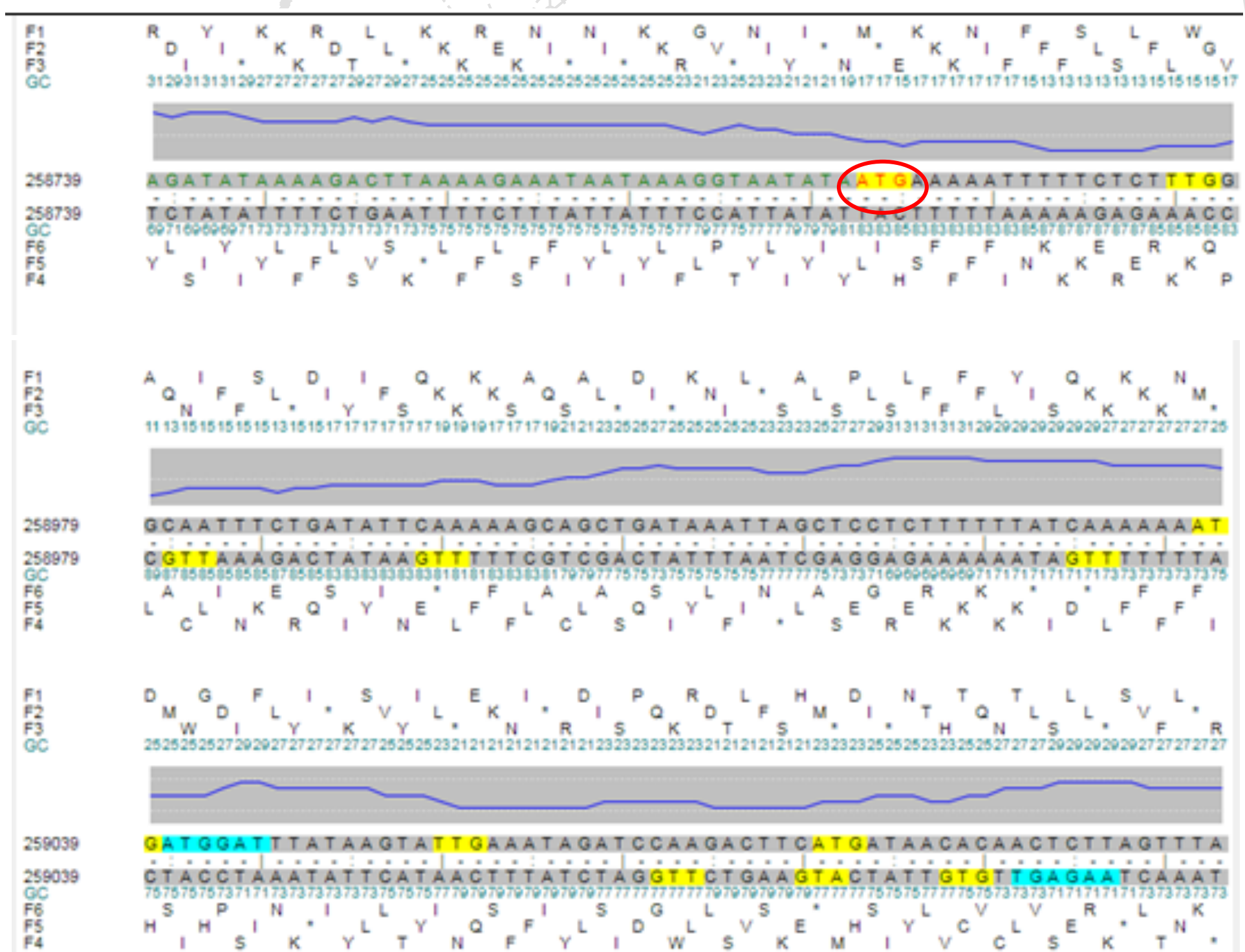


Figure 2: Cj\_0281c results showing absence of Shine-Dalgarno sequence upstream of called start codon. Area of potential new start codon at coordinate 259077.



Figure 3: Cj\_0284c protein crystalline structure of histidine protein kinase

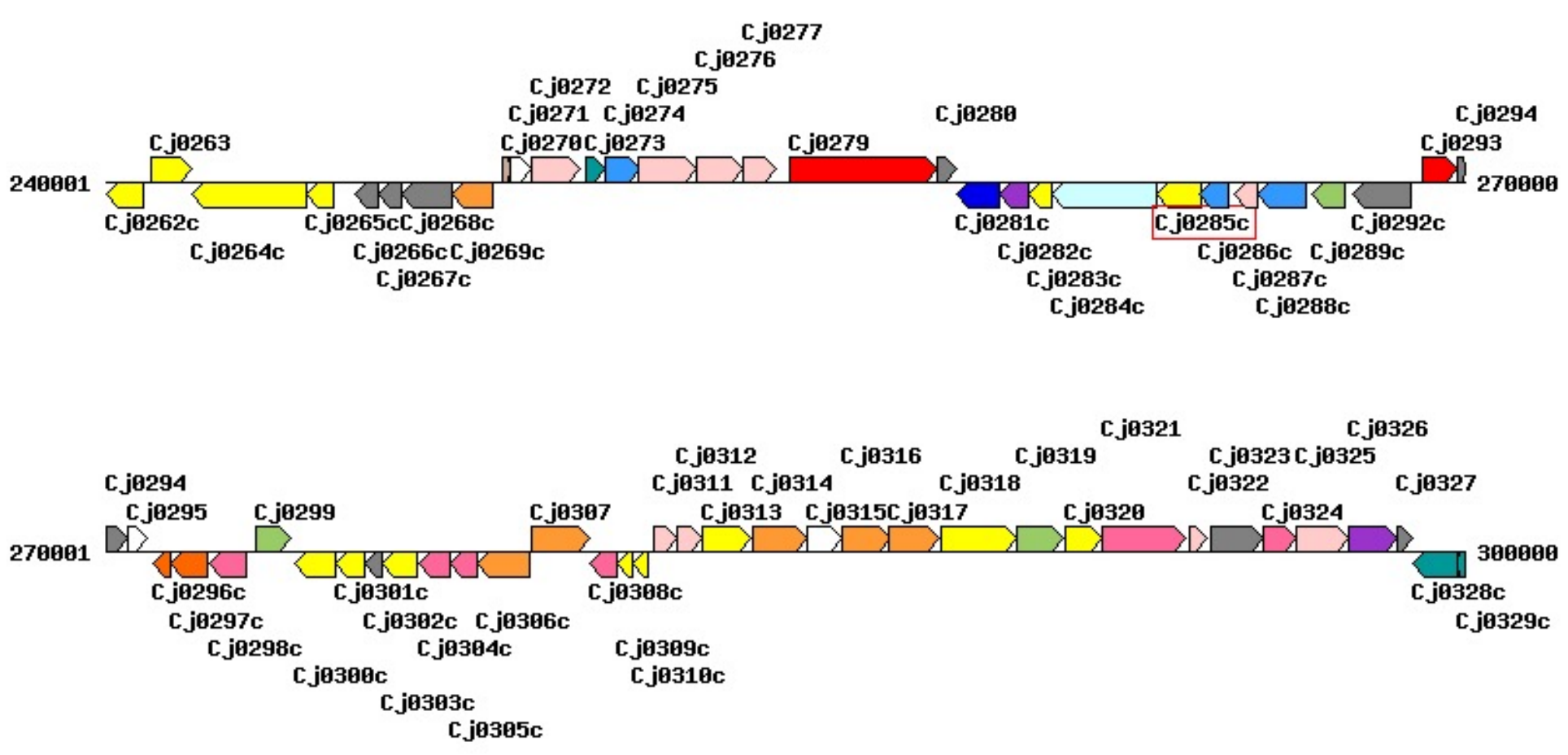


Figure 4: Gene neighborhood of Cj\_0285c.

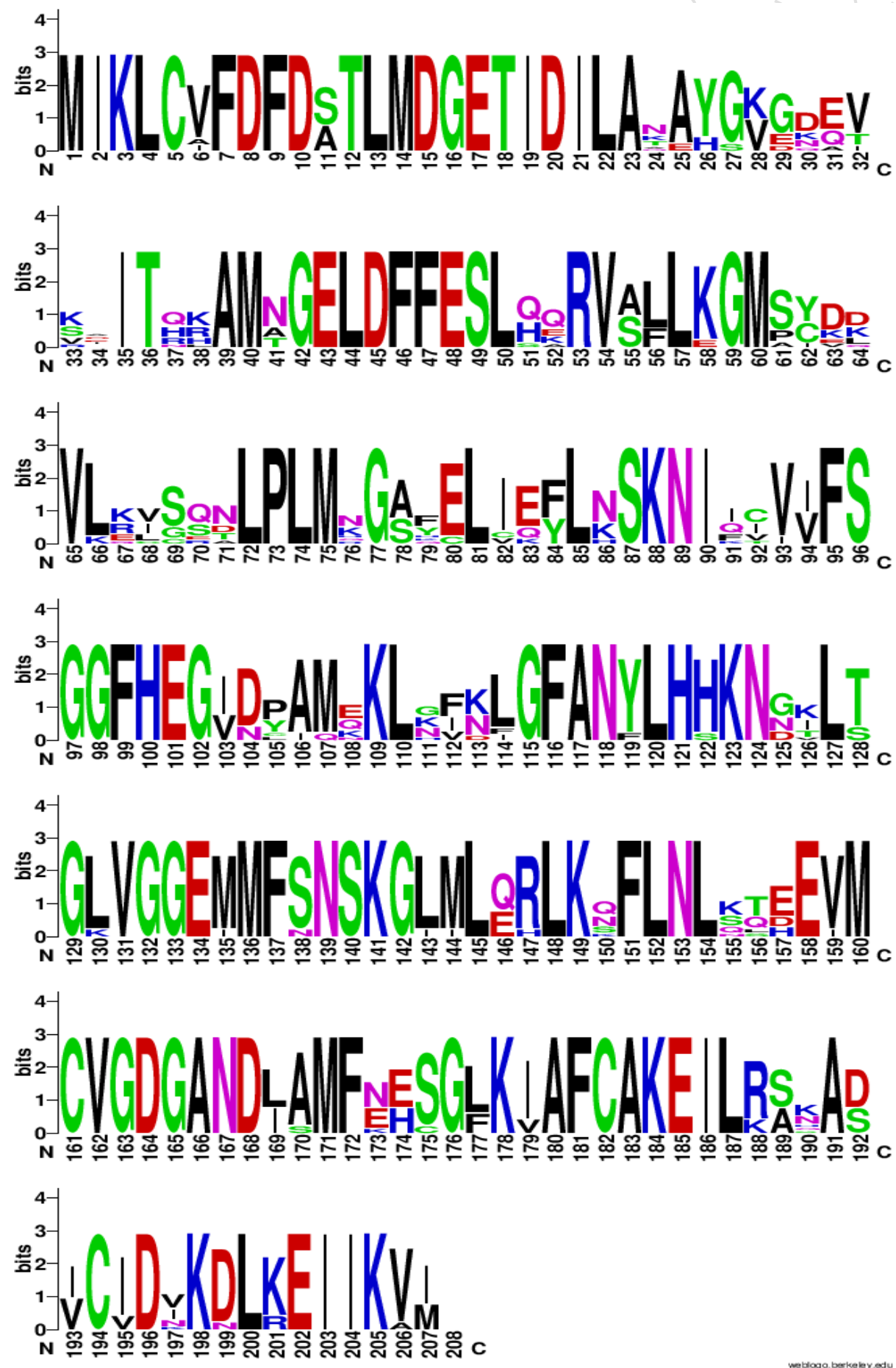


Figure 5: Cj\_0283c partial Weblogo indicating the protein phosphoserine phosphatase SerB is well conserved across a variety of species

## Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation from GENBANK.

### Gene Locus

Cj\_0281c  
Cj\_0282c  
Cj\_0284c  
Cj\_0285c

### Proposed Gene Product

Transaldolase  
Phosphoserine phosphatase SerB  
Histidine Protein Kinase  
Chemotaxis protein CheY Family

## References

"New York: Possible Campylobacter Jejuni Linked to Unpasteurized Raw Milk Prompts Warning." *Outbreak News Today*, 18 Sept. 2017, outbreaknewstoday.com/new-york-possible-campylobacter-jejuni-linked-unpasteurized-raw-milk-prompts-warning-19756/.

"Campylobacter (Campylobacteriosis)." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 2 Oct. 2017, www.cdc.gov/campylobacter/campy-antibiotic-resistance.html.

## Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25OD010536  
Stephen Koury PhD; Department of Biotechnical and Clinical Laboratory Sciences

[www.buffalo.edu](http://www.buffalo.edu)