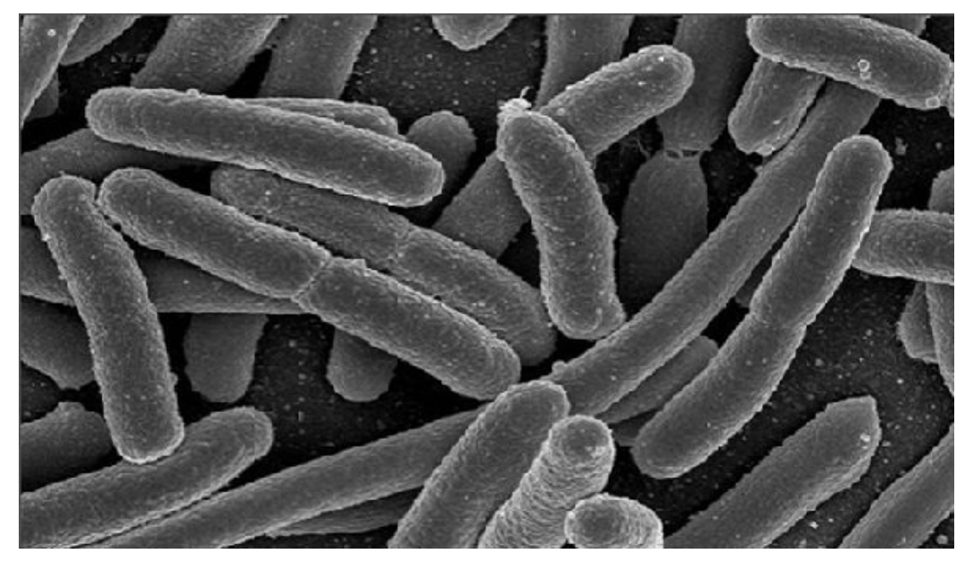
Annotation of the *Propionibacterium acnes* 266 Genome from DNA coordinates 23057 to 27705

Elizabeth Gu*, Ian Shea*, Jordan Marshall*, Caroline Geiger and Kathleen Kramer
Williamsville East High School, 151 Paradise Road, East Amherst, 14051 and The Western New York Genetics in Research Partnership

Introduction

From the microorganism *Propionibacterium acnes (266 CP002409)*, we annotated using the 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, cellular localization data, potential alternative open reading frames, enzymatic function, presence or absence of gene duplication and degradation,the possibility of horizontal gene transfer, and the production of an RNA product. 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 r database.



Methods

To annotate our respective genes, we used Geni-Act to record our findings and annotations. The Basic Local Alignment Search Tool provided by the NCBI and NIH revealed similarities between biological sequences, imparting information on the annotated organism. T-Coffee supplied multiple sequence alignments, which were subsequently used to create sequence logos in WebLogo. PSORT-B and Phobius helped determine the location of the respective gene products based on cellular localization data.

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?

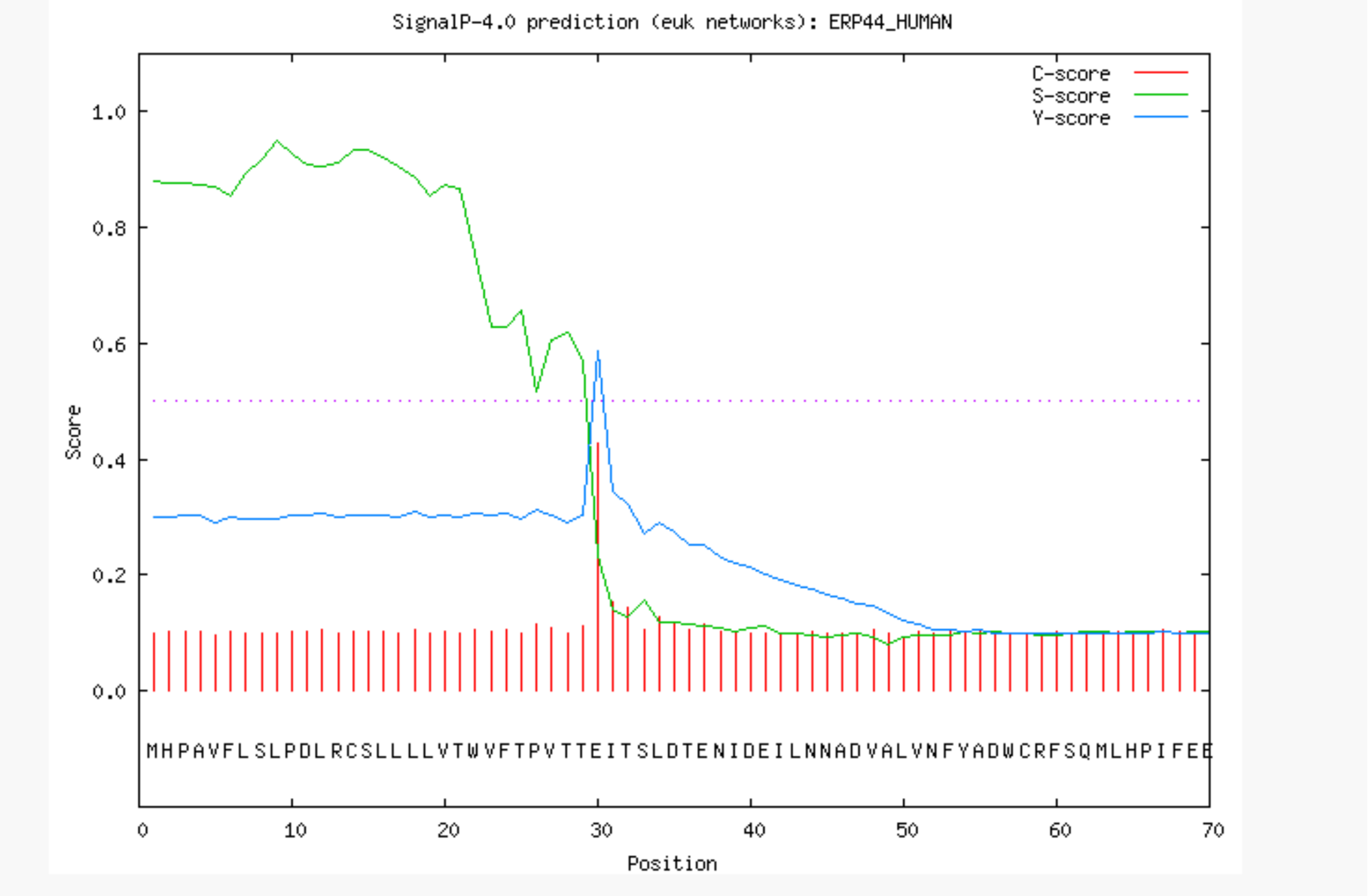
The Phylogeny. fr tool provided insight into organisms with similar gene sequences and genomes.
The collection of the data obtained from the sources brought about the successful annotation of our respective genes.

Data Analysis

Locus PAZ_c00220:
Our results showed that this locus had a strong similarity to the PTS fructose transporter subunit IIA [Cutibacterium acnes] and stains gram-positive. Our signal peptide probability and LipoP predictions revealed that the protein is most likely expected to be located in the cytoplasm of the cell.

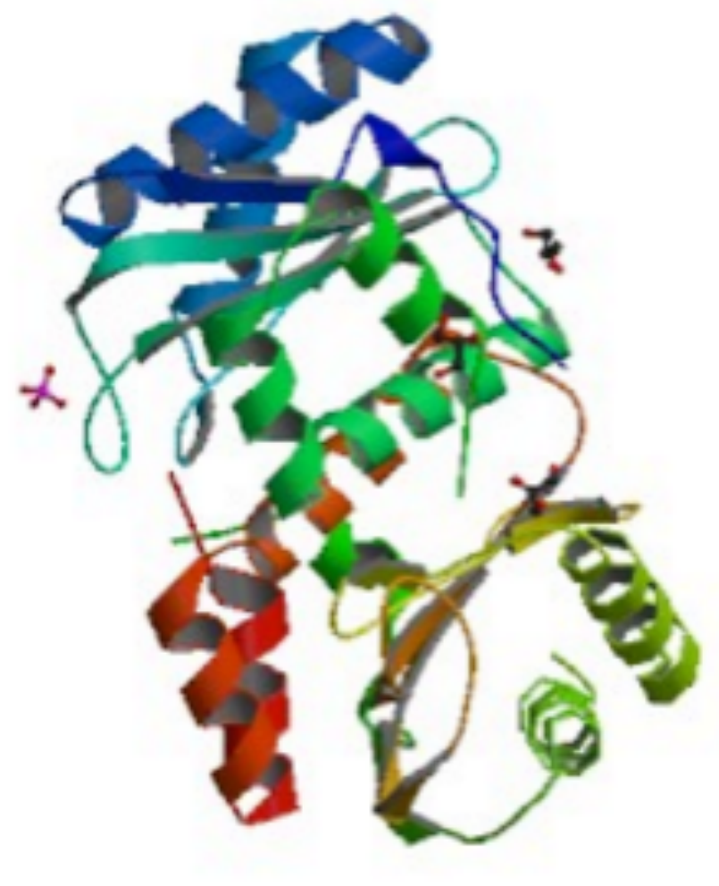
Locus PAZ_c00270:
Our results showed that the organism stains gram-positive and was involved in the PTS system (subunit IIA), which allows for the uptake of fructose into bacteria via an operon. Using the signal peptide probabilities and LipoP predictions, the gene product was revealed to be located in the cytoplasm.

Locus PAZ_c00260:
Our results showed that the gene product at this locus, similar to PAZ_c00220 had a relation to the PTS fructose transporter subunit IIA. The protein in BLAST was not shown to have similarities to the same organism, showing it is not a paralog or product of recent gene duplication.



This graph predicts the presence and location of signal peptide cleavage sites in amino acids for locus PAZ_c00220

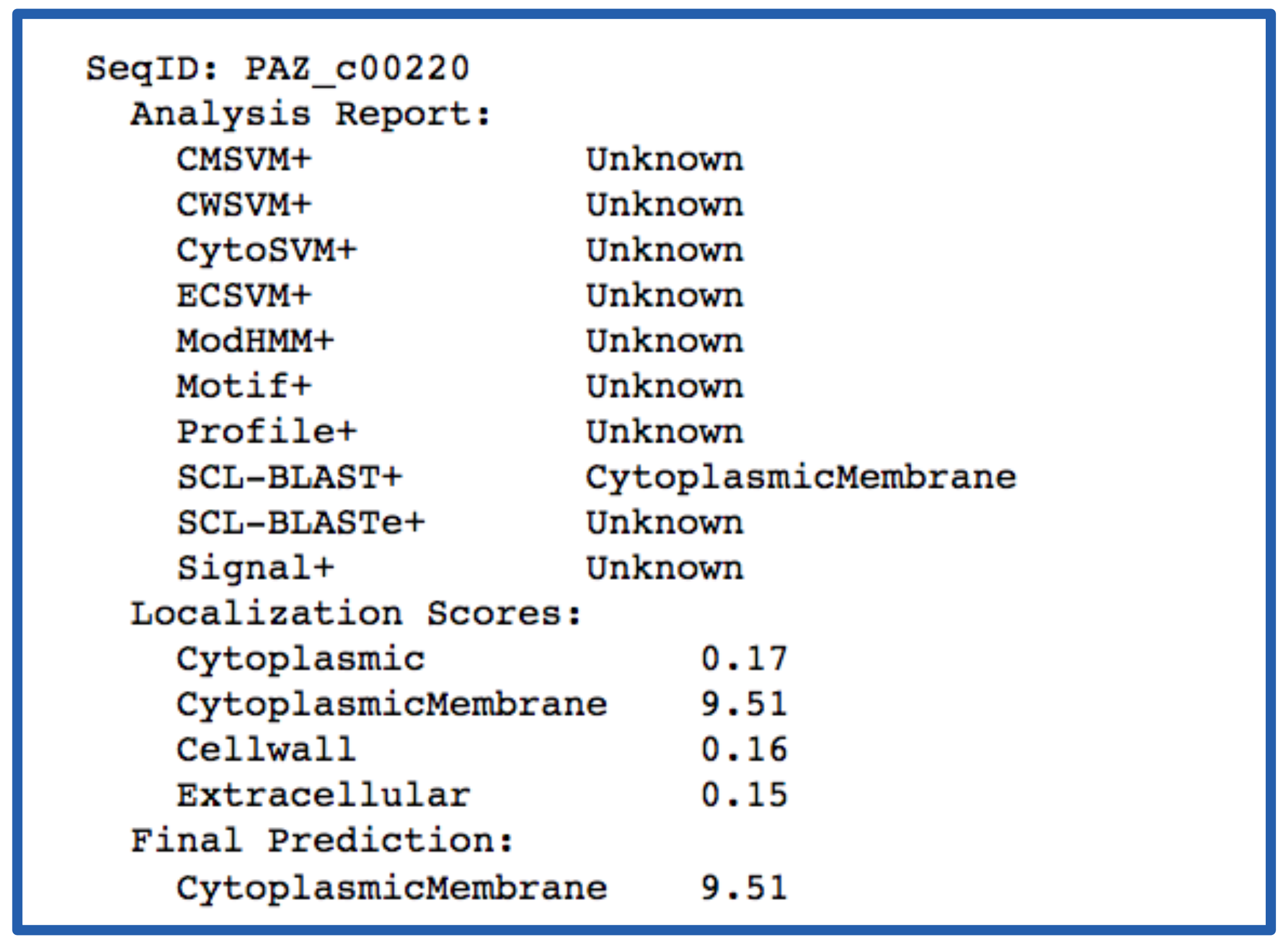
Locus PAZ_c00250:
Our results showed that this locus is located in the cytoplasm of the cell, with no protrusions into the membrane and a lack of a signal peptide. According to the data, this gene was likely to be involved in the PTS fructose transporter subunit IIA. Because of the phylogenetic tree, it is able to be seen that the gene has lost in common with oxidoreductase, an enzyme that serves as an electron transporter in PTS.



3D graphic image of protein shown with locus PAZ_c00270

Results

Through the annotation of our gene, we determined the phylogeny of the bacterium and verified the identity of the bacterium. In our research, we also explored the traits of both the bacterium and our respective gene products, pinpointing the location and function of the gene products. Using resources such as GENI-ACT, WebLogo, PFAM, BLAST, TIGRFAM, Genbank and others, we found that the gene products were all likely to be located in the cytoplasm of the bacterium. We discovered that our respective gene products were all involved in an operon crucial to maintaining homeostasis in *Propionibacterium acnes (266 CP002409)*, the PTS operon, which controls fructose uptake. Though there were some differences in the annotations, the genes appear to be correctly annotated by the r database.



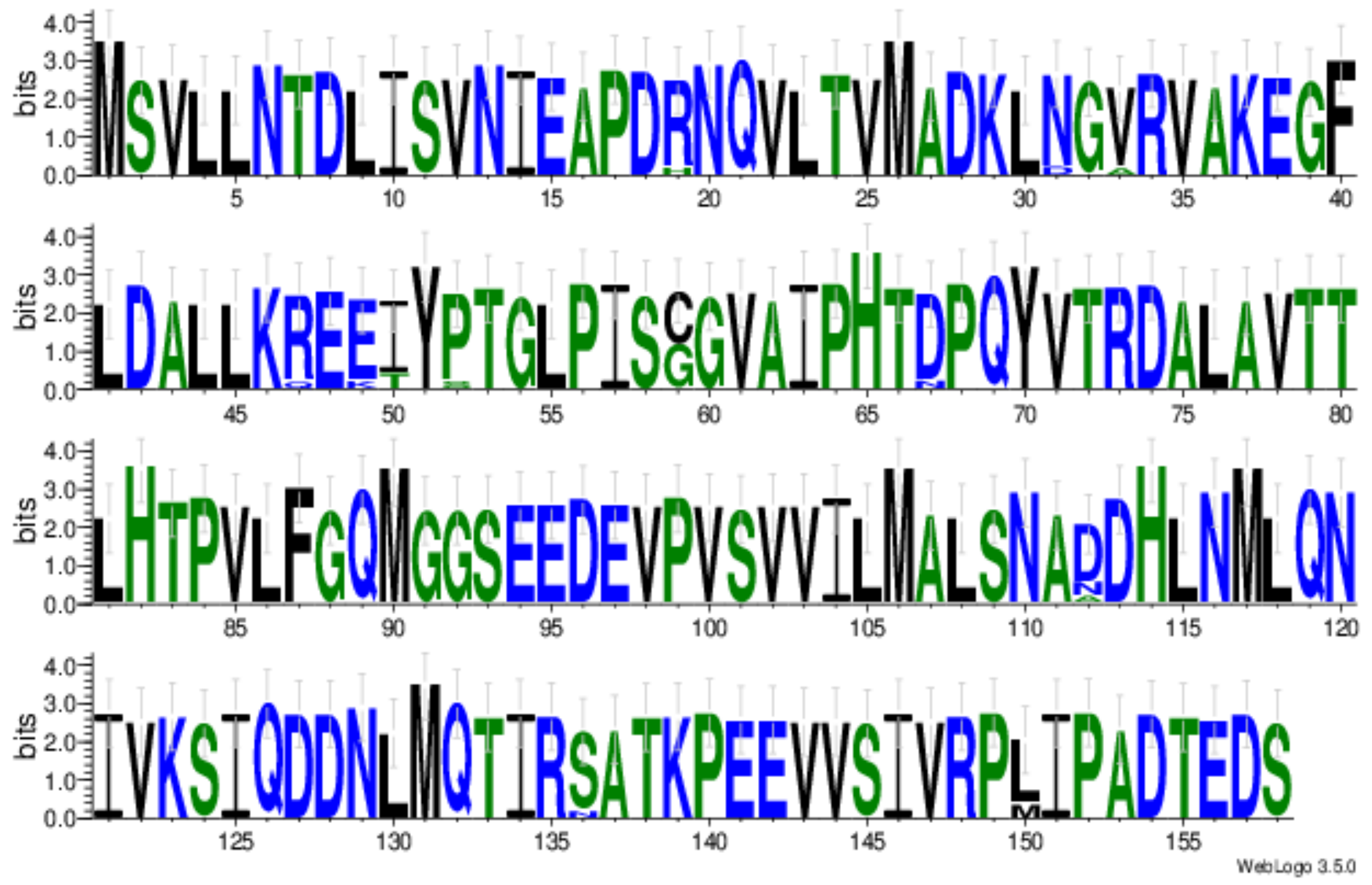
```
SeqID: PAZ_c00220
  Analysis Report:
    CMSVM+              Unknown
    CWSVM+              Unknown
    CytoSVM+            Unknown
    ECSVM+              Unknown
    ModHMM+             Unknown
    Motif+              Unknown
    Profile+            Unknown
    SCL-BLAST+          CytoplasmicMembrane
    SCL-BLASTe+         Unknown
    Signal+             Unknown
  Localization Scores:
    Cytoplasmic              0.17
    CytoplasmicMembrane      9.51
    Cellwall                 0.16
    Extracellular            0.15
  Final Prediction:
    CytoplasmicMembrane      9.51
```

PSORTb results showing location of the protein in the bacterial cell

Conclusion

After completing the annotation of the genes presented to us, the evidence points to our genes having an integral part in the PTS fructose transporter subunit IIA. The BLAST of the amino acid sequence showed that there was a high amount of conservation, which means that there is a lot of similarities between the genes and the bacteria. It can also be concluded that these genes are located in the cytoplasm of the cell because of the signal peptide probability and the LipoP predictions. From this, it is evident that the genes that we annotated are essential in the process of maintaining homeostasis within PTS, possibly coding for some sort of regulatory protein. In the end, we are still left with a lack of knowledge of what these genes code for specifically in the PTS and would need to continue research in order to obtain more information.

Graphic Elements



This WebLogo shows the sequence logos for locus PAZ_c00220

References

1.Geni-Act.org- gene annotation, archive results of genomic testing
2.BLAST - similarity between biological sequences
3.SignalP- signal peptide probabilities
4.LipoP - comparative tool to ascertain protein location
5.T-Coffee - sequence logo maker
6.Weblogo - graphic representation of protein

Acknowledgments

This work was supported by NSF ITEST Strategies Award Number 1311902.
www.buffalo.edu

* indicates equal contribution