

Annotation of the *Klebsiella pneumoniae* Genome from Locus Tags A79E_0001 to A79E_0025

Georgie Alejandro, Shauntrell Bradley, Mariah Burden, Angelina Dentice, Armani Lowmack, Kevon Parson, Meverlyn Toe, Isabella Vargas & Brianna Tucker

McKinley High School- Buffalo, New York and The Western New York Genetics in Research Partnership



University at Buffalo

Abstract

A collection of 25 consecutive genes from the microorganism *Klebsiella pneumoniae* (A79E_0001- A79E_0025) were annotated using the collaborative genome annotation website GENI-ACT. Five of the genes annotated are represented in this study. 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 mutations, the possibility of horizontal gene transfer, and the production of an RNA product were all analyzed for each gene in order to determine the Genbank proposed gene product name. The Genbank proposed gene product name and the proposed gene annotation for each of the genes in the group were incredibly similar. As a result, the genes appear to be correctly annotated in the genetic database.

Introduction

Klebsiella pneumoniae is a strictly anaerobic, non-mobile, non-encapsulated, and non-endospore forming gram negative bacteria (Lin et al., 2012). This organism is classified as a chemoheterotroph, as it requires methionine and several other amino acids for growth. The rod-shaped bacteria ferments lactose and metabolizes glucose to acquire nutrients. The microbe falls into the proteobacteria phylum, the gammaproteobacteria class, the enterobacteriales order, the enterobacteriaceae family (Brisse et al., 2001). *Klebsiella pneumoniae* diverges from *Klebsiella oxytoca*, *Klebsiella terrigena*, *Klebsiella planticola* and *Klebsiella ornithinolytica* on the phylogenetic tree (Brisse et al., 2001).

Klebsiella pneumoniae is a bacteria that usually resides in the intestines of humans. *Klebsiella* can also cause infections in the urinary tract, lower biliary tract, and around the area of surgical wounds. According to Lin et al. (2012), carbapenem-resistant *Klebsiella* is the most common type of strain. This immunity is responsible for approximately 7,900 infections and 520 deaths each year. The bacteria creates a *Klebsiella pneumoniae* carbapenemases which makes the antibiotics ineffective (Lin et al., 2012). The metabolic processes in which *Klebsiella* partakes in have the potential to cause widespread disease.

Methods

The GENI-ACT (<http://www.geni-act.org/>) modules were utilized to complete the genome annotation for *Klebsiella pneumoniae*. The modules are described in the data table.

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

A79E_0010:

The initial proposed product of this gene by GENI-ACT was 2-dehydro-3-deoxy-6-phosphogalactonate aldolase. The BLAST database provides multiple hits that indicate the protein is involved with galactose metabolism. The cellular location of the protein is within the cytoplasm, where it contains several functional protein domains. As a result, the proposed annotation for the gene found within *Klebsiella pneumoniae* is the enzyme 2-dehydro-3-deoxy-6-phosphogalactonate aldolase.

A79E_0012

The initial proposed product of this gene by GENI-ACT was D-galactonate transporter protein. The BLAST database provides several significant hits that indicate the protein is involved with the transport of molecules across the plasma membrane. The protein contains 10 helices that function within the membrane. The multiple sequence alignment shows negligible conservation at the amino terminal. It contains significant conservation in the central to carboxy terminal regions. As a result, the proposed annotation for the gene found within *Klebsiella pneumoniae* is D-galactonate transporter.

A79E_0016

The initial proposed product of this gene by GENI-ACT was heat-shock protein IbpB. The BLAST database provides several significant hits that indicated the protein was effective in protecting against heat shock as well as oxidative stress. The cytoplasmic protein contains several functioning domains. The WebLogo shows significant conservation at the amino, central and carboxy terminal regions throughout the majority of the protein. As a result, the proposed annotation for the gene found within *Klebsiella pneumoniae* is heat-shock protein IbpB.

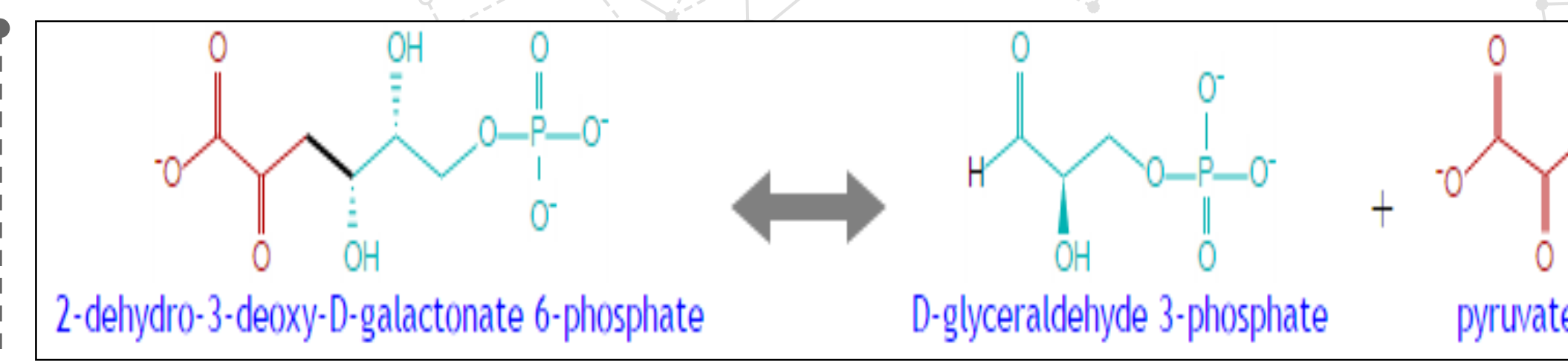


Figure 2 – A79E_0010 coding for an enzyme in the chemical reaction for galactose metabolism.

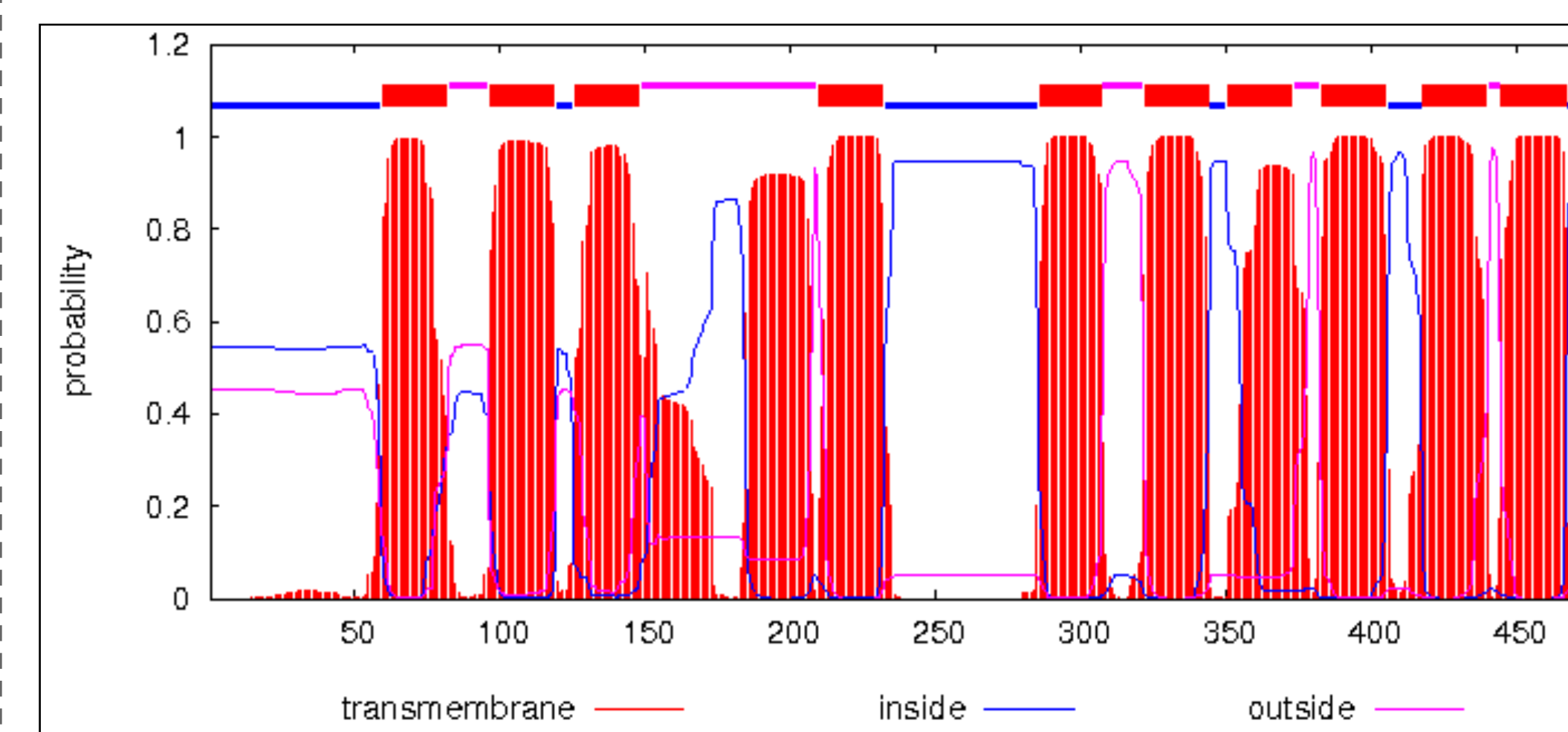


Figure 3 - A79E_012 TMHMM posterior probabilities indicating 10 helices for this transmembrane protein.



Figure 4 – A79E_0019 WebLogo indicates highly conserved amino acids in the amino, central and carboxy terminals.

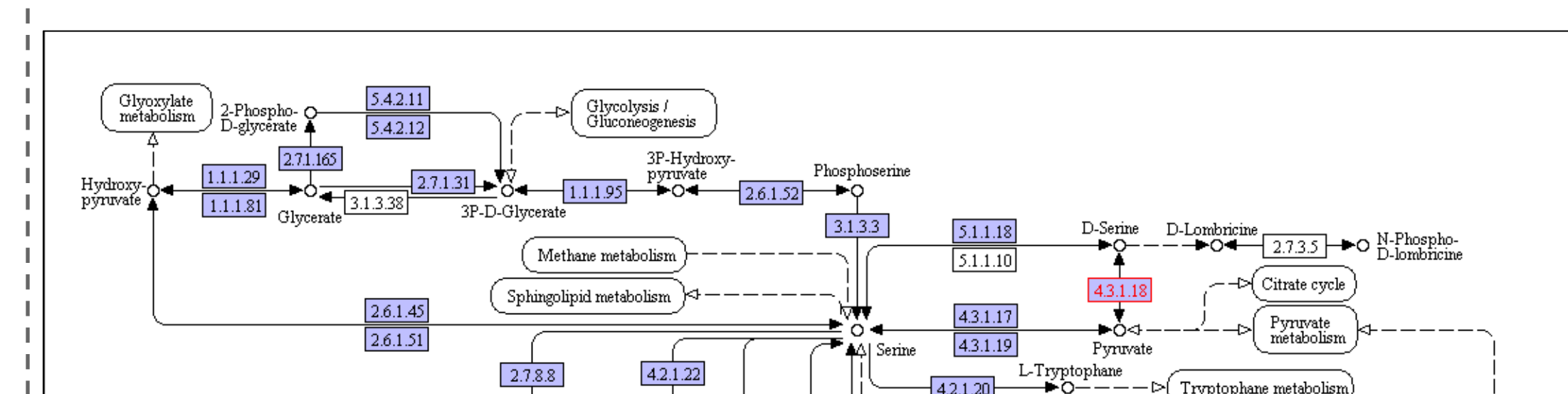


Figure 5 - A79E_0025 enzymatic pathway in which serine is converted into pyruvate utilizing the D-serine dehydratase enzyme.

A79E_0019:

The initial proposed product of this gene by GENI-ACT was GntR family transcription factor. The BLAST database provided multiple hits that indicates the protein is involved with transcriptional regulation in the process of protein synthesis. The protein functions within the cytoplasm in the presence of functional protein domains. This gene contains significant conservation in the amino, central and carboxy terminal regions of the alignment. As a result, the proposed annotation for the gene found within *Klebsiella pneumoniae* is GntR family transcription factor.

A79E_0025 :

The initial proposed product of this gene by GENI-ACT was D-serine dehydratase. The BLAST database provides multiple hits that indicates the protein contained several functional protein domains in which it used for the proper deamination of D-serine. The protein is found within the cytoplasm of the cell. The Pfam hits further support this proteins function as a catalyst. As a result, the proposed annotation for the gene found within *Klebsiella pneumoniae* is D-serine dehydratase.

Conclusion

The proposed gene annotations were relatively similar to the GENI-ACT proposed gene product. As a result, the genes appear to be annotated correctly utilizing several computer databases.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
0010	2-Dehydro-3-Deoxy-6-Phosphogalactonate Aldolase	2-Dehydro-3-Deoxy-6-Phosphogalactonate Aldolase
0012	D-Galactonate Transporter Protein	D-Galactonate Transporter Protein
0016	Heat-Shock Protein IbpB	Heat-Shock Protein IbpB
0019	GntR Family Transcription Factor	GntR Family Transcription Factor
0025	D-Serine Dehydratase	D-Serine Dehydratase

References

- Brisse et al. (2001). Phylogenetic diversity of *Klebsiella pneumoniae* and *Klebsiella oxytoca* clinical isolates revealed by randomly amplified polymorphic DNA, gyrA and parC genes sequencing and automated ribotyping. *International Journal of Systematic and Evolutionary Microbiology*
- Lin et al. (2012). Complete genome sequence of *Klebsiella pneumoniae* 1084, a hypermucoviscosity-negative K1 clinical strain. *Journal of Bacteriology*, 6019-6022.

Acknowledgments

Supported by an NSF Innovative Technology Experiences for Students and Teachers (ITEST) Award - 1311902

www.buffalo.edu

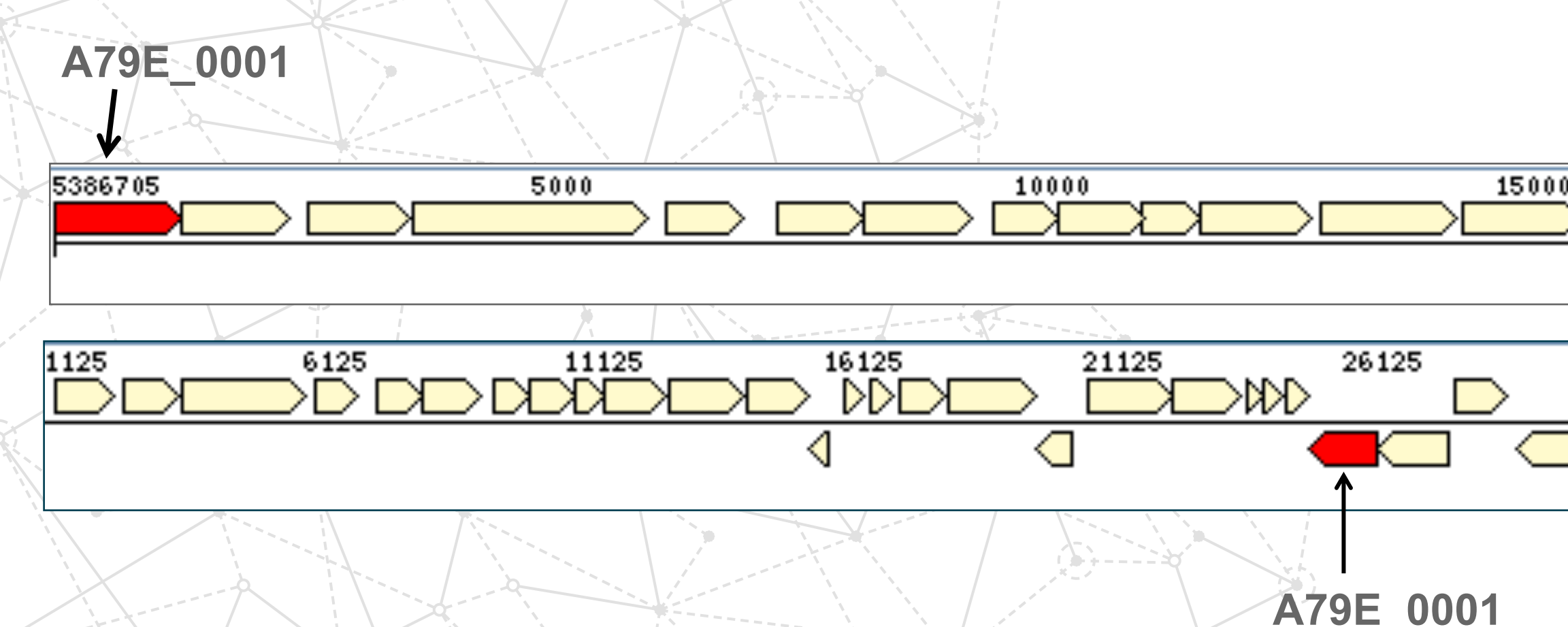


Figure 1. Several of the locus tags and relative position of the genes under investigation in this research project.