Annotation of the *Deinococcus maricopensis* Genome from DNA Coordinates 66120 to 71892 (Locus Tags Deima_RS00310 to Deima_RS00320)

Korina Bassett, Aliyah Brewer, Jaedyn Brewer, Isabel Heavner-Ortiz, Lauren Howell and Advisor Jennifer Clancy
Dundee Junior/Senior High School, Dundee, NY

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
Supported by the National Institutes of Health

Abstract

A group of 3 genes from the microorganism *Deinococcus maricopensis* (Deima_RS00310, RS00315 and RS00330) were annotated using the collaborative genome annotation website GENI-ACT. (Fig. 2) 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 and the possibility of horizontal gene transfer. 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 the database. This organism was chosen for its potential to withstand large doses of radiation and ability to repair itself.

Introduction

Deinococcus maricopensis is a gram positive coccoid bacterium that has the ability to repair DNA damaged by radiation, desiccation and/or oxidative stress. This organism was originally found in 1956 as a contaminant in a can of meat that had been subjected to high doses of ionizing radiation. (Rainey et al. 2005)

According to Sangyong et al. (2019), *Deinococcus* is a microorganism of interest for several reasons. It is one genus of three of the bacterial phylum *Deinococcus-thermus* that is highly resistant to environmental hazards. These kinds of bacterium have thick cell walls that give them gram-positive stains but they include a second membrane and they are closer in structure to gram-negative bacteria.

While the *D. maricopensis* genome has been sequenced (Pukall, et al. 2011), it has been not been as well-studied as its three closest relatives *D. radiodurans, D. thermos and D. geothermalis* which all show promising results in the area of gene repair, resistance to damage and other various traits unique to extremophiles. (Fig. 1)

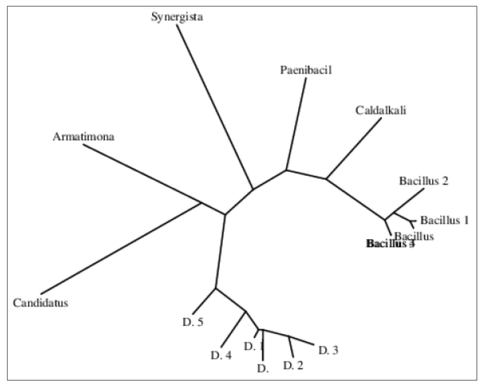

Figure 1 - Deima_RS00330 DrawTree showing no evidence of horizontal gene transfer in both closely and distantly related microorganisms.

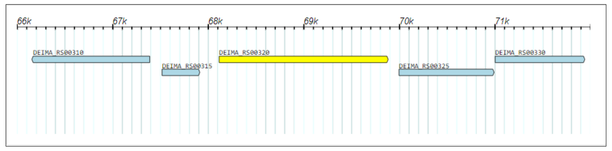

Figure 2. The locus tags and relative position of the genes under investigation in this research

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *D. maricopensis* genome annotation. The modules are described below:

Modules	Activities	Questions Investigated
Basic Information	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of the gene and protein? Where is it located in the genome?
Sequence-Based Similarity	Blast, CDD, T-Coffee, WebLogo	How similar is the protein under investigation to other proteins in GenBank?
Structure-Based Similarity	TIGRfam, Pfam, PDB	What functional domains are present in the protein under investigation?
Cellular Localization	Gram Stain, TMHMM, SignalP, LipoP, Psortb, Phobius	Is the protein under investigation located in the cytoplasm, secreted, in the periplasm or embedded in the cell membrane or cell wall?
Horizontal Gene Transfer	Phylogenetic Tree, Gene Neighborhood, GC Content	Has the protein under investigation co-evolved with the rest of the genome or has it been obtained in a different way?
Final Annotation	Evaluate data from all modules	Has the gene been correctly called by the pipeline annotation?

Results

Deima_RS00310:
The computer pipeline proposed product of this complementary gene was a peptidase M24 protein. (Fig. 3) According to the T-Coffee and WebLogo, this appears to be a highly conserved metallopeptidase protein that is ubiquitous in several kingdoms. TMHMM, SignalP, Psortb and Phobius data all support intracellular localization.

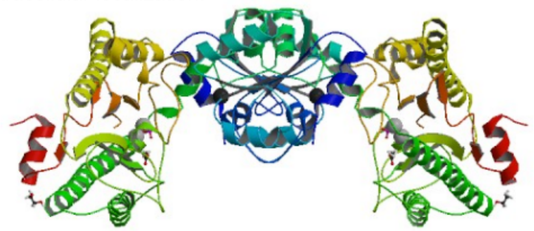

Figure 3. Crystal structure of the amino peptidase protein as coded for by Deima_RS00310.

Deima_RS00315:
The initial proposed product of this gene according to GENI-ACT was an iron-sulfur cluster assembly accessory protein. This was further supported through the BLAST-P data, that showed information matching the data from GENI-ACT. The Pfam results showed that this gene is found in many diverse species, such as in humans and *Haemophilus influenzae*, a bacteria previously thought to have caused influenza. (Fig. 6) The TMHMM data was negative, supporting cytoplasmic localization. The PSORTb results further enforced this by showing that the gene exists within the cytoplasm. Finally, the Web Logo appeared to be mostly well-conserved, (Fig. 4) with green being apparent throughout. This data supports that it contains many regions of polar amino acids, which is consistent with cytoplasmic localization.

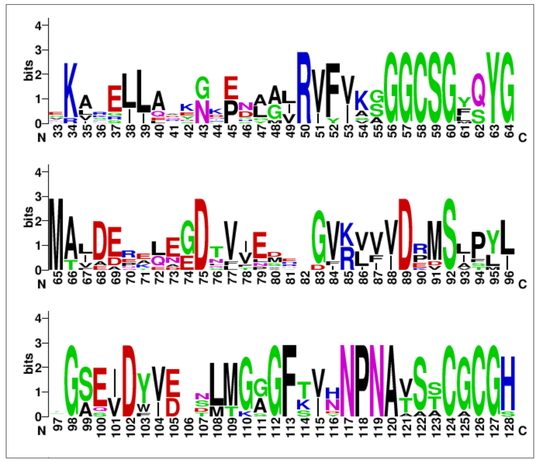

Figure 4. Deima_RS00315 WebLogo delineating the oligobinding peptide protein in both more closely and distantly related microorganisms. The green indicates polar amino acids.

Deima_RS00330:
The computer pipeline proposed product of this gene was a peptide ABC transporter permease. This is supported by the various blast hits that were received, and the significant e values they possessed. The gene is mostly well conserved throughout the related organismal species, and is not a product of horizontal gene transfer. This gene is a transmembrane protein, as evident by the TMHMM graph (Fig. 5), as well as the WebLogo; in this gene's WebLogo, the colors alternate from green and black, which makes sense because the polar amino acids would be located in the cytoplasm and the hydrophobic amino acids would be embedded inside the cell membrane.

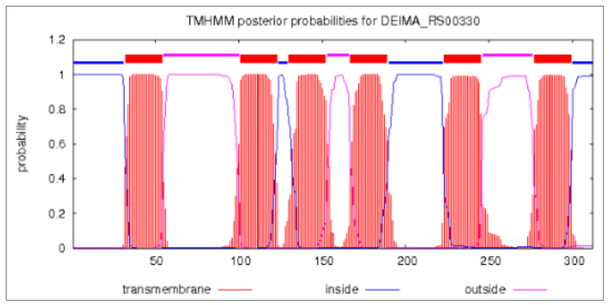

Figure 5. Deima_RS00330 is an ABC transporter permease that crosses the cell membrane 6 times.

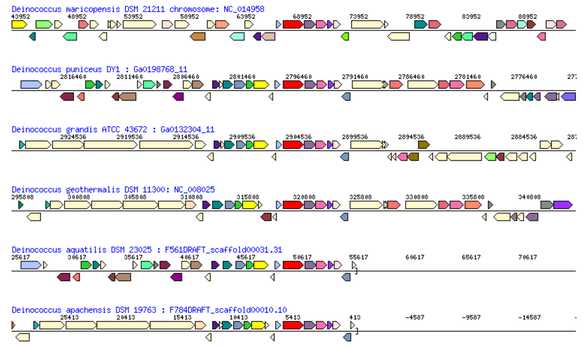

Figure 6. Deima_RS00315 is a highly conserved iron-sulfur cluster assembly protein along with this others in the neighborhood.

Conclusion

The GENI-ACT proposed gene product 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 the computer database.

The current team will continue their work on other genes in this organism due to the potential for application in the fields of biotechnology or cancer research.

Locus Tag	Pipeline Annotation Product Name	Proposed Annotation	Changes Proposed?
DEIMA_RS00310	Peptidase M24	Aminopeptidase P family protein	No
DEIMA_RS00315	Iron-sulfur cluster assembly accessory protein	Fe-S cluster assembly iron-binding protein IscA	No
DEIMA_RS00330	Peptide ABC transporter permease	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	No

References

Pukall et al. (2011). **Complete genome sequence of *Deinococcus maricopensis* type strain (LB-34).** Stand Genomic Sci. 2011 Apr 29;4(2):163-72

Rainey et al. (2005). **Extensive diversity of ionizing-radiation-resistant bacteria recovered from Sonoran Desert soil and description of nine new species of the genus Deinococcus obtained from a single soil sample.** Appl Environ Microbiol. 2005 Sep;71(9):5225-35.

Sangyong, et al. (2019). **Conservation and diversity of radiation and oxidative stress resistance mechanisms in *Deinococcus* species.** FEMS Microbiol Rev. 2019 Jan; 43(1): 19–52.

University at Buffalo The State University of New York
Jacobs School of Medicine and Biomedical Sciences
buffalo.edu

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25GM129209
Dr. Stephen Koury, University at Buffalo