Annotation of the Genome from DNA Coordinates 8160 to 16070 and 50922 to 53393 (or Locus Tags PAZ_c00090 to PAZ_c00130 and PAZ_c00540)

*Zehra Jaffri, *Luna Liu, *Grace Zhong, Paloma Tripathi and Mrs. Kathleen Kramer
Williamsville East High School and The Western New York Genetics in Research Partnership

Abstract

A group of 4 genes from the gram positive rod shaped bacteria *Propionibacterium acnes* (PAZ_c00090, PAZ_c00100, PAZ_c00130, and PAZ_c00540) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene productname 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 the database.

Introduction

Propionibacterium acnes is a slow-growing gram positive bacterium. It is associated with the human skin condition of acne. It is able to colonize in the skin and hair follicles. These bacteria grow very deep inside the pores of skin and feed on the sebum (oil secreted by a gland that waterproofs the skin). Propionibacterium Acnes are not, however the direct cause of skin damage. The real damage is caused by the inflammation by the immune response in reaction to environmental and biological factors.

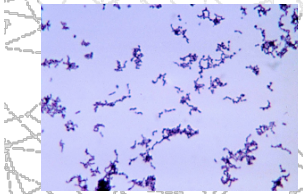


Figure 1: propionibacterium bacterium gram stained under a microscope.

Figure 2: close up of propionibacterium bacterium.

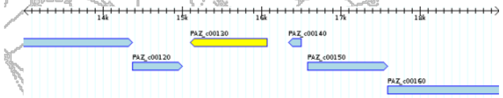


Figure 3: The locus tags and relative position of the genes under investigation in this research

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Propionibacterium acnes* 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?
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

PAZ_c00090

The initial proposed product of this gene by GENI-ACT was **mannose-6-phosphate isomerase**. The results were verified by thorough research using programs mentioned in the methods section. The proposed gene was primarily identified by using BLAST in which we searched nucleotide sequences. This protein was determined to be a cytoplasmic protein based on results from TMHMM and verified by LipoP. Attached below for pictorial reference is a sequence logo that was generated using WebLogo that demonstrates sequence characteristics such as functional units and protein binding sites in the protein (Figure 3). Like other proteins researched that are mentioned in this poster, mannose-6-phosphate isomerase codes for *C.acnes* which interacts with facial oils and causes acne on skin regions such as face, chest, and back (Figure 4).

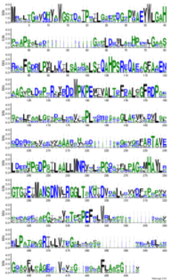
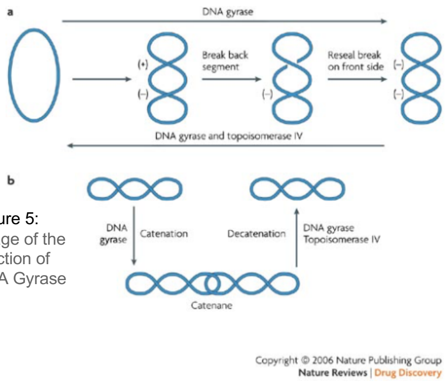


Figure 3

Figure 4

PAZ_c00100

The initial proposed product of this gene by GENI-ACT was a **DNA Gyrase Subunit B**. Through the top BLAST hits using the swissprot database, the gene product has been verified. According to TMHMM, LipoP and PSORT-B, the protein's location is in the cytoplasm. This confirms that this gene product is DNA Gyrase Subunit B.



Copyright © 2006 Nature Publishing Group
Nature Reviews | Drug Discovery

Figure 5: Image of the function of DNA Gyrase

https://www.researchgate.net/profile/Lynn_Silver2/publication/6639635/figure/fig2/AS:281145141940255@1444041633654/Figure-3-Reactions-of-topoisomerasesa-In-eubacteria-DNA-gyrase-imparts-negative.png

PAZ_c00130:

The initial proposed product of this gene by GENI-ACT was a **L-lactate dehydrogenase**. This is supported by the top BLAST hits for the amino acid sequence. Additional information about this gene was done through a series of research.. For example, according to the cellular localization module of GENI_ act this which included Phobius (figure 6) and TMHMM (figure 7) results showed that this protein is most likely non-cytoplasmic and located outside of the membrane. However, this was not supported by LipoP which demonstrated the protein being cytoplasmic.

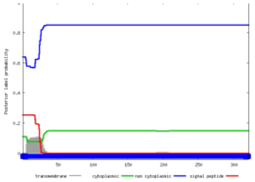


Figure 6

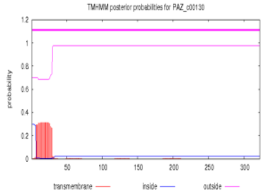
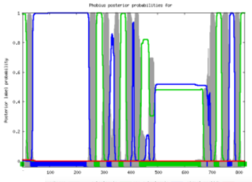


Figure 7

PAZ_c00540:

The initial proposed product of this gene by GENI-ACT was an **ABC transporter associated permease**, meaning that it utilizes the energy of ATP binding and hydrolysis to transport various substances across the membrane. This is supported by the top BLAST hits for the amino acid sequence. The gene is investigated through TMHMM, Phobius, and SignalP. It has about 10 transmembrane helixes, which is supported by the TMHMM results. This prediction is further supported by Phobius, which shows there are about 9 transmembrane helixes. SignalP determined that this protein is not a signal peptide.



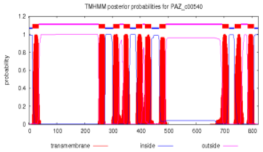


Figure 8. Phobius result (top) and TMHMM results (bottom) show the number of transmembrane helixes in the protein.

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for the members of the group and as such, the genes appear to be correctly annotated by the computer database.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
PAZ_c00090	mannose-6-phosphate isomerase	Cutibacterium acnes
PAZ_c00100	DNA gyrase subunit B	DNA gyrase subunit B
PAZ_c00130	Lactate dehydrogenase	Lactate dehydrogenase
PAZ_c00540	ABC transporter associated permease	ABC transporter associated permease

References

Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (541T). Standards Genomic Sciences,12 - 20.

Acknowledgments

The Williamsville East Science and Technology department would like to thank Dr. Rama Dey-Rao and Dr. Stephen Koury for assisting with the Iformation of this presentation.
Supported by NSF ITEST Strategies Award - 1311902