Annotation of the *Propionibacterium acnes* Genome from Locus Tags PAZ_c00180 to PAZ_c00210

Angela Chen, Pakhi Tripathi, Jessica Meng, Najya Fayyaz and Kathleen Kramer
Williamsville East High School, East Amherst and The Western New York Genetics in Research Partnership

University at Buffalo

Abstract

Four genes in the bacteria *Propionibacterium acnes* were annotated using GENI-ACT, a bioinformatics platform in which information about the gene is compiled. Each gene was manually annotated in terms of its sequence-based information, cellular localization data, a possibility of an alternative open reading frame, structure-based evidence, enzymatic function, and the possibility of duplication, degradation and horizontal gene transfer. All of the genes' initially proposed product matched the results of the annotations. PAZc_00160 produces an ABC transporter protein, PAZc_00180 produces a D-ribose-binding protein precursor, PAZc_00190 produces ribokinase, and PAZc_00210 is a hypothetical protein.

Introduction

Propionibacterium acnes is a gram-positive, non-sporulating, rod-shaped, anaerobic, and slow-growing bacilli. It is most commonly known for its link to causing acne, though it is involved with a wide variety of other illnesses as well. When involved in the pathogenesis of acne vulgaris, the bacterium colonizes deep within the pores and hair follicles of the face, producing various lipases to feed on and digest excess sebum and fatty acids in the pilosebaceous units. The cellular damage, metabolic byproducts, and bacterial debris produced by the rapid growth of *P. acnes* in follicles can trigger inflammation and immune response. *P. acnes* is a normal and usually a common part of the skin's flora, but it is oftentimes found in the digestive tract of humans as well. Metabolic analysis has shown that *P. acnes* has the ability to live in anaerobic and "microaerobic" conditions by utilizing a variety of metabolic pathways with the appropriate catabolic proteins. *P. acnes* is able to undergo a fermentative process generating short-chain fatty acids and propionic acid (where it gets its name from). In addition to fermentation, *P. acnes* can utilize various other anaerobic pathways deriving energy with the help of enzymes such as nitrate reductase, dimethyl sulfoxide reductase and fumarate reductase.

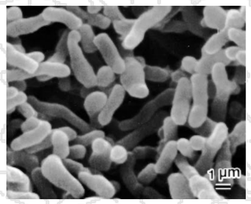

Figure 1 - *Propionibacterium acnes* is a gram-positive human skin commensal that prefers anaerobic growth conditions and is involved in the pathogenesis of acne (Kirschbaum and Kligman, 1963).

Aside from this, Propionibacterium acnes is deservant of further examination as it is also involved in contaminating blood and bodily fluid cultures as well as causing a variety of post-operative and device-related infections. A vast amount of illnesses can be brought on by *P. acnes* such as sarcoidosis, sciatica, endophthalmitis (particularly following intraocular surgery), chronic blepharitis, pulmonary angitis, endocarditis of aortic valves, corneal ulcers, hyperostosis, cholesterol gallstones, allergic alveolitis, and pustulosis. Though antibiotic resistant bacteria are becoming an increasingly dangerous problem globally, *P. acnes* is susceptible to a plethera of antibiotics and natural antimicrobials including erythromycin, clindamycin, doxycycline, minocycline, and to some extent even honey. Interestingly, *P. acnes* is photosensitive; it glows orange when exposed to blacklight, possibly due to the presence of endogenous porphyrins. In addition, ultraviolet radiation, specifically in the range of 405–420 nanometers, is able to kill the bacterium.

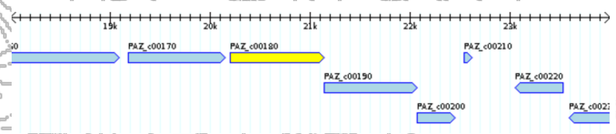

Figure 2 - The locus tags and relative position of the genes under investigation in this research.

Methods

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

Results

PAZ_c00160:

The initially proposed gene product on Geni-Act was ribose ATP-binding cassette (ABC) transporter, which is a type of protein that binds and hydrolyzes ATP in the active transport of substrate molecules such as nutrients. It contains a transmembrane subunit and a membrane-associated subunit. This hypothesis was supported by top BLAST hits and structure based evidence. In addition, it is highly conserved, as evidenced by its weblogo created with the top twenty BLAST results, and it has not undergone recent horizontal gene transfer, as evidenced by its phylogenetic tree. It does not contain any transmembrane helices according to TMHMM, so this particular gene most likely produces the membrane-associated ATPase subunit. However, is is inconclusive where in the cell the protein can be found: LipoP had determined it most likely resided in the cytoplasm, PSORT-B determined it was most likely in the cytoplasmic membrane, and Phobius determined it was most likely non-cytoplasmic.

PAZ_c00180:

The initial proposed gene product on Geni-Act was a D-ribose-binding protein precursor, a periplasmic sugar binding protein with the monosaccharide as the ligand. These proteins are not enzymes and instead are a component of the ABC-type transport systems and are involved in the active transport of water-soluble ligands. The results of

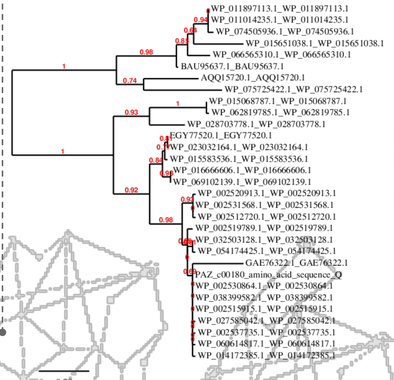

Figure 3 - PAZ_c00180: The phylogenic tree signifies that the protein is clustered with proteins from non-related bacteria. This suggests that horizontal gene transfer is possible.

LipoP signified a high chance of a liposaccharide signal peptide being present. The results of Phobius further supported LipoP by suggesting that the gene product likely included a noncytoplasmic or secreted signal peptide. T-Coffee results suggest that the gene produce could be well-conserved across various speices. TMHMM results indicated a slight chance of roughly the first 25 amino acids are transmembrane and the rest being extracellular. Its phylogenetic tree suggests that horizontal gene transfer is possible, as the protein is clustered with proteins from non-related bacteria. This hypothesis was supported by BLAST hits, cellular localization data, and structure-based evidence Consequently, the proposed annotation persists in its congruence to the initial proposed product.

PAZ_c00190:

The initial proposed product of this gene by Geni-Act was ribokinase, an enzyme catalyzing the first step in ribose catabolism. The rbsK gene encoding ribokinase typically is found with ribose transport genes. According to Phobius it is most likely non-cytoplasmic protein, however, the results from P-SORTB and LipoP found it highly likely to be cytoplasmic . All results concur that it is definitely not a transmembrane protein, nor is it a signal peptide. Ribokinase belongs to the carbohydrate kinase pfkB family (PF00294). All of these enzymes are phosphotransferases that have an alcohol group as an acceptor. This domain is found in a variety of

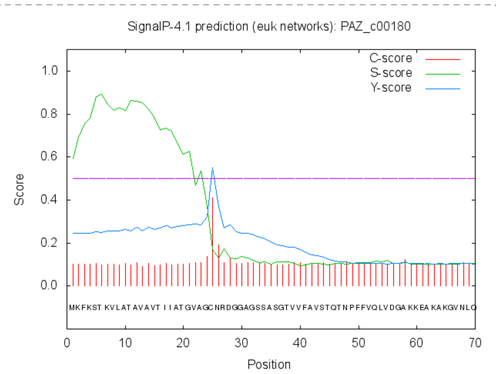

Figure 4 – PAZ_c00180: The SignalP results predict the prescene and location of signal peptide cleavage sites in amino acide sequences. The protein is predicted to have a signal peptide as indicated by the high S-score.

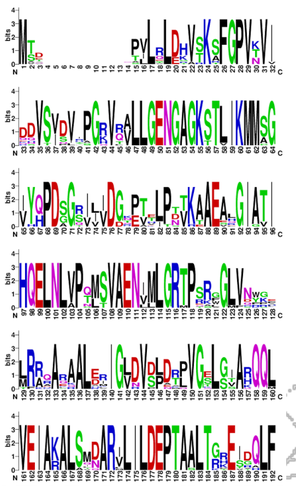

Figure 5 - PAZ_c00160: The WebLogo reveals a species that is well-conserved across domains, as indicated by the abundance of large letters.

carbohydrate and pyrimidine kinases. It is found in phosphomethylpyrimidine kinase (EC), which is part of the thiamine pyrophosphate (TPP) synthesis pathway - TPP being an essential cofactor for many enzymes. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, structure-based evidence, the presence of well-curated protein functional domains within the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence. As such, the proposed annotation is a ribokinase enzyme, matching correspondingly to the product proposed initially.

PAZ_c00210:

The initial proposed product of the gene product on Geni-Act was a hypothetical protein of TA06 bacterium. The BLAST hits results and data of cellular localization supported this hypothesis. The results of PSORT-B and TMHMM provided strong support for the gene's protein to be extracellular, with the highest probability of it being outside the cell. Furthermore, the results of BLAST and WebLogo proved that this genome is not strongly conserved across domains. Therefore, the proposed annotation remains consistent with the initial proposed product.

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the group, thus the genes seem to be correctly annotated by the computer database.

Gene Locus	Geni-Act Products	Proposed Annotation
00190	Ribokinase	Ribokinase
00160	Ribose ABC Transporter	Ribose ABC Transporter
00180	D-ribose-binding Protein Precursor	D-ribose-binding Protein Precursor
00210	Hypothetical Protein	Hypothetical Protein

References

Bhatia, Ajay. "PROPIONIBACTERIUM ACNES AND CHRONIC DISEASES." *The Infectious Etiology of Chronic Diseases: Defining the Relationship, Enhancing the Research, and Mitigating the Effects: Workshop Summary*. U.S. National Library of Medicine, 01 Jan. 1970. Web. 16 May 2017

Acknowledgments

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

www.buffalo.edu