

Introduction to STEM Using *Kytococcus sedentarius* Genes Ksed_05560-Ksed_05610

Brittney Francis, Jacob Squicciarini, Katherine Mann, Adryan Cheeseboro, and Sandra Climenhaga
Albion High School and the Western New York Genetics in Research Partnership

Abstract

Research was undertaken at Albion High School to manually annotate a portion of the *Kytococcus sedentarius* genome. While the purpose of the project was to annotate genes, the main intent was to introduce students to STEM and bioinformatic fields. Following a set of modules set out by the Genomics Education National Initiative Annotation Collaboration Toolkit (GENI-ACT) the gene was analyzed. GENI-ACT allowed students to gather information and draw conclusions based on the data obtained from completing a specific module. This exposure helped high school students be career and college ready. Traditional textbook and lecture learning results only in a pass/fail grade on some test. This hands on approach used real genetic data from a bacteria and allowed students to discover and then contribute to the scientific process. Overall this process was a success.

Introduction

The initiation of the Human Genome Project in 1988 triggered a wave of bioinformatics technology. Bioinformatics is the use of computers and mathematics to analyze complex biological information, such as DNA or protein sequences. This technology can be used to better understand how DNA, amino acids, and proteins are related, and how they work within an organism (Suresh, 2005). With this information scientists can find cures to diseases, trends in evolution, and more. It took scientists 14 years to completely sequence the human genome, but as technology evolves the process is becoming faster. However, no matter how fast computers may become, humans will always be needed to check the accuracy of the computer output. In order to increase the workforce that can analyze such results, researchers at the NSF partnered with the University at Buffalo to work with local high schools, introducing younger students to the field of bioinformatics.

High school students are taught the basics of biology, but with this program, students are gaining a deeper understanding of real world science. STEM stands for science, technology, engineering and math. In 2013 only 44% of U.S. high school graduates were ready for college level math, only 36% were ready for college level science, and 38% of college students that started with a STEM major did not graduate with one (Ditty et al. 2010). Students from Albion High School took part in this gene annotation project organized by the Western New York Genetics in Research Partnership by studying genes from *Kytococcus sedentarius*, (Sims, et al. 2009) a bacterium that can be found on human skin and causes pitted keratolysis. Participants used an online notebook for storing and analyzing data, as well as numerous online tools used for annotating genes such as BLAST, T-Coffee, TMHMM, and WebLogo.

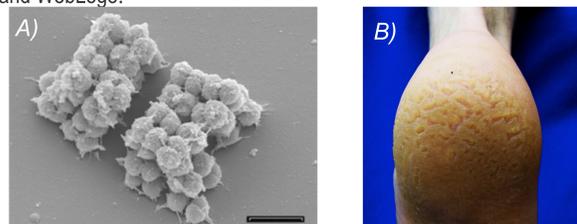


Figure 1. A) Magnified image of *Kytococcus sedentarius*. B) Image of a foot suffering from pitted keratolysis.

Methods and Materials

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

Modules	Activities	Questions Investigated
Module 1- Basic Information Module	DNA Coordinates and 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

The TMHMM tool allows for examination of cellular localization in a graphical format. The small patch of vertical red lines (fig A) suggests a region of hydrophobicity but the data is not enough to suggest a transmembrane helix. In contrast the thick red vertical regions (fig B) predict a transmembrane protein. The red bars indicate the transmembrane helices where it is anchored in the membrane.

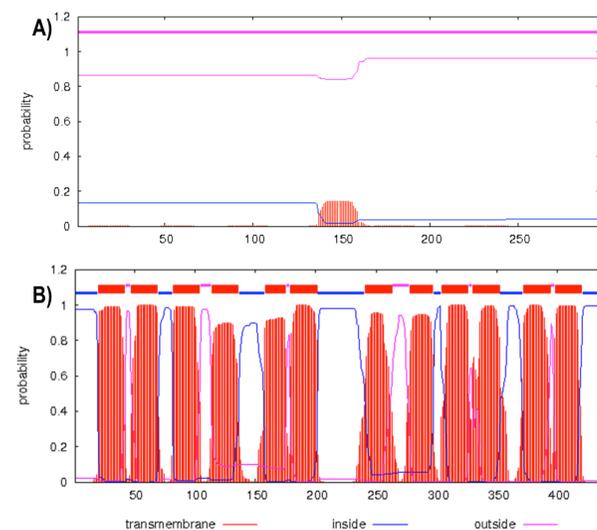


Figure 2. The results of the TMHMM. A) Output showing no predicted transmembrane helices for Ksed_5610. B) Output suggesting 12 transmembrane helices for Ksed_5600.

The T-Coffee tool compares multiple genes to show how well conserved the sequences are. Each line represents a different gene or species and the symbols at the bottom show how well conserved they are. The WebLogo is a graphical representation of the T-Coffee data, and is much easier to read. The different colors of the letters represent the different chemical properties of the official amino acids, and the relative height of each letter shows how frequently the acids occurred.

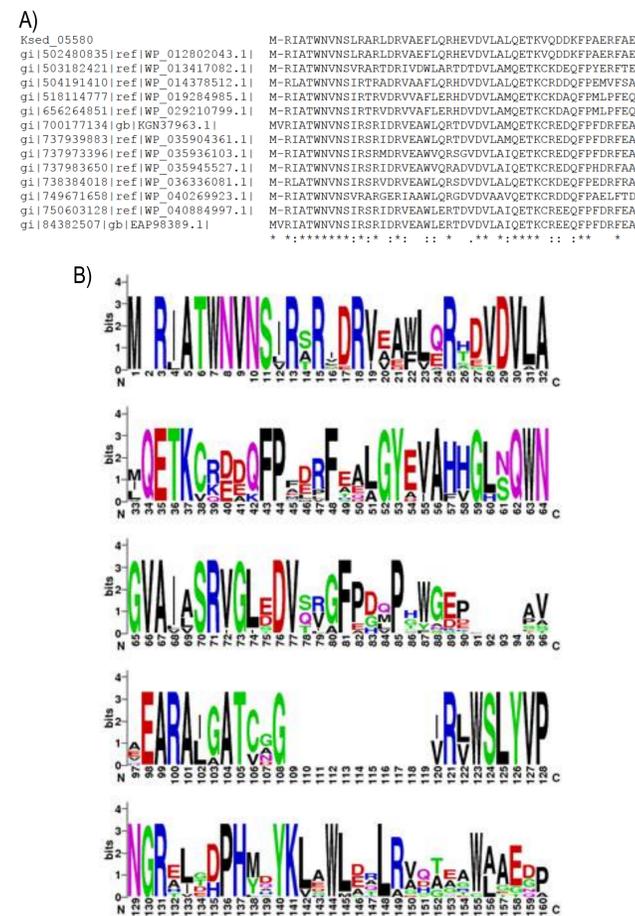


Figure 3. A) T-Coffee results for Ksed_5580. B) WebLogo for Ksed_5580.

The Ksed_05610 gene appears to be well conserved as it appears in several other bacteria. The neighborhood of the gene locus was examined for these bacteria using data from IMG/EDU. This tool helps show the similarities of the gene neighborhood in different species. The results do not show horizontal gene transfer, but instead suggest independent evolution of the gene. The ability to readily access and use this online tool allows the bioinformatics workforce to evaluate microbial relationships. The IMG/EDU tool will only work if researchers exist who can help maintain a curated database which is why it is necessary to bring more people into this field.

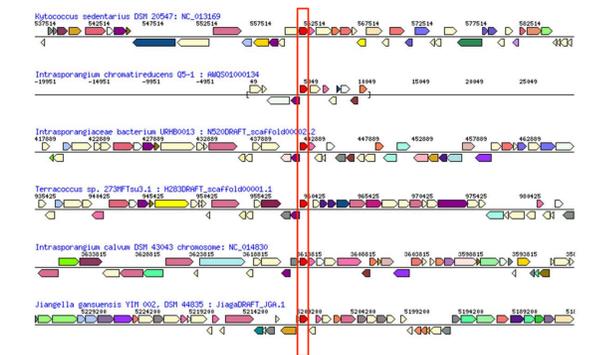


Figure 4. Results from IMG/EDU showing dissimilar neighborhoods of Ksed_05610 and other bacterial. The red box contains the target gene.

Conclusion

The purpose of this project was to introduce high school students to STEM and bioinformatic fields. The students involved in this project were top students at Albion High School and often had trouble finding time to come in after school and work on their gene. Due to this, most participants could not finish analyzing all parts of their genes and missed out on some valuable bioinformatics knowledge. At times, participants felt that they were simply going through the motions of copying and pasting sequences into different online tools and did not really understand what they were doing. It was not until after the module was complete students could take time to reflect on what they had created in a meaningful way. Even though the process seemed confusing, all involved agree that they learned a lot more than what they would have in a typical high school biology class and are glad they set aside time to annotate the *Kytococcus* genes. Overall, the project resulted in the gaining of a basic understanding of bioinformatics and STEM. This project can therefore be considered a success.

References

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

Ditty JL, Kvaal CA, Goodner B, Freyermuth SK, Bailey C, Britton RA, et al. (2010) Incorporating Genomics and Bioinformatics across the Life Sciences Curriculum. PLoS Biol 8(8): e1000448. doi:10.1371/journal.pbio.1000448

Suresh Kumar (2005). Bioinformatics web. Retrieved June 2, 2015 from: <http://www.geocities.com/bioinformaticsweb/>

Acknowledgments

Supported by NSF ITTEST Strategies Award Number 1311902
Dr. Rama Dey-Rao Clinical Assistant Professor, University at Buffalo
Dr. Stephen Koury Research Associate Professor, University at Buffalo