DO NOT CONSIDER FOR TALK

2021 SEA Symposium Abstract

Purdue University

West Lafayette IN

Corresponding Faculty Member: Kari Clase (kclase@purdue.edu)

Using unconventional programs to aid in the annotation of the mycobacteriophage ScoobyDoobyDoo annotation

Anthony P Hegarty, Emily Kerstiens, Lauren Novak, Gillian Smith, Kari Clase

Bacteriophages, often referred to as phages, are the most abundant organism on the Earth and therefore display great genetic diversity. With an estimated population of 1031, the use of powerful bioinformatics tools is essential for the characterization and analysis of individual phages and their relation to one another. Once a phage genome has been auto-annotated, individual phages are clustered with one another; these are similar phage that diverged close to one another in evolutionary time. However, singletons comprising of an individual phage with no close relatives, or clusters with a small number of members, may fall victim to false-positive hits in BLAST databases or programs such as Glimmer and Genemark calling genes with low coding potential. The lack of information pertaining to these phages causes the annotation process to become difficult – forcing the annotator to make decisions based on incomplete or less-than-ideal sets of information. The annotation of phage ScoobyDoobyDoo in cluster C2 is an example of this. Comprised of an estimated 262 genes over 158153bp, Scooby is clustered with two other phages Myrna and Phabba. As mentioned before, the small number of similarly grouped phages causes difficulty when annotating its genome; many BLAST results are poor and orphams are frequent. However, in phages such as this, utilizing programs outside of PECAAN and BLAST databases may help elucidate gene structure and function. Approaching the annotation process in multiple directions will help characterize Phage genomes and assign the correct function to genes that would otherwise be called as NKF. These multiple different approaches are important since they allow a more complete understanding of that phage in addition to the information provided in PECAAN and other standard annotation programs. For example, using phylogenic programs such as Splitstree allow one to understand phage evolutionary relationships. Protein structure and comparison programs such as Phyre2 and I-TASSER allow a spatial understanding of proteins and how they contribute to the organism as a whole. Finally, comparing mutations between sequences in FASTA files gives the annotator a sense of changes on the nucleotide level in closely related phage genomes.

By teaching students how to use these programs while they annotated ScoobyDoobyDoo allowed them a greater understanding of phage structure and lifecycle. This in turn gave additional context and information for use in the annotation process. This presentation hopes to provide an overarching view of these additional techniques and programs in hopes to assist the overall Phage annotation process. Additionally, we hope to contribute to the shared information and unique qualities of Phage in the C2 cluster.