DO NOT CONSIDER FOR TALK

2021 SEA Symposium Abstract

Purdue University

West Lafayette IN

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

New Methods of Phage Protein Study: From Statistical Analysis to Simulations

Emily A Kerstiens, Kari Clase

Phages are being studied for their potential in a variety of applications from food safety to pest control to phage therapy. Every potential phage product must go through extensive testing to determine the safety of a phage product. This is especially true of phage therapy products which must go through the FDA clinical trials. Sending every phage through clinical trials, however, significantly cuts down the number of phages that can be used. A possible method of FDA regulation involves determining therapeutically equivalent phages based on their statistically significant proteins. Then a database of equivalent phages could be built without putting every individual phage through clinical trials. However, with ~70% NKF proteins, there is still too much unknown.

The current standards of phage genome annotation only identify roughly ~30% of protein functions. A new method of protein assignment is proposed here using different alignment programs, hidden Markov models, cellular location predictions, and GROMACS simulations. The phage PotatoSplit was used as a test case, and the NKF proteins dropped from 63% to 37%. Most new proteins are related to phage structure identified using TMHMM and HHBlits. After studying existing programs, a machine learning algorithm is being developed using Python. This algorithm will be trained on phage proteins from PhagesDB and is expected to predict functions or functional categories.