CONSIDER FOR TALK

2021 SEA Symposium Abstract

University of South Florida

Tampa FL

Corresponding Faculty Member: Richard Pollenz (pollenz@usf.edu)



Louis A Otero

Bioinformatic Analysis of the Major Capsid Protein from Bacteriophages between Clusters

Louis A Otero, Richard S Pollenz

The major capsid protein is an integral structural gene involved in creating the capsid of a bacteriophage, where the viral genome is housed. Considering that all bacteriophages must have a capsid as it is a crucial structural protein, one can assume that it would be highly conserved throughout phage clusters. The goal of this research project was to determine what protein structures and domains major capsid proteins were aligning to in HHPRED, and if there was a consensus in said hits that could be used as a better tool to call major capsid proteins. PhagesDB was used to collect the sequences of the major capsid proteins from each cluster, and HHPRED was used to identify which crystal structures the capsid proteins were aligning to and to verify previous annotation calls. 74 phage were used, 1 from every cluster on PhagesDB. 10 phage call their major capsid proteins as major capsid and capsid maturation protease fusion proteins. All of these had an alignment with the domain 1O6E\_A on the N-terminus of the protein, which was not detected on any non-fusion protein. These proteins were also larger than the other major capsid proteins, with a molecular mass of over 50,000. 7 phage, from clusters AS1, AS2, AS3, E, J, O, and Y had the domain 2R9I\_A on the N-terminus of the major capsid protein. 1 phage from Cluster V had the domain 6P3E\_G on the C-terminus of the major capsid protein. The other 47 major capsid proteins did not have domains. In terms of major crystal protein structures that were hit to, the main crystal structure that indicated the presence of a major capsid protein was the hit 3QPR\_F, followed by 1OHG\_D, 3JB5\_F, 6B0X\_G, and 6OMC\_L. However, 22 of the phage that were annotated as having a major capsid protein did not show evidence of it in HHPRED, as either no hits were detected, or very low probability hits to major capsid related proteins were detected, with a common hit being 2R9I\_A, one of the domains found in 7 of the 74 phage. It is intriguing that there are so many phages that called major capsid proteins without evidence from HHPRED confirming that call, especially since they are expected to be highly conserved- this means that while the major capsid proteins may not be uniform in what hits they get, there may be a consensus sequence that allows major capsid proteins to function. For next steps, the major capsid proteins should be placed into Clustal Omega to locate consensus between the proteins that get standard hits, between the proteins that get no hits at all or just 2R9I\_A, between both groups, and between the fusion proteins themselves. A consensus sequence could be used as a better analytical tool for calling major capsid proteins when functional evidence from HHPRED is scarce. Furthermore, doing in vitro assays to create truncated mutations of various major capsid proteins and attempt to grow phage up and see if the capsid forms correctly, if at all, is important.