DO NOT CONSIDER FOR TALK

2021 SEA Symposium Abstract

University of California, Los Angeles

Los Angeles CA

Corresponding Faculty Member: Jordan Parker (jordan.p.parker@kp.org)

Comparative analyses of dissimilarity in Cluster A phages highlighting the non-uniform continuum of diversity

Clara Bang, Sabrina Dunbar, Erick Lopez, Wendy Wong, Andrew Kapinos, Canela Torres, Amanda C Freise, Jordan Moberg Parker

Bacteriophages are viruses that infect bacteria in order to use their cellular processes to reproduce. Similarities in the genetic content of phages are used to categorize them into clusters and subclusters. Phage interactions with bacterial hosts contribute to the development of mosaic genomes, particularly in Cluster A phages, and these phages exhibit high sequence dissimilarity in the region coding for integration cassettes. This research was motivated by the non-uniform diversity of Cluster A and the information gap concerning other regions of dissimilarity across Cluster A genomes. In order to understand what relationships existed between Cluster A subclusters and increase the knowledge of phage and subcluster relationships in the biosphere, regions of dissimilarity were analyzed within subclusters A2, A3, A4, A5 and A6. This paper characterizes dissimilarity based on shared gene content, which illustrated that subclusters A3 and A4 were most highly similar and subcluster A5 between other subclusters and the intra-subcluster dissimilarity of subcluster A2. Dot plot and pham analysis suggested that subclusters A2-A6 feature regions of dissimilarity on the right and left arm of the genomes, specifically in the regions encoding lysin and minor tail proteins. These regions reveal genetic distance when integration cassettes and portal protein genes were compared with Phylogeny.fr. Overall, it was observed that regions of dissimilarities occur across Cluster A bacteriophage genomes, similar to previous work exclusively on integration cassettes. These findings suggest there is non-uniform diversity in phage and subcluster relationships within Cluster A, which disallows many generalizations based on subcluster alone.