CONSIDER FOR TALK

2025 SEA Symposium Abstract

University of California, Los Angeles

Los Angeles CA

Corresponding Faculty Member: Daniel Velez-Ramirez (verde@g.ucla.edu)



Alette Laughton

Uncovering Hidden Relationships: A High-Throughput Double-Loop Comparison Tool for Comprehensive Phage Genome Analysis

Alette Laughton, Krisanavane Reddi, Amanda Freise, Daniel Velez-Ramirez, Tejas Bouklas

Bacteriophages, or phages, are viruses that infect bacteria, playing critical roles in microbial evolution, ecosystem dynamics, and clinical applications like phage therapy. Gene Content Similarity (GCS) is a metric that quantifies the proportion of genes two phages share, with “phams” of genes grouped by nucleotide similarity, calculated by comparing the commonality of gene phams numbers between phages. The classification of phages into clusters relies heavily on GCS analysis, which is essential to revealing evolutionary relationships, guiding taxonomy, and predicting phage function and host interactions. The use of GCS in phage cluster assignment is a relatively new system, with thresholds for cluster assignment evolving from 50% to 35% in recent years. Due to the growing input of sequenced phages into public databases, the need for large-scale comparative genome analysis tools has grown. Here, we introduce a novel computational tool that performs double-loop comparisons of each phages and pham numbers to assess genomic relatedness with high efficiency, boasting over 2 billion gene comparisons in 10 minutes on a typical laptop. This tool features three key functions: (1) an “all sequenced phages to all sequenced phages” analysis, (2) an “all sequenced singletons to all sequenced phages” function to facilitate classification of unclustered phages, and (3) a targeted “single sequenced phage to all sequenced phages” comparison. By leveraging locally stored cached API calls, our approach minimizes strain on the PhagesDB website while enabling more exhaustive genomic comparisons. In utilizing this tool, several reclassifications are proposed: the condensing of clusters CW and DM with additional singletons into one larger cluster, the separation of cluster JB with additional singletons into three distinct sub-clusters, and the addition of various singletons into respective clusters based on GCS. In the future, a tool like this would be essential in a phages toolkit for any new phage discovered, instantly scanning it for GCS relatedness to all other sequenced phages. This tool addresses critical gaps in phage classification by providing a high-throughput method to scan through all available phages, potentially refining cluster assignments and preventing misclassification. Its application extends to evolutionary studies, phage therapy development, and metagenomic analyses, offering a valuable resource for the phage research community.