CONSIDER FOR TALK

2021 SEA Symposium Abstract

University of Nevada Las Vegas

Las Vegas NV

Corresponding Faculty Member: Philippos Tsourkas (philippos.tsourkas@unlv.edu)

Phage Commander, an app for rapid gene identification in bacteriophage genomes using multiple programs

Matt Lazeroff, Geordie Ryder, Sarah Harris, Philippos Tsourkas

Genes in newly sequenced phage genomes are usually identified using one or more freely-available gene identification programs (e.g., Glimmer, GeneMark family, RAST, Prodigal, etc.). No program has been shown to consistently outperform the others; thus, the choice of which program to use is not obvious. Using multiple programs improves accuracy, as multiple programs may identify genes missed by an individual program. However, this requires integrating the results of multiple programs and increases the labor cost of annotation. We present an app developed in our lab named Phage Commander, which runs a phage genome sequence through nine gene identification programs simultaneously, and outputs the results in an easy-to-visualize format. Phage Commander also generates properly formatted output files (.gb) for direct export to NCBI GenBank, or for further processing in DNA Master. Users can select the threshold for determining which genes should be exported (e.g. genes identified by at least one program, genes identified by at least two programs, etc.). Phage Commander was benchmarked using eight high-quality phage genomes whose genes are backed by experimental data. Results show that the best results are obtained by exporting genes identified by at least two or three of nine available programs. Phage Commander is currently being used in our Phage Discovery class and has significantly reduced the labor cost of genome annotation, while also improving accuracy.