CONSIDER FOR TALK

8th Annual SEA-PHAGES Symposium Abstract

Western Kentucky University

Bowling Green KY

Corresponding Faculty Member: Claire Rinehart (claire.rinehart@wku.edu)

PECAAN, a Phage Evidence Collection And Annotation Network

Claire A Rinehart, Bobby L Gaffney, James D Wood, Jason R Smith

The Science Education Alliance PHAGES program has adopted a standard workflow for the identification of gene locations within newly sequence phage genome. This workflow uses several programs to locate and archive the gene locations including DNA Master, Glimmer, GeneMark, and Aragorn. Functional annotations are added by identifying significant matches to proteins with known functions through queries to databases at Phagesdb.org, NCBI, CDD and HHPred. Annotation evidence is typically recorded in electronic files that allow others to check the annotations. The quality review of each annotated genome can be arduous, depending on the quality, and completeness of the annotation evidence.   
 PECAAN was developed to facilitate the collection of gene evidence and implement a consistent and complete presentation during the annotation and the quality review. This database driven web application runs on many web-enabled devices including laptops, tablets and phones with an html5 compliant browser.   
 New phage entries into PECAAN use data derived from a FASTA file, DNA Master, Starterator, and GeneMark. The gene calls and the Glimmer / GeneMark notes are obtained from the DNA Master “Documentation” text. The start site table is also exported from DNA Master. Starterator input comes from the summary table of the starterator output. The host-trained GeneMark output can be submitted to PECAAN as a pdf file and when displayed in PECAAN it scrolls horizontally instead of vertically, thus allowing easy connection of each reading frame between pages. Whenever a new phage is entered and whenever a new start site is selected, PECAAN automatically pulls matches for the gene’s protein from Phagesdb, NCBI protein BLAST, the CDD and HHPred.  
 Annotation of a gene in PECAAN consists of five steps: 1- Choosing the start site from a table of all possible start sites, 2- Entry of a function or NKF, 3- Checking boxes next to the function evidence from Phagesdb BLAST, NCBI BLAST, CDD, and HHPred, 4- Entry of notes. 5- Clicking the “Submit” button to enter the annotator’s name, a time/date stamp, and the changes into a log and database.  
 To identify tRNAs and tmRNAs, the DNA sequence is scanned with Aragorn and tRNA-Scan. The evidence from each is displayed and evidence boxes can be checked to select the correct models.  
 A checkbox at the top of each gene page allows a choice to include the gene or not to include the gene in the final annotation. A button also allows additional genes to be added.   
 PECAAN can export four files to be used in preparation for genome submission. Two can be used to program DNA Master through the “Documentation” “Parse” option. Gene notes and author files can be exported separately. If student annotations are to be graded, the change log can also be exported.  
 PECAAN is written in JAVA and is available on GitHub. We are happy to accommodate additional users and phages from the SEA. We welcome suggestions for extensions to PECAAN.