CONSIDER FOR TALK

10th Annual SEA Symposium Abstract

Western Kentucky University

Bowling Green KY

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



Claire A Rinehart

PECAAN, a Phage Evidence Collection And Annotation Network

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

PECAAN was developed to facilitate the collection of gene evidence and to 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 and GeneMark. The gene calls and the Glimmer / GeneMark notes are obtained from internal PECAAN execution of these programs. Pham information is pulled from Phagesdb.org and Starterator information for the phams comes from Washington University. The host-trained GeneMark output is 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 database matches for the gene’s protein from drawn from Phagesdb, NCBI protein BLAST, the CDD, HHPred and TMHMM.  
 Annotation of a gene in PECAAN consists of five steps: 1- Choosing the start site from a table of all possible start sites after consulting the Glimmer, GeneMark, and Starterator suggestions, 2- Entry of a function or NKF, 3- Checking boxes next to the function evidence from Phagesdb BLAST, NCBI BLAST, CDD, HHPred, and TMHMM, 4- Entry of notes. 5- Clicking the “Save” button to enter the annotator’s name, a time/date stamp, and the changes into a log and the 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 including genes with multiple coding domains, such as the frameshift in the tail assembly chaperones.   
PECAAN can also display live PhamMap results that lets you view your current genome along side one other genome from the same cluster. It is easy to change between phages in the comparison view allowing the user to compare several genomes to the genome that you are annotating.   
PECAAN also allows viewing of the genome along with the six reading frames which are highlighted with the called genes. All potential start (green) and stop (red) sites are highlighted making it easy to find frameshifts.  
 PECAAN can export several files to be used in preparation for genome submission. Two can be used to populate 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.