CONSIDER FOR TALK

2022 SEA Symposium Abstract

University of South Florida

Tampa FL

Corresponding Faculty Member: Richard Pollenz (pollenz@usf.edu)

Annotation of CaiB, a DR Cluster Actinobacteriophage that Infects Gordonia rubripertincta

Nader Abdalla, Esteban Aldana, Veronica Alvarado-Fernandez, Bruna Arenales Salgado De Oliveira, Diane Fakhre, Amelia Haymond, Katelyn Helton, Adi Kanchibhatta, Jahwanza Knight, Sydney Marshall, Maomi Laine Martinez, Arielle Merkher, Savannah Morrow, Katie Nguyen, Jahanvi Patel, Somesh Patel, Aarya Preetam Satardekar, Pravalika Rayala, Kira Ruiz-Houston, Shifa Mohyuddin Shaikh, Adrian E Terron Osorio, Rachel Weitz, Bienna Welsh, Louis Otero, Richard S Pollenz

Phages are the most abundant organism on Earth and are biologically relevant since they have been used as agents to fight infections to antibiotic-resistant bacterial strains. The SEA PHAGES consortium categorizes phage into clusters based on nucleotide identity. CaiB is a siphoviridae phage that was isolated on Gordonia rubripertincta NRRL B-16540 using an enrichment strategy. The CaiB genome is 61,620bp, is 68.6% GC-rich and contains 85 protein coding genes and no tRNAs. The goal of the project was to annotate CaiB and evaluate several repeated nucleotide regions that are upstream of at least 6 genes in the 3’ end of the genome. CaiBN is in the DR cluster that only contains 8 annotated phages and they show a high degree of diversity in the 3’ end of the genome. CaiB is similar in genome organization to other siphoviridae actinobacteriophages in that the 5’ end of the genome with a high number of structural genes upstream of the putative tape measure gene (gp26), followed by several minor tail genes. However, CaiB does not appear to have a canonical slippery sequence in the two genes upstream of the tapemeasure although there is a CCCGGGG that was identified. In addition, the first gene in CaiB is not the terminase and it appears that genes 1-5 make up a metabolic operon containing: a ParB-like nuclease domain protein, two oxidoreductases and a adenylate kinase. Four of these five genes have 4bp ATGA overlaps to the upstream gene. Genes 16-19 also have 4bp overlaps to the upstream genes with genes 16, 17 and 18 all encoding proteins with no known function, but with four transmembrane domains. These TM genes are 21,000bp upstream of the putative lysis cassette (genes 43, 44 and 45) that encodes a split lysin A and lysin B with no identifiable TM proteins immediately upstream or downstream. The majority of the genes in the 3’ end of the genome are of unknown function with a small number of DNA replication enzymes. CaiB has six regions of repeated genome identity upstream of putative ATG start codons that are located within the 80-100bp gaps to the previous genes. All genes are found in the reverse orientation and many of these regions are also conserved in other DR phages. The repeated regions are approximately 60bp long and contain 3-4 subregions of 7-12bp with near 100% conservation. All of the genes associated with these upstream regions are of unknown function. Since phages typically have a verse dense genome with minimal gaps between genes, these regions may be involved in regulation of gene expression and are worthy of future research. We thank the SEA PHAGES consortium and the Hatfull lab for their support of this project.