Apr 17th, 3:00 PM - 4:00 PM

A Software Pipeline for Analyzing Viral Sequences in Bacterial Genomes

Conner J. Copeland
University of Montana, conner.copeland@umconnect.umt.edu

Patrick Secor
University of Montana, patrick.secor@mso.umt.edu

Travis J. Wheeler
University of Montana, travis.wheeler@umontana.edu

A Software Pipeline for Analyzing Viral Sequences in Bacterial Genomes

Conner Copeland[1], Patrick Secor[2], Travis Wheeler[1]

[1]Dept. Of Computer Science, University of Montana, Missoula, MT, USA

[2]Dept. of Biological Sciences, University of Montana, Missoula, MT, USA

Introduction:

- *Pseudomonas aeruginosa* is a bacterial pathogen that infects wounds, burns, airways, and medical hardware
- The majority *of P. aeruginosa* isolates are infected by a filamentous virus called Pf bacteriophage
- Pf bacteriophage integrate into the bacterial chromosome as a prophage and enhance the virulence potential of *P. aeruginosa*[3]
- To better understand the relationship between Pf bacteriophage and *P. aeruginosa*, we have created a software pipeline that identifies viral sequence in *P. aeruginosa* genomes and captures key information about the nature of the insertions. This could translate to better treatment regimens in the future

Fig 1. A chronic *P. aeruginosa* wound infection

How it Works:

- Utilize the ~3,000 *P. aeruginosa* genomes available on public databases

- Acquire Pf bacteriophage reference sequences. We have 50 reference prophages, with a typical length of 12 kilobases per phage

- At least 150,000 operations are required to search for viral sequences in the bacterial genomes. To achieve this, we use Montana Tech's High Performance Computing Cluster

- On the cluster, nhmmer[4] is used to annotate any viral sequences found in the bacterial genomes

- Data are then parsed and viral sequences are annotated using known viral proteins

- Plots are generated to visualize where viral sequences are inserted (**Fig 4**)
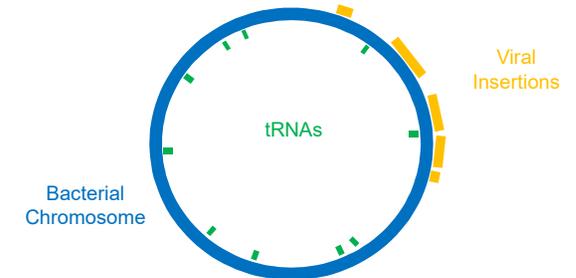
Results: Where Viral Sequences Land

Fig 3. Bacterial genome with several full and partial viral insertions. Interestingly, none are flanked by the canonical tRNA att sites, and all occur near each other. tRNA sites were annotated with tRNAscan-SE[5]. Not to scale.

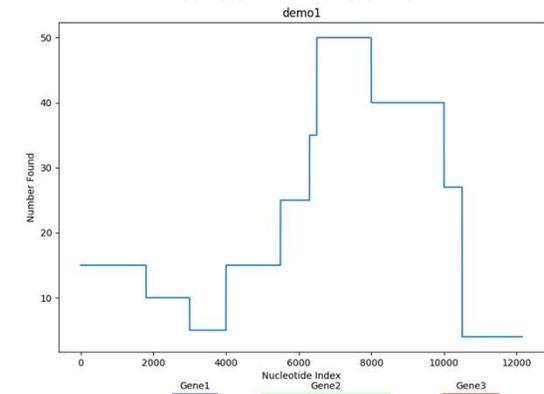## Results: Which Regions of Viral Genomes Get Inserted

Fig 4. This plot shows how often each region of the prophage genome is found inserted into a bacterial genome. Plots like this simplify the visualization of viral genes inserted into (and retained by) bacterial chromosomes. Inserted viral genes may play a role in the severity of *P. aeruginosa* infections.

Diagram of a Viral Chromosomal Insertion:

Before Insertion:

Fig 2. A normal section of bacterial DNA with a tRNA sequence. *P. aeruginosa* prophages likely use tRNA sequences as targets to guide insertion.

After Insertion (Canonical):

As part of the process of inserting viral DNA into the host bacterium, the target tRNA site is copied. This leaves the prophage's sequence flanked by identical tRNA sequence.

[3]Sweere et al., *Science,* 2019; [4]Wheeler et al., *Bioinformatics,* 2013; [5]Lowe et al., Nucl. Acids Res., 1997. Photo is from https://bit.ly/2Gl66Hw