Phage Lab Wiki

"A strong feeling of adventure is animating those who are working on bacterial viruses, a feeling that they have a small part in the great drive towards a fundamental problem in biology".

—Max Delbrück, 1946

Welcome to the VCU Phage Lab Wiki!

We are part of the SEA PHAGES program, a nation-wide initiative sponsored by Howard Hughes Medical Institute and the Hatfull Lab at University of Pittsburgh to change the way we teach introductory science courses. Similar courses are taught at 80+ universities and high schools across the country. We were lucky to host a regional symposium at VCU, with guests from 4 universities and one highschool...and all of them love phages!

In the Phage Lab course (BNFO 251/252 Phage Discovery I/II) you will learn how to undertake scientific research through the discovery and characterization of bacteriophages (viruses infecting bacteria). The students in the fall course (BNFO 251) are ‘phage hunters’. They discover and characterize their own virus using wet lab techniques. The students in the spring course (BNFO 252) learn bioinformatics by exploring bacteriophage genome sequences through genome annotation and comparative genomics projects.

Here are the posters presented by phage students over the last several years:

Fall 2015 Phall Phage Phaire, hosted by VCU!
Summer 2015 7th Annual SEA PHAGES symposium student poster
Summer 2015 7th Annual SEA PHAGES symposium faculty poster
April 2015 VCU Poster Symposium for Undergraduate Research and Creativity
Summer 2014 6th Annual SEA PHAGES symposium student poster

We will use this wiki as our electronic lab notebook.

Students: Please create a child page under your section's parent page and use your name as the title. Pages are added using the + symbol to
the upper right. Each day you are working, you should update your wiki page with your progress, observations, and next steps.

You will be required to keep a detailed record of all that you do in the phage lab. This is REAL research that will one day be published, and you have a responsibility to properly document your experiments. We will use this wiki as an electronic notebook for this semester of the course. You will be able to share your results with your classmates, as well as view their results, making collaboration easier. You should never copy between pages! If you copy your classmates' notebook(s), you will immediately be reported to the honor's council.

Each class or time you work on annotation and research questions, you should document:

- the date,
- objective,
- problem addressed,
- incorporate pertinent data screenshots,
- record name of associated data files, etc.
- Be sure to add a “next session...”!

Example of a notebook update:

**Wednesday January 15, 2014**

Today I will examine Nigalana genome for genome ends.

1. Genome ends

I used “Assembly View” to identify a drop off of reads that might indicate genome ends. This is what it looked like.

![Graph showing genome coverage](image)

Hmmmm.....I’m not seeing a drop off of reads like I’ve seen many times for mycobacteriophages. Instead depth of read coverage in consistent across the genome as indicated by the green line. But there is an interesting "plateau" from around 6000-10,000 bp. I confirmed depth of coverage at several locations using the "Depth of Coverage at Cursor" tool under Misc in the "Show Contig" view:

<table>
<thead>
<tr>
<th>Position</th>
<th>Depth of Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>91</td>
</tr>
<tr>
<td>6456</td>
<td>129</td>
</tr>
<tr>
<td>6457</td>
<td>258</td>
</tr>
<tr>
<td>9323</td>
<td>197</td>
</tr>
<tr>
<td>9324</td>
<td>108</td>
</tr>
</tbody>
</table>

Each repeat end can be visually observed as a drop off of reads. Here is what the end at 6457 looks like:

![End of genome at 6457](image)

Here is what the end at 9323 looks like:

![End of genome at 9323](image)
When many reads end at the same place in a genome assembly, we are suspicious this is in fact a genome end. Therefore, the genome ends could also be identified by searching for these features while scrolling through the reads to examine quality, etc.

The terminal repeat is probably 9232-6457 = 2866 bp long in the phage genomic DNA. Since it is much bigger than the 454 read size (400-500 bp), the assembly software stacks up the reads, resulting in a higher depth of coverage compared to the rest of the genome.

During my next session, I will use Blast to see if there is another closely related phage with a terminal repeat, examine the genome for consensus base calling quality, and make sure I have a complete genome in my assembly.

**Monday January 27: Finishing Nigalana genome**

Today I will incorporate sanger reads into the Nigalana 454 genome assembly to improve the base calls for low quality consensus positions and other positions identified for finishing.

Position 2550 was identified as a position where the assembly software couldn’t tell if there were 9 or 8 As in a polyA run. General quality in the area was low.

<table>
<thead>
<tr>
<th>2530</th>
<th>2540</th>
<th>2550</th>
<th>2560</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>CAAAAAAAA</em>**TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>CAAAAAAAA</em>**ATA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>CAAAAAAAA</em>**ATA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>CAAAAAAAA</em>**ATA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
</tr>
<tr>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>CAAAAAAAA</em>**ATA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>CAAAAAAAA</em>**ATA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>CAAAAAAAA</em>**ATA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>CAAAAAAAA</em>**ATA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
</tr>
<tr>
<td>A<strong>GAATAT<em>Caaa</em></strong>a**TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>A<strong>GAATAT<em>Caaa</em></strong>a**TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>A<strong>GAATAT<em>Caaa</em></strong>a**TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>A<strong>GAATAT<em>Caaa</em></strong>a**TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
</tr>
<tr>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>Caaaaaaaa</em><strong>a</strong>TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>Caaaaaaaa</em><strong>a</strong>TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>Caaaaaaaa</em><strong>a</strong>TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>Caaaaaaaa</em><strong>a</strong>TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
</tr>
<tr>
<td>gtc<em>agttaa</em>gaatat<em>caaaaaaaa</em>**ta<em>taag</em>ggtag*ttac</td>
<td>gtc<em>agttaa</em>gaatat<em>caaaaaaaa</em>**ta<em>taag</em>ggtag*ttac</td>
<td>gtc<em>agttaa</em>gaatat<em>caaaaaaaa</em>**ta<em>taag</em>ggtag*ttac</td>
<td>gtc<em>agttaa</em>gaatat<em>caaaaaaaa</em>**ta<em>taag</em>ggtag*ttac</td>
</tr>
<tr>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>Caaaaaaaa</em><strong>a</strong>TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>Caaaaaaaa</em><strong>a</strong>TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>Caaaaaaaa</em><strong>a</strong>TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
<td>&lt;GTC<em>AGTTAA</em>GAATAT<em>Caaaaaaaa</em><strong>a</strong>TA<em>TAA</em>GTT<em>GGTAG</em>TTAC</td>
</tr>
</tbody>
</table>

Two reads of Sanger data were incorporated into the 454 assembly. We can see 9 As in the Sanger read compared to 8As in the 454 read. The new “A” was added to consensus sequence.
The file was saved as ace.6.

Annotating gene 1 in Nigalana genome: