Rough draft

Rachel Siefring

April 28, 2015

**Introduction:**

Bacteriophages (phages), viruses that infect bacteria, are predicted to outnumber all of the bacteria on the earth by a factor of four. Studies concerning phages continue to point out that phages are extremely diverse. Because of their astounding diversity, further study must be completed to fully understand their biological complexity and importance.

When infecting bacteria, phages inject their DNA or RNA. In order to counteract this effect, bacteria have developed clustered repetitive interspersed palindromic repeats (CRISPRs). Much research has been done to learn about CRISPRs and the mechanisms that they use to halt phage replication and production.

The research group from the Seed et al 2013 article found a CRISPR in five Vibrio ICP1-related phages. Because CRISPRs are a prevention mechanism primarily used by bacteria, this finding is very stunning. If one phage has a CRISPR, there are doubtless others with it also.

**Methods:**

I decided to use "Counts-of-kmers" to find the counts of 23-mers.  I used 23 specifically because it was the smallest nucleotide length of a repeat in Table 1 of He et al.  I could have gone with a higher kmer length.  But, if I'm able to catch the smaller length of a repeat, I'll be able to see the longer portion of the repeat.  This would probably work better than starting with a longer kmer length and missing the shorter repeats.

I'm going to BLAST the [ICP1\_2004\_A](http://www.ncbi.nlm.nih.gov/nuccore/HQ641354.1) genome against other genomes/phages.  I'll use the phages with the closest sequence homology to look for CRISPRs.  [These](http://blast.ncbi.nlm.nih.gov/Blast.cgi) were my results.  None of the phages that came from the BLAST result were different from any of the phages in the Seed, et al paper.

1. **What's my next best option??**
2. I decided to look at the cluster A1 mycobacteriophages.
3. In Figure 1 of the article below, it highlights 20 cluster A phages (these are cluster A1 and A2 phages that were in Figure 1).  Below is a picture of the BioBike code that I used, with each of the following phages inserted into the organism box.  I started the threshold at 10, but bumped it up to 20 after mycobacteriophage DD5.

1. Bethlehem: NIL
2. Bxb1: NIL
3. DD5: NIL
4. Jasper: NIL
5. KBG: NIL
6. Lockley: NIL
7. Solon: NIL
8. U2: NIL
9. Che12: NIL
10. D29: NIL
11. L5: NIL
12. Pukovnik: NIL
13. Peaches: NIL
14. Bxz2:
15. I got the following output below.  I'm not quite sure how to investigate my results.

So, I obviously didn't get anywhere with that.  I looked back at my previous BLAST results to see if there were any other possibilities that I could try.  I tried my BioBike sequence on Vibrio phage [PWH3a-P1](http://www.ncbi.nlm.nih.gov/nucleotide/468642750?report=genbank&log$=nuclalign&blast_rank=13&RID=KV1GZPHH01R).

(COUNTS-OF-K-MERS V\_PWH3A\_P1 23 THRESHOLD 10 :LINEAR)

**I decided to compare GC content between 2 different organisms, just to see if there was a big difference.** The GC content of most mycobacteriophages is above 60% (Table 1 of the paper below):

1. Jacobs-Sera, D., et al. (2012). On the nature of mycobacteriophage diversity and host preference. Virology. 434: 187-201.
2. Most of the mycobacteriophages in Table 1 of this article shows that most mycobacteriophages have a GC content above 60%.

**So, I want to continue with mycobacteriophages…but, maybe the GC fraction of mycobacteriophage doesn't suggest a CRISPR…**

I also searched for the CRISPR repeat from the Vibrio phage in Vibrio-related bacteria in BioBike. No hits were returned.

Additionally, a sequence downstream of the last CRISPR repeat is thymine-rich. I tried looking for a similar region in Vibrio-related bacteria. No hits were returned.

**Discussion:**

Thus far, I haven’t found any phages with a CRISPR, based on the very small sample of phages that I looked at. My next step would be to construct a general algorithm that can be used to search for CRISPRs in all phage. By looking at all phage in Biobike, I will be more likely to find a CRISPR and see the frequency of CRISPRs in phage.

If I can’t find a CRISPR in another phage, I could turn my attention to looking for a possible mechanism that caused the CRISPR to be inserted into the genome.

Right now, it seems very unlikely that I’ll find a CRISPR in another Biobike phage.