**General View of Different Clusters**

Now that you’ve seen how the annotation process works in general, it’s time to look forward to the specific types of genomes you might annotate. Often, specific clusters have shared characteristics that can assist in annotation; annotating a Cluster A genome is different than annotating a Cluster C genome. First, let’s get a view of the different clusters that are out there.

Go to the cluster list on PhagesDB (<http://phagesdb.org/clusters/>).

1. Which cluster has the most members? Which one has the most subclusters?
2. Which clusters have the longest average genome size? Which clusters have the shortest average genome size?
3. Which clusters have the highest and lowest GC content?
4. How many Singletons (phages with no other cluster members) are there? Bonus: why haven’t you found their relatives?
5. What’s the most common cluster for phages that infect Microbacterium? Arthrobacter? Gordonia? Streptomyces?
6. Which cluster (other than Singleton) has members that can infect more than one genus of bacteria?
7. What cluster(s) have we not found any of for several years?

Click on Cluster K to go to its page.

1. What are the most and least common subclusters of Cluster K?

Click on Subcluster K1, then click next/previous subcluster as needed to answer.

1. How do the Cluster K subclusters differ in terms of genome size, GC content, tRNAs, etc.?

**Getting Familiar with Relevant Clusters**

Now let’s look at a few specific clusters in more detail.

1. Go to the cluster list on PhagesDB (<http://phagesdb.org/clusters/>). Take a look through the list of clusters. Which clusters are most commonly found for the host you used? For other hosts?
2. We’re asking you to specifically look at Clusters A, C, EA, and two others of your choosing. The two you choose should have members that infect the host you used for phage isolation. Choose one that is fairly common, and one that is less common. What are the two additional clusters you chose?

**NOTE:** Work through the remaining steps using one cluster at a time (start with Cluster A), then return here to go through again with each of the other clusters you’ll investigate.

1. Choose a phage from your cluster, then BLAST it on PhagesDB. What does the pattern of BLAST hits look like? I.e., are there lots of extremely similar hits, or many hits that are somewhat similar, or some that are similar for part of the genome but not the rest?
2. Write down the names of the two most similar phages to the one you BLASTed. In addition, write down the name of a genome that shows up in the BLAST results but isn’t as closely related.

Go to Phamerator (phamerator.org), find and select the four genomes from Step 4 (your original phage, plus two close hits, plus one less-close hit), and click View Map.

1. Are the genomes you’ve selected Draft annotations or final annotations? (This will be important when doing comparative analyses!)
2. What are some general observations about this group of genomes and genes? How many genes are in these genomes? Are their phamilies mostly conserved? Are their genome lengths fairly consistent? Are their genomes transcribed entirely forwards, or mostly forwards, or a mix? How does the left arm compare to the right arm in terms of shared pham content, nucleotide similarity, gene size, etc.? Anything else that stands out?
3. Look for any genes with functions called. Approximately what percentage of the genes in these genomes have functions called? In what part or parts of these genomes do those genes tend to appear? Are there any “uncommon” functions called? Are there functions called in a subset of the four genomes but not all, even when genes are in the same pham?
4. Can you find any places where one genome appears to have an insertion or deletion relative to another?
5. Do you see any places where the genomes share a pham but nucleotide sequence is not conserved? Or, vice-versa, where phams differ but there is nucleotide sequence similarity?
6. Are there any noticeable gaps where no genes are called?

Repeat steps 3-11 for Clusters C, EA, and the other two clusters you chose, then continue investigating clusters until your curiosity is satisfied. (Ours never is!) When you and your class receive your genome, these are some of the first steps you should take to get a good overview of what your specific annotation process ahead will entail.