Lab 14: Comparing Phage Genomes using Phamerator / Figtree exrcise

Your name:
Your email address:

In our search for inteins we had used phagesdb, blast, and dotplots to compare phage genomes and their genes (see Assignments 4, 6, and 8). In lab 4, exercise 5 you picked a putative intein sequence, used blastp to identify the gene the putative intein belonged to. You also looked for homologous genes in other phages that did not contain the intein. In addition to using blast and dotplots, a nice phage genome comparison tool is PHAMERATOR. To use it, we need the information you gathered in exercise 4: the **name of the phage that contains the intein containing gene, the number of the gene containing the intein, and possibly and the name of a phage and gene number of the intein free homolog.**

Which intein containing gene did you analyze in lab 4? Phage name followed by gene number.
e.g., Bananafence_260 (replace this with the name of your gene).

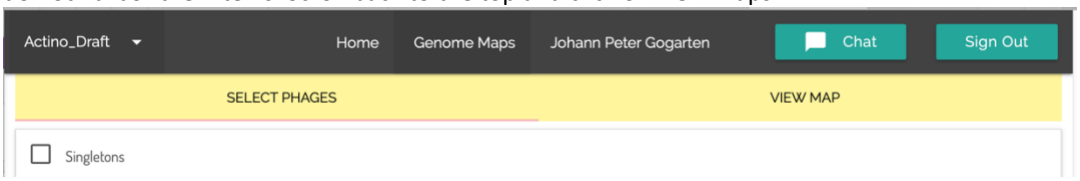
Did you identify an intein free homolog? If yes give the name:
e.g., Willis_262

To use phamerator, you also need to know the cluster and subcluster to which these phages belong. (Go to phagesdb.org and search for the phage (do not enter the gene number in the search-box).

Phage cluster:
e.g., C1

Go to https://phamerator.org, sign in (you need to create an account first!)

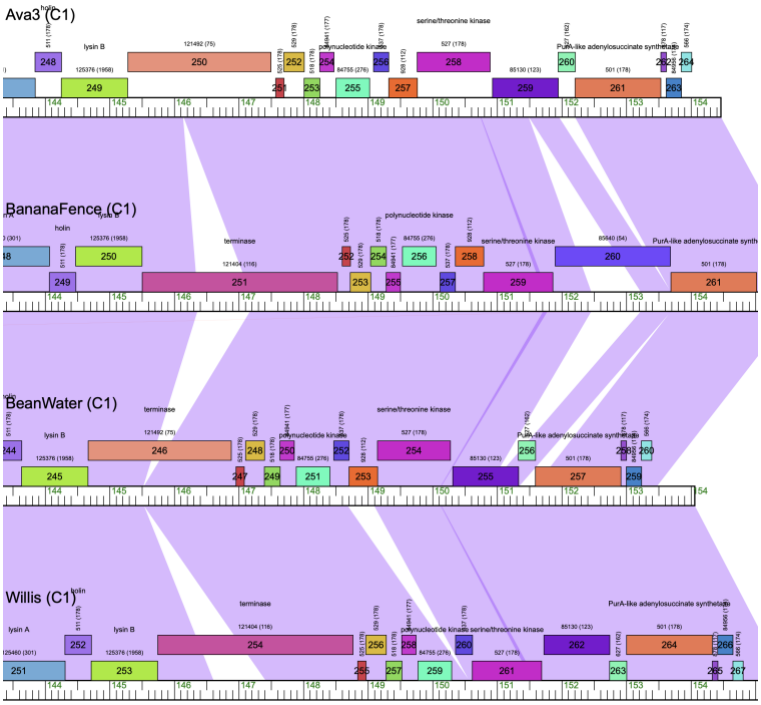
Go to Genome Maps - Select Phages. *I use C1 as example*. (Explanation of the menu items is in the note below. Also, if you move back and forth between the different menus, sometimes one sees only a white page. Use the levers to move to the left.)
Scroll down the list, click on **the line** that says C1 (**do NOT place a checkmark**), in the list of individual genomes that opens, select (place check marks) for genomes that are not in draft stage, and for the phages that harbor or do not harbor the inteins. Scroll back to the top and click on view maps.



Note the forward ORFs are on top, the backwards ones are below the ruler.
Do you have any ORFs on both strands? If yes, what do they encode?

ORFs in the same color belong to the same pham (Phage protein family).
The purple background depicts similar nucleotide sequences (this works in Safari, Chrome gets the coloring wrong).

Go back to select additional phages: select more genomes from cluster C1 and possibly other phages that were picked in blast searches in lab 4.
The image in phamerator might look something like this:



In this example, Bananafence seems to have inteins in gene 251 and 260, Ava3 and BeanWater have none of these inteins, and Willis has the intein in the terminase. Note that for both genes the intein free and the intein containing gene were assigned to different PHAMS. There are other insertions/deletions that are unrelated to inteins, e.g., a homolog to BeanWater_253 is missing in Willis.

In the Map, you can drag the genomes (click on the name) up or down, so that the genomes you are interested in are next to one another.

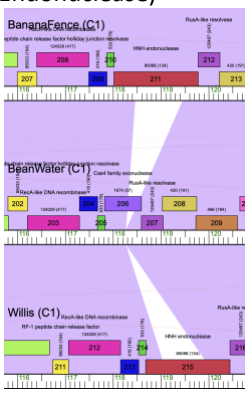
Do you find anything noteworthy?

Provide a list of genes with putative inteins, and their intein free homologs.

intein containg gene (cluster) function	intein free homolog (cluster) function	assigned to the same PHAM?
Willis_254 (C1) Terminase	BeanWater_246(C1) Terminase	No
Bananafence_260 (C1) NKF	BigsWole_267 (C1) NKF	No

Other short comments

copy paste some images. (replace with your image)
E.g., for the above genomes, another intein is suggested in Bananafence_211 (misannotated as HNH Endonuclease)



Notes:
At the bottom of the gene map page is an icon

The linked menus allows you to change setting for coloring and labels, 

and to export the map as a vector graphic  .
For many things (copying items in your notebook), screenshots (command +control + shift + 4) are sufficient, but if you want to put this into a poster and need better resolution, the vector graphics work great. You can directly open them in Powerpoint, or you can edit them (e.g., move labels around) in Adobe illustrator ($$$) or Inkscape (free).

Similarly, on the select phages page is a menu

This allows you to

- move to the top of the screen,
- display all the phages in all of the cluster, or
- remove all phages that you had previously selected.

The latter is nice, because phamerator remembers your previous selection ....

If you have time and did not finish the figtree exercise from last week, do it now (also copied below).

Finished?

Do not forget to email your competed worksheet to gogarten@uconn.edu and daniel.s.phillips@uconn.edu

2) Tree Images in Figtree

Figtree is a very useful program to create publication quality trees. Is also can save them as vector graphics, which makes it easy to edit them in Inkscape or Adobe Illustrator. See slides from Monday's class.

1) Check if figtree is installed on your computer. If not, download and install figtree from the github page

2) The trees we will look at today were calculated for exteins and inteins of the terminase gene in the Actinophages from subcluster C1. The nucleotide sequences were retrieved from phagesDB, renamed to include the state where the phage was isolated, aligned in seaview using muscle, and separate datasets created for intein and extein sequences. For the latter, three different files were created

- a multiple fasta file containing all terminase exteins from cluster C1
- the same as above, but the very divergent sequence from Toneneli removed
- only the extein sequences from phages that contain an intein
- all intein sequences

Each of these phylogenies were calculated using iqtree2. These phylogenies are available in Newick format in this text file

- Open the text file with the tree in a text editor (BBEdit or Notepad++).
- Copy the intein tree onto the clipboard
- Open figtree
- Click in the central figtree window
- Paste the tree (ctrl-V)

3) It is a good idea to start with the intein tree.
Do the following:

- Select "Node" in the central selection field
- click on the long internal branch, select re-root
- move root length lever to the left (we have no idea where this is rooted.
- place check mark into branch labels (in the menus on the left)
- open branch labels menu (on the left)
- select label in the display pull-down menu (these are the bootstrap support values that iqtree wrote into the newick formated file)

- in the selection indicator on top, select taxa
- click on one of the two basal branches
- select a color for the leaves (aka OTUs)
- click on the other basal branch and select a different color
  (Red and green or blue and green work well)

As many of the branches have very low support values, we can display the support in a different way

- Open the "Appearance" menu on the left
- under "colour" by select "label"
- check mark into gradient
- click on setup Colours (sic)
- Adjust the hue levers at the bottom to have the spectrum go from yellow to red (or from yellow to blue); move the other three levers to the right
- The well supported branches now are in blue and the branches that do not have high support values are in green and yellow

- You can modify other things like font sizes, placement of the labels, line width, add a legend, ...

Figtree is intelligent.
The tree we have displayed now was calculated from inteins in terminase genes in genomes of tailed actinophage (viruses that infect actinobacteria) subcluster C1. The terminase is involved in packaging the DNA into the phage heads)
The names are composed of the state (or other indication of from where the phage was isolated). The displayed tree was calculated from the intein only.
The extein sequences have the same names, but obviously the extein tree is expected (and it is) rather different.

Open a new Figtree window (***leave the old tree open in its window***!)

Copy paste one of the extein trees (below) into the empty Figtree window.
There is no harm, if you do this for all three extein trees.

If everything works as expected, the colors selected for the leaves on the intein tree are applied to the new tree.
Under file select new.

Re-root the tree in one of the longer branches (so that it looks balanced). If you use the extein tree with the toneneli sequences you could use this as outgroup. Set the root branch length to zero. "Colour" the branches based on support values (see above).

The clans defined by the intein tree are not reflected in the extein phylogeny. The two intein types were transferred several times between different exteins. Note the nearly identical intein between red and green leaves.

Give two examples for very similar inteins found in divergent exteins clans, i.e., list at least two pairs of actinophages with divergent extein sequences but with similar inteins reflecting recent intein sharing.

Can you find any very similar inteins from a similar geographic distribution, but the exteins are dissimilar?