Thus far our tutorials have really focused on the phage side of things and annotating and understanding the genes present on your particular phage genome(s) of interest. In this tutorial we will take a (quick) look at the host side of things – you may/may not choose to perform all of the following experiments for your particular project, it will all depend on your project hypothesis/aim(s)….

First, to carry out any of the analyses described below, you will probably need to identify a specific bacterial host genome and its accession number (or, in some cases, need to download that genome sequence file).

The best way to find your genome of interest is probably through the NCBI web interface (Figure 1), e.g., by searching for your particular species of interest.

A screenshot of a computer

Description automatically generated

Figure 1. The NCBI genomes resources.

Note that the genomes you find may be of varying quality levels (try to select well-assembled genomes wherever possible). Many *Streptomyces* genomes available in the NCBI databases are incomplete.

Compare the assembly statistics for two different *Streptomyces scabiei* genomes (Figure 2 and Figure 3). Which genome is better assembled, and why? Which should you choose for your analyses?

How many contigs do you expect to find in a *Streptomyces* genome?

A screenshot of a computer

Description automatically generated

Figure 2. Assembly statistics for ASM9130v1

A screenshot of a computer

Description automatically generated

Figure 3. Assembly statistics for ASM3689866v1

1. **Option 1: Search for specific genes/proteins of interest in a host genome.**

Identify host genome(s) of interest and import genomic data into Galaxy – then use BLAST to search for your sequence(s) of interest

Depending on your project aim(s), you may wish to analyse the genome of known (or potential) hosts for your particular bacteriophage(s) of interest.

You can import the host genomes into Galaxy using the **GenBank Accession** number of the genome and the **NCBI Accession Download** tool that you used to download phage genomes.

Once you have imported your genome(s) of interest, you can use **Prokka** to annotate them (similar to the way that you used pharokka to annotate your phage genomes).

A screenshot of a computer

Description automatically generated

Figure 4. Use of Prokka to analyse bacterial genomes in Galaxy.

When you have imported and annotated your genome(s) of interest, you can use the **NCBI BLAST+ makeblastdb** command to make BLAST databases, and the appropriate BLAST searches to search nucleotide or protein databases as required for your particular project aim(s). You could use this approach, for example, to search for particular anti-phage defenses (e.g. Restriction-Modification systems….), for particular toxin genes, etc.

1. **Option 2: Analyse a host genome for presence of CRISPR systems**

You can use CRISPR-Cas++ finder to look for the presence of CRISPR systems in your bacterial host(s) of interest.

1. Go to <https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index>
2. You can input a bacterial genome sequence of interest, but you may very well find that the CRISPRCasdb already contains your genome so it is worth search there first. (You can search by NCBI accession number or binomial species name.)

A screenshot of a computer

Description automatically generated

Figure 5. CRISPRCasFinder program website.

A screenshot of a computer

Description automatically generated

Figure 6. CRISPRCasDB results for Streptomyces albus.

1. Record and analyse your results. You might choose to further analyse these CRISPR-Cas systems (e.g. to look for particular spacer sequences within your phage genomes of interest)

There are also CRISPR prediction tools in Galaxy (or see the defensefinder tool discussed next…)

1. **Option 3: Analyse a host genome for predicted antiphage defense systems using DefenseFinder**

DefenseFinder (Tesson et al. 2012) will predict antiphage systems in bacterial genome sequences (like antiSMASH, it has its limitations….)

1. Open the DefenseFinder website at: <https://defensefinder.mdmlab.fr/>
2. You can browse the RefSeq DB (Figure 6) or upload a particular genome of interest to analyse.

A screenshot of a computer

Description automatically generated

Figure 7. Browsing the RefSEQ DB with a Boolean search query.

To upload a particular genome of interest, you will need to first download it from NCBI (Figure 7) and then unzip it.

A screenshot of a computer

Description automatically generated

Figure 8. Downloading a genome of interest from NCBI.

You should then be able to upload it directly to the DefenseFinder webservice (Figure 8), either by browsing to the file location or by dragging it into the upload box where indicated.

A screenshot of a computer

Description automatically generated

Figure 9. The DefenseFinder webservice.

1. Click on “Results” (Figure 7) and then you can click on the “type” link for each result in turn to learn more about the system (and evaluate the level of evidence supporting a particular defense system).

You can also, if desired, pan and zoom to a desired view of the chromosome that highlights DefenseFinder hits of interest, and export this as a .svg or .png file.

A screenshot of a computer

Description automatically generated

Figure 10. Example DefenseFinder results.

1. **Option 4: Analyse a host genome for its potential to produce putative secondary metabolites**

*Streptomyces* are prolific producers of secondary metabolites – perhaps best-known for their production of antibiotics, but some of these have been shown to act as chemical defenses against phage infection (e.g., c.f. Kronheim et al. 2018). You might therefore, if interested in bacterial-phage interactions, wish to determine whether any of your bacterial host strain(s) of interest can produce these known anti-phage molecules.

One way of predicting the secondary metabolites potentially produced by your strain of interest is through use of a tool such as **antiSMASH** (though note that this tool does have limitations….)

1. Go to the antiSMASH webpage, <https://antismash.secondarymetabolites.org/#!/start>
2. You can input your genome of interest’s NCBI accession number (Figure 4) or else upload the data to antiSMASH directly. You may wish to provide your e-mail address as antiSMASH jobs can take some time to run (depending on the server). [Note that antiSMASH is quite strict about only accepting GenBank accessions – NOT assembly (ASM) numbers!]

A screenshot of a computer

Description automatically generated

Figure 11. Using antiSMASH via the web interface.

1. antiSMASH will predict the biosynthetic gene clusters on your provided genome (Figure 5) and identify the most similar known clusters – keep in mind that these are just predictions, however!

You can click on any identified region of interest to obtain a more detailed view of the genes present, including predicted gene products, their functions, and nucleotide and amino acid sequences.

A screenshot of a computer

Description automatically generated

Figure 12. Example antiSMASH results.

**5. Option 5: Other**

What other phage defense systems might your host strain(s) of interest have? How might you search for them?