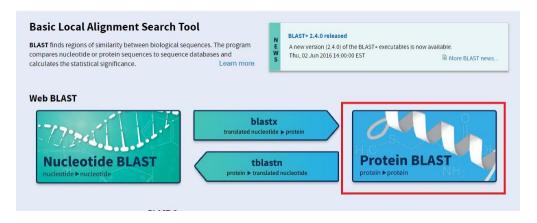
COMMUNITY SCIENCE PROJECT WORKFLOW (with example)

Introduction

How does exchange of genetic material between bacteriophages and bacteria drive bacterial evolution? This is the question we're attempting to answer with the Community Science Project. You can help us gather data to address this question.

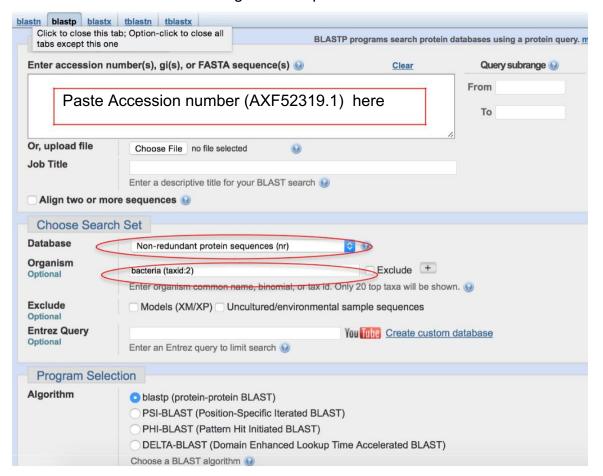
We will start with one of the predicted proteins in the database for a Microviridae bacteriophage which is the **Major capsid protein VP1**. This phage has a small number of genes that are predicted to encode proteins, but our previous work has shown that only VP1 has homologs in bacteria, so that's the only one we'll investigate here.

Part 1: Forward BLAST – looking for bacterial homologs of bacteriophage proteins



- 1. BLAST the Major capsid protein VP1 with Accession ID AXF52319.1
- 2. in the protein database
- 3. Use <black>blastp>, setting the parameters to search the <nr database> but only against <back>bacteria>. This way your top hits will be bacteria, not other bacteriophages.

Use the default settings for blastp.



3. The results of this BLAST will take you to the "Hits" page. Record the **top 10 hits** in an Excel spreadsheet **(provided on the website)** to enter into the Community Science Project database. Make sure to record the hits based on the following cut-off values:

Sequences producing significant alignments:

Select: All None Selected:0

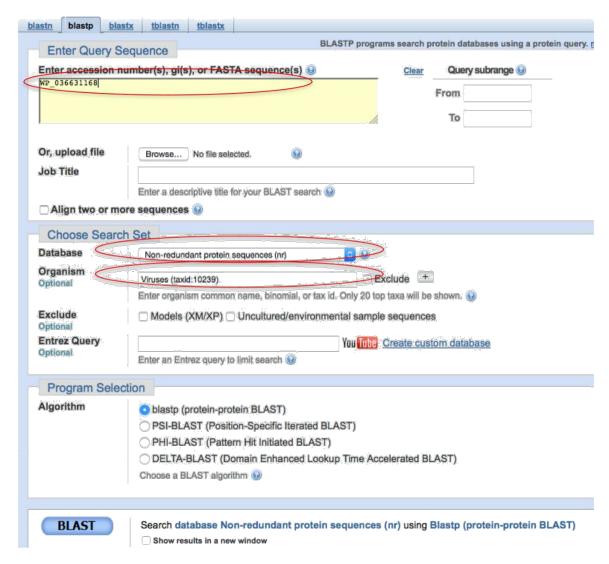


We're looking for things that are relatively closely related, therefore use > 70% query cover and < 1e-50 E-value as the cutoffs for this. If there are fewer than 10 that match these criteria, just capture these in the spreadsheet. If there are cogent reasons to change these cutoffs depending on the system you're looking at, capture this information as well.

Part 2: Reverse BLAST – looking for bacteriophage homologs of bacterial proteins

Now take the top hit from the Forward BLAST result, which in this case is: hypothetical protein [Parabacteroides distasonis] Accession number: TCI01113.1.

1. Perform another BLAST using <blastp> but this time we will be looking for phage proteins. There is no "Bacteriophage" setting, so we'll use the <nr> database and restrict our results to the <virus> sector of the database.



2. The results of the REVERSE BLAST will take you to the "Hits" page and again you will record the top 10 hits (spreadsheet provided on the website). Use the same cutoffs (normally, > 70% query cover and < 1e-50 E-value) as the you did for the forward BLAST.

Part 3: Learning about the protein universe in which the protein of interest resides

Now check to see if your Top Hit from the REVERSE Blast matches the initial phage protein you had searched for.

In our search, the Reverse BLAST top hit is: major capsid protein [Microviridae sp.] with Accession Number: AXF52319.1

This result matches with the initial phage protein we searched for -

Query ID AXF52319.1

Description major capsid protein [Microviridae sp.]

This result conveys that the two proteins are each other's closest homologs. This isn't always the case and suggests that there are other homologs that could be explored. This information can be recorded on the Reverse Blast spreadsheet provided on the website.

Part 4: Submitting your work to Genome Solver

If you want to take part in the Community Science Project, the information in your two spreadsheets should be submitted to the Community Science Project pages on the Genome Solver Qubeshub website (Information for submission to be provided later)

