**Case Study: What to do when more than one gene appears to be terminase, large subunit**

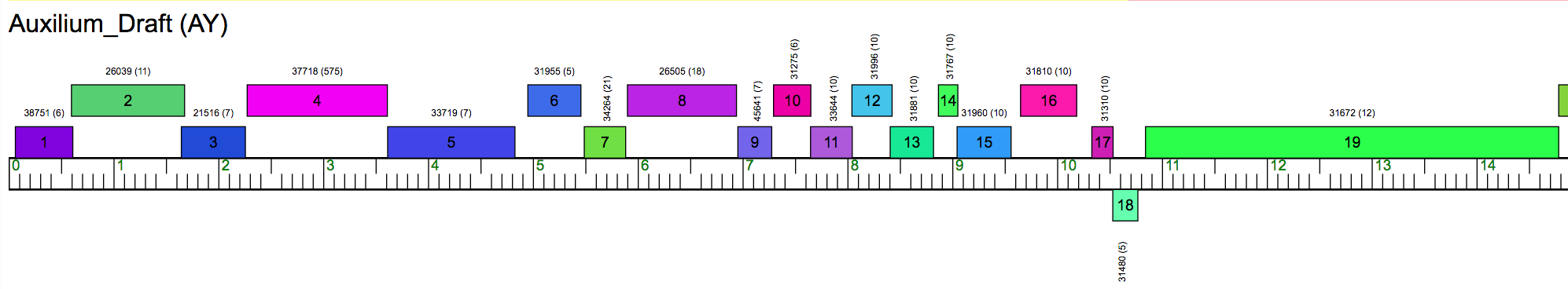
**Objectives:**

* Evaluate synteny of phage genomes and location of terminase related proteins
* Successfully evaluate HHPRED results to determine terminase domains
* Determine regions probable for intron/intein splicing

**Introduction:**

Bacteriophage genomes typically contain 1 -2 genes known as terminase. The terminase usually has ATPase, DNA binding and nuclease domains. It is responsible for the packaging of the DNA into the phage head. This gene is commonly split into two proteins which are labeled *terminase, small subunit* and *terminase, large subunit.* The small subunit of terminase is not easily recognized. When the protein is fully intact or no small subunit can be recognized the function assignment *terminase* is given to the protein.

**Cluster AY Phage Auxilium (Synteny)**

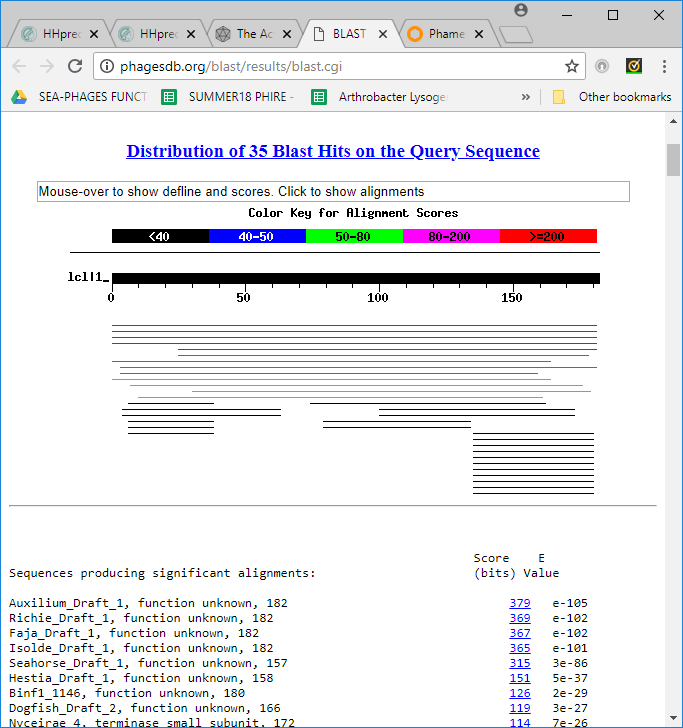


The phamerator map shows the first 19 structural proteins of Auxilium, a Cluster AY genome. Terminase proteins are typically found in the beginning of the genome. In the case of Auxilium gp1 shows a less than stellar hit to a terminase small subunit but gp2 and gp3 show excellent hits to terminase, large subunit. This is different than the typical terminase small subunit/terminases large subunit motif.

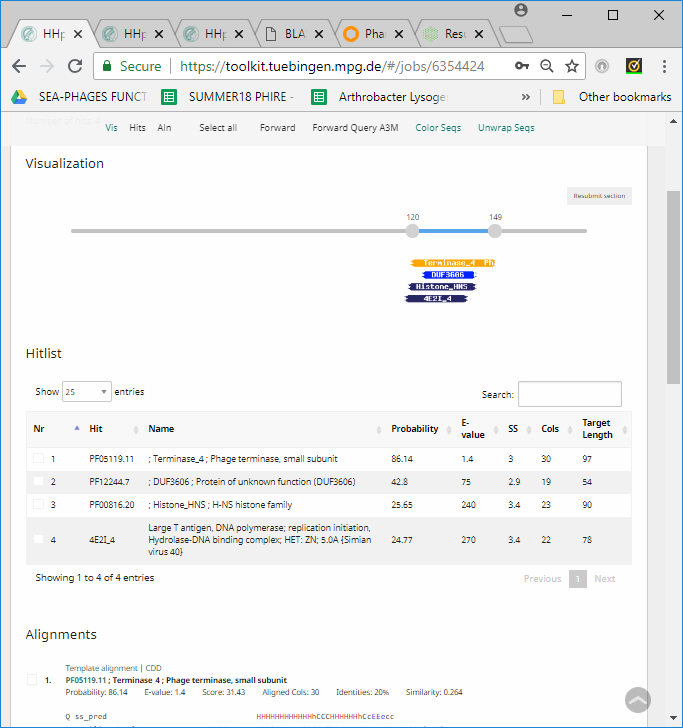
**Functional evaluation of gp1, gp2, and gp3 pf Arthrobacter phage Auxilium**

**Gp1 (64-612)d**

**BlastP at PhageDB:**

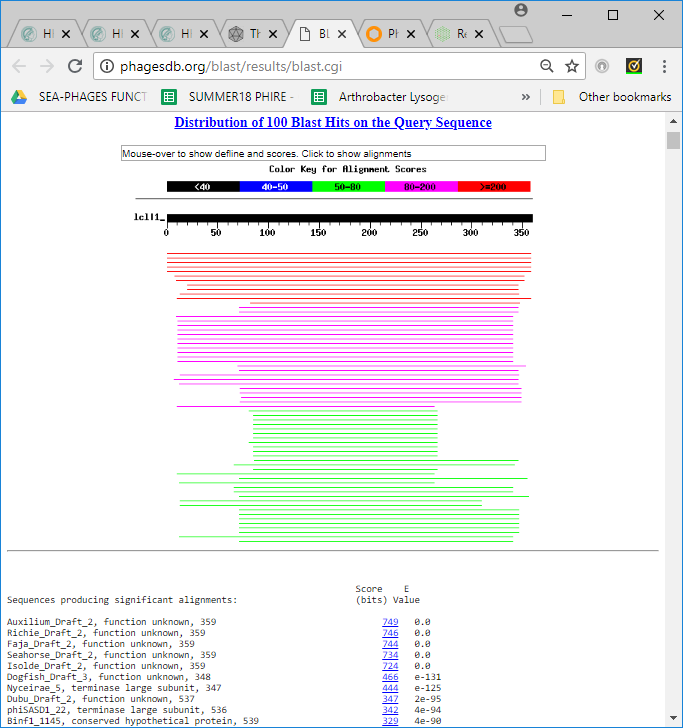


**HHPRED:**

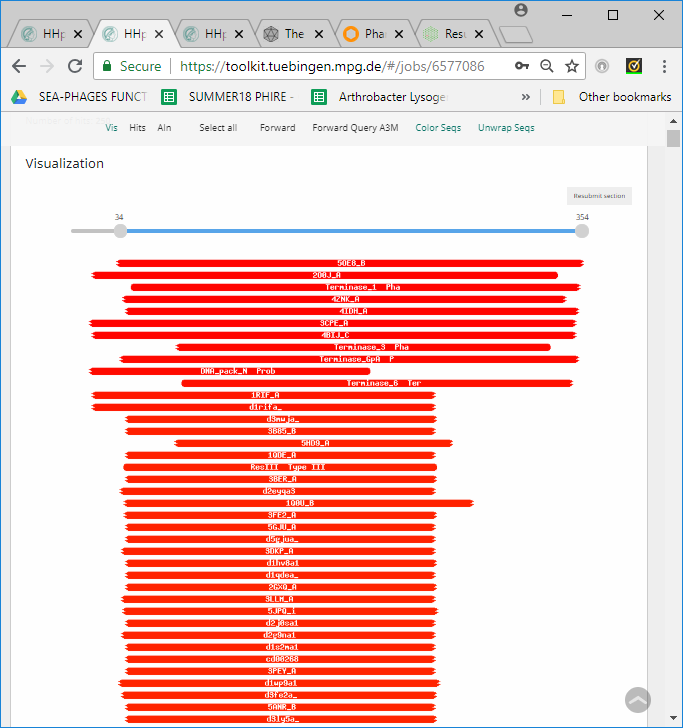


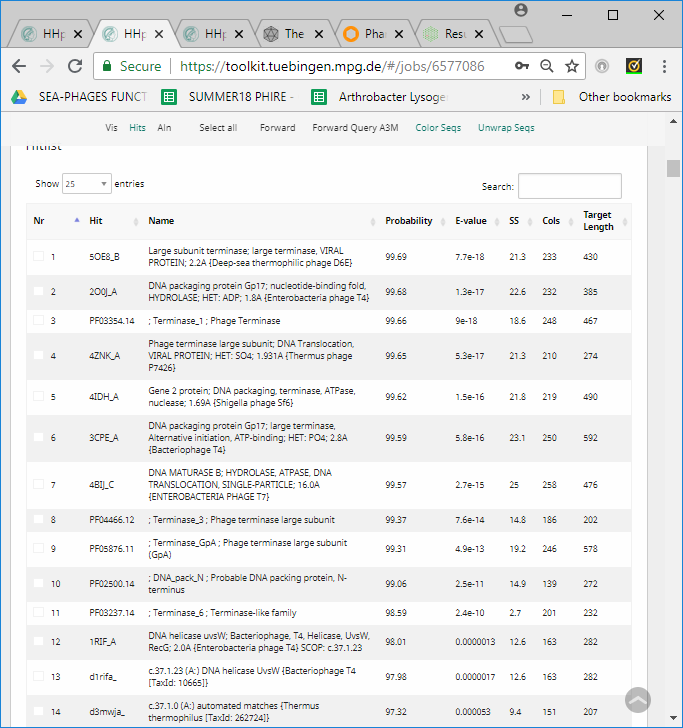
**Gp2**

**BlastP at PhagesDB:**



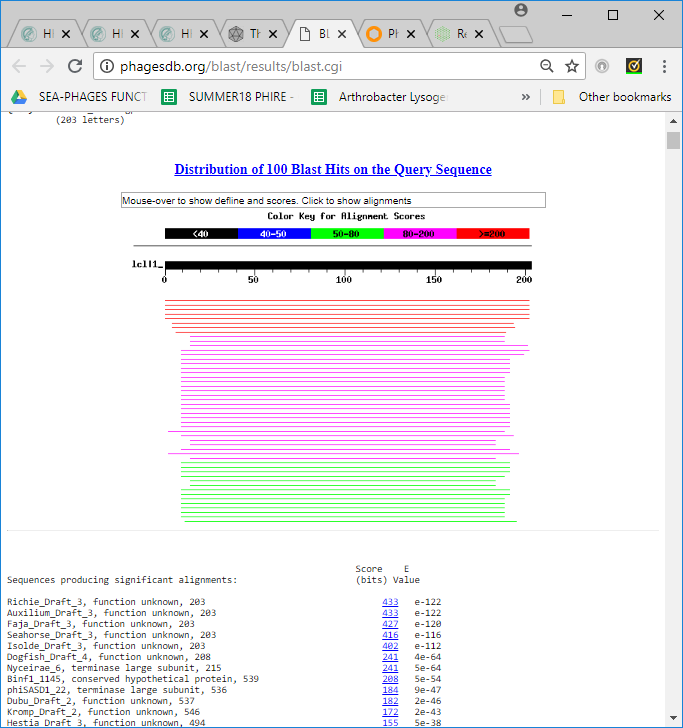
**HHPRED:**



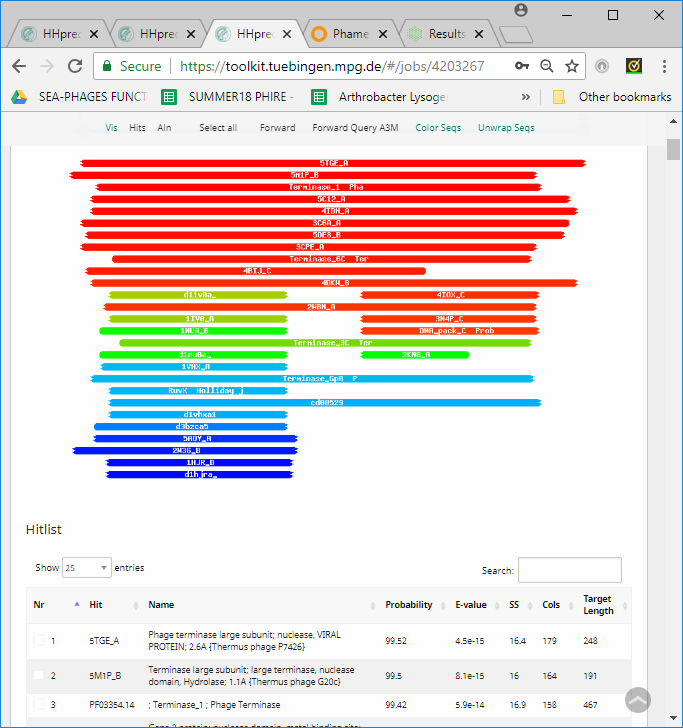


**Gp3**

**BlastP at phageDB:**



HHPRED:



**Analysis of results**

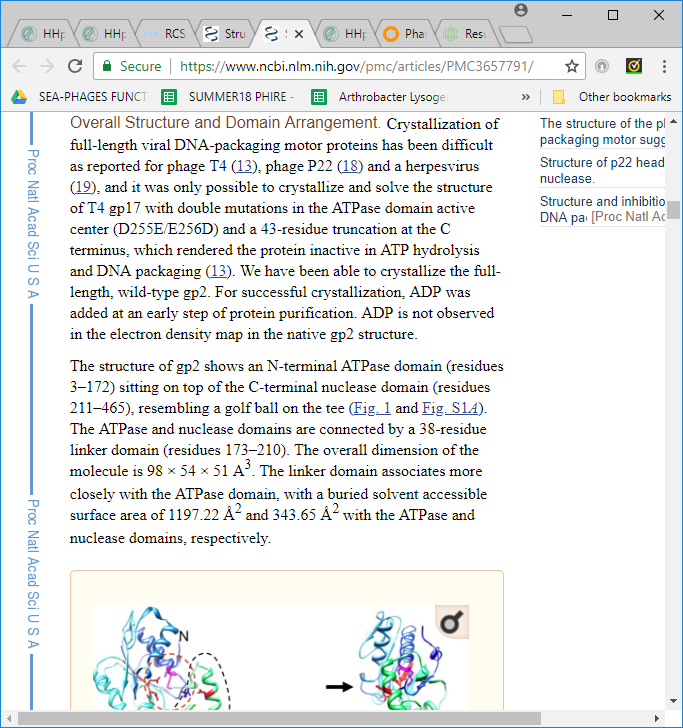
Gp1, gp2 and gp3 all have BlastP results and HHPRED results that support all three being terminase domain containing proteins. Knowing that phages are supposed to contain only 1 terminase, large subunit further analysis is needed. Understanding phage synteny a good guess is that the ATPase domain and the endonuclease domain were split into two different proteins. Let’s look at what this analysis would look like.

**Evaluation of terminase, domains**

A good place to go to get more information is HHPRED. By juxtaposing the HHPRED results I noticed that both genes gp2 and gp3 had a common result 4IDH\_A. Further investigation showed this results had

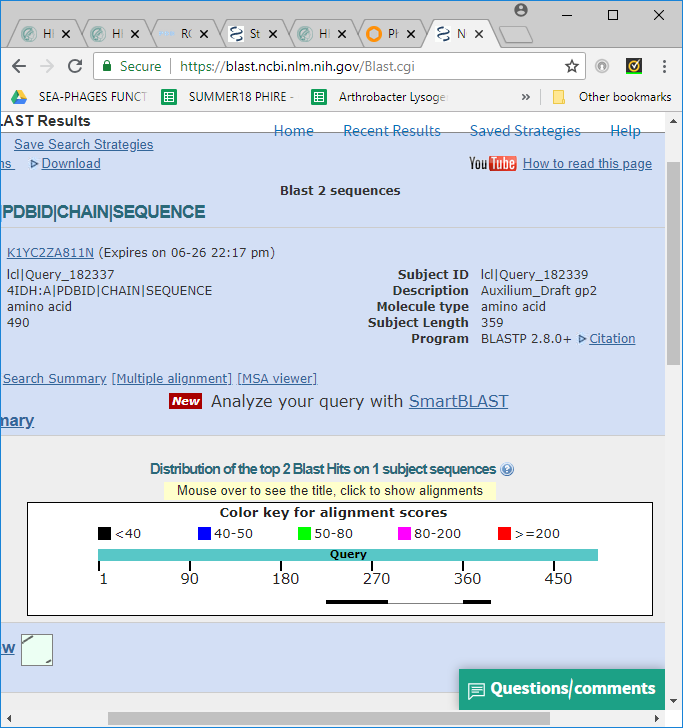
*Zhao H, Christensen TE, Kamau YN, Tang L. Structures of the phage Sf6 large terminase provide new insights into DNA translocation and cleavage. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(20):8075-8080. doi:10.1073/pnas.1301133110.*

In this paper the authors give details into the domains and which residues are responsible for the creation of these domains.



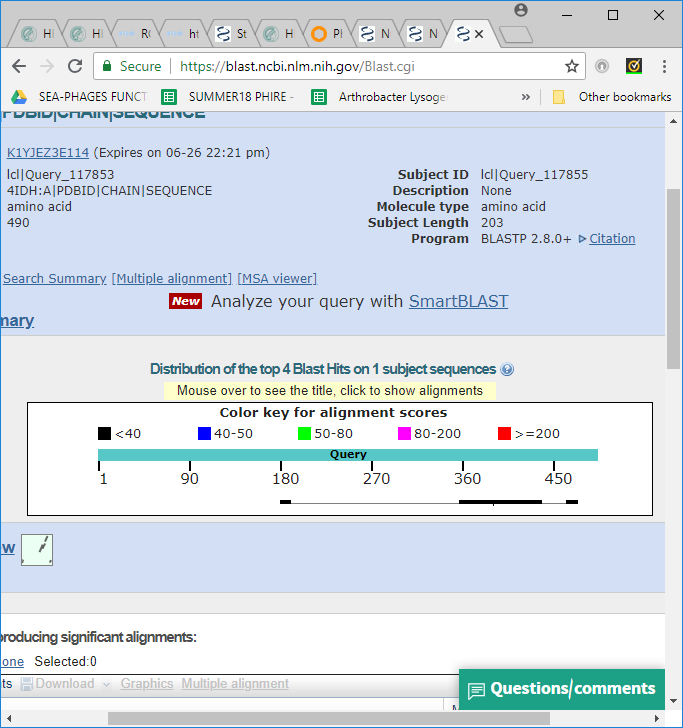
Knowing the domain regions, I was then able to go and retrieve the protein sequence for the phage protein that they isolated and do a NCBI protein-protein pairwise analysis.

**Gp2 Pairwise**



These results show that gp2 is hitting for amino acid 224-284 and then 359-385. According to the paper excerpt this would be the C-terminal nuclease domain. The second region of similarity however is most likely insignificant because the sequence in Auxilium gp2 that is blasting with it is amino acid 21-47 and the results are very poor.

**Gp3 Pairwise**



These results show that gp3 is hitting for amino acid 355-435 and then has regions of similarity that are smaller in the 460-470 region and the 179-189 however these regions of similarity are most likely insignificant. This region would also be the C-terminal nuclease domain.

**Evaluation of results:**

The pairwise analysis did not give a simple answer to the problem. If anything it made it more complicated. Now we have both proteins showing similarity with the C-terminal nuclease domain, and neither showing results with the ATPase domain. We do know though that both gp2 207-274 and gp3 53-123 show similarity with the endonuclease domain present in terminase, large subunits.