Bioinformatic Annotation of Novel Bacteriophage DuncansLeg (L3)

Connor Hadwin and Michael Moore

COASTAL CAROLINA UNIVERSITY

SEA PHAGES

hhmi Howard Hughes Medical Institute

Introduction

Bioinformatic research is increasingly important due to the rise in antibiotic resistance in bacteria. This research focuses on the novel mycobacterium phage **DuncansLeg** (75,593 base pairs). Bioinformatic tools were used to annotate and confirm genes within the DuncansLeg sequence. Functionality was determined for genes by utilizing synteny data, as well as comparing nucleotide and protein products with other published phages.

Analytical Outline

- **DNAMaster**
  - Open Reading Frame
  - Ribosomal Binding Strength
- **GeneMark**
  - Coding potential within genes
  - Start/Stop locations within genes
- **Starterator**
  - Compare called Starts within Pham and Subclusters
- **BLAST ®**
  - Nucleotide and protein database navigation and comparison

Key Annotation: HNH Endonuclease in DuncansLeg

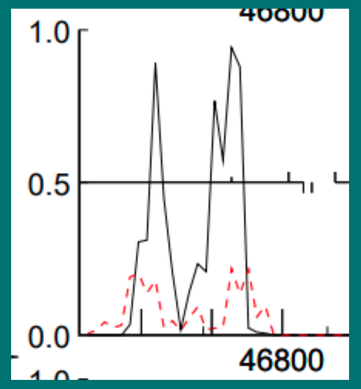

Figure 1: Coding potential from the GeneMark report on DuncansLeg was identified corresponding to a gene that wasn't initially identified. Coding potential strongly correlates to the presence of a gene at that base pair region. Coding potential is a good indicator that an uncalled gene is present.

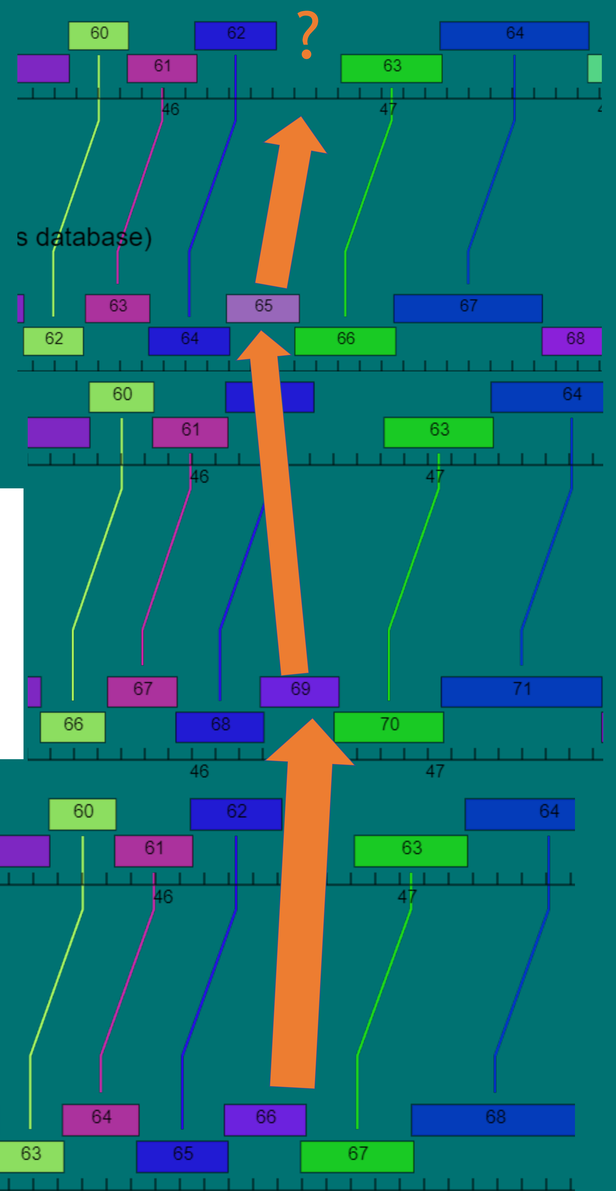

Figure 2: Synteny data shows that this gene is highly conserved throughout the L3 subcluster. This gives strong evidence that the gene should be present in DuncansLeg, even if it wasn't initially called.

```
>Lolly9_65, HNH endonuclease, 110
          Length = 110

 Score =  251 bits (640), Expect = 5e-67
 Identities = 108/110 (98%), Positives = 110/110 (100%)

Query: 1   MASGEAACRRLIKPRSEGFCERCTAWGNLTLHHRKKRSQGGLWTADNCVLLCGHGTTGCH 60
           MASGEAACRRLIKPRSEGFCERCTAWGNLTLHHRKKRSQGGLWTADNCVLLCGHGTTGCH
Sbjct: 1   MASGEAACRRLIKPRSEGFCERCTAWGNLTLHHRKKRSQGGLWTADNCVLLCGHGTTGCH 60

Query: 61  GWIEHHPDLAEAEGWHVRPWQEPSEVPLLWRGNEWVLLTPEGTMNDYHVG 110
           GWIEHHPDLAEA+GWHVRPWQEPSEVP+LWRGNEWVLLTPEGTMNDYHVG
Sbjct: 61  GWIEHHPDLAEAQGWHVRPWQEPSEVPVLWRGNEWVLLTPEGTMNDYHVG 110
```

Figure 3: A pBLAST compares this sequence to genes of other annotated phages, and reveals that the protein products are identical. This is further strong evidence that this gene both exists and has the same HNH endonuclease function.

Tools and Methods Explained

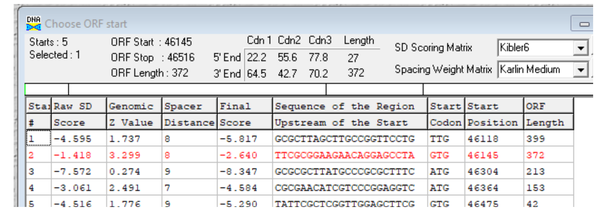

Figure 4: RBS(Ribosomal Binding Site) data provides binding affinity estimates for each gene's potential starts.

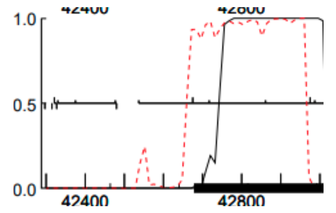

Figure 5: GeneMark indicates regions of coding potential within the genome

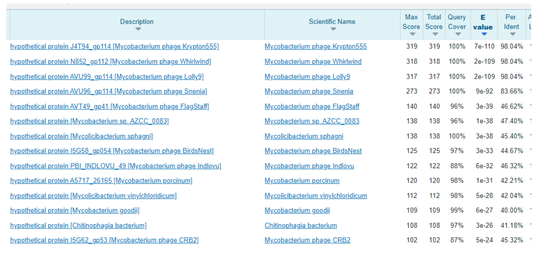

Figure 6: BLAST data shows sequence comparisons and alignment to all sequence data within the NCBI database.

Future Directions

An HNH-endonuclease is called in 75% of L3 phages, yet three L3 phages do not call this gene despite coding potential, amino acid, and positional homology to the DuncansLeg gene. Future research is needed to determine if these phages also contain HNH endonuclease.

Acknowledgements

This research was performed under Dr. M. Cevasco, as part of the HHMI SEA-PHAGES program.