

Bioinformatic Annotation of Novel Bacteriophage DuncansLeg (L3)

Connor Hadwin and Michael Moore

Key Annotation: HNH Endonuclease in DuncansLeg

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

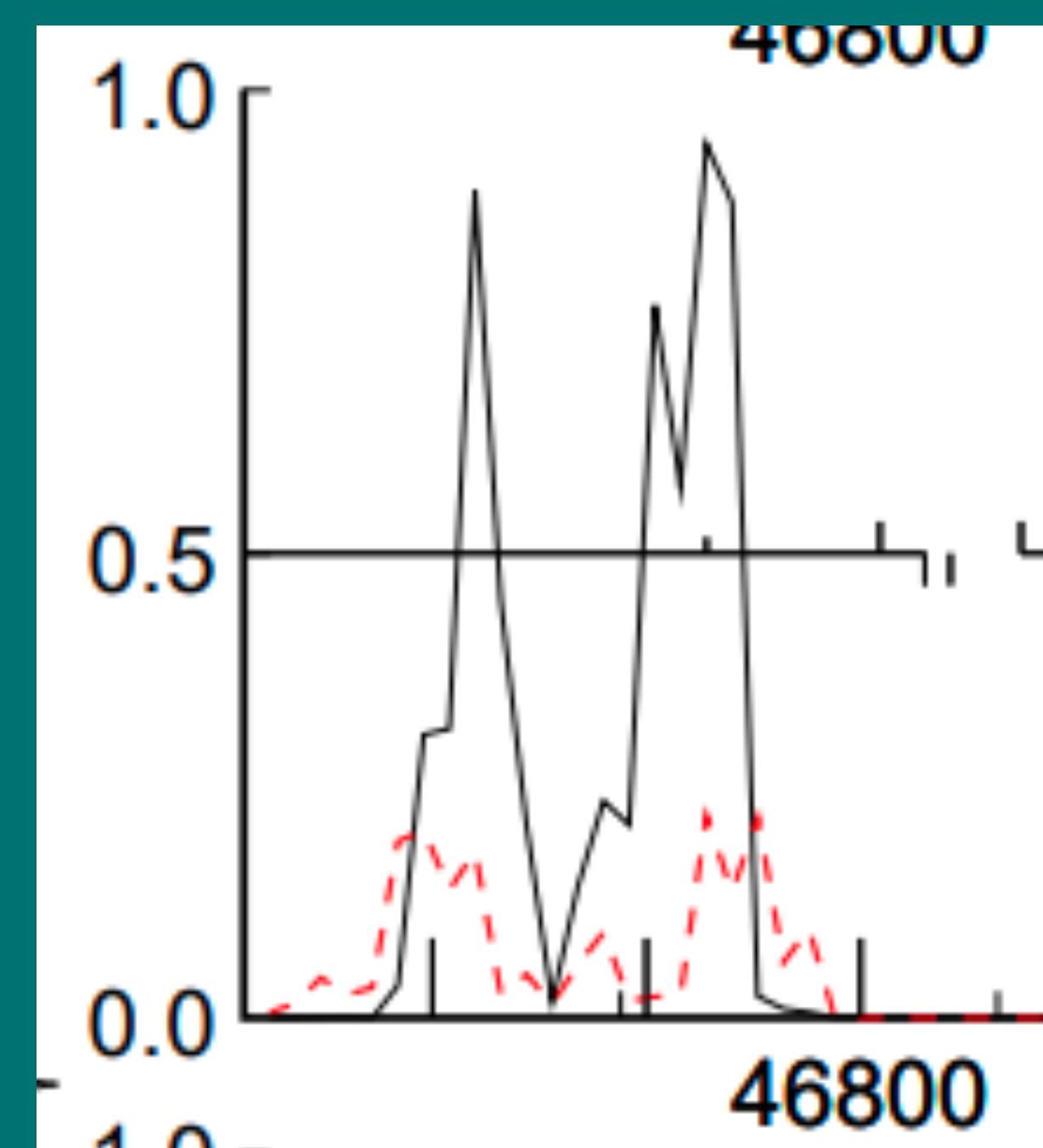
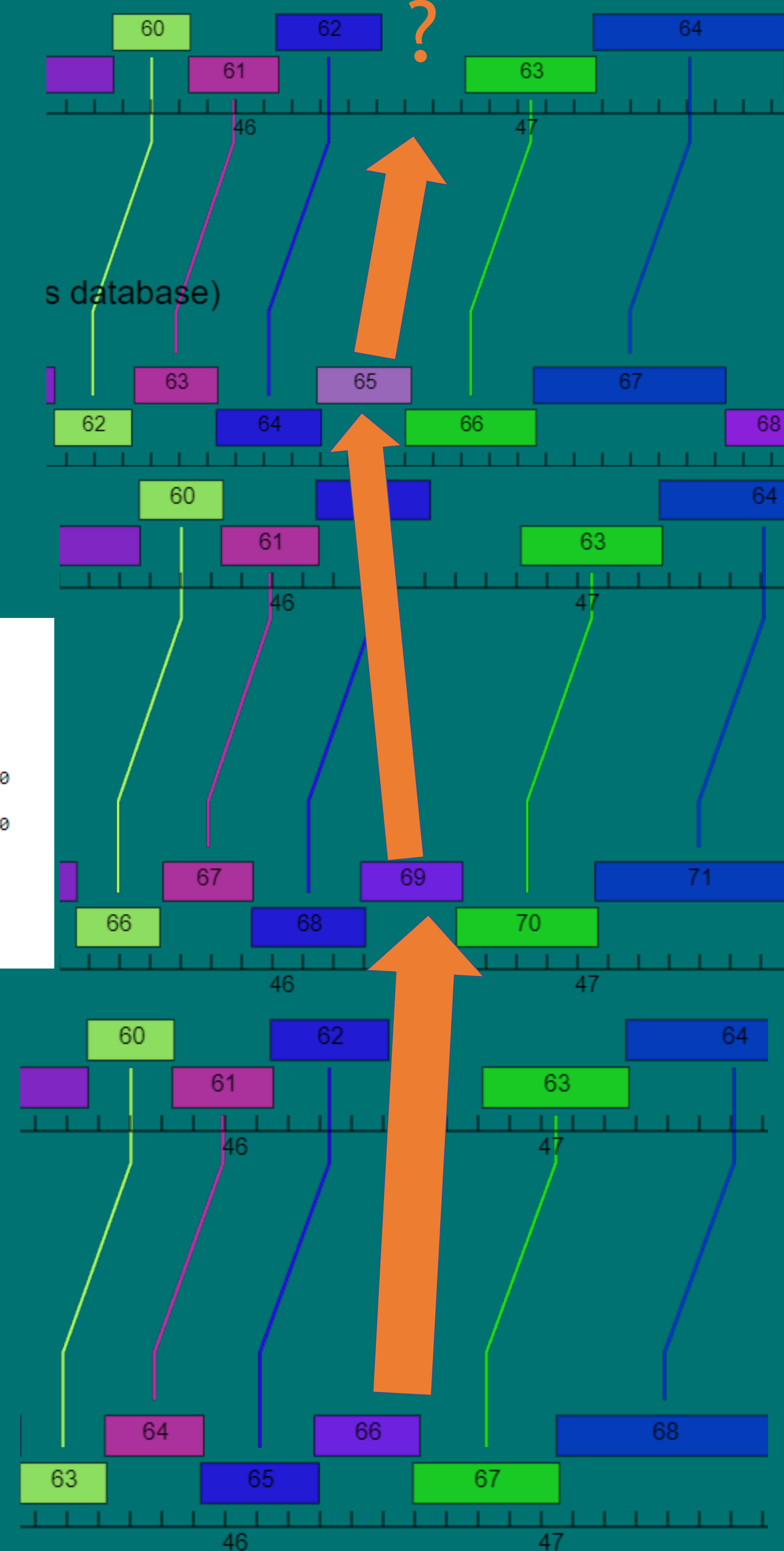


Figure 1: Coding potential from the GeneMark report on DuncansLeg was identified corresponding to a gene that wasn't initially identified. Coding potential strongly correlates to the presence of a gene at that base pair region. Coding potential is a good indicator that an uncalled gene is present.

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 MASGEAACRRLIKPRSEGF CERCTAWGNLTLHHRKKRSQGGLWTADNCVLLCGHGTGCH 60
MASGEAACRRLIKPRSEGF CERCTAWGNLTLHHRKKRSQGGLWTADNCVLLCGHGTGCH
Sbjct: 1 MASGEAACRRLIKPRSEGF CERCTAWGNLTLHHRKKRSQGGLWTADNCVLLCGHGTGCH 60
Query: 61 GWIEHHPDLAEAEAGMHRPWPQEPSEVPLLRGNEWVLLTPEGTMNDYHVG 110
GWIEHHPDLAEAEAGMHRPWPQEPSEVPLLRGNEWVLLTPEGTMNDYHVG
Sbjct: 61 GWIEHHPDLAEAEAGMHRPWPQEPSEVPLLRGNEWVLLTPEGTMNDYHVG 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

Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF
#	Score	2 Value	Distance	Score	Upstream of the Start	Codon	Position	Length
1	-4.595	1.737	8	-5.817	GCGCTTAGCTTGCCTGCTCTG	TTG	46118	399
2	-1.418	3.259	8	-2.640	TTCCGGAGAACAGGAGCCTA	GTG	46145	372
3	-7.572	0.274	9	-8.347	GCGCGCTTAGCGCCGCTTTC	ATG	46304	213
4	-3.061	2.491	7	-4.584	CGCGACATCGTCCCGGAGGTC	ATG	46364	153
5	-4.516	1.776	9	-5.250	TATTGCTCGCTTGCAGCTTCG	GTG	46475	42

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

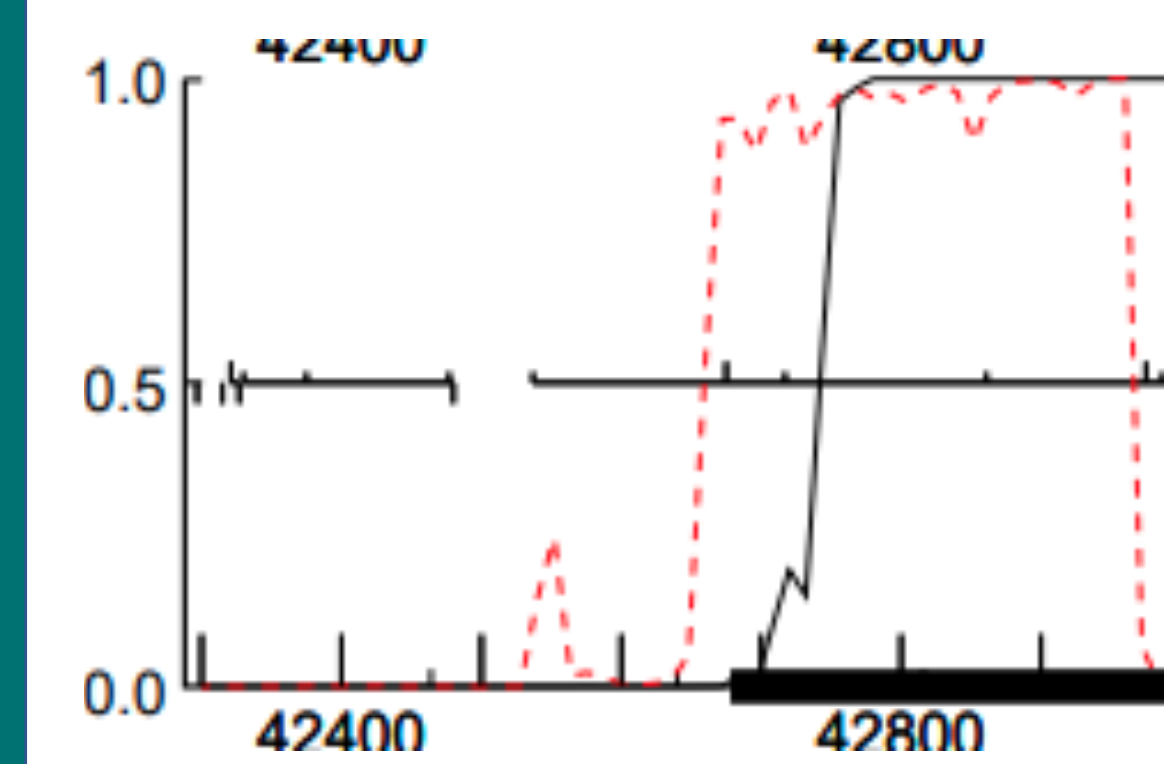


Figure 5: GeneMark indicates regions of coding potential within the genome

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident
hypothetical protein_HTB4_gp114 [Mycobacterium phage Krypton555]	Mycobacterium phage Krypton555	319	319	100%	7e-110	98.04%
hypothetical protein_N852_gp112 [Mycobacterium phage Y301b1e1f]	Mycobacterium phage Y301b1e1f	318	318	100%	2e-109	98.04%
hypothetical protein_AK09_gp114 [Mycobacterium phage Luf9]	Mycobacterium phage Luf9	317	317	100%	2e-109	98.04%
hypothetical protein_AK743_gp111 [Mycobacterium phage Szentia]	Mycobacterium phage Szentia	273	273	100%	9e-92	83.66%
hypothetical protein_AK743_gp111 [Mycobacterium phage FlayStar]	Mycobacterium phage FlayStar	140	140	96%	3e-39	46.62%
hypothetical protein [Mycobacterium sp. AZCC_0883]	Mycobacterium sp. AZCC_0883	138	138	100%	1e-38	47.40%
hypothetical protein [Mycobacterium sebae]	Mycobacterium sebae	138	138	100%	3e-38	45.40%
hypothetical protein_IG58_gp054 [Mycobacterium phage Biddistest]	Mycobacterium phage Biddistest	125	125	97%	3e-33	44.67%
hypothetical protein_PBI_IND00V_49 [Mycobacterium phage Indoox]	Mycobacterium phage Indoox	122	122	88%	6e-32	46.32%
hypothetical protein_A5717_26165 [Mycobacterium porcinum]	Mycobacterium porcinum	120	120	98%	1e-31	42.21%
hypothetical protein [Mycobacterium vivaxhodidum]	Mycobacterium vivaxhodidum	112	112	98%	5e-28	42.04%
hypothetical protein [Mycobacterium goodii]	Mycobacterium goodii	109	109	99%	6e-27	40.00%
hypothetical protein [Chitinophaga bacterium]	Chitinophaga bacterium	108	108	97%	3e-26	41.18%
hypothetical protein_IG67_gp63 [Mycobacterium phage CRB2]	Mycobacterium phage CRB2	102	102	87%	5e-24	45.32%

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.